Connecting Common Ratio and Common Consequence Preferences*

Christina McGranaghan
University of Delaware

Kirby Nielsen
California Institute of Technology

Ted O'Donoghue
Cornell University

Jason Somerville
UC Santa Barbara

Charles D. Sprenger
California Institute of Technology

Abstract

Common ratio and common consequence effects are two foundational deviations from expected utility (EU). Most non-EU models treat them as global preference features, yet there has been little experimental exploration of parameters. Moreover, the two effects are typically analyzed independently despite their natural link through mixture attitudes. We empirically characterize combinations of preferences across the parameter space, documenting new regularities that are inconsistent with leading non-EU models. Finding no prominent model that can explain these regularities, we propose a model of "upside potential" motivated by our results.

JEL: C91, D81, D91.

*For helpful comments and suggestions, we thank Doug Bernheim, Drew Fudenberg, Shengwu Li, and seminar participants at Chapman University, University of Zurich, University of Pittsburgh, Lehigh University, Stanford University, University of California at Berkeley, University of California at Davis, University of California at Los Angeles, University of California at Santa Barbara, University of Southern California, Ohio State University, the Max Planck Institute for Research on Collective Goods, the 2023 Stanford Institute for Theoretical Economics, the 2023 North American Meeting of the Economic Science Association, and the 2023 Spring Behavioral & Experimental Economics Research Conference. The experiment reported in this paper was preregistered in the AEA RCT Registry in August 2022, under the ID AEARCTR-0009974. The experiment was reviewed and granted an exemption by the Institutional Review Board at the California Institute of Technology under protocol number IR21-2073.

1 Introduction

The common ratio effect (CRE) and the common consequence effect (CCE) are two prominent deviations from expected utility (EU). These effects were first proposed as thought experiments by Allais (1953) and served as important counterexamples against EU as a positive model of behavior. They were later popularized by Kahneman and Tversky (1979), who broadened their scope by treating them as foundational features of preferences that motivate the shape of the probability weighting function in prospect theory. Subsequent work found empirical support for both effects, and researchers have developed many non-EU models of decision-making to accommodate them.

Despite the widespread acceptance of the CRE and CCE as stylized facts, the existing evidence suffers from two major limitations. First, while behavioral theories often assume these effects represent global features of preferences, the experimental evidence for each comes from a narrow set of experimental parameters, and thus there is little empirical justification for such global assumptions. Second, previous studies have almost always analyzed the CRE and CCE independently. However, there is a natural connection between them that links both to a third property of risk preferences that has been studied in the literature: attitudes toward probabilistic mixtures of lotteries. Examining these three features of preferences together through connected problems provides a more complete picture of risk attitudes and an improved empirical base for assessing models of risk preferences.

In this paper, we examine connected common ratio (CR), common consequence (CC), and mixture (MX) problems across a broad range of experimental parameters, including many not previously explored in the literature. For parameters close to those in the canonical CR and CC problems, we observe a CRE and a CCE, consistent with conventional findings. However, across much of the parameter space, we find a modest CRE, a pronounced *reverse* CCE, and a robust attraction to probabilistic mixtures. This prominent new regularity in risk attitudes is inconsistent with both EU and leading non-EU models. Because our empirical results conflict sharply with existing models, we propose a theoretical model of "upside potential" that is motivated by our results.

In Section 2, we formally demonstrate the natural connection between CR, CC, and MX problems and how it can yield insight for models of risk preferences. We then leverage this connection to summarize and reanalyze the existing experimental literature. Allais (1953) first introduced the CR problem, but the canonical version comes from Kahneman and Tversky (1979) and presents individuals with the following pair of choice tasks:[1]

***AB* Choice**: Lottery A: 100 percent chance of \$3000 vs. Lottery B: 80 percent chance of \$4000

***CD* Choice**: Lottery C: 25 percent chance of \$3000 vs. Lottery D: 20 percent chance of \$4000

[1] Throughout the paper, we suppress the probability of \$0. For instance, in Lottery B, the remaining probability of 20 percent is on a \$0 outcome.

In this example, lotteries C and D are constructed from lotteries A and B by scaling down the probabilities of the non-zero outcomes by a *common ratio* of 0.25. The canonical version of the CCE is from Allais (1953) and presents participants with the following pair of choice tasks:

$\boldsymbol{AB'}$ **Choice**:	Lottery A:	100 percent chance of \$1M	vs.	Lottery B':	89 percent chance of \$1M 10 percent chance of \$5M
$\boldsymbol{CD}$ **Choice**:	Lottery C:	11 percent chance of \$1M	vs.	Lottery D:	10 percent chance of \$5M

Lotteries C and D are constructed from lotteries A and B' by changing the *common consequence* of an 89 percent chance of \$1M into an 89 percent chance of \$0.

EU predicts that CR and CC manipulations cannot reverse an individual's preference. In contrast, the CRE and CCE describe systematic empirical patterns of people appearing more risk-tolerant in the CD problems relative to the AB and AB' problems, respectively. As a result, the aggregate share of individuals choosing A is larger than the share choosing C in both problems.

The prior literature contains many experiments on the CRE and CCE. Two recent meta-studies by Blavatskyy et al. (2023) on the CRE and Blavatskyy et al. (2022) on the CCE report a total of 224 experiments (143 CRE and 81 CCE) across 48 studies. They document that both effects are sensitive to various design choices, including experimental parameters. We leverage their data to highlight two further points. First, these experiments cover a limited set of experimental parameters that are dominated by the canonical examples in Kahneman and Tversky (1979) and Allais (1953). Second, these experiments have almost always studied the CRE and CCE as independent phenomena—indeed, of the 48 studies, only 10 contain both CRE and CCE experiments, and virtually all of these study the two problems separately.

By studying the CRE and CCE as independent phenomena, the prior literature has neglected a natural connection between them that can provide a more complete picture of the underlying nature of risk attitudes and an improved empirical base for assessing models of risk preferences. While this connection was noted in early theoretical work by Chew and MacCrimmon (1979) and both theoretically and experimentally by Chew and Waller (1986), it remains largely absent from the CRE and CCE literature. To illustrate this connection, consider the following three choice tasks:

$\boldsymbol{AB}$ **Choice**:	Lottery A:	100 percent chance of \$27	vs.	Lottery B:	90 percent chance of \$35
$\boldsymbol{AB'}$ **Choice**:	Lottery A:	100 percent chance of \$27	vs.	Lottery B':	9 percent chance of \$35 90 percent chance of \$27
$\boldsymbol{CD}$ **Choice**:	Lottery C:	10 percent chance of \$27	vs.	Lottery D:	9 percent chance of \$35

The combination of the AB and CD choices constitutes a CR problem while the combination of

the AB' and CD choices constitutes a CC problem. Thus, three tasks are sufficient to study both problems. This formulation also highlights a third comparison: AB versus AB'. Since B' is a probabilistic mixture of lotteries A and B (in this case 90 percent A and 10 percent B), this third comparison reveals attitudes toward probabilistic mixtures. We refer to the combination of the AB and AB' choices as a *mixture (MX)* problem, where EU predicts that people should be neutral to mixtures. We refer to a finding of people appearing more risk-tolerant in the AB' problem relative to the AB problem as a mixture effect (MXE). In other words, the MXE pattern suggests that individuals prefer the probabilistic mixture of A and B to both A and B individually.[2]

We refer to the combination of an AB task, an AB' task, and a CD task as a *connected CR-CC-MX problem*. We characterize such problems with two parameters. The first is the probability of the non-zero outcome in lottery B, which we denote by p (0.9 in the example above). The second parameter, which we denote by r, is both the common ratio that converts A and B into C and D and the mixing weight that generates B' from A and B (0.1 in the example above).[3]

Each connected CR-CC-MX problem allows us to simultaneously identify three features of underlying preferences for a given (p, r) parameter combination: whether people have a *common ratio preference (CRP)*, a *common consequence preference (CCP)*, and a *mixture preference (MXP)*. In each case, a person might have that preference, its reverse (which we label RCRP, RCCP, or RMXP), or be neutral to the manipulation (which we label ⃠CRP, ⃠CCP, and ⃠MXP).[4] Existing models of risk preferences make predictions for the empirical patterns we ought to observe. EU implies neutrality to all three manipulations and therefore pattern ⃠CRP-⃠CCP-⃠MXP. In Section 2.2, we delineate the predictions from several leading parameterized non-EU models. Under each model's common assumptions, all predict a CRP and a CCP, and most further predict either an RMXP or an ⃠MXP; that is, people either dislike or are neutral to mixtures. Furthermore, in most cases these models predict these patterns apply globally and are independent of (p, r). We can test these predictions using connected CR-CC-MX problems.

While it is possible to study CR, CC, and MX problems using binary choice tasks like those above—as most of the prior literature, including Chew and Waller (1986), does—this approach can yield biased conclusions in the presence of choice noise. In McGranaghan et al. (2024), we build on prior work highlighting this potential bias in CRE experiments and demonstrate how paired valuation tasks can provide unbiased inference under these circumstances. Therefore, our primary analysis focuses on valuation tasks in the form of stated indifference points implemented using multiple price

[2]The MX problem relates to tests of betweenness and to work on deliberate randomization; we discuss these literatures in Section 2.4. We also highlight that our connected problems examine one specific type of MX problem in which a binary lottery is mixed with a certain lottery, and thus our analysis will not provide a full characterization of people's attitudes toward mixtures.

[3]The canonical version of the CRE uses $p = 0.8$ and $r = 0.25$, and the canonical version of the CCE uses $p = 10/11$ and $r = 0.11$.

[4]We adopt the "reverse" terminology from the literature, which dates at least to Blavatskyy (2010).

lists. For example, in a CD valuation task analogous to the CD choice task above, a participant selects the amount to be received with a 9 percent chance that makes them indifferent to a 10 percent chance of \$27. To complement the primary analysis, we conduct a secondary analysis using carefully selected choice tasks. Both methods yield the same conclusions about average behavior across the parameter space.[5] Moreover, we document strong links between valuations and choices, indicating that the patterns we find meaningfully reflect underlying preferences.

In Section 3, we describe the details of our experimental design. We recruit 2,102 participants through Prolific for an online experiment. In stage 1 of the experiment, we elicit a series of valuations. Each participant provides AB, AB', and CD valuations for four different (p, r) combinations. Across all of our participants, we implement 20 different (p, r) combinations covering a wide range of the parameter space. In stage 2, the same participants complete binary choice tasks linked to their stage 1 valuation tasks, facilitating direct comparison between the patterns of behavior emerging in valuations and choices.

Section 4 describes our main results. We first study aggregate behavior by analyzing mean valuations across all participants. We find that the three features of preferences react differently to changes in the experimental parameters. CRP is highly sensitive to the common ratio r: It increases as r decreases, and disappears entirely for the two largest values of r in our design. In contrast, CCP is highly sensitive to the high-prize probability p: It decreases as p decreases, transitioning from a small CCP for large p to a substantial RCCP for small p. Finally, we find a robust MXP that increases as either r or p decreases. Taken together, these results yield two striking regularities in the combinations of CR-CC-MX preferences. On one hand, for parameters close to the canonical CR and CC examples (large p and small r), we find CRP and CCP with limited MXP, consistent with the conventional wisdom. On the other hand, across much of the parameter space, we find a robust pattern of CRP-RCCP-MXP, which is inconsistent with both EU and leading non-EU models. The exact same patterns emerge when analyzing the aggregate paired binary choice data, albeit with more noise.

We next analyze behavior at the individual level. Using our valuation data, we examine the distribution of behavior across the 27 possible combinations of CR-CC-MX patterns, where a person might exhibit the effect, its reverse, or no effect for each. We find that mean behavior masks substantial heterogeneity. The modal individual pattern aligns with the CRP-RCCP-MXP combination that frequently appears in mean preferences, but it accounts for only 15 percent of observations. Of

[5]While not the main focus of this paper, we describe in Section 2.5 how the inference problem described in McGranaghan et al. (2024) becomes even more severe when comparing three binary choice tasks, as in a connected CR-CC-MX problem. Section 4.2 presents empirical evidence for the relevance of these biased inference problems for our CR, CC, and MX choice tasks. At the same time, we describe in Sections 4.2 and 4.5 why our choice data yield the same conclusions as our valuations data despite the potential inference problems: Based on pilot data, we were able to select experimental parameters that produced a set of paired choice problems that mitigate aggregate bias by balancing instances in which choice noise would lead to positive and negative bias.

course, some of this variability in behavior at the individual level is likely due to noise. Because our design elicits multiple measures of the same valuation within subject, we can structurally decompose the variability in valuations into an underlying preference component and a noise component. We estimate that roughly half of the variability in valuations is due to preference heterogeneity. The resulting distribution of preferences still yields a modal pattern of CRP-RCCP-MXP, but also involves substantial heterogeneity. Finally, we show that heterogeneity in preferences measured from valuations strongly predicts choice behavior, suggesting that choices and valuations capture some common information about the underlying preferences of interest.

Taken together, our experimental results pose an important challenge to existing models of risk preferences. In simple terms, our results confirm some earlier results that suggest the CRE and CCE are not global phenomena, whereas prominent models of risk preferences either assume or imply that they are. More unique to our data is the prevalence of the pattern CRP-RCCP-MXP, which cannot be accommodated by any of the leading parameterized non-EU models.

Because our experimental results are so at odds with existing models, in Section 5 we attempt to develop a model to rationalize our data. Our approach is motivated by our finding that people have a robust attraction to probabilistic mixtures between a certain payment and a binary lottery with a higher expected value. This finding suggests that, compared to having the certain amount, people prefer to mix in some, but not all, of the binary lottery. To accommodate this preference, we posit that people value lotteries according to their "upside potential," trading off the total probability of winning something against an expected valuation of those winnings—which turns out to be a previously unstudied version of quadratic utility (Machina, 1982; Chew et al., 1991). Encouragingly, this approach to accommodating MXP also explains the modal CRP-RCCP-MXP pattern we observe, and successfully predicts some more nuanced patterns in our data. Expanding beyond our data, we derive more general predictions of the model for indifference curves in a Marschak-Machina triangle—which offers additional testable implications for future research—and show how the model is consistent with the reliable finding of risk tolerance for small probabilities and risk aversion for large probabilities when eliciting certainty equivalents for binary lotteries (as in Tversky and Kahneman, 1992). Though the psychology of upside potential appears a promising way to organize our data, we also discuss the limitations of the model, and hope that future work can expand and improve upon the initial ideas we propose.

Finally, in Section 6 we discuss some broader implications of our analysis, including connections between our results and the recent literature on deliberate randomization and the underlying mechanisms for such phenomena (Cerreia-Vioglio et al., 2015; Agranov and Ortoleva, 2017; Dwenger et al., 2018; Feldman and Rehbeck, 2022; Agranov and Ortoleva, 2023; Agranov et al., 2023).

Our exploration of the connections between CRP, CCP, and MXP is similar in spirit to exercises by Dean and Ortoleva (2019), Chapman et al. (2023), and Stango and Zinman (2023), who study the

correlations between decision-making phenomena across a range of domains including risk, ambiguity, updating, and intertemporal choice. However, our contribution differs in two important ways. First, our work is motivated by structural connections between the three features of risk preferences, and these connections provide a framework for using the empirical patterns that emerge to assess models of risk preferences. In contrast, the correlational analyses above are challenging to interpret because it is unclear how a link between, say, present bias and ambiguity aversion might relate to specific models of choice. Second, one of our primary goals is to study how the three features of risk preferences vary across the parameter space, which is crucial to assessing the performance of different models. In contrast, the papers cited above use a limited number of measures for each phenomena, often at canonical parameterizations. For example, Dean and Ortoleva (2019) have a single measure of the CRE and a single measure of the CCE, each at its canonical parameter values. While such exercises are valuable for documenting broad correlations in classic phenomena across many domains, this approach provides limited information for assessing different models.

Finally, we note the potential benefits of broader exploration of parameter spaces in other domains and the richer appreciation of preferences that might emerge from these exercises. Within the domain of risky choice, our exploration yields a new empirical foundation upon which to theorize. We hope such explorations in other domains will be similarly useful.

2 Background and Prior Literature

The prior literature has almost always studied the CRE and CCE as separate problems, and for each problem, it has focused on a limited set of parameters. In contrast, we study connected problems for a broad set of parameters. In this section, we formalize the connection and highlight the value of investigating connected problems, which is reminiscent of—but expands upon—the exercise in Chew and MacCrimmon (1979) and Chew and Waller (1986). We then describe in detail the limitations of the prior literature.

2.1 Connected CR-CC-MX Problems

For fixed prizes $H > M > 0$, consider three binary choice tasks parameterized by the vector (p, r), where $p, r \in (0, 1)$:

***AB* Choice Task**:	choose Lottery $A \equiv (M, 1)$	or	Lottery $B \equiv (H, p)$
***AB′* Choice Task**:	choose Lottery $A \equiv (M, 1)$	or	Lottery $B' \equiv (H, pr; M, 1 - r)$
***CD* Choice Task**:	choose Lottery $C \equiv (M, r)$	or	Lottery $D \equiv (H, pr)$

The combination of an AB task and a CD task represents a CR Problem. According to EU, scaling down the probabilities of the non-zero outcomes by the common ratio r should leave preferences unchanged: Anyone who prefers A over B should also prefer C over D, and anyone who prefers B over A should also prefer D over C. In contrast, the CRE pattern involves choosing lotteries A and D; that is, people are more risk-tolerant in the CD task than in the AB task.

The combination of an AB' task and a CD task represents a CC Problem. According to EU, shifting the common consequence of a $1-r$ chance of $\$M$ into a $1-r$ chance of \$0 should leave preferences unchanged: Anyone who prefers A over B' should also prefer C over D, and anyone who prefers B' over A should also prefer D over C. In contrast, the CCE pattern involves choosing lotteries A and D; that is, people are more risk-tolerant in the CD task than in the AB' task.

Hence, for a given H and M, each (p, r) generates a connected pair of CR and CC problems. Panel A of Figure 1 illustrates one such connected pair by plotting the five lotteries in a Marschak-Machina probability triangle for a particular (p, r). Each problem separately involves comparing two choice sets on parallel line segments. Connecting the problems involves placing the AB and AB' choice sets on the same line and using the same CD choice for both problems.

