Motivating Mixtures: Stochastic Choice and Convex Preferences*

Jack Adeney

December 4, 2025

Abstract

Expected Utility requires that individuals never strictly prefer a convex mixture of two lotteries over either of the two. However, recent experimental literature has found that these "mixture effects" are pervasive. Decision theoretic models used to rationalize such effects can broadly be split into three categories: 1) decision makers have a strict preference for mixtures, 2) mixing is the realization of strict preferences with noise, and 3) mixing is a representation of uncertainty over preferences. I develop a two-part experiment to identify the dominant mechanism underlying mixture effects. I find that less than 15% of observations are consistent with convex preferences, whereas approximately 75% are consistent with stochastic choice. These findings have important implications for policy development and behavioral welfare analysis when interpreting stochastic choice data.

*California Institute of Technology, HSS Division. *jadeney@caltech.edu*. The experiment reported in this paper was preregistered in the AEA RCT Registry in June 2025, under the ID AEARCTR-0016288. The experiment was reviewed and granted an exemption by the Institutional Review Board at the California Institute of Technology under protocol number 25-1544.

1 Introduction

Decision making under risk is central to many economic choices. The canonical model of expected utility (EU) provides a means of evaluating economic prospects when the outcome is uncertain. Since the development of EU, many counter-examples have surfaced to highlight the inefficacy of this model in simple decision making contexts. One of the most notable is the Allais paradox (Allais, 1953), which demonstrates a frequent violation of EU due to a preference for certain outcomes. While the Allais paradox demonstrates a violation due to "certainty effects", recently there has been a growing interest in exploring another type of deviation from EU—"mixture effects".

To demonstrate exactly how mixture effects violate EU, imagine that an individual is faced with a choice between two lotteries and their 50-50 mixture. The independence axiom of EU ensures that the utility of the mixture always falls between the utility of the most preferred and the least preferred lottery. This implies that the 50-50 mixture is never strictly preferred to both constituent lotteries. Yet, a growing literature has noted that experimental subjects frequently express a preference for mixtures in their choice data (Camerer and Ho, 1994; Butler and Loomes, 2011; Agranov and Ortoleva, 2017, 2020; Feldman and Rehbeck, 2022). These data have been the focus of many new theoretical models attempting to explain mixture effects with different psychological mechanisms and policy implications. Importantly, existing experiments cannot distinguish between these candidate theories, leaving a critical gap in the experimental literature. In this manuscript, I develop a novel experiment designed to disentangle prominent explanations of mixture effects. One of the three explanations emerges as the primary rationalization of mixtures—stochastic choice. Evidence in favor of convex risk preferences and incomplete preferences is present within the data but account for a small proportion of observations when compared to stochastic choice. These findings suggest new ways for evaluating evidence of mixture effects and the models proposed to rationalize them.

Section 2 provides an introduction to the literature documenting mixture effects. Beginning with Mosteller and Nogee (1951), there is overwhelming evidence that mixture effects are present within a variety of decision making environments. Given this fact, more recent research explores whether a willingness to mix in one environment is correlated with a willingness to mix in others (Agranov and Ortoleva, 2017; Feldman and Rehbeck, 2022). Studies of these different environments has generated interest in mixtures as objects that can themselves be used to make inference about preferences. *Section 2* discusses the approaches that different studies have taken in order to do this.

The theoretical literature has responded to this evidence by exhibiting a sharp growth in the number of rationalizing models. *Section 3* introduces some prominent models have have proved foundational to the way in which I interpret mixture effects. These models are often based on similar theoretical foundations, such as a non-degenerate set of utility functions, yet a deeper exploration uncovers some contrasting normative motivations behind constructing mixtures. From these normative motivations I develop three explanations designed to represent each, and give examples of specific models that sit within them. The explanations are as follows; 1) individuals have convex preferences and therefore can have a strict preference for mixtures, 2) mixing can be interpreted as a symptom of the interaction between strict preferences and noise, and 3) mixing occurs as a result of uncertainty over preferences. An analysis of previous experimental research reveals that current data is not equipped to disentangle each explanation. This motivates the necessity for a two-part experimental design where individuals can mix over lotteries, then mix again over one of the original lotteries and the mixture that they previously constructed. This simple addition to otherwise standard experimental protocol provides, to a large extent, the necessary identification.

Section 4 develops a theoretical framework and states the predictions for each explanation within the context of our experiment. The two-part design allows us to directly observe how individuals interact with the mixtures that they previously generated. To our knowledge, this aspect of the design has never been used before and is fundamental for differentiating between explanations. I finish by highlighting the difficulty of comparing each explanation in an unbiased manner and introduce my solution—a bootstrapping technique that compares each explanation against what I refer to as the "naive" and "empirical" null hypotheses.

Section 5 describes the experimental design in greater detail. Data from a total of 900 participants over three treatments were collected using an online sample. Each treatment is designed to mimic a natural environment in which mixing might occur, for example selecting once over a convex choice set or selecting multiple times over a binary choice set. Having preference uncertainty as an explanation creates additional challenges for incentivization because there is no choice prediction when individuals cannot construct a preference. This necessitates an opt-out aspect of the design, which allows participants to completely skip the mixture generating process in exchange for a pre-specified, yet undisclosed substitute lottery. This non-specification decision acts as a proxy for preference uncertainty.

In *Section 6* I state the results of the experiment. The main findings are as follows. Across all treatments, approximately 10% to 15% of *Part 1* answers are left unspecified, meaning that individuals overwhelmingly prefer to generate mixtures themselves. Of those that are specified, approximately 50% are degenerate mixtures, meaning that they place full weight on one of the lotteries. Half of all specified *Part 1* answers therefore result in non-degenerate mixtures, which is substantially more than predicted by EU. Of those specified and non-degenerate observations, one explanation is dominant in explaining the data—stochastic choice. Between 75% and 80% of observations are consistent with stochastic choice, which is significantly more than one would expect from an individual choosing mixture weights uniformly at random. Analysis at the individual level suggests that this is a stable trait within decision makers. On the other hand, convex preferences appear to represent a very small number of observations, and this value is significantly less than the null hypotheses. Evidence of preference uncertainty is also present, but contributes very little to the overall proportion of decisions within our dataset.

Section 7 concludes with a discussion of the findings. Overall, mixtures tend to result from stochastic processes more so than a strict preference for mixtures. Therefore, providing individuals with mixtures generated by aggregating individual decisions is not likely to provide as much welfare as giving individuals the option that they choose most frequently. I discuss the implications of these findings for policy development and behavioral welfare analysis more generally.

2 Evidence of Mixture Effects

Mixture effects have been studied throughout the experimental literature dating back to Mosteller and Nogee (1951), who were the first to note that the empirical frequency of accepting a bet is positively related to the certainty equivalent of that bet.

More recent literature has found that mixing occurs frequently in both repeated choice environments and over convex menus. Agranov and Ortoleva (2017) relate randomization over repeated decisions with models such as Cautious Stochastic Choice (Cerreia-Vioglio et al., 2019), Random Utility (Gul and Pesendorfer, 2006) and incompleteness, finding that 71% of participants choose different lotteries in the same question when repeats are consecutive, and 90% choose different lotteries when repeats are separated. Agranov et al. (2023) find that randomization is an individual trait when the same question is repeated consecutively—17% of participants never mix as opposed to 52% who always mix. The study most similar to ours is that of Feldman and Rehbeck (2022). They directly relate repeated discrete choice tasks with decisions over convex menus. 94.4% of participants have some preference for mixing, and choices from repeated discrete choice tasks are positively correlated with choices from the convex menus. A follow up experiment tests for mixing behavior over two monetary payoffs, which results in a dominance relation (Rehbeck and Stelnicki, 2024). Even with dominated options, they find that more than 70% of created mixtures put a strictly positive probability on the lower monetary payoff. Their studies differ in the sense that they do not have the two part design in which decision makers can directly interact with their previously generated mixtures. Agranov and Ortoleva (2020) test willingness to mix with a novel adaptation of a multiple price list that allows individuals to randomize between options at each row of the list. Over 75% of subjects report non-degenerate mixtures in at least two thirds of questions Agranov and Ortoleva (2022) provides an overview of studies related to revealed preferences for randomization.

The following sections will discuss how preference uncertainty may lead to mixture effects. A specific branch of research relates preference uncertainty to the "preference reversal phenomenon", which suggests that preferences appear contradictory, or mixed, when measured in different domains, for example binary choices versus valuations (Butler and Loomes, 2007, 2011). It's pervasiveness has long since been documented, and summarized in Seidl (2002). Cettolin and Riedl (2019) more explicitly relate preferences under uncertainty and randomization, finding that approximately 16% of participants demonstrate behavior consistent with both incomplete preferences and a preference for randomization. Experimental research that relates mixing with incomplete preferences often require non-traditional forms of incentivization. The benefits and drawbacks of these designs are discussed in *Section 5.4*.

3 Preferences for Mixtures and Rationalizing Theories

Consider a finite set of outcomes $X \subset \mathbb{Z}$ where $|X| = N$. A lottery ℓ can be defined as a probability vector $\ell := \langle p_1^{\ell}, ..., p_N^{\ell} \rangle$ on X. A (binary) menu of two lotteries ℓ and ℓ' is denoted $\{\ell, \ell'\}$. Decision makers are permitted to construct mixtures over each binary menu where a mixture over menu $\{\ell, \ell'\}$ is represented by

$$m(\alpha; \ell, \ell') := \langle \alpha p_1^{\ell} + (1-\alpha) p_1^{\ell'}, ..., \alpha p_N^{\ell} + (1-\alpha) p_N^{\ell'} \rangle$$

for $\alpha \in [0, 1]$. Let the weight placed on lottery ℓ by the decision maker over the menu $\{\ell, \ell'\}$ be denoted as $\alpha^*(\ell, \ell')$. $m^*(\ell, \ell')$ is used as an abbreviation for $m(\alpha^*(\ell, \ell'); \ell, \ell')$.

In addition, each decision maker has a utility function, U, over the finite lottery space. At this stage there are no further assumptions placed on the characteristics of U, other than that it is well defined for each finite lottery. All other characteristics will be stated within the explanations below.

3.1 Potential Rationalizing Theories

A number of theoretical models have been developed to explain how these mixtures are chosen by decision makers. Each of these models imply different normative motivations underlying the mixture generation process, and therefore impose conditions on how decision makers should interact with menus in order to generate $m^*(\ell, \ell')$. In what follows, the models are categorized into three explanations, each of which represents a different type of normative motivation behind generating mixtures.

3.1.1 Individuals strictly prefer mixtures

Allowing individuals to construct mixtures between alternatives increases the menu from the two original alternatives to the entire convex hull of the two. Under this explanation, selecting a mixture represents choosing a mixture that is strictly preferred to all other available mixtures. Formally, there is a mixture weight, α^*, such that

$$\alpha^*(\ell, \ell') = \arg\max_{\alpha}\{U(m(\alpha; \ell, \ell')) : m(\alpha; \ell, \ell') \in Co(\{\ell, \ell'\})\}$$

An example of a model that falls into this explanation is Cautious Expected Utility (CEU; Cerreia-Vioglio et al. 2015, 2019). CEU functions on the premise that decision makers have a non-singular set of utility functions. When a decision is to be made between alternatives, the value of each alternative is defined as the lowest certainty equivalent given by all of the utility functions. The chosen alternative is then the alternative with the highest value. By comparing across the set of utility functions, CEU allows preferences to be convex, meaning that a mixture between two lotteries can be strictly preferred to either of the two original lotteries, unlike traditional EU.

Although this formulation is much more general than suggested by CEU, it highlights the key mechanism behind mixture effects within this explanation class. Namely, the mixture is generated because it is the most preferred option from the set of all possible mixtures.

