Connecting Common Ratio and Common Consequence Preferences*

Christina McGranaghan Kirby Nielsen Ted O'Donoghue
Jason Somerville Charles D. Sprenger

August 20, 2024

Abstract

Common ratio (CR) and common consequence (CC) problems present two foundational deviations from expected utility (EU) but have been studied independently, with little exploration of experimental parameters, and using paired choice tasks that can yield biased inference. We overcome these limitations using valuations and a wide range of parameters to study connected problems, which also capture mixture preferences (MX). We empirically characterize combinations of CR-CC-MX preferences throughout the parameter space, documenting systematic sensitivities and substantial heterogeneity inconsistent with leading non-EU models. We speculate on a model of "upside potential" that explains the modal pattern and captures additional data features.

*McGranaghan: Department of Applied Economics and Statistics, University of Delaware, email: cmcgran@udel.edu; Nielsen: Division of the Humanities and Social Sciences, California Institute of Technology, email: kirby@caltech.edu; O'Donoghue: Department of Economics, Cornell University, email: edo1@cornell.edu; Somerville: Federal Reserve Bank of New York, email: jason.somerville@ny.frb.org; Sprenger: Division of the Humanities and Social Sciences, California Institute of Technology, email: sprenger@caltech.edu. For helpful comments and suggestions, we thank Doug Bernheim, Drew Fudenberg, Shengwu Li, and seminar participants at the University of Zurich, the University of Pittsburgh, Lehigh University, Stanford University, University of California at Davis, 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 views expressed in this paper are those of the authors alone and do not necessarily reflect the views Federal Reserve Bank of New York or the Federal Reserve System. 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 later popularized by Kahneman and Tversky (1979), who treated them as foundational features of preferences that motivate the shape of the probability weighting function in prospect theory. The subsequent literature has 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 has three major limitations. First, the experimental evidence on each effect is based on a limited set of experimental parameters, while behavioral theories motivated by this evidence assume that these effects reflect global features of risk preferences. Second, previous studies have virtually always analyzed the CRE and CCE independently. However, there is an overlooked connection between the two problems that permits studying how the two phenomena relate to each other as well as how they relate to a third key property of risk preferences: attitudes towards probabilistic mixtures of lotteries (MX preferences). Third, the prior literature has almost always studied the CRE and CCE using paired choice tasks, and these tasks can yield biased conclusions in the presence of commonly assumed forms of choice noise. In McGranaghan et al. (2024), we build on prior work highlighting this bias in CRE experiments and demonstrate how paired valuation tasks can provide unbiased inference under these circumstances.

In this paper, we address these three limitations by using valuation tasks to study connected common ratio (CR) and common consequence (CC) problems across a broad range of experimental parameters, including many not previously explored in the literature. While we observe a CRE and CCE for some parameter values, the magnitudes of these effects vary systematically with the parameters to the extent that neither the classic CRE nor CCE generalize across the parameter space. Indeed, we find a large reverse CCE for some parameter configurations. Moreover, we find a robust attraction to mixtures for much of the parameter space we cover. Combining these results for connected problems, we empirically characterize combinations of CR-CC-MX preferences throughout the parameter space, and these patterns are inconsistent with EU and leading non-EU models. We posit a post-hoc theoretical model of "upside potential" designed to explain our modal mean preference. Encouragingly, this model successfully predicts some of the more nuanced data patterns we observe, suggesting it might hold insight for future theoretical work.

In Section 2, we summarize and reanalyze the existing literature on the CRE and CCE. 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]

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

***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

In this example, lotteries C and D are created 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:

***AB′* Choice**: Lottery A: 100 percent chance of \$1M vs. Lottery B': 89 percent chance of \$1M
10 percent chance of \$5M

***CD* Choice**: Lottery C: 11 percent chance of \$1M vs. Lottery D: 10 percent chance of \$5M

Lotteries C and D are created 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 should not impact an individual's relative preference for the two options. In contrast, the CRE and CCE describe a systematic empirical pattern of people appearing more risk-tolerant in the CD problems relative to the AB and AB' problems, respectively, resulting in aggregate choice frequencies where the share of individuals choosing A is larger than the share choosing C in both cases.