Panel A also highlights that lottery B' is a probabilistic mixture of lotteries A and B, with weights $(1-r)$ and r, and suggests a third comparison of two parallel choice sets: AB versus AB'. This comparison reveals people's attitudes toward probabilistic mixtures of a binary lottery and a certain lottery, and thus we refer to the combination of the AB and AB' choices as a *mixture (MX)* problem. According to EU, people should be neutral to probabilistic mixtures, and thus the expected utility of a probabilistic mixture of A and B should be in between the expected utilities of A and B. Hence, anyone who prefers A over B should also prefer A over B', and anyone who prefers B over A should also prefer B' over A. In contrast, we refer to the pattern of choosing A over B and B' over A as a mixture effect (MXE) since people are appearing more risk-tolerant in the AB' problem relative to the AB problem—i.e., they appear to like the probabilistic mixture.

Hence, for a given (p, r), we refer to the combination of an AB task, an AB' task, and a CD task as a *connected CR-CC-MX problem*.

2.2 Why Study Connected Problems?

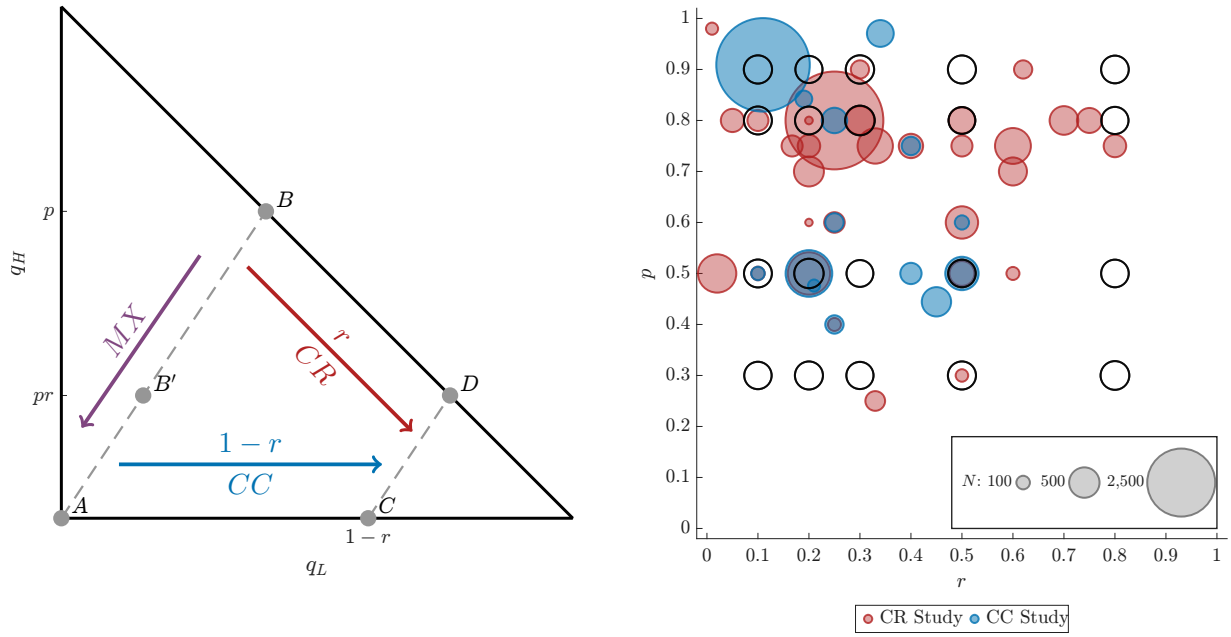
Studying connected problems allows researchers to obtain richer quantitative data that provide a stronger empirical base for assessing models of risk preferences. In particular, it reveals patterns in how people simultaneously react to CR, CC, and MX manipulations, allowing us to jointly study three key characteristics of underlying preferences.

To formalize these preferences, if we assume monotonicity, then for a fixed (p, r, M) there exist three underlying indifference values h^*_{AB}, $h^*_{AB'}$, and h^*_{CD} that satisfy:

Figure 1: Connected Problems and Parameter Coverage in the Prior CR and CC Literature

Panel A: CR, CC, and MX in a MM Triangle

Panel B: CR and CC Parameter Coverage



Note: Panel A depicts a Marschak-Machina (MM) triangle for lotteries with outcomes H, M, and 0 and associated probabilities q_H, q_M, and q_L. Paired binary choices $A \equiv (M, 1)$ vs. $B \equiv (H, p)$ and $C \equiv (M, r)$ vs. $D \equiv (H, pr)$ represent a common ratio (CR) problem. Paired binary choices A vs. $B' \equiv (H, pr; M, 1-r)$ and C vs. D represent a common consequence (CC) problem. Paired binary choices A vs. B and A vs. B' represent a mixture (MX) problem. Panel B depicts the experimental parameters (p, r) used in the prior literature and our paired-choice experiments. Red dots denote the CR experiments reported in Blavatskyy et al. (2023). Blue dots denote the CC experiments reported in Blavatskyy et al. (2022). Black open circles denote paired-choice parameters in stage 2 of our study. The size of each circle reflects the total number of participants who completed a paired choice task with that parameter.

- Prefer A over B if and only if $H < h^*_{AB}$,
- Prefer A over B' if and only if $H < h^*_{AB'}$, and
- Prefer C over D if and only if $H < h^*_{CD}$.

Note that, for each, a smaller h^* implies increased risk tolerance.

A person's reaction to a CR manipulation is captured by $\Delta^*_{CR} \equiv h^*_{AB} - h^*_{CD}$, and their reaction to a CC manipulation is captured by $\Delta^*_{CC} \equiv h^*_{AB'} - h^*_{CD}$. The standard effects suggest greater risk tolerance in the CD choice than in either the AB or AB' choice, which implies $\Delta^*_{CR} > 0$ and $\Delta^*_{CC} > 0$. We label these two features of preferences *common ratio preference (CRP)* and *common consequence preference (CCP)*. We label $\Delta^*_{CR} < 0$ and $\Delta^*_{CC} < 0$ as *reverse common ratio preference (RCRP)* and *reverse common consequence preference (RCCP)*. A person's attitudes toward probabilistic mixtures are captured by $\Delta^*_{MX} \equiv h^*_{AB} - h^*_{AB'}$. If a person likes mixtures, they will have $\Delta^*_{MX} > 0$, which we label a *mixture preference (MXP)*. If they dislike mixtures, they will have $\Delta^*_{MX} < 0$, which we label

a *reverse mixture preference (RMXP).*[6]

With this notation, EU implies $\Delta^*_{CR} = \Delta^*_{CC} = \Delta^*_{MX} = 0$. Over the years, researchers have developed various non-EU models, and these models make predictions for Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX}. In Online Appendix B, we derive these predictions for several prominent parameterized models. In Table 1, we summarize the predictions from each model as a function of the model's key parameter, and also highlight the focal range for that parameter (i.e., the range in which that parameter is typically assumed to lie).

Table 1: Predictions of Leading Non-EU Models for Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX}

Model and Structure	Parameter Range	Predictions
Original Prospect Theory (Kahneman and Tversky, 1979) with $\pi(q) = q^\delta / \left[q^\delta + (1-q)^\delta\right]^{\frac{1}{\delta}}$, $U(B) = \pi(p)v(H)$	$\delta \in (0.279, 1)^\dagger$	$\Delta^*_{CR} > 0,\ \Delta^*_{CC} > 0,\ \Delta^*_{MX} \gtreqless 0$
	$\delta > 1$	$\Delta^*_{CR} < 0,\ \Delta^*_{CC} > 0,\ \Delta^*_{MX} < 0$
Cumulative Prospect Theory (Tversky and Kahneman, 1992) with $\pi(q) = q^\delta / \left[q^\delta + (1-q)^\delta\right]^{\frac{1}{\delta}}$, $U(B) = \pi(p)v(H)$	$\delta \in (0.279, 1)^\dagger$	$\Delta^*_{CR} > 0,\ \Delta^*_{CC} > 0,\ \Delta^*_{MX} \gtreqless 0$
	$\delta > 1$	$\Delta^*_{CR} < 0,\ \Delta^*_{CC} \gtreqless 0,\ \Delta^*_{MX} \gtreqless 0$
Loss Aversion under CPE (Kőszegi and Rabin, 2007)	$\Lambda \in (0, 1)^\dagger$	$\Delta^*_{CR} > 0,\ \Delta^*_{CC} > 0,\ \Delta^*_{MX} < 0$
$U(B) = pu(H) - p(1-p)\Lambda u(H)$	$\Lambda \in (-1, 0)$	$\Delta^*_{CR} < 0,\ \Delta^*_{CC} < 0,\ \Delta^*_{MX} > 0$
Disappointment Aversion (Bell, 1985)	$\beta \in (0, 1)^\dagger$	$\Delta^*_{CR} > 0,\ \Delta^*_{CC} > 0,\ \Delta^*_{MX} < 0$
$U(B) = pu(H) - p(1-p)\beta u(H)$	$\beta \in (-1, 0)$	$\Delta^*_{CR} < 0,\ \Delta^*_{CC} < 0,\ \Delta^*_{MX} > 0$
Disappointment Aversion (Gul, 1991)	$\beta > 0^\dagger$	$\Delta^*_{CR} = \Delta^*_{CC} > 0,\ \Delta^*_{MX} = 0$
$U(B) = \dfrac{p}{1 + \beta(1-p)} u(H)$	$\beta \in (-1, 0)$	$\Delta^*_{CR} = \Delta^*_{CC} < 0,\ \Delta^*_{MX} = 0$

Notes: Table presents predictions of each model for the sign of Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX}; see Online Appendix B for formal derivations. To give a sense of the parametric structure of each model, the first column provides the utility from lottery $B \equiv (H, p)$. For each model, the sign predictions are independent of the utility for outcomes ($v(x)$ or $u(x)$), and thus depend on the single listed parameter. Focal ranges for a model's key parameter are indicated by †, but the table also reports predictions for other ranges where the model is still well-defined but reflects a different psychology. All predictions hold for all $(p, r) \in (0, 1)^2$, except $\gtreqless$ indicates cases where the predictions depend on (p, r). Note that original prospect theory and cumulative prospect theory differ only for lottery B' among our set of five lotteries, and the same applies for Kőszegi-Rabin loss aversion under CPE (Choice-Acclimating Personal Equilibrium) and Bell disappointment aversion.

Table 1 demonstrates how connected CR-CC-MX problems can be used to assess models of non-EU risk preferences. In particular, it reveals three important features. First, each model predicts a

[6]Building on our comment in footnote 2, we note that Δ^*_{MX} only captures attitudes towards mixtures in which a binary lottery is mixed with a certain lottery. The decision-theory literature often considers global mixture attitudes, using the label "quasi-convex preferences" to refer to preferences with an attraction to mixtures of any two lotteries. Note, then, that while quasi-convex preferences imply an MXP (i.e., $\Delta^*_{MX} > 0$), an MXP does not imply a person has quasi-convex preferences.

CRP and CCP for the focal range of its key parameter. This commonality reflects that the CRE and CCE are typically seen as stylized facts that either motivate non-EU models or serve as a litmus test for those models. Second, the models differ in their predictions for mixture preferences. However, for the focal range of each model's key parameter, the prediction is either mixture neutrality or RMXP, except for the two variants of prospect theory where the prediction can go either way depending on the specific (p, r) combination. Third, when we also consider cases where the models' key parameters lie outside their focal ranges, the set of predicted patterns expands, but none of the models permit the pattern CRP, RCCP, and MXP that features prominently in our data.[7]

Two final observations are relevant for linking our data on connected CR-CC-MX problems to non-EU models. First, with a few noted exceptions, the directional predictions in Table 1 hold for any (p, r) combination—that is, these models predict preference patterns that are invariant to (p, r).[8] Second, of the models covered by Table 1, the two prospect theory models have some scope to expand their predictions by permitting alternative functional forms for the probability weighting function $\pi(q)$. We assess in Online Appendix G whether that additional flexibility can explain the patterns in our data.

2.3 Limitations of the Prior Literature on the CRE and CCE

There is a large empirical literature on both the CRE and CCE (and also a smaller empirical literature on the MXE that we discuss in Section 2.4). To illustrate the limitations of the prior evidence on the CRE and CCE, we reanalyze the literature in terms of the (p, r) combinations that researchers have used. Specifically, we merge data from two recent meta-studies: Blavatskyy et al. (2023) identify 143 CR experiments drawn from 39 studies and Blavatskyy et al. (2022) identify 81 CC experiments drawn from 29 studies. We combine the datasets from these meta-studies while converting the probabilities into our (p, r) framework. The resulting dataset yields several insights.

First, the prior literature has almost always studied the CRE and CCE independently. Of the CR and CC studies covered in the two meta-studies, only 10 contain both CR and CC experiments. Among those, only four collected observations for CR and CC problems at the same parameters for the same participants, and only two intentionally studied the connection between the two problems

[7] Table 1 focuses on parameterized non-EU models that make predictions as a function of parameters for any choice set. There are additional non-parameterized models based on axioms that make predictions for one or more of the three features of preferences. Two recent models make predictions for all three: Cautious expected utility (Cerreia-Vioglio et al., 2015) predicts $\Delta^*_{CR} = \Delta^*_{CC} > 0$ and $\Delta^*_{MX} = 0$, and simplicity preferences (Puri, 2025) predict $\Delta^*_{CR} > 0$, $\Delta^*_{CC} > 0$, and $\Delta^*_{MX} < 0$ (see Online Appendix B.6 and B.7). There is also a class of models that impose or satisfy the betweenness axiom, which implies $\Delta^*_{MX} = 0$; this class includes the Gul (1991) model in Table 1 and also models of weighted utility (Chew, 1983; Dekel, 1986). These models are thus inconsistent with the CRP-RCCP-MXP pattern that is prominent in our data.

[8] Beyond the predictions for the pattern of preferences, these models also make predictions about the sensitivities of Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX} to p and r. For probability-weighting models, the comparative statics are often non-monotone at leading parameter values, contrary to our data. For loss aversion and disappointment aversion models, the comparative statics are often directionally opposed to what we observe in our data.

to better understand the nature of risk preferences.[9]

Second, for each problem, the prior literature has used a limited set of parameters. Panel B of Figure 1 illustrates the (p, r) configurations used in this literature: Red circles denote CR experiments, while blue circles denote CC experiments. For each type of experiment, there is a large concentration at the canonical parameters. Of the 143 prior CR experiments, 48 (34%) use the Kahneman and Tversky (1979) values of $(p, r) = (0.80, 0.25)$, depicted by the large red dot in Panel B. Similarly, of the 81 prior CC experiments, 34 (42%) use the Allais (1953) values of $(p, r) = (0.91, 0.11)$, depicted by the large blue dot in Panel B. While there is some variation among the other experiments, Panel B reveals that a substantial portion of the parameter space remains unexplored.

This limited coverage would be of little consequence if the direction of the observed effects were broadly invariant to parameter choices. Indeed, all prominent parametrized non-EU models that we summarize in Table 1 predict, for their focal parameter range, that people have both a CRP and a CCP for all (p, r). However, Blavatskyy et al. (2023) and Blavatskyy et al. (2022) show that both the CRE and CCE are sensitive to parameter choices. Specifically, they find that a CRE is more likely to occur in experiments with smaller r, while the choice of p has no significant impact. Conversely, they find that the RCCE pattern becomes more likely for smaller p (their specification does not include r).[10]

In Table 2, we further explore the sensitivity of these effects to the choice of experimental parameters. We use the following continuous measures as our primary outcome variables:[11]

$$CRE - RCRE \equiv \widehat{\Pr}(A|AB) - \widehat{\Pr}(C|CD)$$
$$CCE - RCCE \equiv \widehat{\Pr}(A|AB') - \widehat{\Pr}(C|CD)$$

In Panel A of Table 2, we estimate OLS regressions with these measures as dependent variables. We replicate the qualitative findings from Blavatskyy et al. (2023) and Blavatskyy et al. (2022) and

[9] Burke et al. (1996) and Loomes and Sugden (1998) have connected CR-CC-MX problems, but this appears incidental. Burke et al. focus on whether the CCE differs for real versus hypothetical incentives; they do not mention that their study includes a connected CR-CC-MX experiment. Loomes and Sugden compare three stochastic specifications of EU, collecting data on roughly 50 binary choices that include five connected CR-CC-MX problems. However, they make no special use of these connected problems, and do not discuss CCE or MXE. Bateman and Munro (2005) and Chew and Waller (1986) intentionally study connected CR-CC-MX problems. Bateman and Munro examine whether couples' joint decision making differs from individual decision making in terms of the CRE, CCE, and MXE, using several connected CR-CC-MX problems. As emphasized in the Introduction, Chew and Waller explicitly highlight how connected CR-CC-MX problems can provide more information about risk preferences. However, both of these papers only cover a small set of parameter configurations.

[10] Blavatskyy et al. (2023) and Blavatskyy et al. (2022) focus much of their analysis on the effects of other features of experiments such as real versus hypothetical stakes, the ratio of high to middle outcomes, and whether lotteries are presented as a probability distributions versus in compound form.

[11] We use $\widehat{\Pr}(X|YZ)$ to denote the share of participants in an experiment that chooses option X from the choice set YZ. Note that $CRE - RCRE$ is equivalent to the share choosing combination A and D minus the share choosing B and C in a CR problem, and $CCE - RCCE$ is equivalent to the share choosing combination A and D minus the share choosing B' and C in a CC problem.

additionally find that CRE and CCE are differentially sensitive to p and r. This result suggests that what we learn about each effect and their connection is contingent on the parameters of the experiment.

In Panel B of Table 2, we divide prior CRE and CCE experiments into two categories: those conducted at the canonical (p, r) values and those conducted at alternative (p, r) values. We find a sizable CRE and CCE for the experiments at their respective canonical parameter values. In contrast, both effect sizes are significantly smaller in magnitude for experiments conducted at non-canonical values. Strikingly, CC experiments conducted away from the canonical (p, r) configuration exhibit an average RCCE.