3.1.2 Mixtures reflect choice proportions of strict preferences with noise

The second interpretation is that decision makers are considering their preference over the two initial lotteries, yet the mixture weights represent a stochastic choice proportion over the lotteries. For example, suppose that the decision maker prefers ℓ from the menu $\{\ell, \ell'\}$ in the sense that $U(\ell) > U(\ell')$. In the absence of noise, one would expect the decision maker to choose lottery ℓ over ℓ' for all decisions made over the menu. However, an introduction of noise could result in the noisy utility of ℓ' occasionally being larger than that of ℓ, and therefore ℓ' is chosen on some occasions. Let $\Gamma : \mathbb{R}^2 \to \mathbb{R}$ represent utility of lottery ℓ after the interaction with noise ϵ. These 'random' utilities are used to make decisions over menus in the stochastic environment. With this notation we can begin to develop our predictions for stochastic choice[1]:

[1] Depending on which assumptions are applied to the distribution of noise, this setup provides the same predictions as a Moderate Utility Model from He and Natenzon (2024), where $P(\ell, \{\ell, \ell'\}) = F\left(\frac{u(\ell)-u(\ell')}{d(\ell,\ell')}\right)$. $P(\ell, \{\ell, \ell'\})$ represents the probability of picking ℓ from the menu $\{\ell, \ell'\}$ and $d(\ell, \ell')$ represents some distance metric on the finite support lottery space.

$$\alpha^*(\ell, \ell') = \mathbb{P}\left[\mathit{\Gamma}(U(\ell), \epsilon_\ell) > \mathit{\Gamma}(U(\ell'), \epsilon_{\ell'})\right]$$

A prominent model within this explanation category is the Random Utility Model (RUM; McFadden (1972); Gul and Pesendorfer (2006)). RUM again assumes a non-singular set of utility functions. The difference between this model and CEU is the way in which these utility functions manifest themselves when a decision is made. Instead of valuing alternatives according to the minimum certainty equivalent across all utility functions, RUM takes a stochastic draw from the set of utility functions. Alternatives are then evaluated according to that utility function and the alternative providing the highest utility is chosen. Each choice is associated with its own stochastic draw of a utility function, meaning that even decisions from identical menus can result in different alternatives being chosen. In this scenario, a mixture can be considered as representing the relative proportions with which each alternative in the menu is chosen. This explanation captures mixture effects as a less deliberate act, and more as an illustration of preference strength over the two original alternatives.

3.1.3 Mixtures represent uncertainty over preferences

Very few models make choice predictions when the preference ordering over alternatives is uncertain. An extreme of this scenario is when a preference is incomplete.[2] Given that there are no predictions as to what a rational decision maker would choose given that the preferred alternative is unknown, selecting a non-degenerate mixture across options should be no more nor less likely than a degenerate mixture (Danan, 2010). Put simply, it seems natural that mixture effects could be a symptom of individuals expressing their uncertainty over which alternative they prefer.

One of the most popular models from the decision theory literature allowing for incomplete preferences is the Multi Utility Model (MUM; Evren (2008); Evren and Ok (2011)). Again, MUM requires a set of utility functions, and again the difference in predictions over choices occurs as a result of the way in which these utility functions are used to generate a choice. In this model, a preference exists between two alternatives if and only if all the utility functions agree on the relative ranking between the two options. If for example all but one prefer ℓ to ℓ' but one prefers ℓ' to ℓ, then the relation between ℓ and ℓ' is deemed incomplete. When this occurs there is no prediction as to what the individual is going to choose.

[2] "Incompleteness" within the decision theory literature refers to a scenario where neither alternative is weakly preferred to the other in the underlying preference relation.

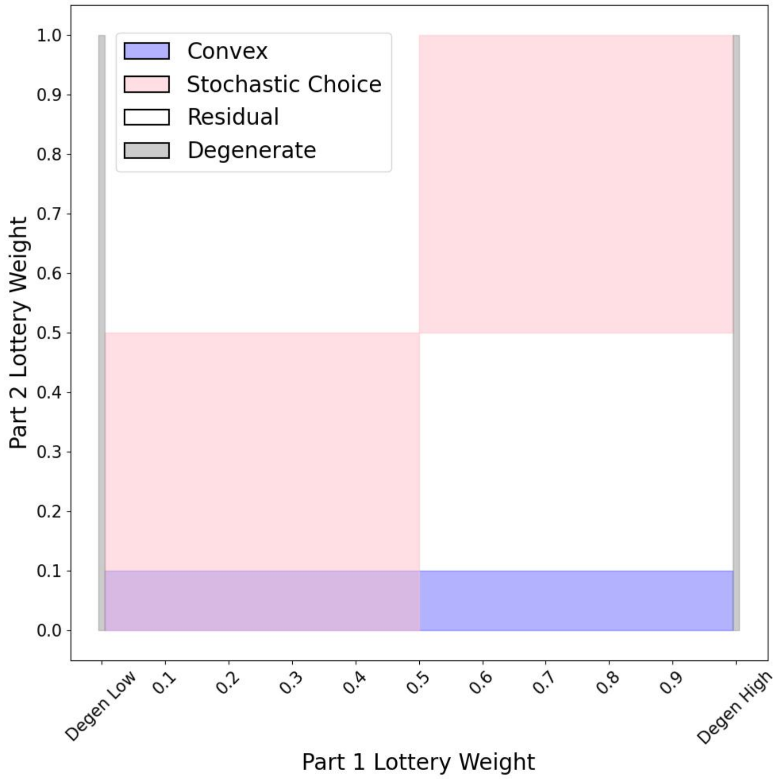
It is worth noting that these explanations are not exhaustive of all theoretical models. For example, models that rely on stochastic consideration, such as (Manzini and Mariotti, 2014) incorporate neither convex preferences, nor do they have noise directly interacting with preference. However, limiting the focus to binary menus in this study implies that one of the two lotteries must not be considered in order for stochastic consideration to be driving the mixture effects. This doesn't seem particularly likely. Drift Diffusion Models (Ratcliff, 1978) and models of bounded rationality are similarly not captured by any of the three explanations.

4 Differentiating Candidate Rationalizations

All of the explanations above provide a rationalization for mixture effects. However, they each suggest different mechanisms underlying the mixture generating process. Current data are unable to dissect these mechanisms, meaning that they cannot shed light on which mechanism is primarily responsible. It is important to understand which of these explanations is the driving force due to the different normative objectives that they each represent. Any policy maker attempting to maximize welfare needs to understand such objectives in order to adequately do so. Given these facts, I introduce a minor addition to the traditional empirical approach for identifying mixture effects—allowing individuals to construct mixtures as they do in standard designs and then allowing them to directly interact with those mixtures in future menus. Specifically, consider a design in two parts where *Part 1* features menus of the form $\{\ell, \ell'\}$ and *Part 2* features menus of the form $\{\ell, m^*(\ell, \ell')\}$, where $m^*(\ell, \ell')$ is the mixture generated from the *Part 1* question. Using this design, one can observe how individuals mix over previously generated mixtures when confronted with them in future questions. The relative weights assigned to these mixtures in *Part 2* menus allow us to disentangle each of the explanations.

Figure 1 depicts the predictions of the first two explanations in the paired mixture weight space. The x-axis represents the weight placed on the lottery in *Part 1*, $\alpha^*(\ell, \ell')$. The y-axis represents the weight placed on the lottery in *Part 2*, $\alpha^*(\ell, m^*(\ell, \ell'))$. As we will see, convex preferences imply that $m^*(\ell, \ell')$ is created in *Part 1* because it is strictly preferred to both ℓ and ℓ', therefore the individual should place full weight on $m^*(\ell, \ell')$ in *Part 2*. For clarity, our interface only allows lottery weights to be accurate to the first decimal place, meaning that there is some coarseness in the sample data. To account for this, we expand the convex consistent region to either full or 90% weight on the mixture. The region consistent with convex preferences can therefore be represented by the purple region in *Figure 1*. On the other hand, stochastic choice implies that the relative weights on lotteries are indicative of preference direction. For example if $m^*(\ell, \ell')$ is made 70% of ℓ and 30% of ℓ' then ℓ is assumed preferred to ℓ'. ℓ should also be preferred to $m^*(\ell, \ell')$ and

Figure 1: Theoretical Predictions



therefore more than 50% of weight should be placed on ℓ in *Part 2*. The reverse should also be true for *Part 1* weights less than 0.5.[3] Regions consistent with this explanation are shaded in pink on *Figure 1*. Finally, preference uncertainty implies that individuals do not know how to specify their mixture. If this is true, then they would likely prefer to bypass the question in exchange for a pre-determined mixture. If they bypass the question in *Part 1*, then it also seems likely that they will bypass the related question in *Part 2*. This prediction cannot be observed in the space of *Figure 1* and so necessitates the opt out property of the experimental design. Further details as to why this hypothesis might be reasonable are given in *Section 4.3*.[4]

4.1 Convex preferences

If the individual has convex preferences then they construct the *Part 1* mixture because they strictly prefer it to all other mixtures. Given that the available set of mixtures in

[3]Additional assumptions are required for this prediction to hold, and are explained in more detail in *Section 4.2*.

[4]These intuitive hypotheses are actually satisfied by some common theoretical models of incomplete preferences, for example the Expected Multi-Utility Model (Evren and Ok, 2011).

Part 2 is a subset of those available in *Part 1*, the individual should place full weight on the generated mixture in the *Part 2* question in order to remain consistent. This provides our first testable hypothesis:

Hypothesis 1: Convex Preferences

$$\alpha^*(\ell, m^*(\ell, \ell')) = 0$$

4.2 Stochastic Choice

Take the same two part setup as before but suppose the second explanation is driving mixture effects. Under this interpretation, we are less concerned with the mixture lottery as an object of focus but rather the relative weights placed on each of the constituent lotteries. As such, there is no reason to assume that the mixture generated in *Part 1* will be selected again in *Part 2*. The stochastic choice interpretation suggests that the relative weights on each of the lotteries are indicative of preference direction, and this direction should be consistent across *Part 1* and *Part 2* questions. We require two additional assumptions on the noise distribution and utility structure in order to identify this within the experimental setup.

Assumption 1:

- *1a.* $\Gamma(\cdot, \cdot)$ is strictly monotone increasing in both arguments and $\Gamma(x, 0) = x$
- *1b.* $(\epsilon, \epsilon') \sim F_{\epsilon,\epsilon'}(\cdot, \cdot)$ where $F_{\epsilon,\epsilon'}$ is continuous and symmetric around the vector $(0, 0)$.

Assumption 1a. implies that for any fixed realization of noise, $|\Gamma(U(\ell), \epsilon) - \Gamma(U(\ell'), \epsilon)|$ is increasing with the difference of the underlying utilities. The condition that $\Gamma(x, 0) = x$ is not necessary for the purposes of the stochastic choice explanation, but is useful as it nests the deterministic scenario by returning the true underlying utility in the absence of noise. *Assumption 1b.* implies that the shocks being attributed to each lottery utility are from a distribution with median zero, which is necessary to ensure that the distribution of $\Gamma(\cdot, \cdot)$ is centered around the underlying utility of the lottery. Together these assumptions generate two main observations. First, if $U(\ell) = U(\ell')$ then the probability of choosing lottery ℓ from the menu $\{\ell, \ell'\}$ is 0.5. Second, if $U(\ell) > U(\ell')$ then the probability of choosing ℓ over ℓ' is greater than 0.5, meaning that the proportion of times an option is expected to be chosen corresponds directionally with the ranking of that option versus the alternative. Because this explanation suggests that the relative weight placed on a lottery provides information about the ranking of options, the values $\alpha^*(\ell, \ell')$ should respect the direction of preference ranking over the alternatives ℓ and ℓ'. Put concretely, if $U(\ell) > U(\ell')$ then $\alpha^*(\ell, \ell') > 0.5$.

In order to make predictions over *Part 2* questions we require some understanding of where the mixture utility sits with respect to the constituent lotteries. For example, if the specified mixture over the menu $\{\ell, \ell'\}$ places more weight on lottery ℓ, then this suggests that ℓ is preferred to ℓ'. However, there is currently no means of making predictions over the menus $\{m^*(\ell, \ell'), \ell\}$ and $\{m^*(\ell, \ell'), \ell'\}$. For that, one further assumption on U is required;

Assumption 2: $\min\{U(\ell), U(\ell')\} < U(m^*(\ell, \ell')) < \max\{U(\ell), U(\ell')\}$

Assumption 2 is referred to as the "mixture-betweenness" axiom by Camerer and Ho (1994), and is equivalent to quasi-convexity and quasi-concavity in preferences. Assuming that the mixture is strictly preferred to both alternatives, as in *explanation 1* results in a violation of this axiom, and therefore can be used to identify differences between choices in terms of *explanations 1* and *2*.[5]

Hypothesis 2: Stochastic Choice

$$(\alpha^*(\ell, \ell') - 0.5)(\alpha^*(\ell, m^*(\ell, \ell')) - 0.5) > 0$$

The lottery with the most weight given to it in *Part 1* is preferred to the alternative lottery. Given that the utility of the mixture must sit in between the utilities of the original lotteries, it follows that the utility of the mixture is less than the utility of the lottery with greater *Part 1* weight, and greater than the utility of the lottery with lesser *Part 1* weight. The proportion of weight (specifically, whether more or less than 50%) placed on the lotteries in *Part 2* should reflect the same ordering of utilities.