The prior literature has conducted many experiments on the CRE and CCE. Two recent meta-studies by Blavatskyy et al. (2022) on the CRE and Blavatskyy et al. (2023) on the CCE report a total of 224 experiments (143 CRE and 81 CCE) across 48 studies. They document that the CRE and CCE are sensitive to various experimental design choices, including experimental parameters. We leverage their data to highlight three further points. First, these existing experiments cover a limited set of experimental parameters. Second, these existing experiments have almost always studied the CRE and CCE as independent phenomena. Third, this literature relies almost exclusively on paired choice tasks.[2]

By studying the CRE and CCE as independent phenomena, the prior literature has neglected a natural connection between the two problems.[3] To illustrate the connection, consider the following three choice tasks:

[2] In fact, the two meta-studies cover *only* experiments that use paired choice tasks. McGranaghan et al. (2024) provide a summary of CRE experiments that use the valuations approach that we describe below, and report only 10 experiments across four studies.

[3] Of the 48 studies covered by the two meta-studies, only 10 contain both CRE and CCE experiments, and only two of those discuss the CRE and CCE as connected phenomena (Bateman and Munro, 2005; Chew and Waller, 1986).

***AB* Choice**:	Lottery A:	100 percent chance of \$27	vs.	Lottery B:	90 percent chance of \$35
***AB′* Choice**:	Lottery A:	100 percent chance of \$27	vs.	Lottery B':	9 percent chance of \$35 90 percent chance of \$27
***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, and thus three tasks suffice to study both problems. But this formulation also highlights that there is a third comparison: the combination of the AB and AB' choices. Since lottery B' is a mixture of lotteries A and B (in this case 90 percent A and 10 percent B), this third comparison reveals people's attitudes toward 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.[4]

We refer to the combination of an AB task, an AB' task, and a CD task as a *connected CR-CC-MX problem*. We can characterize a connected CR-CC-MX problem using 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 refers to both the common ratio that converts lotteries A and B into lotteries C and D and the probabilistic mixture of A and B that generates lottery B', which we denote by r (0.1 in the example above).[5]

Each connected CR-CC-MX problem permits 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 might be neutral to the manipulation (for which we use $\oslash$CRP, $\oslash$CCP, and $\oslash$MXP). We can compare these empirical patterns to the predictions of various theories of risk preferences. EU implies that people are neutral to all three manipulations and, therefore, that they have $\oslash$CRP-$\oslash$CCP-$\oslash$MXP. In Section 2.2, we delineate the predictions from leading non-EU models. All predict a CRP and a CCP for the focal suggested range of their key parameters, and most further predict either an RMXP or an $\oslash$MXP; that is, people either dislike or are neutral to mixtures. Furthermore, in most cases, these models predict these patterns apply globally and that they are independent of (p, r). These are predictions that we can test using connected CR-CC-MX problems.

[4]The MX problem relates to tests of betweenness and to work on deliberate randomization; we discuss these literatures in Section 2.4.

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

By relying almost exclusively on paired choice tasks, the existing literature may be subject to biased conclusions if there is differential choice noise across tasks. In Section 2.5, we build on prior work in McGranaghan et al. (2024) to describe how paired choice tasks do not reliably reveal the underlying preference components, and how the inference problem is even worse when we want to compare three binary choice tasks, as in a connected CR-CC-MX problem. [6] Hence, our primary analysis focuses on valuation tasks in the form of stated indifference points implemented using multiple price lists. For instance, in a CD valuation task analogous to the CD choice task above, a participant selects the amount received with a 9 percent chance that makes them indifferent to a 10 percent chance of \$27. Unlike choices, these valuations can deliver credible measures of CRP, CCP, and MXP that permit robust inference under common assumptions on the form of noise.

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 a total of 20 different (p, r) combinations that cover a wide range of the parameter space. In stage 2 of the experiment, we present the same participants with binary choice tasks linked to the valuation tasks they saw in stage 1. These additional choice data allow us to validate our first-stage results, assess whether paired choice tasks yield biased inference, and connect our findings to the prior literature.

Section 4 describes our main results. We first analyze mean preferences 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, where it gets larger for smaller r and disappears entirely for the two largest values of r that we consider. In contrast, CCP is highly sensitive to the high-prize probability p, where it gets smaller for smaller p, and there is substantial RCCP for the smallest value of p that we consider. Finally, we find a robust MXP that gets larger for smaller r and smaller p. Together, these results yield combinations of CR-CC-MX preferences with two striking regularities. On one hand, for parameters close to their values in the canonical CR and CC examples (i.e., for large p and small r), we find CRP and CCP with limited MXP, consistent with the conventional wisdom. On the other hand, for small p and small r, we find a robust pattern CRP-RCCP-MXP, which is inconsistent both with EU and leading non-EU models.