Table 2: Sensitivity of Results to Experimental Parameters in the Prior Literature

Panel A. Sensitivity to Experimental Parameters

	(1) CR Study	(2) CC Study
H Probability: p	6.10	50.70
	(11.27)	(14.01)
Common Ratio: r	−57.82	−45.16
	(8.01)	(19.10)
Sample	Blavatskyy et al. (2023)	Blavatskyy et al. (2022)
Outcome Mean	19.1	8.64
Observations	143	81

Panel B. Canonical vs. Non-Canonical

	(3) Canonical	(4) Other	(5) Difference
	(i): KT Parameters		
CRE − RCRE	27.58	19.00	−8.58
	(14.78)	(24.04)	[−2.29]
Experiments	48	95	143
	(ii): Allais Parameters		
CCE − RCCE	17.40	−3.05	−20.46
	(17.34)	(22.95)	[−4.55]
Experiments	34	47	81

Notes: Panel A presents linear regressions that assess the sensitivity of experimental results from CR or CC studies based on the probability of the high outcome (p) and the common ratio (r). Column (1) presents the results for the 143 experiments reported in Blavatskyy et al. (2023), where the outcome is the net share of participants displaying a CRE relative to an RCRE, $CRE - RCRE$. Column (2) presents the results for the 81 experiments reported in Blavatskyy et al. (2022), where the outcome is the net share of participants displaying a CCE relative to an RCCE, $CCE - RCCE$. Specifications also include the value of the high outcome (H), the relative stakes ($M/(pH)$), and a dummy for whether the experiment had real stakes. Standard errors are in parentheses. Panel B presents the average of these outcomes based on whether the experiments reported in Blavatskyy et al. (2023) and Blavatskyy et al. (2022) were conducted at the canonical parameters in Kahneman and Tversky (1979) (KT; $p = 0.8$, $r = 0.25$) or Allais (1953) ($p = 0.91$, $r = 0.11$), respectively. Standard deviations are in parentheses, and t-statistics are in brackets. Regressions and Panel B summaries weight experiments by the number of participants that took part in a given study. Panel A outcome means are unweighted.

Recent work has shown that researcher decisions related to experimental design, data processing, and empirical analysis can have a substantial impact on the results (Steegen et al., 2016; Simonsohn et al., 2020; Huber et al., 2023; Menkveld et al., 2024). Table 2 reinforces this point for CR and CC studies—the choices of p and r, along with other design parameters, are critical researcher decisions that can play a large role in determining the outcome. This further highlights the need for broad exploration of the parameter space when analyzing patterns of CR-CC-MX preferences.

This sort of broad exploration is largely absent from the prior literature. Most prior studies

have not focused on testing whether the CRE and CCE are robust to the choice of experimental parameters, in part because many do not have these effects as their main object of interest. Of the 39 CRE studies identified by Blavatskyy et al. (2023), 21 (53%) use only one or two parameter configurations, and the average number of parameter configurations per study is less than four. Of the 29 CCE studies identified by Blavatskyy et al. (2022), 18 (62%) use only one or two parameter configurations, and the average number of parameter configurations per study is less than three. The limited use of parameters within and across prior studies highlights a need for more scrutiny in establishing whether CRE and CCE reflect global features of preferences.[12]

2.4 Prior Literature on Mixture Preferences

Separate from the literature on the CRE and CCE, two distinct strands of work speak to mixture preferences. The first is a small literature from the 1980s through the early 2000s on direct tests of the betweenness axiom. Two papers provide nice reviews of this literature: Camerer and Ho (1994) summarize and contribute to the early literature on betweenness, and Blavatskyy (2006) provides an update. Both papers conclude that individuals frequently violate betweenness, but the evidence is mixed on the direction of violation, with some studies finding more mixture loving and others finding more mixture aversion. These tests of betweenness are almost never in the context of connected CR-CC-MX problems.

Second, there is an emerging literature on deliberate randomization (see, for example, Agranov and Ortoleva, 2017; Dwenger et al., 2018; Feldman and Rehbeck, 2022; Agranov et al., 2023; Agranov and Ortoleva, 2023). These studies typically present individuals with the same decision problem repeated multiple times, either interspersed throughout the study or explicitly repeated in a row. Evidence from these studies suggests that individuals often prefer to generate mixtures through randomization across task repetitions. However, this type of design does not allow for the measurement of *aversion* to randomization (i.e., RMXP). We discuss in Section 6 how our analysis relates to this literature.

We emphasize that our goal is not to investigate mixture preferences in isolation; if it were, connected problems would be unnecessary, and designs more directly connected to the work surveyed by Camerer and Ho (1994) and Blavatskyy (2006) would be better suited. Rather, our goal is to identify combinations of CR, CC, and MX preferences, providing the richer data required for a more holistic model assessment.

[12]There are a few recent papers that feature broader parameter coverage, including Kobayashi and Lucia (2023), Jain and Nielsen (2024), Lucia (2024), and McGranaghan et al. (2024).

2.5 Choices and Valuations

The prior literatures on the CRE, CCE, and MXE have relied almost exclusively on paired binary choice tasks of the form described in Section 2.1.[13] If there is decision noise, however, then even if responses to individual binary choice tasks are unbiased noisy reflections of underlying preferences, a comparison of responses across a pair of binary choice tasks can yield biased conclusions. McGranaghan et al. (2024) build on prior work to demonstrate this issue in the context of the CRE, and show how analogous paired *valuation tasks* can avoid the problem.

Motivated by McGranaghan et al. (2024), our primary analysis focuses on three valuation tasks:

$$\begin{array}{lllll} \textbf{\textit{AB}} \textbf{ Valuation Task:} & \text{state an } h_{AB} \text{ such that} & (M,1) & \sim & (h_{AB},p), \\ \textbf{\textit{AB}}' \textbf{ Valuation Task:} & \text{state an } h_{AB'} \text{ such that} & (M,1) & \sim & (h_{AB'},pr;M,1-r), \\ \textbf{\textit{CD}} \textbf{ Valuation Task:} & \text{state an } h_{CD} \text{ such that} & (M,r) & \sim & (h_{CD},pr). \end{array}$$

To see how such valuation tasks can yield unbiased conclusions, consider a person who reports valuations with noise. Specifically, for a fixed (p, r, M), the person has underlying indifference values h^*_{AB}, $h^*_{AB'}$, and h^*_{CD} as defined in Section 2.2, but reports valuations $h_{AB} = h^*_{AB} + \varepsilon_{AB}$, $h_{AB'} = h^*_{AB'} + \varepsilon_{AB'}$, and $h_{CD} = h^*_{CD} + \varepsilon_{CD}$, where ε_{AB}, $\varepsilon_{AB'}$, and ε_{CD} are random variables that reflect noise. The person's empirically measured preferences are then

$$\begin{aligned} \Delta_{CR} &\equiv h_{AB} - h_{CD} &&= \Delta^*_{CR} + \varepsilon_{AB} - \varepsilon_{CD}, \\ \Delta_{CC} &\equiv h_{AB'} - h_{CD} &&= \Delta^*_{CC} + \varepsilon_{AB'} - \varepsilon_{CD}, \\ \Delta_{MX} &\equiv h_{AB} - h_{AB'} &&= \Delta^*_{MX} + \varepsilon_{AB} - \varepsilon_{AB'}. \end{aligned}$$

As long as $E[\varepsilon_{AB}] = E[\varepsilon_{AB'}] = E[\varepsilon_{CD}] = 0$, Δ_{CR}, Δ_{CC}, and Δ_{MX} are unbiased measures of the three objects of interest Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX}.[14]

In contrast to comparing behavior across valuation tasks, comparing behavior across binary choice tasks can lead to biased inference. To illustrate, consider an EU maximizer with no CRP who faces a CR paired choice task. If the parameters of the problem are such that EU preferences favor A and C (e.g., for small H), then with unbiased choice noise, they will choose both A and C with probabilities greater than 1/2. However, if the noise has a differential impact on the two choice tasks, then the two choice probabilities will not be the same. In particular, if the noise has a larger impact on the CD choice task, then we would predict $1 > \Pr(A|AB) > \Pr(C|CD) > 1/2$; that is,

[13] All 224 experiments from the two meta-studies by Blavatskyy et al. (2022) and Blavatskyy et al. (2023) use paired choices. In McGranaghan et al. (2024), we identified only 10 CRE experiments across four studies that instead use valuations.

[14] In McGranaghan et al. (2024), we also demonstrate that for some formulations of non-mean-zero noise—in particular, noise expressed in terms of utility—one can still test the null of $\Delta^*_{CR} = 0$ using a sign test. In Online Appendix C.1, we develop an analogous argument in the context of connected CR-CC-MX problems, and we conduct both means and sign tests in Section 4.

we would predict $CRE - RCRE > 0$.[15] An analogous argument applies when the parameters are such that EU preferences favor B and D (e.g., for large H): If the noise has a larger impact on the CD choice task, then we would predict $1/2 > \Pr(C|CD) > \Pr(A|AB) > 0$; that is, we would predict $CRE - RCRE < 0$. Hence, the bias can be upward or downward depending on the choice of the experimental parameter H. Further, aggregating across heterogeneous EU individuals with different preferred alternatives can easily generate aggregate choice proportions $Pr(A|AB) > 1/2 > Pr(C|CD)$ or the reverse.

In McGranaghan et al. (2024), we formalize these arguments in more detail in the context of CR problems. In Online Appendix C.2, we provide the analogous formal argument for the three tasks that constitute a connected CR-CC-MX problem. We further demonstrate that the potential for misleading conclusions is even greater when attempting to identify preference patterns by comparing behavior across three binary choices.

Given the shortcomings of comparing behavior across binary choice tasks, our primary analysis focuses on identifying patterns of Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX} using valuation tasks. An additional benefit of valuation tasks is that Δ_{CR}, Δ_{CC}, and Δ_{MX} provide direct measures of the *magnitudes* of their corresponding underlying preferences Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX}. In contrast, with binary choice tasks, we only observe $\widehat{\Pr}(A|AB) - \widehat{\Pr}(C|CD)$, $\widehat{\Pr}(A|AB') - \widehat{\Pr}(C|CD)$, and $\widehat{\Pr}(A|AB) - \widehat{\Pr}(A|AB')$. Even if the signs of these differences were unbiased, their magnitudes have no clear interpretation because they represent a combination of the magnitude of the underlying Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX} and the magnitude of the differential noise.

While our discussion above (and in McGranaghan et al., 2024) assumes that individual (as opposed to paired) valuation and choice tasks both provide unbiased measures of underlying preferences, some researchers (e.g., Brown and Healy, 2018 and Freeman et al., 2019) have argued that individual valuation tasks might themselves be biased—for example, due to menu effects that draw individuals toward the middle of a multiple price list.[16] Given this potential concern, our experiment also conducts binary choice tasks that are linked to the valuation tasks. In fact, the analysis in McGranaghan et al. (2024) points to an approach for mitigating the potential bias when comparing behavior across choice tasks: select multiple payment values—in this case, multiple values for H—such that upward and downward biases roughly offset. Hence, in our paired binary choice tasks, we intentionally select values of the high payment H based on pilot data to accomplish this balance. Given this approach, the main empirical patterns that we document below are the same for both our valuation and choice data.

[15] We use $\Pr(A|AB)$ to denote the theoretical prediction for the observed $\widehat{\Pr}(A|AB)$.

[16] In our design, each of our three key measures Δ_{CR}, Δ_{CC}, and Δ_{MX} represents a *difference* in responses across two valuation tasks with the same menu. Hence, even if menu effects distort valuation levels, they may not distort valuation differences.

3 Experimental Methodology

We study connected CR-CC-MX problems using both paired valuation tasks and paired choice tasks at 20 different (p, r) combinations.[17] We consider four possible values of $p \in \{0.3, 0.5, 0.8, 0.9\}$ and five possible values of $r \in \{0.1, 0.2, 0.3, 0.5, 0.8\}$. These values cover a wide range of (p, r) configurations while still including some of the most popular parameter choices in the prior literature. Open circles in Panel B of Figure 1 indicate our choice of (p, r) combinations.

Figure 2 provides an overview of the experiment timeline. The experiment consists of two stages. In stage 1, we randomly assign each participant to four of our 20 (p, r) combinations, and each participant then completes 20 valuation tasks. Each valuation task fixes M and uses a multiple price list to elicit the value of H that makes the participant indifferent between two lotteries. In stage 2, the same participant completes 24 binary choice tasks for the same four (p, r) combinations as in stage 1. Participants complete all 20 valuations before the 24 binary choices, and we randomize the order of questions within each stage.

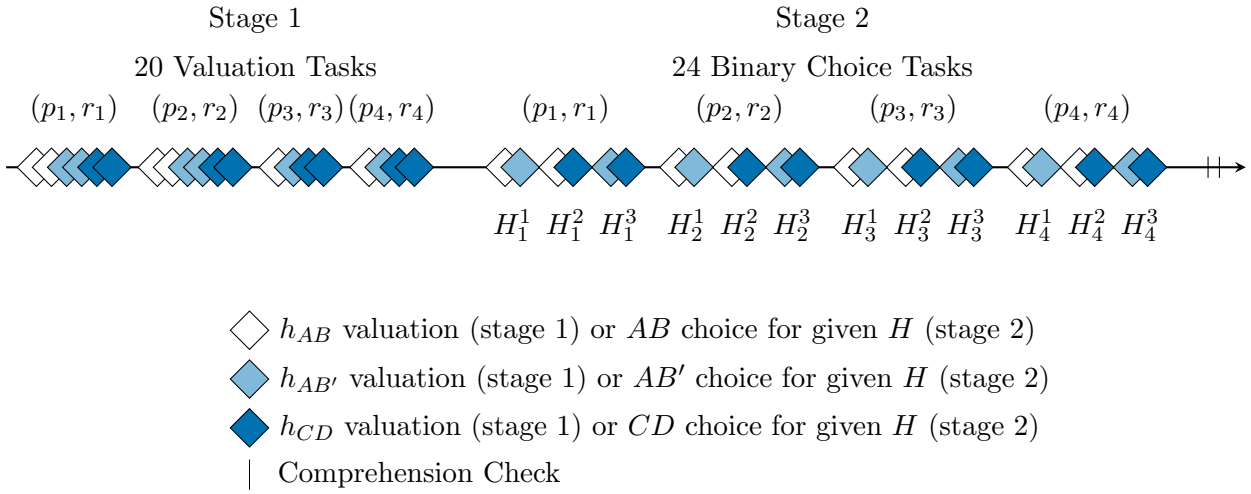


Figure 2: Experiment Timeline

3.1 Stage 1: Valuation Tasks

In stage 1, each participant completes 20 valuation tasks across four randomly drawn combinations of (p, r). We draw the values of p without replacement so that each participant sees all four possible values of p. Conversely, we draw the values of r with replacement, so a participant might face the same r across multiple values of p.[18]

[17] We conducted the experiment in October 2022 and preregistered it in the AEA RCT Registry under the ID AEARCTR-0009974 prior to data collection.

[18] We draw p without replacement to avoid duplicate elicitations of the same AB task, since the AB task does not depend on r.

For each (p, r) combination, we elicit the following indifference points:

$$\begin{aligned} h_{AB} \quad &\text{such that} : (M, 1) \sim (h_{AB}, p) \\ h_{AB'} \quad &\text{such that} : (M, 1) \sim (h_{AB'}, pr; M, 1-r) \\ h_{CD} \quad &\text{such that} : (M, r) \sim (h_{CD}, pr) \end{aligned}$$

We fix $M = \$(p \cdot 30)$ and elicit the relevant valuations using a multiple price list.[19] The left-hand option in each price list is fixed at either lottery $(M, 1)$ or (M, r). The right-hand option is either lottery (H, p), $(H, pr; M, 1-r)$, or (H, pr), with H varying in \$1 increments from $\$M$ to $\$(M + 50)$. For example, for $p = 0.3$, we fix $M = \$9$ and vary H from \$9 to \$59. We take the average value of H at the switching rows as our measure of the indifference valuation.[20]

For each price list, we enforce a single switching point. Clicking a row in the left column selects the left-hand option in that row and all rows above; clicking a row in the right column selects the right-hand option in that row and all rows below. Participants can adjust their selections as much as they want before submitting their final choices for that valuation task. Online Appendix Figures H.1 to H.3 provide example screenshots of all three variants of the valuation task.

For two randomly drawn (p, r) combinations that a given participant sees, we elicit each valuation twice, generating observations $\{h_{AB}, h'_{AB}, h_{AB'}, h'_{AB'}, h_{CD}, h'_{CD}\}$ where the repeat elicitations are denoted by h'. For the remaining two (p, r) combinations, we only elicit the CD valuation twice, generating observations $\{h_{AB}, h_{AB'}, h_{CD}, h'_{CD}\}$.[21] Therefore, we have duplicate observations for all CD valuations but only half of AB and AB' valuations. We discuss the purpose of collecting multiple elicitations of the same valuation in Section 4. In total, participants complete 20 valuation tasks: six valuations each for two (p, r) combinations and four valuations each for the remaining two (p, r) combinations. We randomize the order of valuation tasks subject to the constraint that repeat elicitations of the same valuation are separated from each other by at least three other valuation tasks.

[19] We set M in this way for practical reasons. We did not want M to be too large when $p = 0.3$ to avoid very large H values at indifference. We also did not want M to be too small when $p = 0.9$ to avoid being unable to detect preference variations given our \$1 increments in the price list. Our approach means that a risk neutral person would choose $H = \$30$ for all price lists.

[20] In prior work, we also implement m-valuation tasks in which we hold H fixed and elicit the M that makes participants indifferent between two lotteries (McGranaghan et al., 2024). For AB' valuations, varying M would lead to changes in both columns of the price list, which adds a layer of complexity to the valuation process. For this reason, we focus on h-valuation tasks in this paper.

[21] For each participant and (p, r) pair, we randomly label one of the two elicitations h_{XY} and the other h'_{XY} for $XY \in \{AB, AB', CD\}$. In other words, the superscript prime does not indicate the order in which the valuations were elicited.

3.2 Stage 2: Paired Choice Tasks

In stage 2, each participant completes 24 binary choices across the same four (p, r) combinations they saw in stage 1. These 24 binary choices comprise 12 paired choice tasks, three for each (p, r) combination: a CR pair (an AB choice and a CD choice), a CC pair (an AB' choice and a CD choice), and an MX pair (an AB choice and an AB' choice). Online Appendix Figures H.4 to H.6 present examples of each type of binary choice task.