It is worth noting that this hypothesis is weaker than the one suggested in *Section 3.1.2*. Rather than the relative weights being identical to the choice proportions over lotteries, *Hypothesis 2* states that the relative weights should be the same side of 0.5 as the choice proportions, providing a notion of "directional consistency". Although not fully equivalent, this notion of preference direction implied by choice proportions is related to stochastic choice rules satisfying weak and moderate transitivity (He and Natenzon, 2024). A strengthening of this hypothesis requires further assumptions to be made on the distribution of noise. One possible strengthening allows us to condition on the magnitude

[5] The Random Expected Utility model of Gul and Pesendorfer (2006) implies that $\alpha^*(\ell, \ell') = \alpha^*(\ell, m^*(\ell, \ell'))$. Predictions differ as this model is not sensitive to the distance between utilities, only the ranking. For discussion on the equivalence between discrete choice and Random Utility, see Block and Marschak (1959)

of relative weights rather than just direction. This extension is discussed in *Section A1*.

4.3 Preference Uncertainty

Finally, suppose that the third explanation holds, and that the majority of mixture effects are an expression of individuals not having a preference over alternatives within the menu. Intuitively, if the individual cannot construct the preference over menu $\{\ell, \ell'\}$, then it seems likely that they will not be able to construct the preference over a menu containing two lotteries weakly "in between" ℓ and ℓ', for example over the menu $\{\ell, m^*(\ell, \ell')\}$. To state this prediction formally, suppose that an incomplete preference between ℓ and ℓ' is denoted $\ell \bowtie \ell'$, then the third hypothesis can be written as follows,

Hypothesis 3:

If $\ell \bowtie \ell'$, then $\ell \bowtie m(\ell, \ell')$ for any $m(\ell, \ell') \in Co(\{\ell, \ell'\}) \backslash \{\ell, \ell'\}$

4.4 Developing a Test

The first two explanations—convex preferences and stochastic choice, both make predictions over weight placed on the lottery in *Part 1* and *Part 2* questions. However, both explanations make predictions over correlation of observations and spatial location of observations. Additionally, the regions considered consistent with each of the explanations are not of the same area. To elaborate, convex preferences allow observations to place either full or almost full weight on the mixture in *Part 2*. This results in a region covering approximately 18% of the total available space. Conversely, stochastic choice consistent observations can sit within a region covering approximately 60% of the overall space. Therefore, most standard econometric tools are not equipped to test for significance of the two explanations in an unbiased manner.

Some spatial tests from the economics literature have been suggested for problems such as these. Selten (1991) for example develops a test of area theories that are normalized by the amount of space that would be covered given some benchmark prediction. Assuming completely random choices over the *Part 1* and *Part 2* lottery weight space provides a benchmark that is unbiased towards both explanations. This null is hereafter referred to as the *naive* null hypothesis.

Given that the marginal distribution of *Part 1* lottery weight is determined by the decision makers, it might not be reasonable to assume uniformity across *Part 1* weights. Fudenberg et al. (2023) develop a test to balance the restrictiveness of a model with its accuracy of predicting empirical behavior. In doing so, they develop a "completeness" metric that corrects for this difference in marginal distributions from uniformity. The second null, referred to as the *empirical* null hypothesis, is inspired by their metric.

These null hypotheses are generated by first simulating a dataset according to i.i.d draws of *Part 1* and *Part 2* lottery weights. The distribution is uniform over the support of *Part 2* weights for both the *naive* and *empirical* null hypotheses. As mentioned previously, the distribution of *Part 1* lottery weights is uniform for the *naive* null hypothesis, and taken from the empirical marginal distribution of *Part 1* weights for the *empirical* null hypothesis. Next, the proportion of observations that are consistent with convex preferences, stochastic choice, and residual are computed and then compared to the empirical proportion observed in our dataset. We repeat this data simulation process for a number of iterations, then, akin to a standard parametric bootstrapping technique, we use these simulated proportions to determine confidence intervals for each of our explanations against the null hypotheses.

Two minor adjustments have to be made when moving from the continuous space of theoretical predictions, as in *Figure 1*, to the more coarse space of observations provided by our experimental design. As mentioned previously, participants can set mixture weights accurate to the first decimal place: $\{0.0, 0.1, ..., 1.0\}$. Additionally, placing full weight on one of the lotteries in *Part 1* results in the mixture being identical to one of the two original lotteries, meaning that the *Part 2* question is either identical to the *Part 1* question or contains two identical lotteries. As these observations are not useful to us, we simulate over the truncated *Part 1* lottery weight space, $\{0.1, ..., 0.9\}$ while still simulating over the full *Part 2* lottery weight space, $\{0.0, 0.1, ..., 1.0\}$.

5 Experimental Design

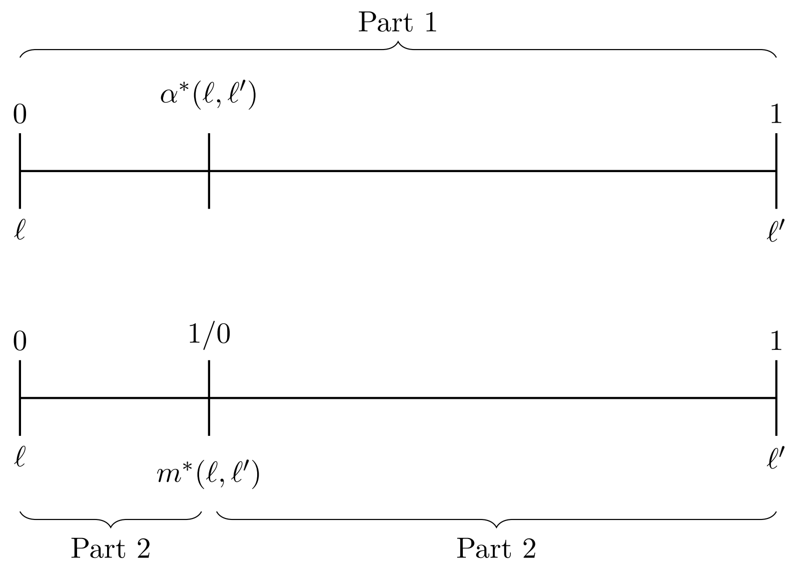
There are three treatments, all of which are split into two parts. Each part features 12 binary comparisons between simple lotteries. *Part 1* contains binary comparisons between lotteries taken from an initial set, $\{\ell, \ell'\}$, whereas *Part 2* comparisons include lotteries from the initial set and lotteries that are equivalent to mixtures provided in *Part 1*, $\{\ell, m^*(\ell, \ell')\}$.

For each comparison, decision makers have the opportunity to specify their preferred lottery choices, or move on to the next question without specifying (opt out). If choices are not provided, then a pre-specified, yet undisclosed lottery is used as default. Further details as to how this is determined are given in *Section 5.4*.

5.1 Part 1

There are a total of 24 initial lotteries from which the binary comparisons in *Part 1* are comprised. These lotteries are split into two groups, where each group contains lotteries that are approximate mean preserving spreads of each other. These groups have an approximate expected value of \$12 and \$14 respectively, while the support size of lotteries vary from 2 to 5.

Figure 2: Part 1 and Part 2 Questions



Menus are then constructed both within group and between group. Three comparisons are constructed containing lotteries from only group 1, and three are constructed containing lotteries from only group 2. A further six comparisons are constructed containing one lottery from each group. This makes the total of 12 binary comparisons in *Part 1*. All binary comparisons are shown in random order to each participant and no lottery appears in more than one comparison.

5.2 Part 2

Every comparison in *Part 2* contains one mixture from *Part 1*, and one of the initial lotteries from which the mixture was constructed. This means that for each of the 12 comparisons in *Part 1*, there are two possible comparisons to be shown in *Part 2*. *Figure 2* illustrates how *Part 2* questions are constructed using the mixture generated in *Part 1*. Each of the under braces over the lower horizontal line represent separate *Part 2* questions, where the left brace shows the *Part 2* question concerning the menu $\{\ell, m^*(\ell,\ell')\}$ and the right brace represents the question concerning $\{m^*(\ell,\ell'),\ell'\}$.

The binary comparisons for *Part 2* are then chosen as follows. The interface randomly selects up to four mixtures that were specified by the decision maker in *Part 1*, and asks both binary comparisons for each mixture in *Part 2*. This makes a total of up to eight questions. The remaining four questions are taken from two randomly chosen mixtures that were not specified by the decision maker. If there were less than four specified mixtures, or less than two unspecified mixtures, the interface randomly selects questions in order to get as close to that proportion as possible. These proportions are selected such that we have sufficient data to make comparisons between *Part 1* and *Part 2* both for questions where mixtures were specified, as well as for questions where mixtures were not specified.

5.3 Treatments

There are three main treatments. Each treatment is designed to capture a different setting in which we might consider mixing to be prevalent. The experiment takes a between-subject design, meaning that each participant only participates in a single treatment.

Treatment 1 (Slider) provides an illustration of the two lotteries in the menu at the top of the screen, and a third box in the middle titled "Your Preferred Lottery". Participants specify their preferred lottery using a slider that ranges from 0 to 10. As they move this slider, the box titled "Your Preferred Lottery" shows the mixture lottery associated with a convex combination of the two original lotteries. The relative weights used to construct the mixture correspond to the position of the slider, and the mixture lottery adjusts dynamically as the slider moves. The slider is used for both *Part 1* and *Part 2* questions in *Treatment 1*, and the *Slider* treatment interface is depicted in *Figure 3*.

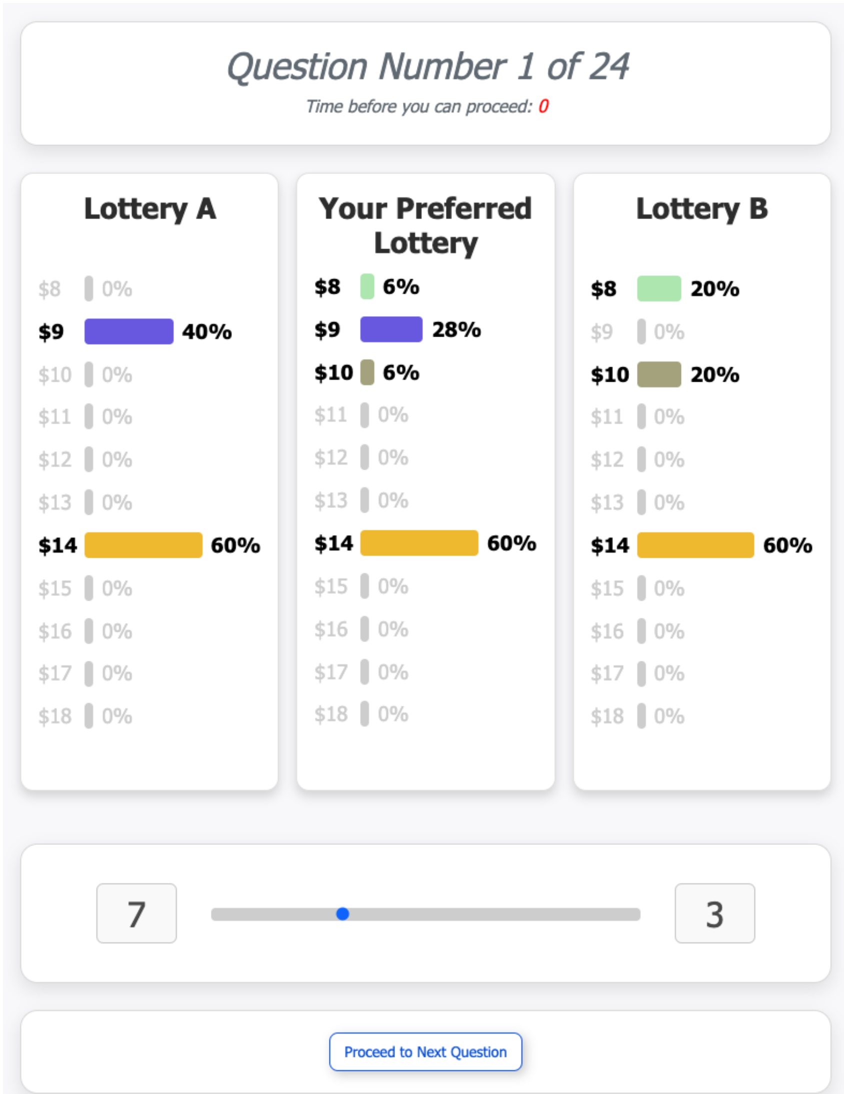
Treatment 2 (Repeated Choice) speaks more directly to the repeated choice representation of mixing. Instead of having a slider, participants are shown the two original lotteries and are asked to provide 10 answers. Each answer is a forced choice between "Lottery A" and "Lottery B". Participants are informed that, if they are eligible for bonus payment, the lottery that they answered in one of their 10 answers for one random question will be simulated. The mixtures in *Part 2* are constructed according to the proportion of the ten "Lottery A" answers versus "Lottery B" answers in *Part 1*. *Figure 4* illustrates the *Repeated Choice* treatment interface.