We next analyze behavior at the individual level. Specifically, we investigate the distribution of behavior across the 27 possible combinations of CR-CC-MX patterns, where a person might exhibit the effect, the reverse effect, or no effect for each. We find that mean behavior masks substantial heterogeneity. The modal pattern at the individual level aligns with the CRP-RCCP-MXP combination that frequently appears in mean preferences, but it accounts for only 15 percent of observations. Of course, much of the variability in behavior at the individual level could be due to noise. Because our experiment elicits multiple measures of the same valuation within-subject,

[6] While these inference challenges are not the primary focus of this paper, in Section 4.3 we document empirical evidence demonstrating their relevance for our CR, CC, and MX choice tasks.

we are able to structurally decompose the variability in valuations into an underlying preferences component and a noise component. We estimate that roughly half of the variability in valuations is due to preference heterogeneity. The underlying distribution of preferences that we estimate still yields a modal pattern of CRP-RCCP-MXP, but it also still implies substantial heterogeneity. Hence, any non-EU model of risk preferences must be able to explain the combination CRP-RCCP-MXP while also having the flexibility to permit different patterns across individuals and across experimental parameters.

Given that both the modal pattern of CRP-RCCP-MXP and the more general prevalence of MXP violate both EU and leading non-EU models, in Section 5 we speculate on a possible model that might be able to explain them. Our results suggest that many individuals gravitate towards lotteries with more potential upside of winning a non-zero prize while limiting the chance of getting nothing. We operationalize this idea formally by proposing a simple model of "upside potential" in which people trade off the total probability of winning something versus the expected winnings conditional on winning something. Encouragingly, in addition to the modal preference pattern we set out to capture, this model successfully predicts some of the more nuanced data patterns we observe. Given the post-hoc nature of our model and its predictions, we do not view this evidence as formal tests of our theory, but rather as a promising first step that might inform new non-EU models that better capture the patterns that emerge from our analysis.

In Section 6 we discuss some broader implications of our analysis. In particular, we connect our results to a recent literature identifying purposeful randomization and the underlying mechanisms for such phenomena (Agranov and Ortoleva, 2023, 2017; Agranov et al., 2023; Dwenger et al., 2018; Feldman and Rehbeck, 2022; Cerreia-Vioglio et al., 2015). We also highlight how our work reinforces the fundamental benefits of identifying patterns of preferences using valuation tasks like our own relative to binary choice tasks (Bernheim and Sprenger, 2020; McGranaghan et al., 2024; Carrera et al., 2022). We conclude by calling for the development of new models of risk preferences alongside the development of a more complete empirical foundation for risky choice.

2 Background and Prior Literature

The prior literature has virtually always studied the CRE and CCE as separate problems, and for each problem, the literature has focused on a limited set of parameters. In contrast, we study connected CR and CC problems for a broad set of parameters. In this section, we formally demonstrate the connection between the two problems, highlight the value of investigating connected problems, and 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. For this pair, EU predicts that preferences are invariant to scaling down the probabilities of the non-zero outcomes by the common ratio r. Thus, an individual should prefer either lotteries A and C or lotteries B and D. 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.

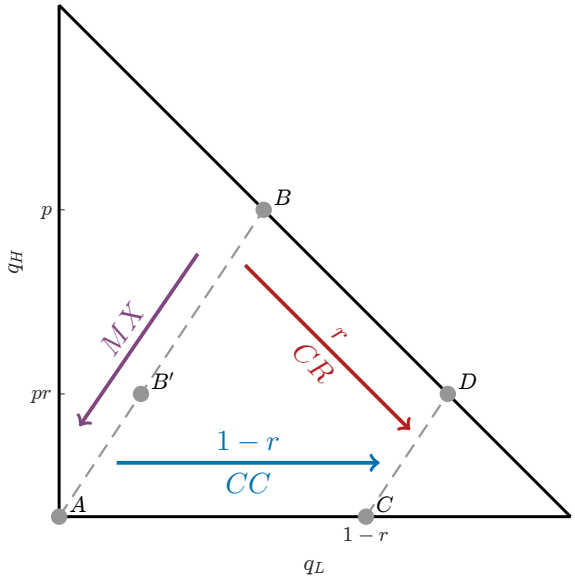
The combination of an AB' task and a CD task represents a CC Problem. For this pair, EU predicts that preferences are invariant to shifting the common consequence of a $1 - r$ chance of $\$M$ into a $1 - r$ chance of \$0. Thus, an individual should prefer either lotteries A and C or lotteries B' and D. 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.