Each binary choice that a participant sees in stage 2 is equivalent to a single row from one of the price lists they saw in stage 1. Specifically, for each (p, r) combination, we again fix $M = \$(p \cdot 30)$. For each of the three types of paired choice tasks, we then draw a random value of H from the relevant row in Table 3 without replacement. Hence, for a given (p, r) combination, we conduct three paired choice tasks (CR, CC, or MX) with a different value of H for each pair (but with the same H for both choices within a given pair). We randomize the order in which these choices appear, with no restrictions on the order. The values of H in Table 3 were selected based on pilot data to roughly balance the number of cases with upward versus downward bias (see our discussion in Section 2.5). Specifically, the goal was to have three smaller values intended to produce a majority that prefer A over B, A over B', and C over D, and three larger values where the majority exhibits the opposite pattern.

Table 3: Stage 2 H Parameter Values by p

	(1)	(2)	(3)	(4)	(5)	(6)
$p = 0.9$	31	36	41	46	51	56
$p = 0.8$	29	34	39	44	49	54
$p = 0.5$	28	33	38	43	48	53
$p = 0.3$	27	32	37	42	47	52

Notes: Table presents values of H used in stage 2 for each p. For each (p, r) that a participant saw, they are presented with three paired choice tasks (one CR, one CC, and one MX); we randomly selected three of the six H values for the relevant p, and assigned one to each of these paired choice tasks.

It is useful to benchmark the scale of our stage 2 paired choice tasks relative to the prior CR and CC literature. We can view stage 2 as conducting many paired-choice "experiments," where one experiment consists of responses from multiple participants for a fixed set of parameters (p, r, H). With 20 (p, r) combinations and six values of H for each, we effectively run 120 CR experiments, 120 CC experiments, and 120 MX experiments, with a total of 8,408 observations for each type of experiment. In comparison, the two meta-studies discussed in Section 2.3 report 143 CR experiments with 14,909 total observations and 81 CC experiments with 8,947 observations.

3.3 Additional Design Details

Before stage 1, participants complete an unincentivized attention check and a quiz about the payment mechanism. After completing stage 2, they complete two incentivized comprehension checks to gauge their understanding of the multiple price list format and the binary choice tasks. The first tests whether individuals can correctly fill out a price list given a specified indifference value. The second tests whether participants can correctly answer a binary choice question when given another person's responses to a multiple price list. Online Appendix Figures H.7 and H.8 provide example screenshots of these comprehension checks.[22] Finally, to break up the tasks and reduce fatigue, we present participants with an unincentivized visual puzzle after every fifth question in both experiment stages. Online Appendix Figure H.9 provides an example of one of these puzzles.

3.4 Recruiting

We recruited 2,102 participants through Prolific who met the following eligibility criteria: at least a high school education; between the ages of 18 and 31; living in the United States or Western Europe; minimum Prolific approval rating of 99 percent; fluent in English; and 50 to 1,000 previous Prolific submissions. We targeted this group of participants to approximate the typical undergraduate sample recruited in prior CC and CR studies. We also recruited a gender-balanced sample with an even split of male and female participants. Online Appendix Table A.1 presents summary statistics for our sample.

We paid each participant a fixed $5 fee for completing the experiment. In addition, we randomly selected one in five participants to receive a bonus based on their responses. Specifically, we randomly chose one of their 46 decisions (20 valuations, 24 choices, and two incentivized comprehension checks) to be the decision that counts. If the decision that counts was a valuation task, then we randomly selected one row of the price list and paid the participant according to their choice in that row. If the decision that counts was a binary choice task, then we paid the participant based on the option they selected. If the decision that counts was a comprehension-check question, then we paid the participant $5 if they answered correctly. The average completion time was 27.3 minutes, and the average bonus payment for selected participants was $15.76.

[22] For the first comprehension check, 85 percent of participants answer correctly, and for the second, 79 percent answer correctly. See Online Appendix Table A.1. The qualitative patterns in Figure 3 are unchanged if we restrict the sample to those who pass both comprehension checks.

4 Results

4.1 Aggregate Behavior in Valuation Tasks

We first study aggregate behavior across all participants using the AB, AB', and CD valuations collected in stage 1 to measure CR, CC, and MX preferences—that is, to measure Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX}.[23]

As described in Section 3.1, we collect some valuations twice and randomly label them h_{XY} and h'_{XY} for $XY \in \{AB, AB', CD\}$.[24] We collect repeat valuations for three main reasons. First, we use the repeat valuations to get an initial assessment of the reliability of the stage 1 data. We compute the correlation between repeated elicitations of the same valuation—e.g., h_{CD} and h'_{CD}—for each (p, r). Online Appendix Table A.3 reports these correlations; all are positive, ranging from 0.389 to 0.731. Given that the repeat elicitations are separated from each other by at least three other tasks, these strong correlations suggest that our valuation data capture meaningful information about the underlying preferences of interest.

Second, we use the repeat valuations to construct independent measures of our three main outcomes of interest. Specifically, we define the following measures:

$\Delta_{CR} \equiv h_{AB} - h_{CD}$, a (noisy) measure of an individual's CRP, Δ^*_{CR},

$\Delta_{CC} \equiv h_{AB'} - h'_{CD}$, a (noisy) measure of an individual's CCP, Δ^*_{CC}, and

$\Delta_{MX} \equiv h'_{AB} - h'_{AB'}$, a (noisy) measure of an individual's MXP, Δ^*_{MX}.

If we had only a single elicitation for each valuation, then measurement error in the valuations would induce mechanical correlations between Δ_{CR}, Δ_{CC}, and Δ_{MX}. For example, using the same measure h_{CD} to construct both Δ_{CR} and Δ_{CC} would create a mechanical positive correlation between Δ_{CR} and Δ_{CC} if there is any measurement error in h_{CD}. By eliciting valuations twice and using independent observations to construct our difference measures, we avoid this problem.[25]

Third, we use the repeat valuations to disentangle noise from preferences, which we describe in detail in Section 4.4.

To investigate aggregate behavior, we analyze the mean values of Δ_{CR}, Δ_{CC}, and Δ_{MX} across all participants for each (p, r) combination. Figure 3 presents these quantities by p, separately for

[23] We pre-registered the construction of our experimental outcomes and the analyses that we describe throughout Section 4 in our pre-analysis plan. We note any deviations and additional exploratory analyses.

[24] Online Appendix Table A.2 reports the means for all six valuations $(h_{AB}, h'_{AB}, h_{AB'}, h'_{AB'}, h_{CD}, h'_{CD})$ for each (p, r) combination.

[25] Recall that for two randomly drawn (p, r) combinations that a participant sees, we elicit each valuation twice, generating observations $\{h_{AB}, h'_{AB}, h_{AB'}, h'_{AB'}, h_{CD}, h'_{CD}\}$ and allowing us to construct independent measures of each of the three main objects of study. For the remaining two (p, r) combinations, we only elicit the CD valuation twice, generating observations $\{h_{AB}, h_{AB'}, h_{CD}, h'_{CD}\}$ and allowing us to construct independent measures of Δ_{CR} and Δ_{CC} but not Δ_{MX}. This means that we have twice as many observations for Δ_{CR} and Δ_{CC} than for Δ_{MX}.

each r. EU predicts all of these values to be zero, while many non-EU models predict a systematic $\Delta_{CR} > 0$ and $\Delta_{CC} > 0$. Both of these general predictions are markedly inconsistent with the data, which show clear distinct patterns for Δ_{CR}, Δ_{CC}, and Δ_{MX} across the parameter space. Panel A summarizes CR preferences: Δ_{CR} is largely invariant to changes in p but declines sharply in r, leading to $\Delta_{CR} \approx 0$ at higher r values. Panel B summarizes CC preferences: In contrast to Δ_{CR}, Δ_{CC} is largely invariant to changes in r but is increasing sharply in p, with $\Delta_{CC} < 0$ at lower p values. The structural connection between CR and CC preferences—and their distinct patterns throughout the parameter space—implies that people cannot have a neutral attitude to probabilistic mixtures. This again emerges as a distinct systematic pattern in Δ_{MX} shown in Panel C: Mean Δ_{MX} is positive for most parameters and becomes substantially larger as p and r fall. Online Appendix Tables A.4, A.5, and A.6 provide information on the statistical significance of the patterns in Figure 3 through mean and sign tests for the EU null of zero along with regression results for how Δ_{CR}, Δ_{CC}, and Δ_{MX} change with p and r.[26] These tables indicate that quantities in excess of roughly 1.5 in absolute value in Figure 3 are statistically different from zero and that the noted sensitivities to p and r are statistically significant.[27]

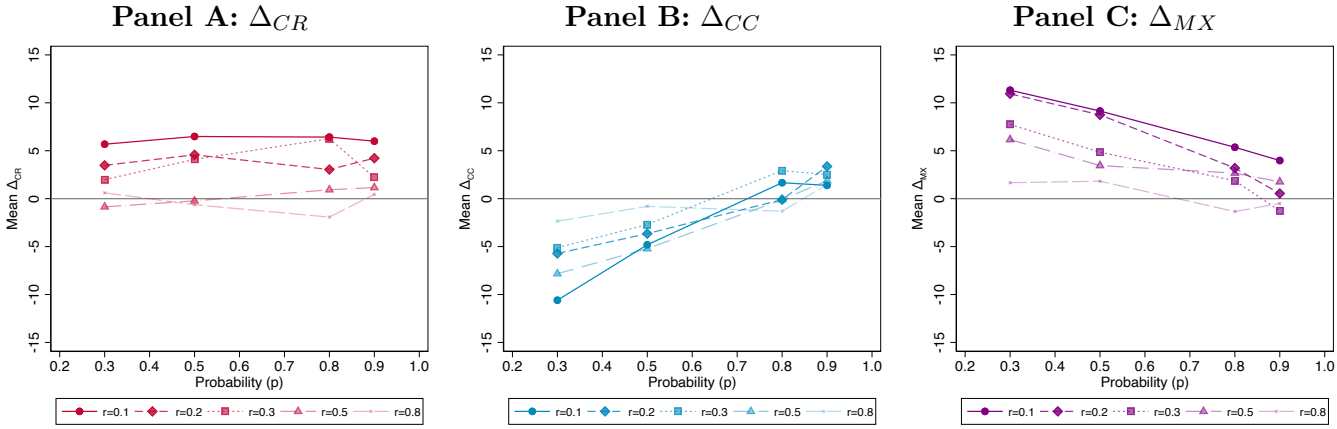
Looking across the three panels in Figure 3, two striking regularities emerge. First, near the canonical parameterizations for CR and CC problems (i.e., at $p = 0.8$ or 0.9 and $r = 0.1$, 0.2, or 0.3), we find both $\Delta_{CR} > 0$ and $\Delta_{CC} > 0$. Thus, CRP and CCP appear at canonical parameter values, consistent with the findings from the paired-choice paradigm. At the same time, Δ_{MX} tends toward zero at these values such that the generally positive attitude toward mixtures seen in Panel C would be missed by focusing exclusively on the canonical parameterizations. Second, for small p and small r, we find a robust pattern of $\Delta_{CR} > 0$, $\Delta_{CC} < 0$, and $\Delta_{MX} > 0$, which contrasts sharply with the predictions of the leading parametrized non-EU models in Table 1.[28]

[26] In McGranaghan et al. (2024), we similarly tested for the existence of CRP using paired valuation tasks. While most of McGranaghan et al. (2024) presents results based on choose-m tasks in which participants report the middle value that makes them indifferent between Options A and B and C and D respectively, we also collected choose-h tasks for validation and robustness that are more directly comparable to the measures in this paper. McGranaghan et al. (2024) explored a smaller range of r ($r \in \{0.2, 0.4, 0.6\}$ compared to $r \in \{0.1, 0.2, 0.3, 0.5, 0.8\}$ in this paper) and also captures some lower values of p ($p \in \{0.1, 0.2, 0.5, 0.8, 0.9\}$ compared to $p \in \{0.3, 0.5, 0.8, 0.9\}$ in this paper). In both cases, we find no evidence of a systematic CRP as the direction and magnitude of the effects depend on experimental parameter choices, especially the choice of r. Relative to McGranaghan et al. (2024), the parameters considered in this paper yield slightly more instances of CRP, particularly for cases in which $r = 0.1$. This underscores the role low values of r play in generating evidence of CRP.

[27] An alternative, though less direct, measure of MXP is $\Delta_{CR} - \Delta_{CC}$. Online Appendix Tables A.4 and A.5 also provide results for this measure, which are qualitatively and quantitatively similar to the results for our direct measure Δ_{MX}.

[28] The aggregate patterns for Δ_{CR}, Δ_{CC}, and Δ_{MX} in Figure 3 are qualitatively reproduced when we construct corresponding mean measures using only each participant's first observation. Though noisier, similar sensitivities to problem parameters remain along with frequent observations of $\Delta_{CR} > 0$, $\Delta_{CC} < 0$, and $\Delta_{MX} > 0$.

Figure 3: Stage 1 Valuations: Mean Δ_{CR}, Δ_{CC}, and Δ_{MX} by p for each r



Notes: Figure depicts mean values of $\Delta_{CR} = h_{AB} - h_{CD}$, $\Delta_{CC} = h_{AB'} - h'_{CD}$, and $\Delta_{MX} = h'_{AB} - h'_{AB'}$. Panels A and B aggregate over 8,408 observations, with each point corresponding to roughly 420 observations. Panel C aggregates over the 4,204 observations for which we elicit h'_{AB} and $h'_{AB'}$, with each point corresponding to roughly 210 observations. Expected utility predicts $\Delta = 0$ in all three panels.

4.2 Aggregate Behavior in Choice Tasks

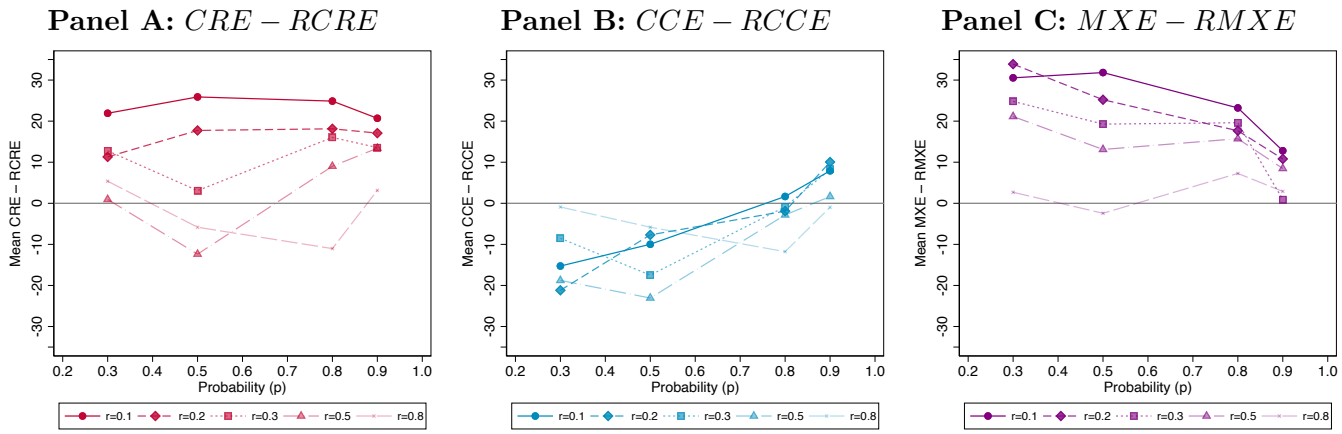
As a secondary analysis, we examine aggregate behavior using stage 2 choices. For each (p, r) combination, we pool data across all payment values for H to generate the measures $CRE - RCRE$, $CCE - RCCE$, and $MXE - RMXE$. The first two measures are defined in Section 2.3, and the third is defined analogously.

Figure 4 is the analog of Figure 3 when we use choices instead of valuations. The same qualitative conclusions emerge for our three main preferences of interest: The CR pattern is largely invariant to p, with a CRE at small r but not at large r; the CC pattern is largely invariant to r, with a mild CCE at large p but a strong $RCCE$ at small p; and the MX pattern depends on both p and r, with a general MXE that is substantially stronger for smaller p and smaller r. Online Appendix Table A.9 is analogous to Table A.5 and confirms the statistical significance of these patterns.

Although our stage 2 choices yield the same qualitative message as our stage 1 valuations, Section 2.5 highlights a drawback of using binary choice tasks: When comparing behavior across such tasks, conclusions can be biased if noise has a differential impact across decisions. In McGranaghan et al. (2024), we demonstrate the existence of differential noise within the context of the CRE, and in particular, that noise has a larger impact on the CD decision than on the AB decision. In Online Appendix C.3, we conduct a similar analysis using the data from the current experiment. We find evidence of differential noise in all three problems. For CR problems, we again find that noise has a larger impact on the CD decision than the AB decision. For CC problems, we find that noise has a larger impact on the AB' decision than the CD decision. For the MX problems, we find that noise has a larger impact on the AB' decision than the AB decision.[29] Given that we find evidence

[29]Interestingly, these conclusions differ from the predictions of EU with additive i.i.d utility noise, which predicts

Figure 4: Stage 2 Choices: Mean $CRE - RCRE$, $CCE - RCCE$, $MXE - RMXE$ by p for each r



Notes: Figure depicts average values of $CRE - RCRE$, $CCE - RCCE$, and $MXE - RMXE$. Each panel aggregates over 8,408 observations. Each point in each panel aggregates over the six payment values for H noted in Table 3 and corresponds to approximately 420 observations. The line for zero choice difference is indicated in all three panels but only serves as a reference.

of differential noise for all three problems, it may seem surprising that our choice data yield the same aggregate conclusions as our valuations data. However, as noted in Section 2.5, this successful replication stems from our deliberate selection of experimental parameters to induce offsetting biases in stage 2; specifically, it reflects our choice of a balanced set of H payment values at each (p, r, M).