Finally, *Treatment 3 (Slider Info)* is identical to *Treatment 1*, except that participants are informed at the beginning of *Part 2* that the specified or non-specified preferred lotteries will be shown again in *Part 2* questions. The exact wording states, "...in every question, one of the lotteries (either Lottery A or Lottery B) will be a preferred lottery that either you specified in *Part 1* or was chosen for you." A concern while designing the interface was that individuals won't recognize their mixtures in *Part 2* questions. *Treatment 3* is designed to address this.

5.4 Incentives and Payments

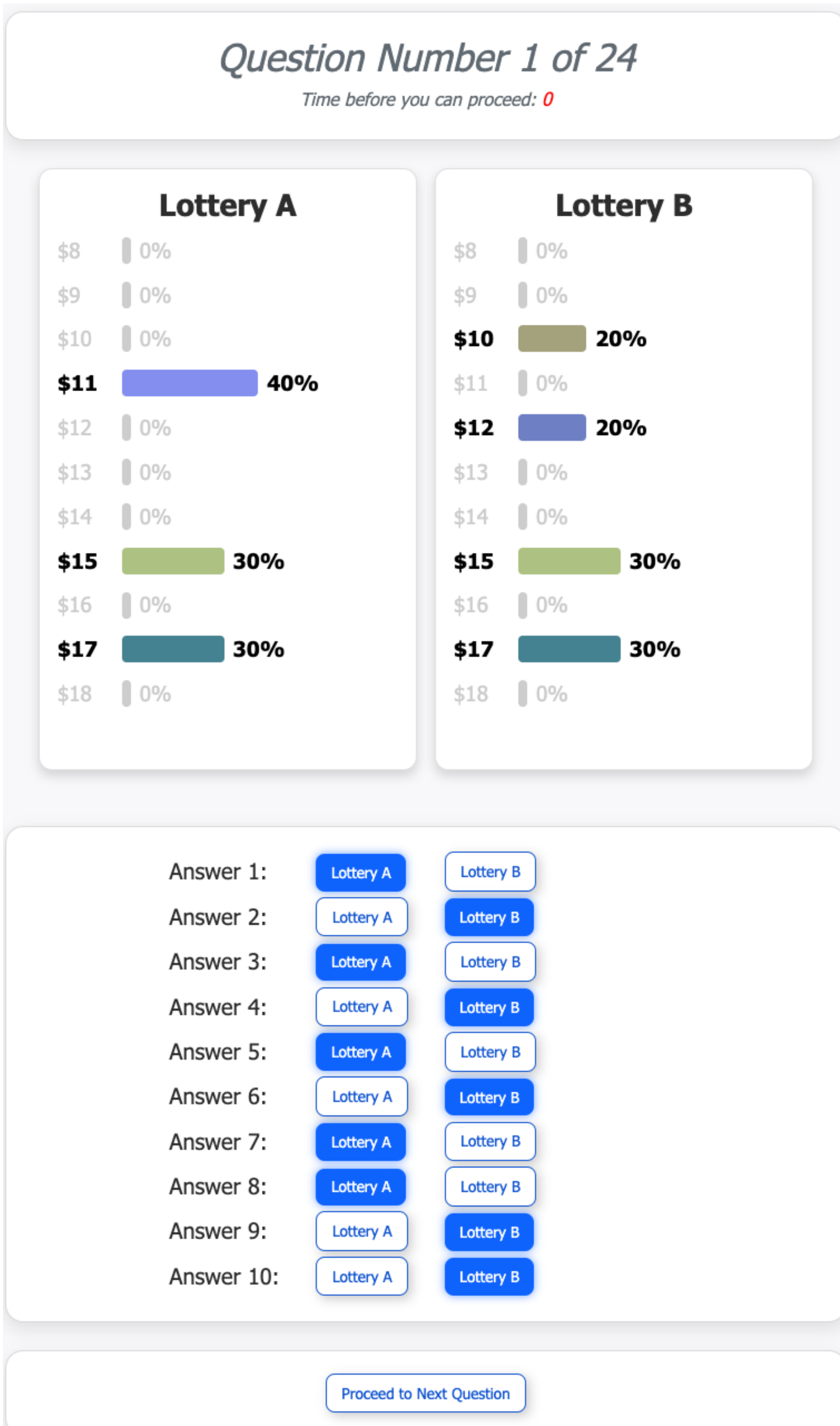
Participants are provided a participation payment of $5.50. One in five participants will also be selected for bonus payment. If they are selected, a random question will be selected as the bonus question.

Figure 3: Slider Treatment Decision Screen



Notes: Lottery A and Lottery B are the constituent lotteries over which the decision maker can generate a mixture The lottery labelled "Your Preferred Lottery" represents the mixture corresponding to the slider position. In this case the preferred lottery is mixture constituting of 70% *Lottery A and* 30% *Lottery B. The preferred lottery dynamically changes as the position of the slider changes.*

Figure 4: Repeated Choice Treatment Decision Screen



Notes: Lottery A and Lottery B are the constituent lotteries over which the decision maker can generate a mixture. Individuals choose between Lottery A and Lottery B a total of 10 times. The mixtures for Part 2 questions are then generated according to the proportion of times Lottery A was chosen over Lottery B.

In this question, the participant may or may not have chosen to specify their preferred lottery. If they did specify, then for *Treatments 1* and *3*, the reduced lottery associated with that mixture will be simulated by the computer and a payoff is provided according to the outcome. In *Treatment 2*, one of the answers will be drawn at random and the preferred lottery for that answer will be simulated. The bonus payment will then be the simulated outcome of that lottery.

If the participant did not choose their preferred lottery, then in *Treatment 1* and *3*, the computer resorts to a pre-specified mixture over the two lotteries within the menu. This mixture is generated uniformly at random across the convex combination of the two lotteries. The bonus payment is then equal to the simulated outcome of that lottery. In *Treatment 2*, a number between 0 and 10 is drawn at random to denote the number of Lottery A choices (10 minus that number denotes the number of Lottery B choices). These are then shuffled, and the lottery corresponding to the previously designated bonus answer is simulated. This methodology ensures that the payment mechanism when the lotteries are not specified is equivalent across treatments.[6][7]

Participants also answered two comprehension questions at the end of the study, both of which could have been selected as the bonus question. If this was the case then they receive a fixed bonus of $5 if answered correctly, and $0 otherwise.

6 Results

6.1 Data Quality and Filtering

Data from a total of 300 participants per treatment were collected in July and August 2025. Treatment interfaces where coded in *oTree* (Chen et al., 2016) and submissions were collected via the online platform Prolific. The median time taken was 22m 38s and the average payment was $8.27, which includes the $5.50 participation fee.

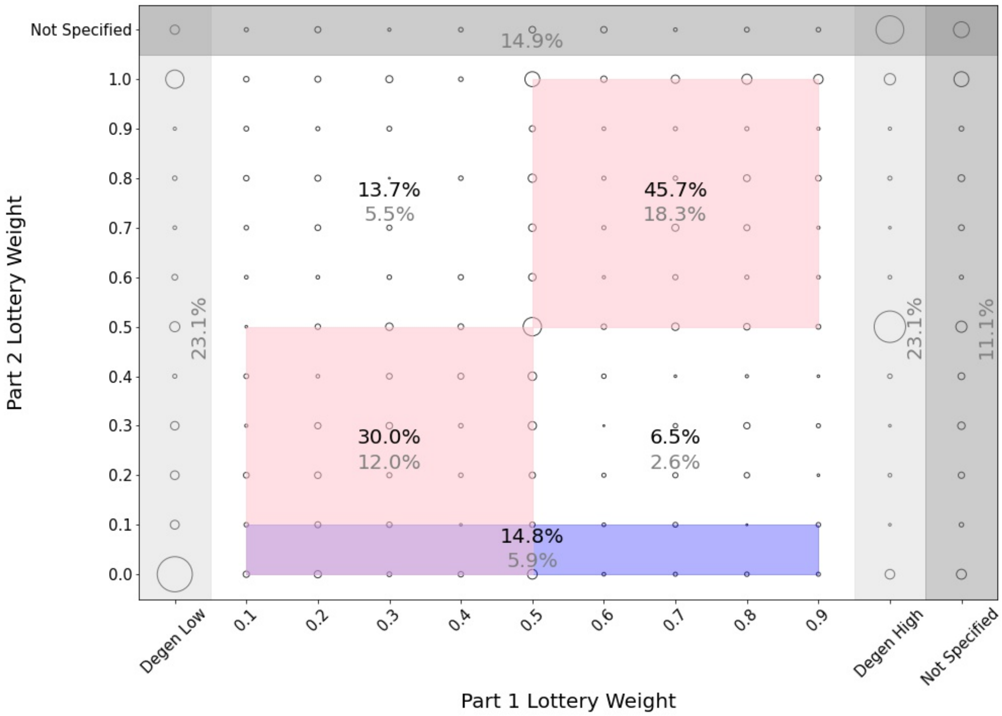
A total of 224 participants have been dropped from the original dataset due to failing comprehension questions or not specifying a sufficient number of mixtures. Specifically,

[6]Participants also have to wait 10 seconds before they can proceed to the next question. This helps prevent a preference for non-specification due to speed of completing the study.

[7]Incentivizing *incompleteness* is a challenging problem, and our methodology of allowing participants to skip answering questions acts as a well incentivized proxy for this. It is worth re-iterating that our proxy is likely a lower bound of preference uncertainty—firstly because participants tend to prefer answering a question rather than not, and secondly because an ambiguity averse decision maker would have a weak preference for constructing a mixture as opposed to the randomly generated substitute lottery. On the other hand, we do not attempt to further categorize the residual observations, many of which may have been chosen for no other reason than preference uncertainty. Other papers with novel methods of incentivizing uncertainty or incompleteness include Danan and Ziegelmeyer (2006), Nielsen and Rigotti (2022), Halevy et al. (2023), and Feldman and Zhou (2024).

206 participants failed at least one of the two comprehension questions, and 24 chose not to specify their answers for every question in either *Part 1* or *Part 2* of the study.

Figure 5: Part 1 and Part 2 Mixture Weights: Slider Treatment



Notes: Larger points represent a greater number of observations. Light pink regions in the upper right and lower left quadrant represent regions that are directionally consistent. The blue region given by $y \leq 0.1$ *represents the region that is consistent with convex preferences. Not Specified refers to observations where weights were not specified. Degen Low (resp. Degen High) refers to cases where the Part 1 lottery weight was* 0 *(resp.* 1*). Values in black represent the number of explanation consistent regions as a proportion of fully specified and Part 1 non-degenerate observations. Light grey values represent the same number but as a proportion of all paired observations. Proportions sum to more than* 1 *as some observations can be consistent with both convex preferences and stochastic choice.*

6.2 Part 1

This section discusses the main findings from *Part 1* answers in the *Slider* treatment. Due to the experimental design selecting only a subset of *Part 1* answers to be paired with *Part 2* questions, we begin by looking at all *Part 1* answers, regardless of whether they were later paired with a *Part 2* question.