For a given H and M, each (p, r) generates a connected pair of CR and CC problems. Panel A of Figure 1 visually illustrates one such connected pair by plotting the five lotteries in a Marschak-Machina probability simplex for one particular (p, r). Each problem separately involves comparing two choice sets on parallel line segments in a Marschak-Machina triangle. Connecting the problems puts the AB and AB' choice sets on the same line and then uses the same CD choice for both problems. Panel A highlights that lottery B' is a probabilistic mixture of lotteries A and B, with relative proportions $(1 - r)$ and r, and suggests a third relevant comparison of two parallel choice sets: comparing the AB and AB' tasks. This comparison reveals people's attitudes towards probabilistic mixtures, and thus we refer to the combination of the AB and AB' choices as a *mixture (MX)* problem. EU predicts that people are neutral to mixtures, and thus should prefer lotteries A and A or lotteries B and B'. In contrast, we refer to the pattern of choosing lotteries A and B' as a mixture effect (MXE) since people appearing more risk-tolerant in the AB' problem relative to the AB problem implies that they like mixtures.

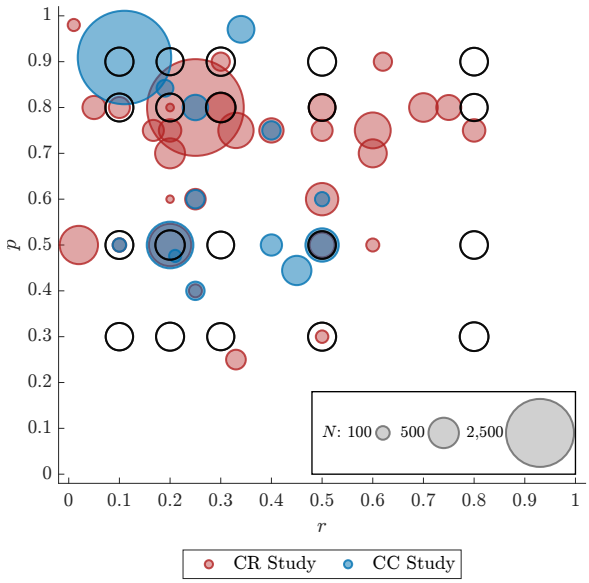
Hence, for a given (p, r), we refer to the combination of an AB task, an AB' task, and a CD task as a *connected CR-CC-MX problem.*

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

Panel A: CC, CR, 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.

2.2 Why Study Connected Problems?

By studying connected problems, researchers can uncover a more comprehensive picture of individual risk preferences than by studying each problem in isolation. As a simple initial illustration, consider the additional implications of studying connected CR-CC-MX problems for EU. For each individual problem, EU makes a particular prediction: For instance, for a CR problem, EU predicts that individuals prefer A and C or B and D. However, for a connected CR-CC-MX problem, EU makes a stronger prediction across the three choice tasks, specifically requiring either combination A, A, and C or combination B, B', and D. More generally, by studying connected CR-CC-MX problems, we can uncover patterns in how people react to CR manipulations, how people react to CC manipulations, and how people react to MX manipulations—that is, researchers can study three characteristics of underlying preferences.

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

- 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 $H < h^*_{CD}$.

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 that people are more risk tolerant 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 attitude toward probabilistic mixtures is 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)*.

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 the patterns of individual preferences, Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX}. In Appendix B, we derive these predictions for several leading non-EU 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).[7]

Table 1 highlights how connected CR-CC-MX problems can be used to assess various models of non-EU risk preferences in the literature. Indeed, it reveals three important features common to this set of prominent models. First, each model predicts a CRP and a CCP for the focal range of its key parameter. This commonality reflects the fact 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, and perhaps most interestingly, when we also consider cases where the models' key parameters lie outside their focal ranges, the set of predicted patterns expands, but it does not include all possibilities. In particular, none of the models permit the pattern CRP, RCCP, and MXP that features prominently in our data.