4.3 Heterogeneity in Behavioral Patterns

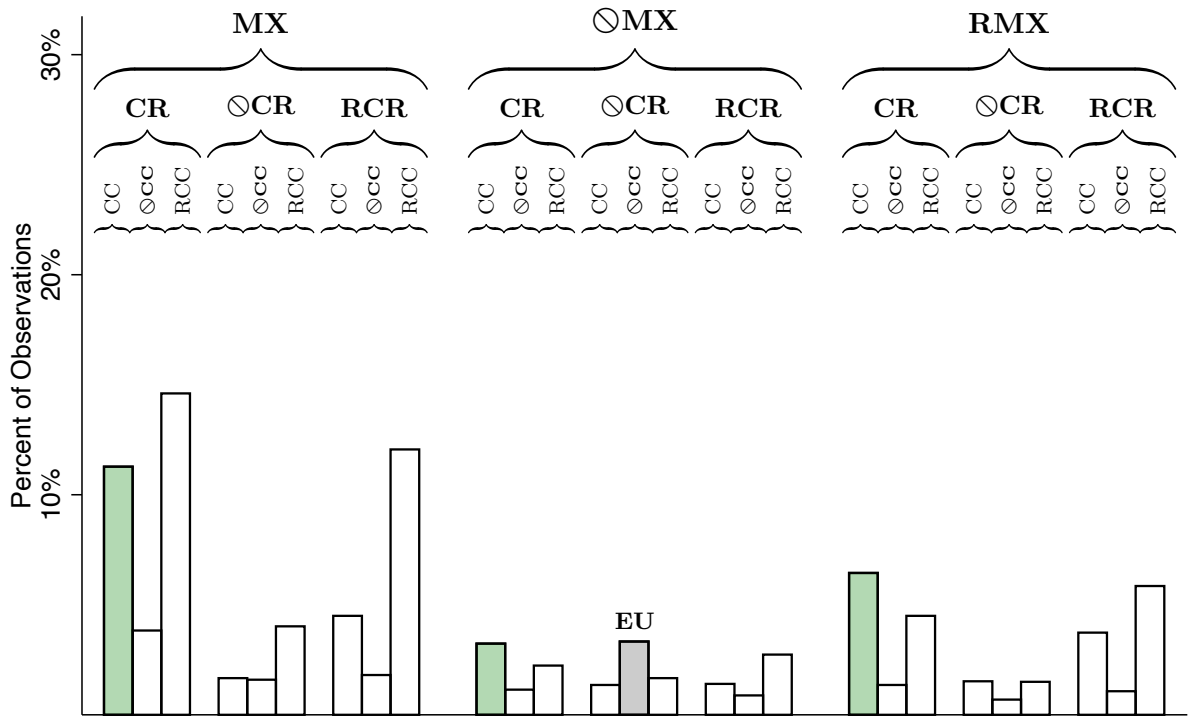
Whereas Figures 3 and 4 characterize mean responses, it is also important to investigate behavior at the individual level.[30] To relate our data to the directional model predictions in Table 1, we focus on participants' directional responses in valuation tasks, which map directly to those predictions. For example, we create indicators for whether a person exhibits $\Delta_{CR} > 0$, $\Delta_{CR} = 0$ (denoted $\oslash$CR), or $\Delta_{CR} < 0$ in their valuations for CR problems. With three possible directional responses for each of the three measures of interest, there are 27 possible combinations. Figure 5 presents a histogram across these 27 combinations using the 4,204 observations for which we have independent measures of Δ_{CR}, Δ_{CC}, and Δ_{MX}. Note that the variation in Figure 5 is across (p, r) and both across and within participants.

Figure 5 reveals that the mean behavior in Figures 3 and 4 masks substantial heterogeneity. Patterns that are consistent with the standard behavioral hypothesis of $\Delta_{CR} > 0$ and $\Delta_{CC} > 0$ are highlighted in green. Such observations constitute only a small fraction of the observed patterns (21.0 percent). As foreshadowed by Figures 3 and 4, the most frequent single pattern is $\Delta_{CR} > 0$, $\Delta_{CC} < 0$,

that noise has a smaller impact on the AB choices than either the AB' or CD choices, with equal impact on the latter two. See Online Appendix C.3 for more details.

[30] We did not pre-register any specific hypotheses or analyses with regards to these individual preference patterns. This heterogeneity analysis is therefore exploratory.

Figure 5: Histogram of Response Patterns



Notes: Figure presents histogram of $(sign(\Delta_{CR}), sign(\Delta_{CC}), sign(\Delta_{MX}))$ combinations, where $\Delta_{CR} = h_{AB} - h_{CD}$, $\Delta_{CC} = h_{AB'} - h'_{CD}$, and $\Delta_{MX} = h'_{AB} - h'_{AB'}$. The histogram covers the 4,204 observations for which we elicit h'_{AB} and $h'_{AB'}$, with each participant contributing two observations. Each variable can have three potential signs, leading to 27 possible patterns. These signs correspond to the named patterns (e.g., CR to $\Delta_{CR} > 0$, RCR to $\Delta_{CR} < 0$, and ⊘CR to $\Delta_{CR} = 0$). Patterns marked in light green are ones with both $\Delta_{CR} > 0$ and $\Delta_{CC} > 0$, a frequent prediction of leading parametrized behavioral models (see Table 1).

and $\Delta_{MX} > 0$, though even this modal pattern accounts for only 14.6 percent of observations.[31]

To the extent that some of the variability in responses reflects heterogeneous preferences, it will be important for models of risk preferences to accommodate this variability. Before reaching that conclusion, however, we must assess the extent to which the variability in responses reflects heterogeneous preferences versus noise.

4.4 Decomposing Variability in Responses into Preference and Noise

In this section, we estimate the population distribution of underlying preferences and the magnitude of decision noise.[32] We then use these estimates for two purposes. First, we assess how much of the

[31] Online Appendix Figures A.1 to A.3 depict these distributions for various (p, r) subsets. Near the canonical parameterizations ($p = 0.8$ or 0.9 and $r = 0.1$, 0.2, or 0.3), the combination $\Delta_{CR} > 0$ and $\Delta_{CC} > 0$ is indeed more prevalent (28.9 percent). Nonetheless, there is still substantial variability, and the combination $\Delta_{CR} > 0$, $\Delta_{CC} < 0$, and $\Delta_{MX} > 0$ still accounts for 13.9 percent of observations. Alternatively, for parametrizations where Figure 3 suggests the $\Delta_{CR} > 0$, $\Delta_{CC} < 0$, and $\Delta_{MX} > 0$ combination should be strongest ($p = 0.3$ or 0.5 and $r = 0.1$, 0.2, or 0.3), that combination constitutes 21.4 percent. However, there is substantial variability at all problem parameters.

[32] Our pre-registration outlined a potential exploratory exercise to recover the distributions of preferences incorporating both valuations and choices. In this exercise we focus on only valuation data and discuss the connection between choices and valuations in Section 4.5.

variability in our data is due to preference heterogeneity versus noise. Second, we derive what the distribution of preference patterns from Figure 5 would look like in the absence of noise.

To disentangle noise from preferences, we leverage the fact that we collect repeat measures of participants' valuations. Intuitively, if the noise is independent across repeat elicitations, then the covariance between the two elicitations reflects only heterogeneity in preferences. We provide an overview of our approach here. Online Appendix D provides a more complete description of our decomposition.

We begin with some notation for the population distribution of the underlying indifference values for a fixed (p, r, M). We define the expectation $E(h^*_{AB}, h^*_{AB'}, h^*_{CD}) \equiv (\mu^*_{AB}, \mu^*_{AB'}, \mu^*_{CD})$, and the variance-covariance matrix

$$\mathbf{V}\begin{pmatrix} h^*_{AB} \\ h^*_{AB'} \\ h^*_{CD} \end{pmatrix} \equiv \begin{pmatrix} \theta^2_{AB} & \theta_{AB,AB'} & \theta_{AB,CD} \\ \theta_{AB,AB'} & \theta^2_{AB'} & \theta_{AB',CD} \\ \theta_{AB,CD} & \theta_{AB',CD} & \theta^2_{CD} \end{pmatrix}. \tag{1}$$

For $XY \in \{AB, AB', CD\}$, we assume an individual's repeated XY valuations are

$$h_{XY} = h^*_{XY} + \varepsilon_{XY} \qquad \text{and} \qquad h'_{XY} = h^*_{XY} + \varepsilon'_{XY},$$

where $E(\varepsilon_{XY}) = E(\varepsilon'_{XY}) = 0$, $var(\varepsilon_{XY}) = var(\varepsilon'_{XY}) = \sigma^2_{XY}$, and ε_{XY} and ε'_{XY} are independent of each other, underlying preferences, and all other noise draws.

Under these assumptions, we can derive theoretical predictions for the empirical moments:

$$\begin{aligned} var(h_{XY}) &= \theta^2_{XY} + \sigma^2_{XY}, \\ cov(h_{XY}, h'_{XY}) &= \theta^2_{XY}, \text{ and} \\ cov(h_{XY}, h_{WZ}) &= \theta_{XY,WZ}. \end{aligned}$$

The first equation highlights that with only a single elicitation of an XY valuation, we could not decompose its variance into the separate preference and noise components. The second equation captures our central intuition: If the noise draws for the two elicitations are independent, then their covariance identifies the variance of the preference heterogeneity.[33]

[33] If, instead, ε_{XY} and ε'_{XY} were not independent from each other (but still independent from preferences), then $cov(h_{XY}, h'_{XY}) = \theta^2_{XY} + cov(\varepsilon_{XY}, \varepsilon'_{XY})$. In such a case, the covariance between the two elicitations provides a biased estimated of preference heterogeneity. Though the bias is theoretically unsigned, it would be natural to assume $cov(\varepsilon_{XY}, \varepsilon'_{XY}) > 0$, which would lead us to overestimate the extent of preference heterogeneity, attributing to preferences behavioral heterogeneity that should be attributed to noise. If features of our elicitations led to ε_{XY} and ε'_{XY} being independent from each other but not independent from preferences, perhaps due to substantial boundary effects, then $cov(h_{XY}, h'_{XY}) = \theta^2_{XY} + cov(\varepsilon_{XY}, h^*_{XY}) + cov(\varepsilon'_{XY}, h^*_{XY})$, again leading to a biased estimate of preference heterogeneity. Though this bias is also unsigned, it would be natural to assume negative correlations between errors and preferences if boundary effects caused small values of h^*_{XY} to be inflated and large values to be deflated. This would lead us to underestimate the extent of preference heterogeneity. While we cannot assess the magnitude of such effects, one piece of suggestive evidence is that only approximately 10% of observations lie at the boundaries of our price lists.

Given this formulation, there are twelve model parameters to estimate for each (p, r): three μ^*_{XY} terms capturing the population average preference, three θ^2_{XY} terms capturing preference heterogeneity, three $\theta_{XY,WZ}$ terms capturing the covariance between preferences, and three σ^2_{XY} terms capturing the impact of noise. We derive estimates of these terms by calculating the relevant sample moments. For example, $\hat{\theta}^2_{AB} = cov_s(h_{AB}, h'_{AB})$ and $\hat{\sigma}^2_{AB} = var_s(h_{AB}) - cov_s(h_{AB}, h'_{AB})$, where cov_s and var_s denote a sample covariance or variance (see Online Appendix D.1 for details on how we calculate these moments given the structure of our data). Online Appendix Table A.7 reports our estimates of the 12 calculated terms for each (p, r) combination.[34]

Given these estimates, we can quantify how much of the variability in our data is due to heterogeneity in preferences versus noise. Consider first the variability in the indifference values h_{AB}, $h_{AB'}$, and h_{CD}. The last three columns of Online Appendix Table A.7 report the proportion of the variability for each indifference value that is due to preferences; that is, the ratio $\widehat{var}(h^*_{XY})/\widehat{var}(h_{XY}) = \hat{\theta}^2_{XY}/(\hat{\theta}^2_{XY} + \hat{\sigma}^2_{XY})$ for each $XY \in \{AB, AB', CD\}$. Averaging across the 20 (p, r) combinations, preference heterogeneity accounts for 61 percent of the variation in h_{AB}, 58 percent of the variation in $h_{AB'}$, and 48 percent of the variation in h_{CD}.

Next we consider variability in the preference measures Δ_{CR}, Δ_{CC}, and Δ_{MX}. For $\Delta_{CR} \equiv h_{AB} - h_{CD}$, it is straightforward to derive that

$$\begin{aligned} var(\Delta_{CR}) &= var(\Delta^*_{CR}) + \sigma^2_{AB} + \sigma^2_{CD}, \quad \text{and} \\ var(\Delta^*_{CR}) &= \theta^2_{AB} + \theta^2_{CD} - 2\theta_{AB,CD}, \end{aligned}$$

along with the analogous expressions for Δ_{CC} and Δ_{MX}. Online Appendix Table A.8 uses the estimates in Online Appendix Table A.7 to calculate these six variance terms for each (p, r) combination.[35] The last three columns of Online Appendix Table A.8 report the proportion of the variability for each preference measure that is due to preferences; that is, the ratio $\widehat{var}(\Delta^*_Z)/\widehat{var}(\Delta_Z)$ for each $Z \in \{CR, CC, MX\}$. Averaging across the 20 (p, r) combinations, preference heterogeneity accounts for 31 percent of the variation in Δ_{CR}, 30 percent of the variation in Δ_{CC}, and 25 percent of the variation in Δ_{MX}.

We next construct what the distribution of response patterns from Figure 5 would look like in the absence of decision noise. To do so, we make the additional assumption that the underlying preferences $(h^*_{AB}, h^*_{AB'}, h^*_{CD})$ have a joint normal distribution. For each (p, r) combination, we use the parameter estimates in Online Appendix Table A.7 to simulate 100,000 draws from this joint normal distribution and then convert each draw to a Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX}.[36] This approach allows

[34] Online Appendix D.5 describes a more sophisticated approach using maximum-likelihood estimation (MLE). Because that approach requires additional distributional and implementation assumptions, we prefer the moment-based approach in the text. However, the MLE approach yields virtually identical conclusions.

[35] When using this approach, nothing guarantees that the calculated $var(\Delta^*_Z) > 0$, and indeed there is one instance in which we estimate a negative variance term (for Δ_{MX} when $(p, r) = (0.3, 0.5)$). We exclude this case from our analysis.

[36] Specifically, we round each h^* draw to the midpoint of its two closest integers (e.g., any draw strictly between \$2

us to isolate the preference patterns underlying the observed response patterns depicted in Figure 5.

Figure 6 reproduces the distribution of response patterns in Figure 5 but adds black dots to denote the simulated distribution of preference patterns generated using the approach above. Note that any pattern that implies an intransitivity in the underlying indifference values $(h^*_{AB}, h^*_{AB'}, h^*_{CD})$ cannot emerge from this simulation.[37] Of the 27 possible response patterns, 14 imply an intransitivity. Neither our decomposed preferences nor any transitive theory of preferences can accommodate such patterns, which are marked in gray in Figure 6.[38]

Of the remaining 13 possible preference patterns, six are "strict" in the sense that they do not include a $\Delta^*_{CR} = 0$, $\Delta^*_{CC} = 0$, or $\Delta^*_{MX} = 0$, while the other seven are "weak" in the sense that they include a null preference. Our simulation permits both strict and weak patterns, and indeed, all 13 patterns have a positive share of simulated preferences. Figure 6 shows that some patterns arise more frequently in preferences than in the raw responses while the reverse is true for others. The four most prominent strict patterns (marked P1, P2, P3, P4, and indicated in blue) account for approximately 73 percent of simulated preferences, but only 44 percent of raw responses. In other words, our simulation shows that these patterns would be even more frequent in the absence of decision noise. The three most prominent weak patterns (marked P12, P23, and P34, to denote the strict preference patterns that they lie between and indicated in light blue) account for eight percent of simulated preferences and 11 percent of raw responses. These patterns would therefore be slightly less frequent in the absence of decision noise. Together, the seven marked patterns account for 81 percent of simulated preferences and 55 percent of raw responses, representing the majority of both. As we show in Section 5, these marked patterns emerge as the seven patterns permitted by the model of upside potential that we posit to capture the observed attraction to probabilistic mixtures between a certain payment and a binary lottery with a higher expected value.

4.5 Connecting Choices and Valuations

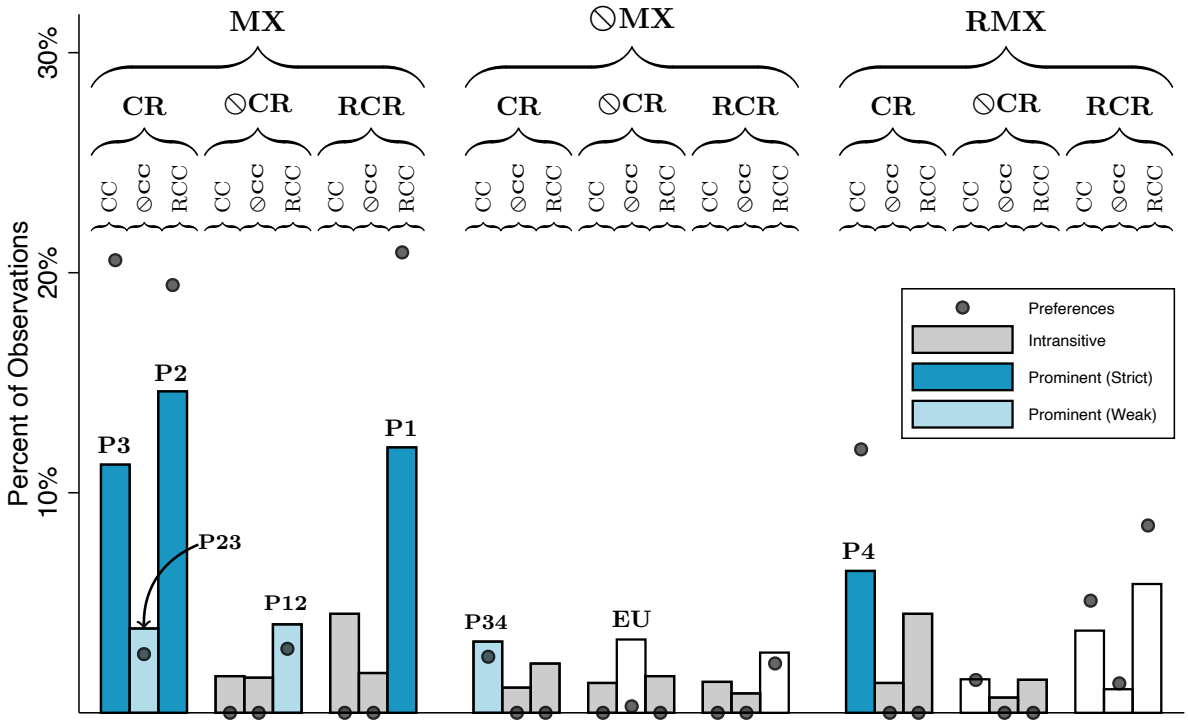
Before turning to our rationalization, we return to the connection between choices and valuations, but now we focus on connections at the *individual* level. Our design directly links each stage 2 binary choice to a corresponding stage 1 valuation. More precisely, if in stage 1 a participant provided valuations for a particular (p, r, M), then in stage 2 they made binary choices for a CR problem, a CC problem, and an MX problem for that same (p, r, M) and a randomly chosen payment value H.

and \$3 is converted to \$2.50). We then use these adjusted h^* terms to generate the Δ^* terms. This approach permits the simulated Δ^* terms to be zero. See Online Appendix D.3 for complete details.