6.2.1 Specification rates are high

As mentioned in *Section 5*, participants were given the opportunity to opt out of constructing their mixtures in exchange for a pre-specified and undisclosed lottery. Overall, only 8.7% of *Part 1* answers are unspecified, meaning that decision makers have a strong preference for constructing their own preferred lottery given the opportunity. 56.2% of participants chose to specify their answers for all *Part 1* questions, while approximately 1.3% of participants chose to specify less than half of their answers. The willingness to specify appears to be independent of whether the lotteries had the same or different expected values—91.4% of questions with different expected values for each lottery and 90.1% for those where the lotteries had the same expected value (z-stat: 0.571, p-value: 0.568).

6.2.2 Approximately half of specified answers are degenerate

50.5% of specified *Part 1* answers place full weight on one of the lotteries, meaning that approximately half of observations involved decision makers constructing a non-degenerate mixture. The modal non-degenerate weight placed on lotteries is an equal split of weight across the two lotteries. This equal split accounts for 14.0% of all *Part 1* specified answers.

Overall, participants tend to specify the majority of their *Part 1* answers, and approximately half of those specified answers are non-degenerate. There is no strong relationship between characteristics of the lotteries within a question and the probability of specification. Additionally, there is no evidence that participants are less likely to specify their mixtures as they progress through *Part 1* questions (see *Figures A2 and A3* for more detail). The proportion of degenerate mixing conditional on specification is significantly smaller than would be predicted by EU, as this claims that the proportion of non-degenerate mixing should be vanishingly small and only occur under complete indifference between the two lotteries. As such, these levels of mixing demonstrate clear violations of EU, and suggest that giving decision makers the opportunity to select their "preferred lottery" from the convex hull of alternatives results in them doing so frequently.

6.3 Part 2

The following sections focus on paired (*Part 1*, *Part 2*) observations. *Table 1* provides a breakdown of *Part 1* specification, *Part 1* degenerate, and *Part 2* specification proportions, as well as the proportion of observations associated with each of the explanations. These values differ slightly from the results provided in the previous section due to the experimental interface selectively sampling between *Part 1* specified and *Part 1* unspecified answers. "Fully specified and *Part 1* non-degenerate" hereafter refers to observations in the last

Table 1: Proportions and Null predictions for Slider Treatment

Part 1 Specified	Part 1 Degenerate	Part 2 Specified	Explanation Type	Proportion Total	Proportion Subset	Null Naive	Null Empirical
False				0.111	–	–	–
True	True			0.462	–	–	–
True	False	False		0.027	–	–	–
True	False	True	Convex	0.059	0.148	0.182***	0.182***
True	False	True	Stochastic Choice	0.303	0.757	0.596***	0.685***
True	False	True	Residual	0.081	0.202	0.323***	0.253***

Notes: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. *Part 1 Specified refers to the proportion of Part 1 observations that are specified. Part 1 Degenerate refers to the proportion of Part 1 observations that are degenerate mixtures. Part 2 Specified refers to the proportion of observations that are specified in Part 2. Type refers to the different explanations. Proportion Total refers to the proportion of overall paired observations. Proportion Subset refers to the proportion of fully specified and Part 1 non-degenerate observations. Null Naive and Null Empirical refer to the proportion of observations expected to be consistent under the naive null hypothesis and the empirical null hypothesis respectively. Stars represent the significance of difference between Proportion Subset and each of the Null Predictions. Proportions may sum to more than* 1 *as observations can be consistent with both convex preferences and stochastic choice.*

three rows of *Table 1*.

6.3.1 Consistency with convex preferences is low

The overall proportion of observations placing full or 90% weight on the mixture in *Part 2* is 5.9%. This is equal to 14.8% when we consider this value as a proportion of fully specified and *Part 1* non-degenerate observations. Given that consistency with convex preferences requires this number to be equal to 100%, there is very little evidence of participants having a strict preference for the mixture in *Part 2*. *Figure 5* illustrates the empirical proportions of observations consistent with the different explanations. Values in black represent the number of observations as a proportion of all paired observations. Values in grey represent the number of observations as a proportion of fully specified and *Part 1* non-degenerate observations.

Of those observations that are convex consistent, 48.2% occur when the lottery weight in *Part 1* is strictly less than 0.5. 24.1% occur when the weight on the lottery in *Part 1* is equal to 0.5. This suggests that the percentage of convex consistent observations is driven largely by cases where the lottery was not the majority constituent of the generated mixture in *Part 1*.

Although the proportion of convex preference consistent observations is clearly smaller than what is predicted in *Section 3.1.1*, it is also smaller than what is predicted under both of the null hypotheses. *Table 1* states the predictions under the *naive* null hypothesis where observations are distributed uniformly at random across the (*Part 1*, *Part 2*) non-degenerate lottery weight space. This space corresponds to the regions not shaded in grey of *Figure 5*. Under this null, an average of 18.2% of observations should be consistent with convex preferences. The parametric bootstrapping technique described in *Section 4.4* implies that the actual proportion of convex consistent observations is significantly smaller than the naive null hypothesis at the 1% level. Not only is there very little evidence in support of convex preferences, but the empirical proportion is also significantly lower than what would be observed if decision makers were selecting mixture weights at random. This provides strong support against convex preferences being the main underlying rationalization for mixture effects.

6.3.2 The proportion of stochastic choice consistent observations is high

The set of observations consistent with stochastic choice is depicted by the pink regions in *Figure 5*. 30.3% of all paired observations fall within this region, which equates to 75.7% of fully specified and *Part 1* non-degenerate observations.

Unlike the convex consistent observations, stochastic choice consistent observations are relatively balanced across *Part 1* lottery weights—28.0% occur when *Part 1* lottery weight is strictly less than 0.5 and 31.6% occur when the weight is strictly more than 0.5. The remaining proportion occurs when *Part 1* lottery weight is distributed equally across both lotteries.

This proportion is significantly larger than the *naive* null prediction at the 1% level. The average proportion of uniformly distributed points falling within the consistent region is 59.6%, which is 16.1 percentage points less than the sample proportion. *Table 1* also shows the prediction under the *empirical* null hypothesis. The *empirical* null hypothesis is designed to account for the lack of uniformity across *Part 1* lottery weights. This is particularly important for the stochastic choice explanation as it corrects for the mass of observations distributing *Part 1* weights equally across constituent lotteries. Despite this adjustment, the sample proportion is again significantly larger than the empirical null hypothesis at the 1% level.

6.3.3 The proportion of residual observations is low

Residual observations are fully specified and *Part 1* non-degenerate observations that are consistent with neither convex preferences nor stochastic choice. Residual observations account for 20.2% of fully specified and *Part 1* non-degenerate observations, which is significantly less than both the *naive* and *empirical* null hypothesis predictions at the 1% level. The proportion of residual observations with *Part 1* lottery weight less than 0.5 is the main contributor to total residual observations. This is natural as the space consistent with neither of the other explanations is smaller when *Part 1* lottery weight is greater than 0.5.

6.3.4 Non-specification in *Part 1* does not imply non-specification in *Part 2*

To re-iterate, 91.3% of all (unpaired) *Part 1* mixtures are specified on aggregate. Of the *Part 2* questions that did not involve a degenerate mixture, 90.6% were specified. This difference is not significant (z-stat: 0.700, p-value: 0.484) and suggests that there is little variation in the aggregate differences in specification proportions across parts.

However, *Hypothesis 3* makes a specific prediction over *Part 1*, *Part 2* pairs, suggesting that non-specification for the *Part 1* question implies non-specification for the related *Part 2* question. The intuition is that if a decision maker is unable to decide which lottery they prefer between ℓ and ℓ', then they should be unable to decide between ℓ and $m^*(\ell, \ell')$. When looking at paired observations, the probability of specification in *Part 2* when the *Part 1* mixture was specified is 93.6%. The probability of specification in *Part 2* when the *Part 1* mixture is not specified is 76.0% (z-stat: 8.632, p-value < 0.001). Although the probability of specification is significantly lower when the *Part 1* mixture has not been specified, the hypothesis states that specification rates should be close to zero, and this is clearly not the case. Consequently, there is not sufficient evidence to suggest that preference uncertainty is the main contributor to mixture effects.

In summary, stochastic choice consistent observations dominate the space over (*Part 1*, *Part 2*) lottery weights. Both convex consistent and residual observations make up a significantly lower proportion of observations than would be predicted if decision makers were behaving randomly, yet no explanation captures all observations. Degenerate mixing occurs often but with a much lower frequency than EU predicts. There is also little evidence of preference uncertainty.

6.4 Differences Across Treatments

Although the findings above are with respect to the *Slider* treatment, they tend to be robust across all three of the decision making interfaces. *Table 6.4* specifies the proportions of observations falling into each explanation category for all treatments. The *Slider Info* treatment is effective in increasing the number of observations that are consistent with convex preferences to the extent that it is significantly larger than both null hypotheses. The reverse is true for the two other treatments. The proportion of residual observations remain relatively constant across treatments, while the proportion of stochastic choice consistent observations is slightly higher in the *Repeated Choice* treatment at 78.4% compared to either of the *Slider* treatments.

Table 2: Proportion of Observations Consistent with Explanation by Treatment

	Slider Info			Repeated Choice		
	Proportion	Naive	Empirical	Proportion	Naive	Empirical
Convex	0.237	0.182^{***}	0.182^{***}	0.114	0.182^{***}	0.182^{***}
SC	0.757	0.596^{***}	0.679^{***}	0.784	0.596^{***}	0.681^{***}
Residual	0.184	0.323^{***}	0.257^{***}	0.19	0.323^{***}	0.254^{***}

Notes: $^{*}p < 0.1$*;* $^{**}p < 0.05$*;* $^{***}p < 0.01$. *Proportion refers to the proportion of fully specified and Part 1 non-degenerate answers that are consistent with each of the explanations. Naive and Empirical refer to the predicted proportions under the naive and empirical nulls. Refer to Section 4.4 for further details. Convex refers to convex preferences, SC refers to stochastic choice, and Residual refers to residual observations. Equivalent values for the Slider treatment are given in Table 1.*

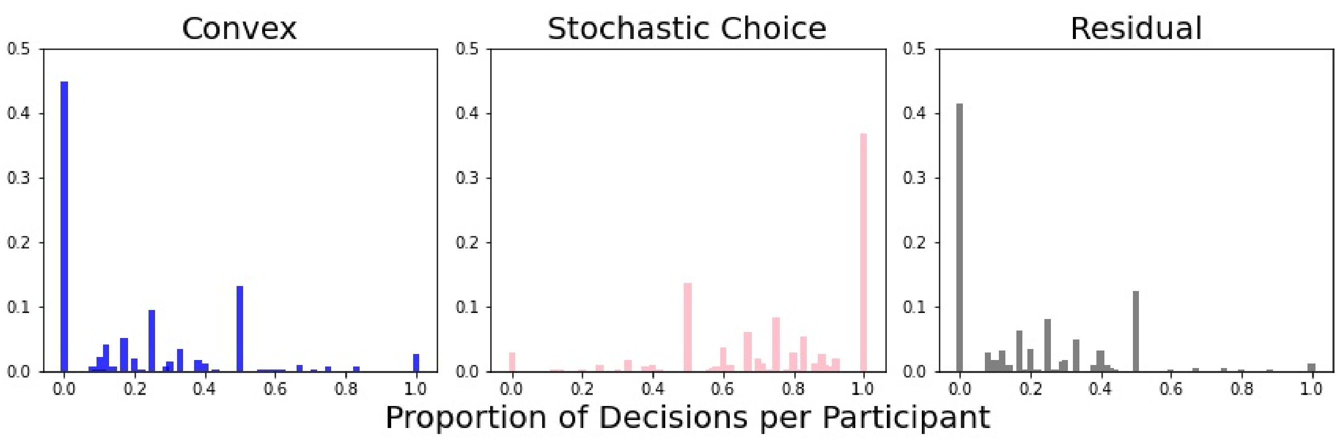
The other noticeable difference between treatments occurs in the specified and *Part 1* degenerate answers. 13.4% of observations in the *Repeated Choice* treatment have *Part 1* weights not specified, which is approximately 5 percentage points more than the other two treatments. Additionally, 65.6% of *Part 1* specified answers are degenerate in the *Repeated Choice* treatment, as opposed to approximately 50.5% in the two *Slider* treatments. This suggests that making repeated decisions across a fixed menu results in a slightly higher rate of non-specification and a slightly higher rate of degenerate mixing as opposed to constructing convex mixtures directly.

Given that the proportions of explanation consistent observations are relatively similar across treatments, the following two sections looks at observations pooled across all treatments.