[37] For example, pattern RCRP-CCP-MXP would require $h^*_{CD} > h^*_{AB}$, $h^*_{AB'} > h^*_{CD}$, and $h^*_{AB} > h^*_{AB'}$, which cannot occur from a single realization of $(h^*_{AB}, h^*_{AB'}, h^*_{CD})$.

[38] Such patterns are empirically possible in our experiment given that we use independent noisy measures of the three Δ terms. Moreover, the frequencies of such patterns give an indication of the prevalence of noise relative to preferences. In the extreme, if all patterns were equally likely, the majority (i.e., $14/27 = 52\%$) would have inconsistencies. Instead, we observe that intransitive patterns represent 26% of overall response patterns, averaging 1.9% of observations per pattern and exceeding 3% in only two cases.

Figure 6: Preference Patterns vs. Data Patterns

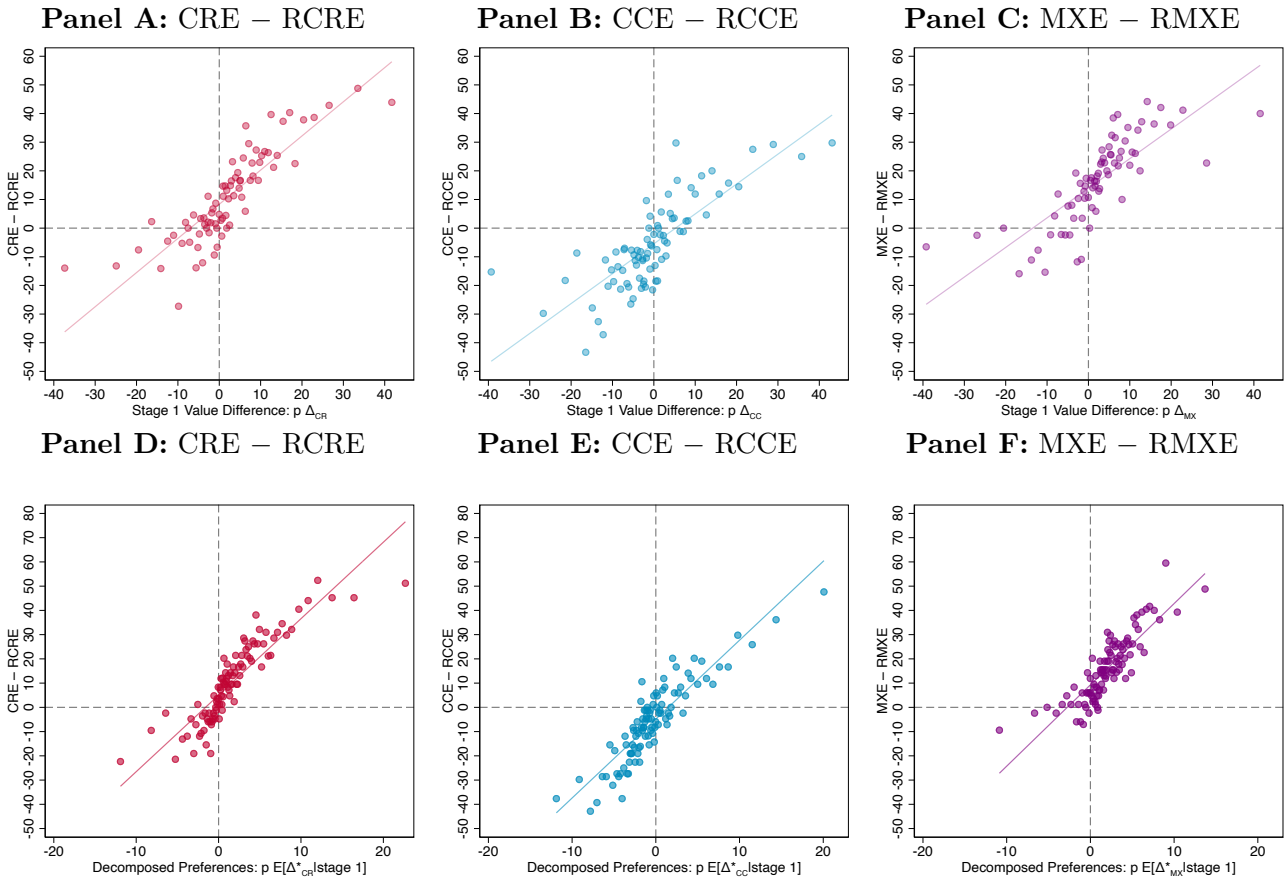


Note: Figure presents histograms of $(sign(\Delta_{CR}), sign(\Delta_{CC}), sign(\Delta_{MX}))$ combinations, where $\Delta_{CR} = h_{AB} - h_{CD}$, $\Delta_{CC} = h_{AB'} - h'_{CD}$, and $\Delta_{MX} = h'_{AB} - h'_{AB'}$, for the 4,204 observations for which we elicit h'_{AB} and $h'_{AB'}$. Each variable can have three potential signs, leading to 27 possible patterns. These signs correspond to the named patterns (e.g., CR to $\Delta_{CR} > 0$, RCR to $\Delta_{CR} < 0$, and $\oslash$CR to $\Delta_{CR} = 0$). Bars represent the share of raw responses and are identical to those in Figure 5. Black dots represent the share of simulated preferences following the decomposition and simulation described in Section 4.4. Patterns marked in gray denote patterns that would imply an intransitivity absent decision noise. Patterns marked in blue and denoted P1, P2, P3, and P4 are the four most frequent strict patterns. Patterns marked in light blue and denoted P12, P23, and P34 are weak patterns in between these frequent strict patterns. The strict and weak patterns labeled and marked in blue are those that are consistent with our theoretical development in Section 5.

In other words, each stage 2 binary choice corresponds to a specific row from a corresponding stage 1 price list.

Panels A-C of Figure 7 illustrate that there is a strong connection between participants' stage 1 valuations and their stage 2 choices. In Panel A, the horizontal axis reflects participants' stage 1 Δ_{CR} measures grouped into percentile bins, and the vertical axis denotes the average stage 2 $CRE - RCRE$ within that bin for the corresponding stage 2 binary choice tasks. Panels B and C repeat this exercise for Δ_{CC} and Δ_{MX}. Across panels, stage 1 preference measures are highly predictive of stage 2 choice patterns. The correspondence is not perfect, and in particular, participants are more likely to exhibit the CRE pattern in choices than their Δ_{CR} measure predicts, less likely to exhibit the CCE pattern in choices than their Δ_{CC} measure predicts, and more likely to exhibit the MXE pattern in choices than their Δ_{MX} measure predicts. However, there is clearly a strong link between participants' valuations and their choices.

Figure 7: Predicting Stage 2 Results using Stage 1 Valuations



Notes: Figure relates individual measures of stage 1 Δ_{CR}, Δ_{CC}, and Δ_{MX} to stage 2 measures of $CRE - RCRE$, $CCE - RCCE$, and $MXE - RMXE$, respectively. Panels A-C use raw stage 1 responses. Panels D-F use the estimated population distribution of preferences from the decomposition in Section 4.4 combined with a participant's raw stage 1 valuations to generate posterior preference measures $E[\Delta^*_{CR}|\text{stage 1}]$, $E[\Delta^*_{CC}|\text{stage 1}]$, and $E[\Delta^*_{MX}|\text{stage 1}]$ for that participant. For each x-axis, one hundred equally sized bins are constructed, with approximately 84 observations per bin for CR and CC measures and 42 observations per bin for MX measures. Within each bin, the value of stage 2 choice differences is calculated to construct the y-axes. Due to a large number of observations at some values (particularly zero), there are 78, 82, and 74 unique bins in Panels A, B, and C, respectively. To make valuations comparable across (p, r), all stage 1 measures are scaled by p to control for the fact that a fixed value of the measure is predicted to yield a larger stage 2 effect the larger is p (see Online Appendix C.3 for details).

Panels D-F of Figure 7 shows that the link between stage 1 valuations and stage 2 choices becomes even tighter when we use our decomposition from Section 4.4 to reduce the noise in participants' stage 1 valuations. Specifically, we combine our estimate for the population distribution of preferences with each participant's stage 1 valuations to generate posterior expectations for each participant's $(h^*_{AB}, h^*_{AB'}, h^*_{CD})$ (see Online Appendix D.4 for details). We then convert those valuations into posterior expectations for each participant's $(\Delta^*_{CR}, \Delta^*_{CC}, \Delta^*_{MX})$, and use those in Panels D-F.[39]

[39]Online Appendix Figure A.4 illustrates the link between stage 1 elicited indifference values and the corresponding stage 2 choice probabilities; specifically, how the likelihood of choosing A in an AB binary choice task depends on how the randomly selected H for that choice task compares to a participant's stage 1 indifference value h_{AB}. That link is also strong and provides further validation of our approach, but it is less directly related to our conclusions about CRP, CCP, and MXP.

Figure 7 highlights that valuations and carefully selected choices both capture common information about the underlying preferences. As such, it is not clear that either is more reliable for eliciting preferences—contrary to claims that choices are more reliable due to their simplicity and transparency (e.g., Brown and Healy, 2018 and Freeman et al., 2019). Indeed, our analysis demonstrates how using both choices and valuations in tandem can deliver stronger conclusions than either approach alone.

5 Proposed Rationalization: A Preference for Upside Potential

The modal pattern of CRP-RCCP-MXP and the broader prevalence of MXP violate both EU and the leading parameterized non-EU models in Table 1. Lacking an existing theory that accommodates these findings, we sketch a tentative model that might help to rationalize them; this model turns out to be a previously unstudied version of quadratic utility (Machina, 1982; Chew et al., 1991). While preliminary, we hope our analysis can spur the development of future models of risk preferences.

5.1 A Preference for Upside Potential

To motivate our approach, consider a concrete example from our data. When $p = 0.5$ (and hence $M = p \cdot 30 = 15$) and $r = 0.2$, the mean valuations are $h_{AB} = 38$, $h_{AB'} = 29$ and $h_{CD} = 33$. These valuations imply:

$$A = (15, 1) \sim (38, 0.5) = B, \text{ while } B' = (38, 0.1; 15, 0.8) \succ A, B.$$

The indifference $A \sim B$ reflects risk aversion: Relative to lottery A, lottery B offers higher expected value but also more risk. Now consider what it means to have an MXP where lottery B' is preferred to both A and B: Relative to A, the person wants some of the additional expected value in B at the cost of some of the additional risk, but not all of the additional expected value in B at the cost of all of the additional risk. At first glance, this intuition resembles the classic EU intuition from insurance and finance settings for a person deciding how much risk to accept. However, the EU logic does not hold in this example because the person is considering probabilistic mixtures between lotteries rather than hedging payoff amounts across states. Under EU, probabilistic mixtures between lotteries do not offer hedging benefits.

We consider an alternative psychology in which a person thinks in terms of *upside potential (UP)*: They first identify a set of outcomes that they consider to be "winning" outcomes, and then trade off (i) the total probability of winning versus (ii) an expected valuation of those winnings. Formally, if we let W denote the set of winning outcomes, then the person evaluates a lottery $X \equiv (x_1, q_1; ...; x_N, q_N)$

as

$$U(X) = \underbrace{\left(\sum_{i=1}^{N} q_i u(x_i)\right)}_{EU} + \underbrace{\left(\sum_{j=1}^{N} q_j I(x_j \in W)\right)}_{\substack{\text{probability of} \\ \text{winning something}}} \underbrace{\left(\sum_{k=1}^{N} q_k I(x_k \in W)\kappa(x_k)\right)}_{\substack{\text{expected valuation} \\ \text{of winnings}}}. \tag{2}$$

The first term is standard expected utility. The second term captures preferences for UP as the total probability of winning multiplied by the expected valuation of those winnings.[40] Returning to our motivating example, even when $A \sim B$, the UP model permits B' to improve on both. Relative to B, B' is attractive because it increases the total probability of winning something; relative to A, B' is attractive because it increases the expected valuation of the winnings.

The UP model falls within the broader class of "quadratic utility" models (Machina, 1982; Chew et al., 1991). For discrete lotteries of the form $X \equiv (x_1, q_1; ...; x_N, q_N)$, quadratic utility includes any model that can be written as $U(X) = \sum_{i=1}^{N} \sum_{j=1}^{N} \phi(x_i, x_j) q_i q_j$ for some symmetric function ϕ. For $\phi(x, y) = (u(x) + u(y))/2 + (v(x)w(y) + v(y)w(x))/2$, quadratic utility yields

$$U(X) = \left(\sum_{i=1}^{N} q_i u(x_i)\right) + \left(\sum_{j=1}^{N} q_j v(x_j)\right)\left(\sum_{k=1}^{N} q_k w(x_k)\right),$$

and setting $v(z) = I(z \in W)$ and $w(z) = I(z \in W)\kappa(z)$ gives equation (2).

While this subcase of quadratic utility has not been studied before (to the best of our knowledge), it combines two previously proposed subcases, neither of which can explain our empirical patterns. Machina (1982) explores the case where $U(X) = \left(\sum_{i=1}^{N} q_i u(x_i)\right) + \left(\sum_{j=1}^{N} q_j v(x_j)\right)^2$; that is, the sum of an expected-utility functional and a second expected-utility functional squared (with $v(x) \geqslant 0$ for all x). Chew et al. (1991) discuss Machina's case and also $U(X) = \left(\sum_{i=1}^{N} q_i v(x_i)\right)\left(\sum_{j=1}^{N} q_j w(x_j)\right)$; that is, the product of two different expected-utility functionals (with $v(x) \geqslant 0$ and $w(x) \geqslant 0$ for all x). The former can generate a CRP and a CCP when $v(x) \neq u(x)$, but—as noted by Chew et al. (1991)—it also generates an aversion to mixtures of any two lotteries. Hence, in our context, the Machina specification would predict a global RMXP, which is inconsistent with our data. Chew et al. (1991) show that the latter generates preferences with an attraction to mixtures of any two lotteries, and thus in our context, it would predict a global MXP. We show it further generates a global ⃠CRP and RCCP. Because the vast majority of our data cover patterns other than ⃠CRP-RCCP-MXP, this case is also inconsistent with our data.[41]

Before deriving predictions for the UP model in equation (2), we must specify which outcomes the

[40] To keep this model well behaved, we assume that $\kappa(x) \geqslant 0$ for all $x \in W$.

[41] See Online Appendix E.2 for details, including the ϕ functions that correspond to the two cases described in the text, and for a formal derivation of their predictions. Masatlioglu and Raymond (2016) show that the Kőszegi and Rabin (2007) model of loss aversion under CPE also falls within the class of quadratic utility models; its predictions (Table 1) also do not fit our data.

person considers "winning" outcomes. For our setting in which people face only binary and trinary lotteries that yield either zero or a positive amount, a natural assumption is that all strictly positive outcomes are in the set of winning outcomes W while zero is not (we discuss this assumption more in Section 5.3). Applying this assumption, and normalizing $u(0) = 0$,[42] equation (2) implies that the triplet $(h^*_{AB}, h^*_{AB'}, h^*_{CD})$ solves

$$\begin{aligned} u(M) + \kappa(M) &= pu(h^*_{AB}) + p^2\kappa(h^*_{AB}) && (3)\\ u(M) + \kappa(M) &= pru(h^*_{AB'}) + (1-r)u(M) + (pr + 1 - r)\left[pr\kappa(h^*_{AB'}) + (1-r)\kappa(M)\right] && (4)\\ ru(M) + r^2\kappa(M) &= pru(h^*_{CD}) + (pr)^2\kappa(h^*_{CD}). && (5) \end{aligned}$$

In Online Appendix E.1, we formally derive predictions from these equations, including the following proposition:

Proposition 1. Suppose that $(h^*_{AB}, h^*_{AB'}, h^*_{CD})$ is derived from equations (3), (4), and (5). For any $u(x)$ that is strictly increasing in x and $\kappa(x)$ that is strictly increasing in x and has $\kappa(x) \geqslant 0$ for all $x > 0$, a person must exhibit one of the following seven patterns of behavior:[43]

P1: $0 > \Delta^*_{CR} > \Delta^*_{CC}$ and $\Delta^*_{MX} > 0$ (RCRP−RCCP−MXP)
P12: $0 = \Delta^*_{CR} > \Delta^*_{CC}$ and $\Delta^*_{MX} > 0$ ($\oslash$CRP−RCCP−MXP)
P2: $\Delta^*_{CR} > 0 > \Delta^*_{CC}$ and $\Delta^*_{MX} > 0$ (CRP−RCCP−MXP)
P23: $\Delta^*_{CR} > \Delta^*_{CC} = 0$ and $\Delta^*_{MX} > 0$ (CRP−$\oslash$CCP−MXP)
P3: $\Delta^*_{CR} > \Delta^*_{CC} > 0$ and $\Delta^*_{MX} > 0$ (CRP−CCP−MXP)
P34: $\Delta^*_{CR} = \Delta^*_{CC} > 0$ and $\Delta^*_{MX} = 0$ (CRP−CCP− $\oslash$MXP)
P4: $\Delta^*_{CC} > \Delta^*_{CR} > 0$ and $\Delta^*_{MX} < 0$ (CRP−CCP−RMXP).

Proposition 1 establishes the empirical content of the UP model: Only seven of the 13 possible transitive patterns of behavior can occur. Recall that Figure 6 in Section 4.4 plots histograms of raw response patterns and decomposed preference patterns that emerge in our data, and we label the seven patterns in Proposition 1. In the histogram of decomposed preference patterns, the seven patterns from Proposition 1 account for 81 percent of observations, with the other six possible patterns accounting for only 19 percent. Moreover, in the histogram of raw response patterns—which also admits 14 additional intransitive patterns that can emerge only due to noise—the seven patterns from Proposition 1 still account for 55 percent, with the other six possible transitive patterns accounting for only 18 percent. Thus, the restrictions imposed by the UP model seem to accord well with the

[42] This normalization merely simplifies these equations; in Online Appendix E.1, we present these equations and prove all results without making this normalization.

[43] Patterns including only strict preferences are denoted by single numbers, while patterns including a weak preference (e.g., $\oslash$CRP) are denoted by two numbers referring to the two corresponding strict-preference patterns.

patterns we observe in our data.[44]

Proposition 2 refines Proposition 1 for the special case where $u(x) = x$, allowing us to isolate the role of the UP term in generating novel predictions.[45]

Proposition 2. Suppose that $(h^*_{AB}, h^*_{AB'}, h^*_{CD})$ is derived from equations (3), (4), and (5). If $u(x) = x$, then:

(1) If $\kappa(x)$ is strictly increasing in x and has $\kappa(x) \geqslant 0$ for all $x > 0$, then:

   (a) If $ph^*_{AB} > M$ (i.e., if the person is risk averse in the AB decision), then $\Delta^*_{CR} > 0$ and thus the person must exhibit one of patterns P2, P23, P3, P34, or P4.

   (b) If $ph^*_{AB} = M$ (i.e., if the person is risk neutral in the AB decision), then $\Delta^*_{CR} = 0$ and thus the person must exhibit pattern P12.

   (c) If $ph^*_{AB} < M$ (i.e., if the person is risk loving in the AB decision), then $\Delta^*_{CR} < 0$ and thus the person must exhibit pattern P1.

(2) If $\kappa(x) = \phi x$ for some $\phi > 0$, then $ph^*_{AB} > M$ (i.e., the person is risk averse in the AB decision), $\Delta^*_{CR} > 0$, and $\Delta^*_{MX} > 0$, and thus the person must exhibit one of patterns P2, P23, or P3.

Part 1 of Proposition 2 provides additional testable predictions of the UP model that hold if the EU term is approximately linear. First, whether people exhibit risk aversion or risk tolerance in their AB decision is a sufficient statistic for whether they exhibit CRP or RCRP. Second, whether people exhibit risk aversion or risk tolerance in their AB decision also predicts a partial separation into preference patterns. We test these predictions using an approach analogous to that described in Section 4.5 (see in particular the notes for Figure 7) in which we use our posterior valuation estimates to construct an individual-level risk aversion measure in the AB task ($E[ph^*_{AB} - M|\text{stage 1}]$) and a corresponding CRP measure ($E[\Delta^*_{CR}|\text{stage 1}]$). We then analyze the relationship between the signs of these variables. Roughly 13 percent of observations exhibit risk tolerance, $E[ph^*_{AB} - M|\text{stage 1}] < 0$; of these, the majority (55 percent) exhibit RCRP. The remaining 87 percent of observations exhibit risk aversion; of these, the majority (69 percent) exhibit CRP. Thus, risk aversion is significantly correlated with CR preferences as the UP model predicts (Fisher's exact test $p < 0.001$). Extending

[44] In the histogram of raw response patterns, five of the six transitive patterns that UP rules out occur at roughly the same rate as the intransitive patterns that arise only due to noise: The two strict transitive patterns each occur as often as each of the two strict intransitive preferences, and three of the four weak transitive preferences occur roughly as often as each of the 12 weak intransitive preferences. The exception is the (weak) EU pattern ruled out by the UP model, which occurs at a rate similar to the weak transitive patterns permitted by the model.

[45] For the reasons articulated in Rabin (2000) and Rabin and Thaler (2001), to the extent that $u(x)$ is meant to capture a standard EU motivation, it might be natural assume approximate linearity for u given the small stakes on our experiment. Such an assumption would be analogous to the suggestion in Kőszegi and Rabin (2007) to assume concave utility for large stakes but linear utility for small stakes, where the latter can be interpreted as a local approximation for the former.

this analysis to the posterior calculations for broader CR-CC-MX preference patterns, we find that for risk tolerant observations with $E[ph^*_{AB} - M|\text{stage 1}] < 0$, 48% exhibit P1 as predicted by Proposition 2, while 21% of risk averse observations exhibit P1 (Fisher's exact test $p < 0.001$). In contrast for risk averse observations, 69% exhibit P2, P3, or P4 as predicted by Proposition 2 (no posterior calculations indicate the intermediate P12, P23, or P34 patterns), while 45% of risk tolerant observations exhibit P2, P3, or P4 (Fisher's exact test $p < 0.001$). Thus, risk aversion is also significantly correlated with the nature of CR-CC-MX preferences as the UP model predicts.

For the interested reader, Proposition 2 also helps to clarify some of the mathematical mechanics in the UP model. In particular, Part 2 of Proposition 2 highlights that the basic structure of the UP model—specifically, how probabilities enter the UP term in equation (2)—creates a propensity to exhibit risk aversion in the AB decision and both a CRP and an MXP. The source of risk aversion in the AB decision is straightforward: A linear u means that the EU term argues for risk neutrality; when κ is also linear, the UP term argues for risk aversion in the AB choice, as the risky option B has its probability squared, thus making it less valuable. The combined impact of EU and UP terms is thus risk aversion in the AB problem.

The logic for why a person exhibits a CRP is related. Relative to the AB decision, scaling down the winning probabilities by a common ratio r scales down the EU term by r but scales down the UP term by r^2, and thus makes the UP term relatively less important. Since the UP term is the source of risk aversion, the net impact is that scaling down the winning probabilities by a common ratio r causes the person to become less risk averse—that is, they have a CRP.

The logic for why an MXP arises is more nuanced. Suppose $A \sim B$, recall $B' = (1-r)A + rB$, and consider how $U(B')$ varies with r. As r changes, there are two effects on the UP term. First, an increase in r leads to a reduction in the probability of winning, which creates a motive to make r small. Second, an increase in r leads to a change in the expected valuation of winnings. If there is risk aversion in the AB decision, increasing r leads to an increase in the expected valuation of winnings, which creates a motive to make r large. When κ is linear, the latter effect dominates for r small, while the former effect dominates for r large, hence why an intermediate r can be better than both $r = 0$ and $r = 1$—i.e., there is an MXP.

Of course, Part 1 of Proposition 2 highlights how, even if u is linear, nonlinearity in κ can generate all seven possible patterns from Proposition 1, and the logic for linear κ helps to explain when different patterns emerge for nonlinear κ. As long as there is risk aversion in the AB decision (which relies on κ not being too convex over the relevant range), the person will exhibit a CRP. But unlike for linear κ, Part 1a permits that there might be ⃠MXP or RMXP, which occur when κ is sufficiently concave over the relevant range. Parts 1b and 1c highlight that with a nonlinear κ there could be risk neutrality or risk loving in the AB decision (which occur when κ is sufficiently convex over the relevant range), in which case the person would exhibit ⃠CRP or RCRP, respectively. For

these cases, the person must exhibit an MXP and a RCCP.

5.2 The Shape of the $\kappa(x)$ Function

We observe all seven response patterns predicted by Proposition 1 in our data, which, by Proposition 2, suggests a nonlinear κ function. To gain some initial insight into the possible shape of κ, we estimate it on our aggregate valuation data, identifying it from experimental variation in (p, r) combinations. Because our model is post-hoc and the experiment was not designed specifically to identify or estimate the shape of the κ function, we urge caution in interpreting this estimation. Nonetheless, we hope this analysis offers useful guidance for future modeling.

Our data include 60 aggregate valuations: the mean responses for h_{AB}, $h_{AB'}$, and h_{CD} for each of the 20 (p, r) combinations. Equations (3) to (5) implicitly define h^*_{AB}, $h^*_{AB'}$, and h^*_{CD} as a function of the experimental parameters (p, r, M) and the functions u and κ. To isolate the role of UP, we set $u(x) = x$, making the κ function the only object to estimate.

We parameterize the κ function as $\kappa(z; \theta)$ where θ is a vector of parameters to be estimated, and estimate θ using non-linear least squares on the 60 observations assuming

$$h_{XY} = h^*_{XY}(p, r, M, \theta) + \epsilon \qquad \text{for } XY \in \{AB, AB', CD\}.$$

Online Appendix F describes the estimation details and provides the complete parameter estimates.

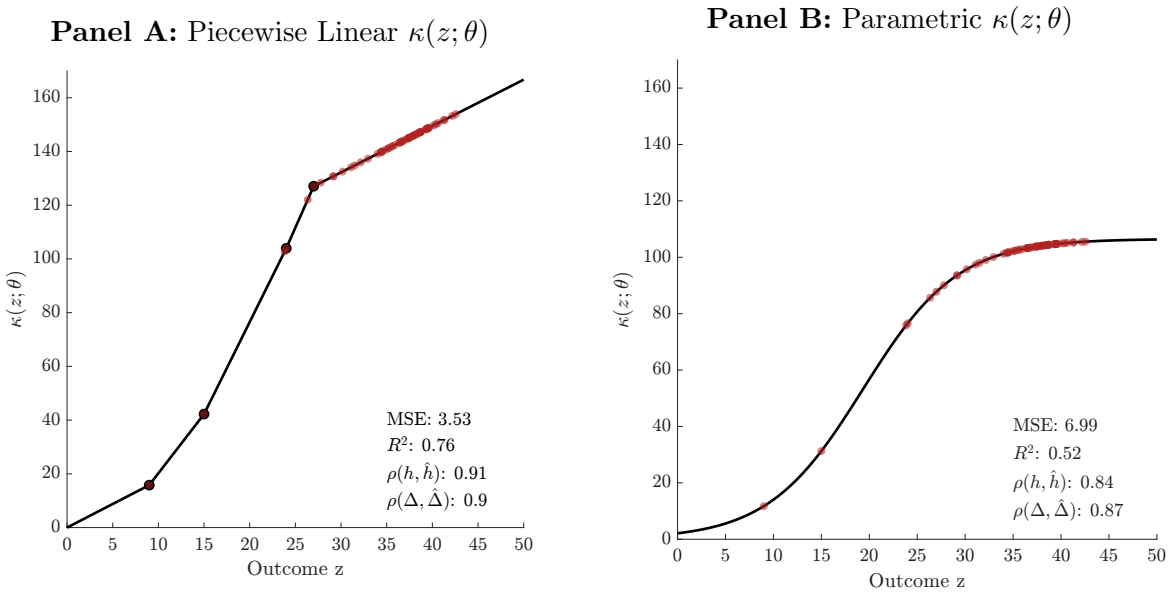
Lacking a strong a priori sense of the shape of κ, we begin with a flexible specification. In our 60 observations, $\kappa(z; \theta)$ is evaluated at $M = \{9, 15, 24, 27\}$ as well as values of h ranging from approximately 24 to 45. We therefore specify a five-parameter, continuous piecewise-linear form for κ with knots at 9, 15, 24, and 27.[46] The resulting estimate, shown in Panel A of Figure 8, exhibits an S-shape: convex for low values of z and concave for higher values. The estimated model captures a large share of the variation in h_{AB}, h'_{AB}, and h_{CD} across values of (p, r), yielding $R^2 = 0.76$ and an $MSE = 3.53$.[47] The correlation between predicted and observed h is 0.91, and the correlation for the corresponding Δ terms is 0.90 (see Online Appendix Figure A.5 Panel A). Figure 8 also clarifies the data limitations of our current design for estimation: κ is evaluated at only 4 points below $x = 30$ (the values of M fixed in the design) and then a cluster of points around $x = 35$.

Motivated by the S-shape in panel A of Figure 8, Online Appendix F.2 explores several parame-

[46] In Online Appendix F, we report a version with an additional knot at 36. Since that model performs similarly, we retain the five-parameter version as our preferred specification.

[47] $R^2 = 1 - RSS/TSS$, where RSS is the sum of squared residuals between the estimated model and the data, and TSS is the sum of squared deviations to the average h across all 60 observations. Negative values indicate that predicting the mean for every observation would yield a better fit than the estimated model.

Figure 8: Estimates for $\kappa(z;\theta)$ under Upside-Potential Model



Notes: Figure presents estimates for $\kappa(z;\theta)$ under the UP model. The estimation is conducted using non-linear least squares with 60 observations of mean responses for h_{AB}, $h_{AB'}$, and h_{CD} (20 observations for each). Panel A presents estimates for $\kappa(z;\theta)$ when using a five-part piecewise linear formulation, and Panel B presents estimates for $\kappa(z;\theta)$ when using a three-parameter functional form for κ. Light red points in each panel correspond to locations where the function is evaluated in estimation. In panel A, black dots denote kinks in the piecewise linear formulation; in the estimation, the function is evaluated 15 times at each kink point. Each panel also presents fit values of mean squared error (MSE), in-sample R^2, correlation between predicted and actual h values, and correlation between predicted and actual Δ values. The in-sample R^2 is given by $1 - RSS/TSS$, where TSS is the sum of squared deviations to the average h among the 60 observations, and RSS is the sum of squared residuals between the estimated model to the data. See Online Appendix F.2 for details.

terized functional forms for κ. Among these, the best fit is a three-parameter scaled logistic:

$$\kappa(z;\boldsymbol{\theta}) = \theta_1 \left[\frac{1}{1+\exp(-\theta_2(z-\theta_3))} \right], \quad \theta_1, \theta_2, \theta_3 > 0. \tag{6}$$

Panel B of Figure 8 shows the estimated κ under this specification. Despite its parsimony, this parameterization retains substantial predictive accuracy with $R^2 = 0.52$, $MSE = 6.99$, a correlation between predicted and actual h of 0.84, and a correlation between predicted and actual Δ of 0.87 (see Online Appendix Figure A.5 panel B).

The three-parameter sigmoid performs reasonably well in-sample, suggesting that it could be used to generate predictions that shed light on the implications and limitations of the UP model. However, the fact that it is bounded above makes it ill-suited for global predictions. Moreover, because it substantially underperforms the piece-wise linear function in panel A, there is clear scope for identifying better parameterized functional forms. Such an exercise would be better pursued with alternative data that are intentionally designed to estimate κ.[48]

[48] In Online Appendix G.1, we estimate both original prospect theory (OPT) and cumulative prospect theory (CPT) with a flexible functional form for the probability weighting function $\pi(q)$ specified as a six-part piece-wise linear

5.3 Other Implications and Limitations

Although several predictions of the UP model align well with our data, it is a post-hoc rationalization meant to stimulate future theorizing. To help inform such work, we outline some additional testable predictions of the model and discuss several of its limitations.

In Online Appendix E.3, we derive predictions of the UP model for indifference curves in a Marschak-Machina triangle. The model predicts that five patterns are possible, and these are presented in Figure 9.[49] In Panel A, a single linear indifference curve (shown in bold) separates convex (mixture-loving) indifference curves to the left and concave (mixture-averse) indifference curves to the right. In this case, the average slopes of the indifference curves must fan out everywhere.[50] Panel B shows the case where all indifference curves are concave, which also requires fanning out everywhere. Panels C, D, and E show cases where all indifference curves are convex: For all three cases, indifference curves fan out to the left of the AB indifference curve, but the cases differ in whether the average slopes of indifference curves to the right of the AB indifference curve fan out, remain constant, or fan in.

The indifference curves in Figure 9 suggest two testable predictions. First, although the UP model allows average slopes to fan out, remain constant, or fan in to the right of the AB indifference curve (permitting CRP, $\oslash$CRP, or RCRP), it requires fanning out to the left of the AB indifference curve. Hence, when comparing an AB decision to the EF decision depicted in each panel, the UP model predicts greater risk aversion in the latter.[51] Second, in any given triangle, indifference curves can transition from convex to concave at most once, and if there is a transition it must be from convex in the upper left to concave in the lower right. Hence, for instance, if a person exhibits an aversion to mixtures of lotteries A and B, then they must also exhibit an aversion to mixtures of any two lotteries to the right of A and B, such as mixtures of C and D. Analogously, if a person exhibits an attraction to mixtures of A and B, then they must also exhibit an attraction to mixtures of any two lotteries to the left of A and B, such as mixing lotteries E and F. Our data cannot speak to either of these predictions; collecting data along these lines might help guide future theorizing.

Beyond proposing new testable predictions, it is also useful to ask whether the UP model can

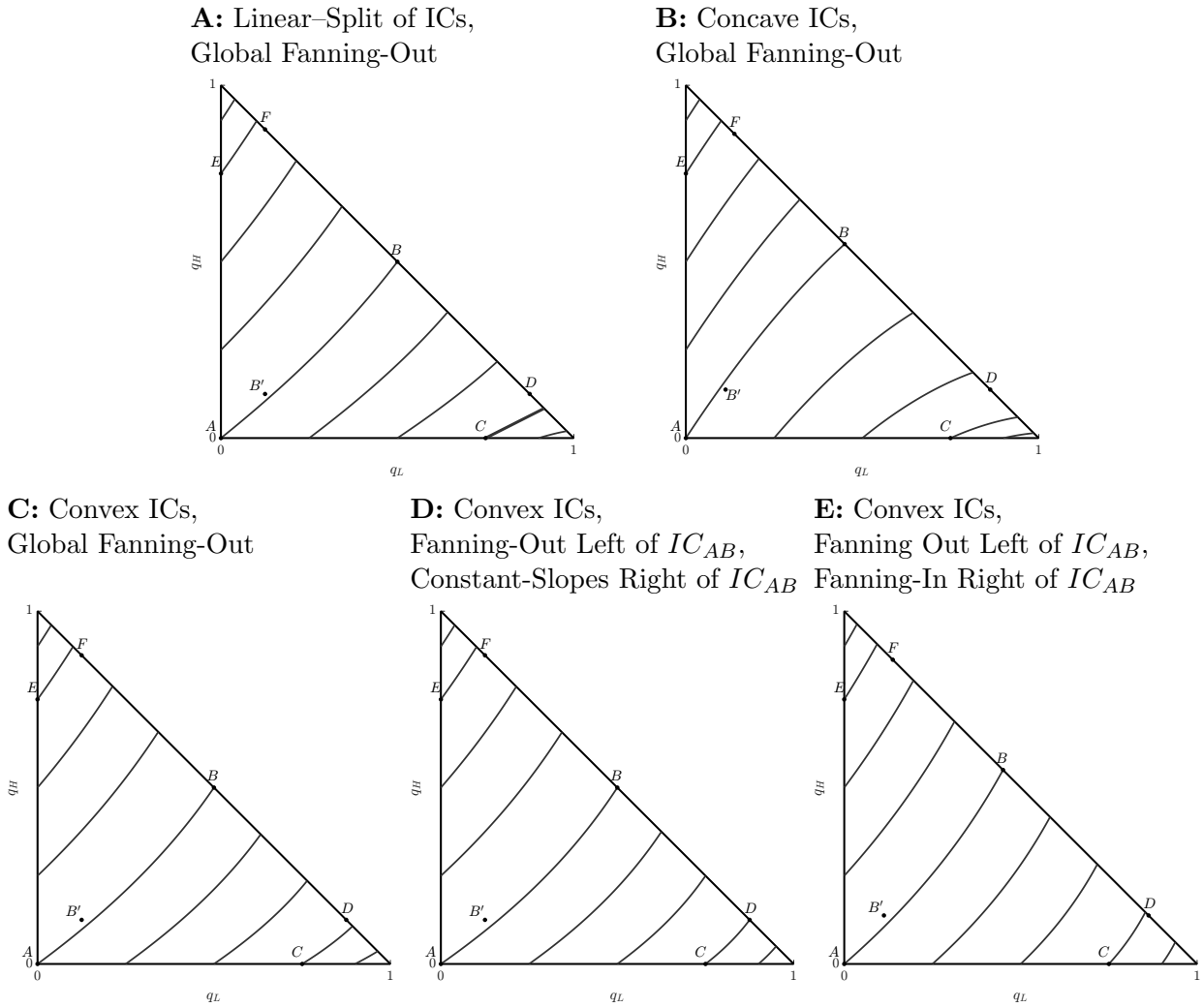
functional form that permits (but does not require) discontinuities at $q = 0$ and $q = 1$. Both models substantially underperform the UP estimates reported in the text.