6.5 Individual Level Heterogeneity

Convex and stochastic choice consistent observations are both present within the data, yet it is not clear whether either of these explanations are stable decision maker traits. *Figure 6* shows the proportion of observations per individual that are consistent with each of the explanations. The vast majority of individuals have weakly less than 50% of observations consistent with convex preferences (93.1%). Similar to convex preferences, the proportion of residual observations is relatively low within participant.

Figure 6: Individual Level Heterogeneity for All Treatments



Notes: The x-axis shows the proportion of fully specified and Part 1 non-degenerate answers, within participant, that are consistent with each of the explanations. The y-axis shows the proportion of participants.

On the other hand, stochastic choice consistent decisions are frequent within participant as well as across—77.4% of participants have strictly more than 50% of observations consistent with stochastic choice. The modal proportion of choices consistent with stochastic choice per participant is 1, whereas it is 0 for both convex consistent observations and residuals.

6.6 Determinants of Mixture Weights

Table 3 shows OLS regressions for lottery weight on lottery and lottery pair characteristics. Regression (1) shows that the weight on lottery B is significantly increasing in the expected value of lottery B and significantly decreasing in the expected value of lottery A, demonstrating that changes in weights are sensitive to the difference in expected value. Second, weight on lottery B is decreasing in the variance of lottery B, significant at the 10% level. Although neither variance of lottery A or difference in variance is significant, it demonstrates that individuals are at very least taking the variance of lotteries into consideration when selecting mixture weights.

Table 3: OLS Regression of Lottery Weight on Menu Characteristics

	Weight On Lottery B					
	(1)		(2)		(3)	
Intercept	0.023	(0.064)	0.021	(0.063)	0.017	(0.01)
LA EV	-0.122^{***}	(0.005)	-0.124^{***}	(0.005)	–	
LB EV	0.123^{***}	(0.005)	0.125^{***}	(0.005)	–	
LA Var	0.002	(0.001)	-0.0	(0.001)	–	
LB Var	-0.004^{***}	(0.001)	-0.001	(0.001)	–	
Part	0.003	(0.007)	0.006	(0.009)	0.005	(0.007)
LA Support	–		0.03^{***}	(0.004)	–	
LB Support	–		-0.031^{***}	(0.004)	–	
EV Diff.	–		–		0.124^{***}	(0.004)
Var Diff.	–		–		-0.0	(0.001)
Support Diff.	–		–		-0.03^{***}	(0.003)
N Obs.	10748		10748		10748	
R Squared	0.162		0.18		0.18	

Notes: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. *Li EV, Li Var, and Li Support refers to the Expected Value, Variance, and Support for lotteries* $i \in \{A, B\}$ *respectively. Variables with Diff. is the lottery characteristic for lottery B minus the lottery characteristic for lottery A. Standard errors are show in parentheses and clustered at the individual level.*

Whether the question is a *Part 1* or *Part 2* question is not a significant determinant of lottery weight when controlling for other lottery characteristics. Finally, LA Support and

LB Support in regression (2), and support difference in regression (3) are all significant at the 1% level. The larger the support of lottery B, the less weight is placed on lottery B. Given that the support of the (non-degenerate) mixture must be at least as large as the support of both constituent lotteries, this aversion to large supports may be capturing some aversion to mixtures in *Part 2* questions.[8]

7 Discussion

This study attempts to identify the motivating mechanisms underlying mixture effects. In order to do so, I categorize rationalizing models from the theoretical literature into three normative explanations, and design a two-part experiment to disentangle them. I find that less than 15% of paired observations are consistent with convex preferences. This value is significantly less than what would be expected from an individual constructing mixtures uniformly at random, implying not only a lack of evidence in favor of convex preferences, but significant evidence against. On the other hand, approximately 75% of observations are consistent with our notion of stochastic choice, and, although no single explanation is congruent with all the data, stochastic choice appears to be the predominant driver of mixture effects. There is also little evidence of residual observations and mixtures due to preference uncertainty.

In terms of contribution to the theoretical literature, this study suggests that models allowing for stochasticity over preferences tend to outperform more modern models allowing for preference uncertainty or convexity of preferences. The additional complication of allowing for non-standard preferences may not be necessary, and instead we should focus on structuring our models around standard, well-behaved preferences, but allowing for noise to produce a variation over decisions when implemented in traditional choice frameworks.

Practically speaking, these findings have important implications for policy design and behavioral welfare more generally. Analysts are frequently tasked with making inference about preferences from "mixed" datasets. Our findings suggest that mixture effects are not due to convex preferences, and mixing should not be considered as a deliberate act designed to construct a larger, maximally preferred alternative. Consequently, responding by offering the decision maker a mixture lottery aligning with the mixture in the dataset is not likely to be a satisfactory welfare maximizing solution. In fact, the results suggest that the information about preferences is mainly contained within the relative weights assigned to constituent lotteries within the mixture. This suggests that a policy designer

[8]This finding is closely related to a finding in Puri (2018), where an agent "assesses a lottery less favorably if it contains more outcomes".

wishing to maximize decision maker welfare might be better off providing the alternative that is chosen most frequently, or, equivalently, the alternative that has the largest weight assigned to it within the mixture.

References

M. Agranov and P. Ortoleva. Stochastic Choice and Preferences for Randomization. *Journal of Political Economy*, 125(1):40–68, Feb. 2017. ISSN 0022-3808. doi: 10.1086/689774. URL https://www.journals.uchicago.edu/doi/full/10.1086/689774. Publisher: The University of Chicago Press.

M. Agranov and P. Ortoleva. Ranges of Preferences and Randomization. *Report, Princeton Univ.*, Mar. 2020. URL https://agranov.caltech.edu/documents/13172/Ranges_5.pdf.

M. Agranov and P. Ortoleva. Revealed Preferences for Randomization: An Overview. *AEA Papers and Proceedings*, 112:426–430, 2022. ISSN 2574-0768. URL https://www.jstor.org/stable/27174432. Publisher: American Economic Association.

M. Agranov, P. J. Healy, and K. Nielsen. Stable Randomisation. *The Economic Journal*, 133(655):2553–2579, Oct. 2023. ISSN 0013-0133. doi: 10.1093/ej/uead039. URL https://doi.org/10.1093/ej/uead039.

M. Allais. Le Comportement De L'Homme Rationnel Devant Le Risque: Critique Des Postulats et Axiomees de L'Ecole Americaine. *Econometrica*, 21(4), Oct. 1953. URL https://doi.org/10.2307/1907921.

H. D. Block and J. Marschak. Random Orderings and Stochastic Theories of Response. *Cowles Foundation Discussion Papers*, page 289, Mar. 1959. URL https://elischolar.library.yale.edu/cowles-discussion-paper-series/289?utm_source=elischolar.library.yale.edu%2Fcowles-discussion-paper-series%2F289&utm_medium=PDF&utm_campaign=PDFCoverPages.

D. Butler and G. Loomes. Imprecision as an account of violations of independence and betweenness. *Journal of Economic Behavior & Organization*, 80(3):511–522, Dec. 2011. ISSN 01672681. doi: 10.1016/j.jebo.2011.05.008. URL https://linkinghub.elsevier.com/retrieve/pii/S0167268111001284.

D. J. Butler and G. C. Loomes. Imprecision as an Account of the Preference Reversal Phenomenon. *The American Economic Review*, 97(1):277–297, 2007. ISSN 0002-8282. URL https://www.jstor.org/stable/30034395. Publisher: American Economic Association.

C. F. Camerer and T.-H. Ho. Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, 8(2):167–196, Mar. 1994. ISSN 0895-5646, 1573-0476. doi: 10.1007/BF01065371. URL http://link.springer.com/10.1007/BF01065371.

S. Cerreia-Vioglio, D. Dillenberger, and P. Ortoleva. Cautious Expected Utility and the Certainty Effect. *Econometrica*, 83(2):693–728, 2015. ISSN 0012-9682. doi: 10.3982/ECTA11733. URL https://www.econometricsociety.org/doi/10.3982/ECTA11733.

S. Cerreia-Vioglio, D. Dillenberger, P. Ortoleva, and G. Riella. Deliberately Stochastic. *The American Economic Review*, 109(7):2425–2445, 2019. ISSN 0002-8282. URL https://www.jstor.org/stable/26739277. Publisher: American Economic Association.

E. Cettolin and A. Riedl. Revealed preferences under uncertainty: Incomplete preferences and preferences for randomization. *Journal of Economic Theory*, 181:547–585, May 2019. ISSN 00220531. doi: 10.1016/j.jet.2019.03.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0022053119300262.

D. L. Chen, M. Schonger, and C. Wickens. oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9: 88–97, 2016. ISSN 2214-6350. doi: https://doi.org/10.1016/j.jbef.2015.12.001. URL https://www.sciencedirect.com/science/article/pii/S2214635016000101.

E. Danan. Randomization vs. Selection: How to Choose in the Absence of Preference? *Management Science*, 56(3):503–518, Mar. 2010. ISSN 0025-1909, 1526-5501. doi: 10.1287/mnsc.1090.1116. URL https://pubsonline.informs.org/doi/10.1287/mnsc.1090.1116.

E. Danan and A. Ziegelmeyer. Are preferences incomplete? An experimental study using flexible choices. *Max Planck Inst. of Economics, Strategic Interaction Group*, 2006. URL https://www.ericdanan.net/research/preferences.pdf.

Evren. On the existence of expected multi-utility representations. *Economic Theory*, 35 (3):575–592, June 2008. ISSN 0938-2259, 1432-0479. doi: 10.1007/s00199-007-0252-5. URL http://link.springer.com/10.1007/s00199-007-0252-5.

Evren and E. A. Ok. On the multi-utility representation of preference relations. *Journal of Mathematical Economics*, 47(4-5):554–563, Aug. 2011. ISSN 03044068. doi: 10.1016/j.jmateco.2011.07.003. URL https://linkinghub.elsevier.com/retrieve/pii/S0304406811000693.

P. Feldman and J. Rehbeck. Revealing a preference for mixtures: An experimental study of risk. *Quantitative Economics*, 13(2):761–786, 2022. ISSN 1759-7323. doi: 10.3982/QE1694. URL http://qeconomics.org/ojs/index.php/qe/article/view/QE1694.

P. Feldman and N. Zhou. Experimental Tests of Incomplete Preferences, 2024. URL https://www.ssrn.com/abstract=4920948.

D. Fudenberg, W. Gao, and A. Liang. How Flexible is that Functional Form? Quantifying the Restrictiveness of Theories, Aug. 2023. URL http://arxiv.org/abs/2007.09213. arXiv:2007.09213 [econ].

F. Gul and W. Pesendorfer. Random Expected Utility. *Econometrica*, 74(1):121–146, Jan. 2006. ISSN 0012-9682, 1468-0262. doi: 10.1111/j.1468-0262.2006.00651.x. URL http://doi.wiley.com/10.1111/j.1468-0262.2006.00651.x.

Y. Halevy, D. Walker-Jones, and L. Zrill. Difficult Decisions. July 2023. URL https://yoram-halevy.faculty.economics.utoronto.ca/wp-content/uploads/DD.pdf.

J. He and P. Natenzon. Moderate Utility. *American Economic Review: Insights*, 6(2): 176–195, June 2024. ISSN 2640-205X, 2640-2068. doi: 10.1257/aeri.20230085. URL https://pubs.aeaweb.org/doi/10.1257/aeri.20230085.

P. Manzini and M. Mariotti. Stochastic Choice and Consideration Sets. *Econometrica*, 82(3):1153–1176, 2014. ISSN 0012-9682. doi: 10.3982/ECTA10575. URL http://doi.wiley.com/10.3982/ECTA10575.

D. McFadden. Conditional Logit of Qualitative Choice Behavior. *WP*, Nov. 1972.

F. Mosteller and P. Nogee. An Experimental Measurement of Utility. *Journal of Political Economy*, 59(5):371–404, Oct. 1951. ISSN 0022-3808. doi: 10.1086/257106. URL https://www.journals.uchicago.edu/doi/abs/10.1086/257106. Publisher: The University of Chicago Press.

K. Nielsen and L. Rigotti. Revealed Incomplete Preferences, Oct. 2022. URL http://arxiv.org/abs/2205.08584. arXiv:2205.08584 [econ].

I. Puri. Preference for Simplicity. *SSRN Electronic Journal*, 2018. ISSN 1556-5068. doi: 10.2139/ssrn.3253494. URL https://www.ssrn.com/abstract=3253494.