[49] To help connect these five patterns of indifference curves to the seven patterns of behavior in Proposition 1, each panel in Figure 9 depicts a set of five points that would correspond to one (p, r) combination in our experiment (A, B, B', C and D), where we have chosen B such that $A \sim B$. See Online Appendix E.3 for details.

[50] We label an indifference curve as convex versus concave based on its shape in a Marschak-Machina triangle with q_L and q_H on the axes. Given this labeling, convex (concave) indifference curves correspond to locally quasi-concave (quasi-convex) preferences. Each indifference curve connects a point on either the horizontal axis or vertical axis to a point on the hypotenuse; "average slope" refers to the slope of the line segment that connects these two points. See Online Appendix E.3 for formal statements.

[51] The asymmetry around the AB indifference curve arises because lotteries along the vertical axis (such as E) guarantee a winning outcome, whereas lotteries on the hypotenuse and along the horizontal axis have some possibility of not winning.

Figure 9: Upside Potential Indifference Curves



Notes: Figure depicts the five possible patterns of indifference curves (ICs) that can arise in a Marschak–Machina (MM) Triangle under the UP model (see Online Appendix E.3). Each triangle depicts lotteries with outcomes H, M, and 0 and associated probabilities q_H, q_M, and q_L. Lotteries $A = (M, 1)$, $B = (H, p)$, $C = (M, r)$, $D = (H, rp)$, and $B' = (H, rp; M, 1 - r)$ correspond to one (p, r) combination in our experiment, where we have chosen p such that $A \sim B$; lotteries $E = (H, s; M, 1 - s)$ and $F = (H, s + (1 - s)p)$ are the analogues to lotteries C and D but located in the upper-left corner. In Panel A, there is one linear indifference curve (shown in bold), where indifference curves are convex to its left and concave to its right, and where average slopes fan out throughout. The linear indifference curve could be anywhere in the triangle. In Panel B, all indifference curves are concave, and average slopes fan out throughout. In Panels C, D, and E, all indifference curves are convex, and average slopes fan out to the left of the AB indifference curve which we denote as IC_{AB} for convenience; the panels differ in that average slopes to the right of the AB indifference curve can fan out (Panel C), be constant (Panel D), or fan in (Panel E). See Online Appendix E.3 for a discussion of how these five patterns of indifference curves connect to the seven possible patterns of behavior in Proposition 1.

accommodate established empirical regularities. For instance, one reliable empirical finding is that certainty equivalents for binary lotteries—i.e., an m such that $(m, 1) \sim (H, p)$—tend to exhibit risk aversion ($m < pH$) for larger p but risk tolerance ($m > pH$) for smaller p (see, e.g., Tversky and Kahneman, 1992). In Online Appendix E.4, we demonstrate that the UP model with an S-shaped κ function similar to the one we estimate can naturally generate this pattern.

While the UP model seems to perform well in some domains, it has some important limitations. First, not all prior empirical findings are well accommodated by the UP model. For example, some prior papers find a CRE for paired choice tasks when each choice task involves options with equal expected values.[52] The UP model can accommodate this pattern, but only if we assume that the expected-utility term in the model has a convex u (see Online Appendix E.5 for details). Because this is contrary to the more usual EU assumption that u is concave and (perhaps) approximately linear for small stakes, our formulation of the UP model may not accommodate this pattern.

Second, although we have developed the UP model to in principle be portable to other contexts, there are open questions in how to do so. Perhaps most notable is the question of what determines the set W of winning outcomes. In our setting, we adopt the natural assumption that positive outcomes are winning outcomes while zero is not. To apply the model more broadly, we need a more principled definition of this set—and data designed to assess it. This issue is analogous to the definition of the reference point in prospect theory and other reference-dependent models. As in that literature, one might assume some exogenous rule, or allow a person's assessment of winning outcomes to depend on features of the choice set (in the spirit of Kőszegi and Rabin, 2007).

A related issue is that a person could have an analogous concern for *downside potential*. In other words, a person might also identify a set of outcomes that they consider to be "losing" outcomes, and then trade off the total probability of losing versus an expected (negative) valuation of those losses. A better understanding of when and how individuals focus on the upside or downside of a lottery is crucial for applying the UP model (or related models) more broadly.

Finally, and more speculatively, the κ function itself may be context dependent. In particular, it is not at all clear that a κ function estimated on small stakes would extrapolate well for making predictions on large stakes. Indeed, an outcome of \$30 might generate substantial "upside" utility in the context of our experiment, but if we re-ran our experiment with all stakes increased by a factor of 10, the same outcome of \$30 might generate very little "upside" utility.

The UP model that we develop is merely a first attempt to rationalize our data given that they

[52] In Table 1 in Blavatskyy et al. (2023), the 11 CR experiments with "Probability of highest outcome in risky lottery in the first question" equal to "Ratio of middle to highest outcome" involve choices between lotteries with equal expected values. Another prominent example is a more general form of common ratio effect, subproportionality, demonstrated in problems 7 and 8 in Kahneman and Tversky (1979); while this latter example is outside the domain of our study in that lottery A is not a certain lottery, an analogous result to that in Online Appendix E.5 still applies. While our discussion here assumes that these observed CRE patterns for equal-expected value lotteries have some preference component, we remind the reader that, because they are found using paired choice tasks, they could be due to choice noise (as discussed in Section 2.5).

are inconsistent with leading parameterized non-EU models. By delineating the ways in which it does well while also highlighting its limitations, we hope our analysis can spur the development of future models of risk preferences.

6 Discussion

In this paper, we study connected CR-CC-MX problems across a broad range of experimental parameters. While our empirical findings for mean preferences match the common wisdom of CRP, CCP, and near-zero MXP when using canonical experimental parameters, we find a robust pattern of CRP-RCCP-MXP at other experimental parameters. More generally, our empirical results highlight how mean preference patterns can be sensitive to experimental parameters and that there can be substantial heterogeneity in these patterns. Because these empirical results are inconsistent with both EU and leading non-EU models, we posit a theoretical model motivated by the prevalence of MXP in our data. Encouragingly, this approach to accommodating MXP also explains the modal CRP-RCCP-MXP pattern we observe, and successfully predicts some more nuanced patterns in our data, suggesting it might provide useful direction for future theoretical work. We conclude by discussing some broader implications of our analysis.

Our analysis highlights the benefits of studying connected phenomena over a broad range of the parameter space to deliver both a richer picture of preferences and sharper tests of existing theories. Many prominent non-standard models of preferences—in domains such as risk, ambiguity, intertemporal choice and beyond—are built on data from a few canonical examples, with limited empirical insights of how behavior might be sensitive to relevant features of the decision problem. Moreover, the prior literature has often studied non-standard choice problems in isolation rather than together. A more comprehensive empirical foundation for evaluating theories emerges by jointly studying multiple preference features across a broad range of parameters.

Our analysis also reinforces a key methodological point from McGranaghan et al. (2024) that cautions against using only binary choice data when making comparisons across decisions and instead encourages also using valuation data for such comparisons. Many behavioral "effects" are based on evidence from comparisons of paired binary choice tasks. In our data, the CR, CC, and MX problems all show hallmarks of the confounds induced by differential noise, indicating that binary choice data could present a biased picture of the underlying preferences unless problem parameters are carefully chosen. Our binary choice data ultimately generate the same qualitative conclusions as our valuations data due to our deliberate selection of balanced problems with offsetting bias—a selection made possible by the fact that we had pilot data on valuations. In the future, researchers analyzing behavioral effects should consider the potential for bias in isolated choice problems, mitigate

this bias through careful selection of offsetting problems where possible, and complement choice data with valuation data to deliver stronger conclusions than either approach alone.

Our analysis relates to a nascent literature that explores deliberate randomization (i.e., mixing) between lotteries (Agranov and Ortoleva, 2017; Dwenger et al., 2018; Feldman and Rehbeck, 2022; Agranov et al., 2023; Agranov and Ortoleva, 2023). These studies typically present individuals with the same decision problem multiple times, either mixed throughout the study or repeated explicitly in a row, and randomization manifests as making different choices across repetitions. A notable limitation of such designs is that they cannot reveal an RMXP. Nonetheless, deliberate randomization between two successive presentations of an AB choice is clearly consistent with exhibiting an MXP in our context. Future work could explore whether there is an empirical connection between these behaviors by testing whether the people who exhibit an MXP in paired decision tasks are the same as those who exhibit deliberate randomization in two successive presentations of an AB choice.[53] Exploring the empirical connection seems particularly apt given the different theoretical explanations for mixture preferences here and elsewhere. In particular, a common rationalization of purposeful randomization in the prior literature has been the theory of cautious expected utility (Cerreia-Vioglio et al., 2015) (CEU), but CEU maintains betweenness in comparisons with degenerate lotteries like lottery A in our study, and thus it cannot accommodate the type of mixture preferences we study and observe. Additionally, the UP model permits conditions under which individuals are averse to mixtures, rather than always exhibiting a systematic preference for them.

On the surface, our empirical finding of an MXP and the proposed rationalization in the form of our UP model may seem to contradict recent work suggesting that decisionmakers penalize lotteries with greater numbers of outcomes (Bernheim and Sprenger, 2020; Fudenberg and Puri, 2022; Puri, 2025). However, that work focuses on different types of lotteries and, therefore, could be consistent with our results. For example, Bernheim and Sprenger (2020) find that people prefer binary lotteries to nearby trinary lotteries that are constructed by splitting one of the binary lottery payments into two equally likely payments while retaining the same expected value. This comparison is very different from our MX problem. Moreover, the UP model can accommodate both phenomena. As we discuss in Section 5.1, our model generates an MXP even with a linear κ function. At the same time, whether a person is averse to the type of lottery splitting studied by Bernheim and Sprenger (2020) is related to the local concavity or convexity of κ around the split outcome, with local concavity implying an aversion to splits (see Online Appendix E.6 for details). Investigating these connections is a promising direction for future work.

Finally, and most importantly, our results call for the development of new models of risk preferences alongside the development of a more complete empirical foundation for risky choice. Our

[53] In the spirit of such exercises, Feldman and Rehbeck (2022) link non-constant responses in successive presentations to preferences for mixtures elicited directly as a choice of convex combinations between lotteries.

empirical findings are at odds with both EU and leading non-EU models. We have taken a first step in postulating a post-hoc model that performs well on the narrow domain of our data, but it is clearly not a complete account of risky choice. CR-CC-MX problems represent a tiny fraction of risky choices and extending the domain through principled exploration of problems away from this structure is a critical next step. We hope the psychology of upside potential, and its novel perspective on the potential source of risk attitudes, will prove useful for guiding such empirical explorations and the theories that will follow.

Data Availability

Data and code are available at https://doi.org/10.7910/DVN/OGJENQ.

References

Agranov, M., Healy, P. J., and Nielsen, K. (2023). Stable randomisation. *Economic Journal*, 133(655):2553–2579.

Agranov, M. and Ortoleva, P. (2017). Stochastic choice and preferences for randomization. *Journal of Political Economy*, 125(1):40–68.

Agranov, M. and Ortoleva, P. (2023). Ranges of randomization. *Review of Economics and Statistics*.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica*, 21(4):503–546.

Bateman, I. and Munro, A. (2005). An experiment on risky choice amongst households. *The Economic Journal*, 115(502):C176–C189.

Bell, D. E. (1985). Disappointment in decision making under uncertainty. *Operations research*, 33(1):1–27.

Bernheim, B. D. and Sprenger, C. (2020). On the empirical validity of cumulative prospect theory: Experimental evidence of rank-independent probability weighting. *Econometrica*, 88(4):1363–1409.

Blavatskyy, P., Ortmann, A., and Panchenko, V. (2022). On the experimental robustness of the Allais paradox. *American Economic Journal: Microeconomics*, 14(1):143–63.

Blavatskyy, P., Panchenko, V., and Ortmann, A. (2023). How common is the common-ratio effect? *Experimental Economics*, pages 253–272.

Blavatskyy, P. R. (2006). Violations of betweenness or random errors? *Economics Letters*, 91(1):34–38.

Blavatskyy, P. R. (2010). Reverse common ratio effect. *Journal of Risk and Uncertainty*, 40(3):219–241.

Brown, A. L. and Healy, P. J. (2018). Separated decisions. *European Economic Review*, 101:20–34.

Burke, M. S., Carter, J. R., Gominiak, R. D., and Ohl, D. F. (1996). An experimental note on the Allais paradox and monetary incentives. *Empirical Economics*, 21(4):617–632.

Camerer, C. F. and Ho, T.-H. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, 8(2):167–196.

Cerreia-Vioglio, S., Dillenberger, D., and Ortoleva, P. (2015). Cautious expected utility and the certainty effect. *Econometrica*, 83(2):693–728.

Chapman, J., Dean, M., Ortoleva, P., Snowberg, E., and Camerer, C. (2023). Econographics. *Journal of Political Economy Microeconomics*, 1(1):115–161.

Chew, S. H. (1983). A generalization of the quasi-linear mean with applications to the measurement of income inequality and decision theory resolving the Allais paradox. *Econometrica*, 51(4):1065–1092.

Chew, S. H., Epstein, L. G., and Segal, U. (1991). Mixture symmetry and quadratic utility. *Econometrica: Journal of the Econometric Society*, pages 139–163.

Chew, S. H. and MacCrimmon, K. (1979). Alpha utility theory, lottery composition, and the allais paradox. *Working paper*, 686.

Chew, S. H. and Waller, W. S. (1986). Empirical tests of weighted utility theory. *Journal of Mathematical Psychology*, 30(1):55–72.

Dean, M. and Ortoleva, P. (2019). The empirical relationship between nonstandard economic behaviors. *Proceedings of the National Academy of Sciences*, 116(33):16262–16267.

Dekel, E. (1986). An axiomatic characterization of preferences under uncertainty: Weakening the independence axiom. *Journal of Economic Theory*, 40(2):304–318.

Dwenger, N., Kübler, D., and Weizsäcker, G. (2018). Flipping a coin: Evidence from university applications. *Journal of Public Economics*, 167:240–250.

Feldman, P. and Rehbeck, J. (2022). Revealing a preference for mixtures: An experimental study of risk. *Quantitative Economics*, 13:761–786.

Freeman, D. J., Halevy, Y., and Kneeland, T. (2019). Eliciting risk preferences using choice lists. *Quantitative Economics*, 10(1):217–237.

Fudenberg, D. and Puri, I. (2022). Simplicity and probability weighting in choice under risk. *American Economic Review: Papers & Proceedings*, 122:421–425.

Gul, F. (1991). A theory of disappointment aversion. *Econometrica*, 59(3):667–686.

Huber, C., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Weitzel, U., Abellán, M., Adayeva, X., Ay, F. C., Barron, K., et al. (2023). Competition and moral behavior: A meta-analysis of forty-five crowd-sourced experimental designs. *Proceedings of the National Academy of Sciences*, 120(23):e2215572120.

Jain, R. and Nielsen, K. (2024). A systematic test of the independence axiom near certainty. *Working Paper*.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.

Kobayashi, S. J. and Lucia, A. (2023). Robust estimation of risk preferences. *Working Paper*.

Kőszegi, B. and Rabin, M. (2007). Reference-dependent risk attitudes. *American Economic Review*, 97(4):1047–1073.

Loomes, G. and Sugden, R. (1998). Testing different stochastic specifications of risky choice. *Economica*, 65(260):581–598.

Lucia, A. (2024). What drives violations of the independence axiom? the role of decision confidence. *Working Paper*.

Machina, M. J. (1982). "expected utility" analysis without the independence axiom. *Econometrica*, 50(2):277–323.

Masatlioglu, Y. and Raymond, C. (2016). A behavioral analysis of stochastic reference dependence. *American Economic Review*, 106(9):2760–2782.

McGranaghan, C., Nielsen, K., O'Donoghue, T., Somerville, J., and Sprenger, C. D. (2024). Distinguishing common ratio preferences from common ratio effects using paired valuation tasks. *American Economic Review*, 114(2):307–347.

Menkveld, A. J., Dreber, A., Holzmeister, F., Huber, J., Johannesson, M., Kirchler, M., Neusüss, S., Razen, M., Weitzel, U., Abad-Díaz, D., et al. (2024). Nonstandard errors. *The Journal of Finance*, 79(3):2339–2390.

Puri, I. (2025). Simplicity and risk. *Journal of Finance*, 80(2):1029–1080.

Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*, 68(5):1281–1292.

Rabin, M. and Thaler, R. H. (2001). Anomalies: risk aversion. *Journal of Economic perspectives*, 15(1):219–232.

Simonsohn, U., Simmons, J. P., and Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, 4(11):1208–1214.

Stango, V. and Zinman, J. (2023). We are all behavioural, more, or less: A taxonomy of consumer decision-making. *Review of Economic Studies*, 90(3):1470–1498.

Steegen, S., Tuerlinckx, F., Gelman, A., and Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5):702–712.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.