R. Ratcliff. A Theory of Memory Retrieval. *Psychological Review*, 85(2), Mar. 1978. URL https://bpb-us-w2.wpmucdn.com/u.osu.edu/dist/6/60429/files/2018/07/psychrev78-1ohb1a5.pdf.

J. Rehbeck and S. Stelnicki. An Experiment on the Integration of Monetary and Subjective Value. *SSRN Electronic Journal*, 2024. ISSN 1556-5068. doi: 10.2139/ssrn.4702878. URL https://www.ssrn.com/abstract=4702878.

C. Seidl. Preference Reversal. *Journal of Economic Surveys*, 16(5):621–655, 2002. ISSN 1467-6419. doi: 10.1111/1467-6419.00184. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-6419.00184. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-6419.00184.

R. Selten. Properties of a measure of predictive success. *Mathematical Social Sciences*, 21(2):153–167, Apr. 1991. ISSN 01654896. doi: 10.1016/0165-4896(91)90076-4. URL https://linkinghub.elsevier.com/retrieve/pii/0165489691900764.

Appendix

A1 Size Consistency

It could be argued that our notion of stochastic choice consistency is a weak one. This may be true, but further assumptions are required for strengthening.

For example, allowing ϵ to be drawn from an identical distribution across questions results in a condition on the relative size of weights across parts.

Assumption 3: $(\epsilon, \epsilon') \sim iid\ F_{\epsilon,\epsilon'}(\cdot,\cdot)$

This allows for the distribution of noise to be fixed across all *Part 1* and *Part 2* questions. As a result, the fact that the utility of the mixture sits between the utility of the most preferred and least preferred lotteries results in the effect of noise being larger in *Part 2*. Consequently, the stochastic choice model predicts that weights placed on lotteries in *Part 2* questions will be less extreme than the weights placed on the same lotteries in *Part 1*. This formalizes our next hypothesis;

Hypothesis 2': Stochastic Choice and Size Consistency

1. $0.5 < \alpha^*(\ell, m^*(\ell,\ell')) < \alpha^*(\ell,\ell')$ if $\alpha^*(\ell,\ell') > 0.5$
2. $\alpha^*(\ell,\ell') < \alpha^*(\ell, m^*(\ell,\ell')) < 0.5$ if $\alpha^*(\ell,\ell') < 0.5$

Intuitively, allowing noise to be drawn from the same distribution means that the proportion of weight placed on the most preferred lottery must increase with the difference in utilities. Because the mixture utility sits between the utilities of the two original lotteries, the relative utility difference between the two original lotteries is larger than the difference between one of the lotteries and the mixture lottery. This implies that noise has a greater effect in *Part 2* questions than in *Part 1* questions, resulting in weights being less extreme in *Part 2*. I refer to this as *Size Consistency*, and *Hypothesis 2'* reflects the combination of stochastic choice and *Size Consistency*.

There is empirical support for size consistency, although it is not as definitive compared to stochastic choice. *Table A4* shows the proportion of size consistent observations both as a proportion of all fully specified and *Part 1* non-degenerate observations, and as a proportion of stochastic choice consistent observations. Size consistent and stochastic choice consistent observations as a proportion of fully specified and *Part 1* non-degenerate observations are significantly larger than the *empirical* null across all treatments. Size consistent observations as a proportion of stochastic choice consistent observations are also significantly larger than the *empirical* null hypotheses across two of the three treatments.

Slider Info has this value as significantly smaller at the 10% level. *Naive* null hypothesis values are larger than *empirical* null values due to the fact that size consistency is most restrictive when *Part 1* lottery weight is equal to 0.5. The *empirical* null therefore has a smaller value than the *naive* null given the mass of observations at 0.5 in the sample data.

A2 Proofs

A2.1 Hypothesis 2 (Stochastic Choice)

$a > b \implies P(\Gamma(a, \epsilon) > \Gamma(b, \epsilon')) > 0.5$

Define the following two sets:

$$A := \{(\epsilon, \epsilon') : \epsilon > \epsilon'\}$$

and

$$B := \{(\epsilon, \epsilon') : \epsilon < \epsilon' \ \& \ \Gamma(a, \epsilon) > \Gamma(b, \epsilon')\}$$

The event $\Gamma(a, \epsilon) > \Gamma(b, \epsilon')$ is equivalent to either events A or B occurring. By definition they are disjoint, and so $P(A \cup B) = P(A) + P(B)$. Therefore, we must prove that $P(A) + P(B) \geq 0.5$.

$P(A) = 0.5$ follows from symmetry of the joint distribution $F_{\epsilon,\epsilon'}$. Define $\Delta(\epsilon, \epsilon') := \Gamma(a, \epsilon) - \Gamma(b, \epsilon')$ for some (ϵ, ϵ') on the interior of the support of $F_{\epsilon,\epsilon'}$. By strict monotonicity of Γ, $\Delta(\epsilon, \epsilon) > 0$. By continuity of $F_{\epsilon,\epsilon'}$, it must be the case that there exists some (ϵ, ϵ') in the support of $F_{\epsilon,\epsilon'}$ where ϵ' is larger than ϵ and such that $\Delta(\epsilon, \epsilon') > 0$. This proves that the set B is non-empty and contains (ϵ, ϵ') pairs that occur with strictly positive probability, meaning that $P(B) > 0$. $P(A \cup B) > 0.5$ implies $P(\Gamma(a, \epsilon) > \Gamma(b, \epsilon')) > 0.5$.

$P(\Gamma(a, \epsilon) > \Gamma(b, \epsilon')) > 0.5 \implies a > b$

Towards a contrapositive suppose $a \leq b$. By monotonicity of Γ in the first argument we know that $\Gamma(a, \epsilon) \leq \Gamma(b, \epsilon)$. By monotonicity of the second argument we also know that the event $\Gamma(a, \epsilon) > \Gamma(b, \epsilon')$ is contained within the event $\epsilon > \epsilon'$. The probability of this event occurring is 0.5 by symmetry of $F_{\epsilon,\epsilon'}$. Therefore, the probability that $\Gamma(a, \epsilon) > \Gamma(b, \epsilon')$ is bounded above by 0.5 when $a \leq b$. We reach a contrapositive.

Given what has been proven above, $\alpha^*(\ell, \ell') > 0.5$ implies $U(\ell) > U(\ell')$. *Assumption 2* implies that $U(\ell) > U(m^*(\ell, \ell')) > U(\ell')$. $U(\ell) > U(m^*(\ell, \ell'))$ implies $\alpha^*(\ell, m^*(\ell, \ell')) > 0.5$. Hence we have *Hypothesis 2*.

A2.2 Hypothesis 2’ (Stochastic Choice and Size Consistency)

Suppose WLOG that $U(\ell) > U(\ell')$. By assumption 2 this implies $U(\ell) > U(m^*(\ell, \ell')) > U(\ell')$. Define the following two sets:

$$A := \{(\epsilon, \epsilon') : \Gamma(U(\ell), \epsilon) > \Gamma(U(\ell'), \epsilon')\}$$

and

$$B := \{(\epsilon, \epsilon') : \Gamma(U(\ell), \epsilon) > \Gamma(U(m^*(\ell, \ell')), \epsilon')\}$$

and note that the occurrence of the event B implies the occurrence of event A for any given realization of (ϵ, ϵ'). The identical distribution of shock pairs implies that $P(B) \leq P(A)$. If follows that $P(\Gamma(U(\ell), \epsilon) > \Gamma(U(\ell'), \epsilon')) \geq P(\Gamma(U(\ell), \epsilon'') > \Gamma(U(m^*(\ell, \ell')), \epsilon'''))$. If the mixture weights are proportional to the probability of choosing one lottery over another, then the previous statement implies $\alpha^*(\ell, \ell') \geq \alpha^*(\ell, m^*(\ell, \ell'))$. The more general result can therefore be stated as $|\alpha^*(\ell, \ell') - 0.5| \geq |\alpha^*(\ell, m^*(\ell, \ell')) - 0.5|$.

A3 Supplementary Tables and Figures

Table A1: Observation Counts for Filtration

Treatment	Part	Total	After Degen. Drop	After Non-specified Drop
Slider	1	2796	2796	2552
	2	2796	1440	1305
Slider Info	1	2772	2772	2548
	2	2772	1458	1337
Repeated Choice	1	2544	2544	2202
	2	2544	1064	804

Notes: Total refers to the total number of observations by treatment and part that satisfied the initial comprehension check filters. After Degen. Drop refers to the number of observations that also where not Part 1 degenerate answers. After Non-specified Drop refers to the number of observations that satisfy previous filters and also are were specified.

Table A2: Summary Statistics

Treatment	Part	Specified		Degenerate Mixture		Weight on Higher EV		Weight on Mixture	
Slider	1	0.913	(0.005)	0.505	(0.01)	0.657	(0.007)	—	
	2	0.906	(0.008)	0.304	(0.013)	0.58	(0.009)	0.446	(0.009)
Slider Info	1	0.919	(0.005)	0.506	(0.01)	0.668	(0.007)	—	
	2	0.917	(0.007)	0.39	(0.013)	0.603	(0.01)	0.475	(0.01)
Repeated Choice	1	0.866	(0.007)	0.656	(0.01)	0.715	(0.008)	—	
	2	0.756	(0.013)	0.34	(0.017)	0.571	(0.012)	0.425	(0.012)

Notes: Summary statistics by treatment and part of unpaired observations. Only Part 2 observations where the mixture is degenerate from Part 1 are excluded. Specified refers to the proportion of observations that are specified. Degenerate Mixture refers to the proportion of specified observations that are degenerate. Weight on Higher EV refers to the weight on the higher expected value lottery conditional on expected values not being equal. Weight on Mixture refers to the weight on the mixture. This only applies for Part 2 observations. Standard errors are given in parentheses.

Table A3: Proportions of Part 1 and Part 2 Lottery Weights by Location

	Lottery Weight Part 2											
	Slider				Slider Info				Repeated Choice			
	> 0.5	$= 0.5$	< 0.5	Total	> 0.5	$= 0.5$	< 0.5	Total	> 0.5	$= 0.5$	< 0.5	Total
> 0.5	0.194	0.045	0.106	0.345	0.188	0.042	0.121	0.351	0.234	0.032	0.091	0.357
$= 0.5$	0.124	0.094	0.088	0.306	0.125	0.072	0.097	0.294	0.157	0.056	0.085	0.298
< 0.5	0.137	0.038	0.174	0.349	0.122	0.033	0.2	0.355	0.125	0.048	0.173	0.346
Total	0.455	0.176	0.369	—	0.435	0.147	0.418	—	0.515	0.136	0.349	—

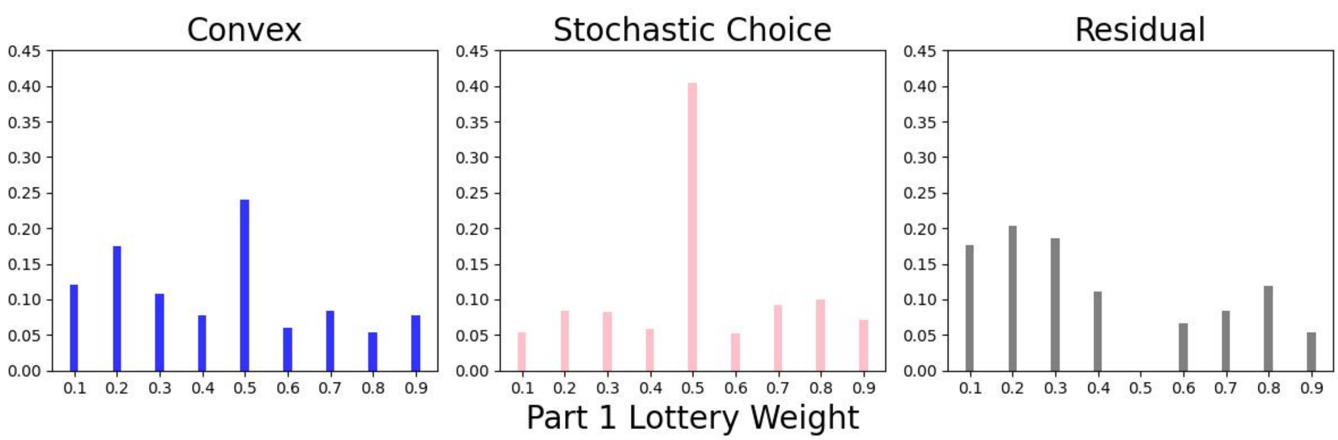
Notes: Column values represent proportions of Part 2 lottery weights that are greater than, equal to, or less than 0.5*. Rows represent the proportion of Part 1 lottery weights that are greater than, equal to, or less than* 0.5*. These are proportions of fully specified and Part 1 non-degenerate observations from the paired dataset.*

Table A4: Size Consistency and Stochastic Choice across Treatments

	Slider			Slider Info			Repeated Choice		
	Prop.	Naive	Emp.	Prop.	Naive	Emp.	Prop.	Naive	Emp.
Size and SC	0.342	0.293***	0.253***	0.275	0.293	0.247**	0.317	0.293*	0.229***
Size as SC	0.452	0.529***	0.389***	0.363	0.529***	0.39*	0.404	0.529***	0.368**

Notes: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. *Prop. refers to the sample proportion of consistent observations in the paired, fully specified and Part 1 non-degenerate dataset. Naive refers to the Naive null hypothesis proportion. Emp. refers to the Empirical null hypothesis proportion. Asterisks represent level of significance of difference between Prop. and the respective hypothesis. Size and SC refers to the proportion of size consistent and stochastic choice consistent observations as a proportion of fully specified and Part 1 non-degenerate observations. Size as SC refers to the the number of size consistent observations as a proportion of stochastic choice consistent observations.*

Figure A1: Distribution of Explanation Consistent Observations (*Slider* treatment)



Notes: Figures show a breakdown of explanation consistent observations by Part 1 lottery weight. The x-axis shows the Part 1 lottery weight and the y-axis shows the proportion of observations.

Table A5: Proportion of *Part 2* Paired Observations Consistent with Explanations

	Slider			Slider Info			Repeated Choice		
	0%	50%	100%	0%	50%	100%	0%	50%	100%
Convex	0.71	0.277	0.013	0.6	0.321	0.079	0.799	0.179	0.022
Stochastic Choice	0.081	0.318	0.6	0.086	0.305	0.609	0.051	0.337	0.612
Residual	0.646	0.311	0.044	0.685	0.271	0.045	0.63	0.348	0.022

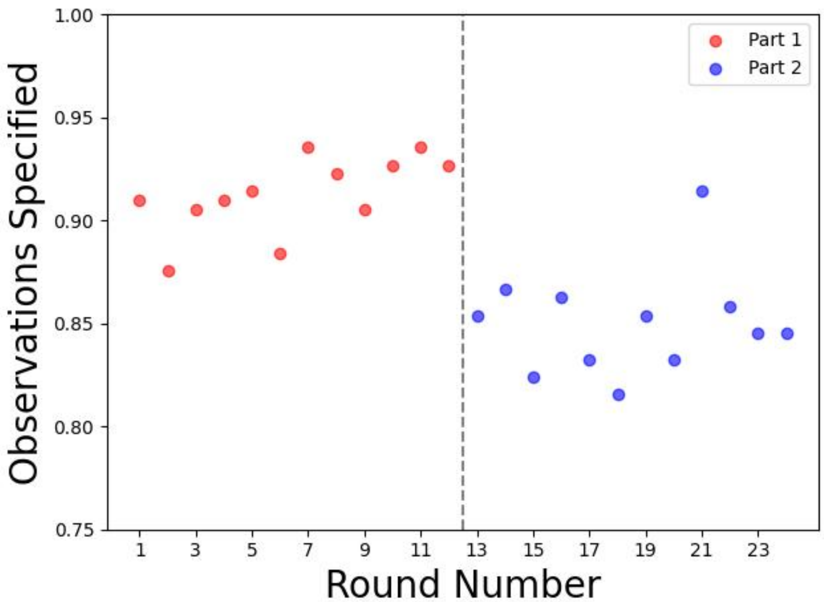
Notes: Part 2 paired observations: $\{\ell, m^*(\ell,\ell')\}$ *and* $\{\ell', m^*(\ell,\ell')\}$*, consistent with Convex Preferences and Stochastic Choice as a proportion of fully specified and Part 1 non-degenerate observations. Both Part 2 questions are specified.*

Table A6: Observations Consistent with Explanations by *Part 1* Lottery Weight

	Slider			Slider Info			Repeated Choice		
	< 0.5	$= 0.5$	> 0.5	< 0.5	$= 0.5$	> 0.5	< 0.5	$= 0.5$	> 0.5
Convex	0.041	0.072	0.036	0.059	0.114	0.064	0.026	0.062	0.026
Stochastic Choice	0.239	0.212	0.306	0.23	0.233	0.294	0.266	0.221	0.298
Residual	0.065	0.137	—	0.062	0.122	—	0.066	0.125	—

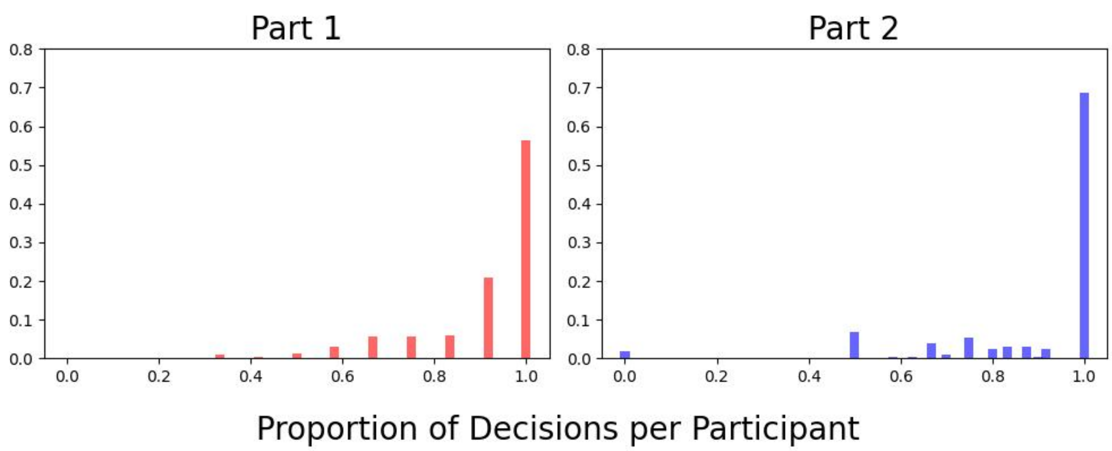
Notes: Number of observations consistent with convex preferences and stochastic choice by Part 1 lottery weight as a proportion of fully specified and Part 1 non-degenerate observations. < 0.5*,* $= 0.5$*, and* > 0.5 *refer to Part 1 lottery weights less than* 0.5*, equal to* 0.5 *and greater than* 0.5*.*

Figure A2: Proportion of Specified Observations by Round Number (*Slider* treatment)



Notes: The proportion of observations specified versus not specified by round number. Data from the unpaired dataset. Cases in Part 2 where the mixture is degenerate are dropped.

Figure A3: Proportion of Specified Observations per Participant (*Slider* treatment)



Notes: The x-axis refers to the proportion of observations that are specified per each participant in each part of the experiment (the weight at 1 *is the proportion of participants that specified all of their mixtures in that part. The weight at* 0 *is the proportion of participants that specified none of their mixtures in that part). There is no weight at* 0 *for Part 1 because these participants were filtered out of the main dataset. There is weight at* 0 *in Part 2 because some participants have specified for degenerate mixtures and not for non-degenerate mixtures, meaning that they were not filtered out of the main dataset.*

Table A7: Specification and Consistency Rates for Observation Triplets

	Slider	Slider info	Repeated choice
Specification Rate	0.881	0.894	0.742
All Specified	0.733	0.765	0.513
Part 2 Consistent (Weak)	0.619	0.616	0.59
Part 2 Consistent (Strong)	0.43	0.444	0.392
Convex Consistent	0.1	0.09	0.095
Stochastic Choice Consistent	0.462	0.434	0.425

Notes: This table provides statistics for the triplets $\alpha^*(\ell,\ell'), \alpha^*(\ell, m^*(\ell,\ell'))$ *and* $\alpha^*(\ell', m^*(\ell,\ell'))$*. Triplets containing Part 1 degenerate mixtures are excluded. Specification Rate refers to the average number of the three mixture weights that were specified. All Specified refers to the proportion of triplets in which all three weights are specified. The remaining statistics concern observation triplets that are fully specified. Part 2 Consistent (Weak) refers to the proportion of observations where* $(\alpha^*(\ell, m^*(\ell,\ell')) - 0.5)(0.5 - \alpha^*(\ell', m^*(\ell,\ell'))) \geq 0$*. Part 2 Consistent (Strong) refers to the proportion of triplets where* $(\alpha^*(\ell, m^*(\ell,\ell')) - 0.5)(0.5 - \alpha^*(\ell', m^*(\ell,\ell'))) > 0$ *or* $\alpha^*(\ell, m^*(\ell,\ell')) = \alpha^*(\ell', m^*(\ell,\ell')) = 0.5$*. Convex Consistent refers to the proportion of triplets where* $\alpha^*(\ell, m^*(\ell,\ell')) \leq 0.1$ *and* $\alpha^*(\ell', m^*(\ell,\ell')) \leq 0.1$*. Stochastic Choice Consistent refers to the proportion of triplets that satisfy* $(\alpha^*(\ell,\ell') - 0.5)(\alpha^*(\ell, m^*(\ell,\ell'))) \geq 0$ *and* $(\alpha^*(\ell,\ell') - 0.5)(\alpha^*(\ell', m^*(\ell,\ell'))) \geq 0$

Table A8: Lotteries

Group	\$8	\$9	\$10	\$11	\$12	\$13	\$14	\$15	\$16	\$17	\$18	EV	Variance	Support Size
1	0.0	0.0	0.5	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	12.0	4.0	2
1	0.0	0.25	0.0	0.25	0.0	0.0	0.5	0.0	0.0	0.0	0.0	12.0	4.5	3
1	0.0	0.0	0.5	0.0	0.0	0.25	0.0	0.25	0.0	0.0	0.0	12.0	4.5	3
1	0.0	0.25	0.0	0.25	0.0	0.25	0.0	0.25	0.0	0.0	0.0	12.0	5.0	4
1	0.0	0.4	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	12.0	6.0	2
1	0.2	0.0	0.2	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	12.0	6.4	3
1	0.0	0.4	0.0	0.0	0.0	0.3	0.0	0.3	0.0	0.0	0.0	12.0	6.6	3
1	0.2	0.0	0.2	0.0	0.0	0.3	0.0	0.3	0.0	0.0	0.0	12.0	7.0	4
1	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	12.0	9.0	2
1	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0	12.2	11.76	2
1	0.3	0.0	0.1	0.0	0.0	0.1	0.0	0.5	0.0	0.0	0.0	12.2	9.76	4
1	0.4	0.1	0.0	0.0	0.0	0.0	0.0	0.2	0.3	0.0	0.0	11.9	13.89	4
2	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.5	0.0	0.0	14.0	4.0	2
2	0.0	0.0	0.0	0.25	0.0	0.25	0.0	0.0	0.5	0.0	0.0	14.0	4.5	3
2	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.25	0.0	0.25	0.0	14.0	4.5	3
2	0.0	0.0	0.0	0.25	0.0	0.25	0.0	0.25	0.0	0.25	0.0	14.0	5.0	4
2	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.6	0.0	0.0	14.0	6.0	2
2	0.0	0.0	0.2	0.0	0.0	0.2	0.0	0.0	0.6	0.0	0.0	14.2	5.76	3
2	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.3	0.0	0.3	0.0	14.0	6.6	3
2	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.3	0.0	0.3	0.0	14.0	7.0	4
2	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.5	0.0	14.0	9.0	2
2	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	14.2	11.76	2
2	0.0	0.0	0.3	0.1	0.0	0.0	0.0	0.1	0.0	0.5	0.0	14.1	10.29	4
2	0.0	0.0	0.3	0.1	0.1	0.0	0.0	0.0	0.0	0.3	0.2	14.0	12.0	5

Notes: Table shows the structure and characteristics of each of the 24 lotteries used in the experiment. Group refers to the group that the lottery belongs to (either EV12 or EV14). \$i shows the probability of receiving \$i in that lottery. EV stands for expected value. Support Size refers to the number of unique outcomes that may occur with strictly positive probability.