Two final observations will be relevant for linking our data on connected CR-CC-MX problems to non-EU models. First, except in the few cases noted, the directional predictions in Table 1 hold for any (p, r) combination—that is, these models predict directional patterns that are invariant to (p, r). 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 discuss in Section 5.3 whether that additional flexibility could be used to explain

[7] While we focus on parameterized models, one could instead frame this discussion in terms of axioms of choice. In particular, the CRE and the CCE are classic violations of the independence axiom—i.e., it is the independence axiom that implies $\Delta^*_{CR} = \Delta^*_{CC} = \Delta^*_{MX} = 0$. Some have suggested replacing the independence axiom with a betweenness axiom that permits Δ^*_{CR} and Δ^*_{CC} to be nonzero but requires that they equal each other and thus that $\Delta^*_{MX} = 0$. Gul's 1991 model of Disappointment Aversion, among others, falls in this class.

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 $\pi(q) = q^\delta / \left[q^\delta + (1-q)^\delta\right]^{\frac{1}{\delta}}$ $U(B) = \pi(p)v(H)$	$\delta \in (0.279, 1)^\dagger$	$\Delta^*_{CR} > 0,\ \Delta^*_{CC} > 0,\ \Delta^*_{MX} \gtreqless 0$
	$\delta > 1$	$\Delta^*_{CR} < 0,\ \Delta^*_{CC}$ and $\Delta^*_{MX} \gtreqless 0$
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)	$\beta > 0^\dagger$	$\Delta^*_{CR} = \Delta^*_{CC} > 0,\ \Delta^*_{MX} = 0$
$U(B) = \dfrac{p}{1+\beta(1-p)} u(H)$	$\beta \in (-1, 0)$	$\Delta^*_{CR} = \Delta^*_{CC} < 0,\ \Delta^*_{MX} = 0$
Cautious Expected Utility (CEU) (Cerreia-Vioglio et al., 2015) (see Appendix B.6 for structure)	not applicable	$\Delta^*_{CR} = \Delta^*_{CC} > 0,\ \Delta^*_{MX} = 0$
Simplicity Preferences (Puri, 2024) (see Appendix B.7 for structure)	not applicable	$\Delta^*_{CR} > 0,\ \Delta^*_{CC} > 0,\ \Delta^*_{MX} < 0$

Notes: Table presents predictions of each model for the sign of Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX}; see Appendix B for formal derivations. Each of the first five models is a parameterized model. To give a sense of the parametric structure, 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 well-defined. 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.

the patterns in our data.

2.3 Limitations of the Prior Literature on the CRE and CCE

There is a large prior 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 prior empirical evidence on the CRE and CCE, we reanalyze the prior literature in terms of the (p, r) combinations that researchers have used. To do so, we merge the data from two recent meta-studies: Blavatskyy

et al. (2022) identify 143 CR experiments drawn from 39 studies and Blavatskyy et al. (2023) identify 81 CC experiments drawn from 29 studies. All 224 experiments noted in these meta-studies rely on paired choices which, as we discuss in Section 2.5, suffer from substantial inferential problems both when considering each problem in isolation and when considering their connection. We combine the data from these meta-studies while converting the probabilities into our (p, r) framework. The resulting dataset yields several insights.

First, the prior literature has virtually always studied the CRE and CCE as separate problems. Of the prior CR and CC studies covered by the two meta-studies, only 10 contain both CR and CC experiments. Of these 10, 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 to better understand the nature of risk preferences.[8]

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. The red circles correspond to prior CR experiments, while the blue circles correspond to prior CC experiments. For each type of experiment, there is a large mode at the canonical versions of each problem. Of the 143 prior CR experiments, 48 (34%) used 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%) used 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 others, 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, the leading non-EU models that we summarize in Table 1 predict that people have both a CRP and a CCP for all (p, r). However, Blavatskyy et al. (2022) and Blavatskyy et al. (2023) 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).[9]

In Table 2, we further explore the sensitivity of these effects to the choice of experimental

[8] Burke et al. (1996) and Loomes and Sugden (1998) have connected CR-CC-MX problems, but it seems incidental. The focus of Burke et al. is whether the CCE differs for real versus hypothetical incentives; the fact that their study actually includes a connected CR-CC-MX experiment is not mentioned by the authors. The focus of Loomes and Sugden is comparing three stochastic specifications of EU; to do so, they collect data on roughly 50 binary choices that include within them five connected CR-CC-MX problems, however, they do not make any special use of these connected problems, and they do not discuss the CCE or mixture preferences. Bateman and Munro (2005) and Chew and Waller (1986) intentionally study connected CR-CC-MX problems. Bateman and Munro studies whether couples' joint decision making differs from individual decision making in terms of the CRE, CCE, and mixture preferences, and they use several connected CR-CC-MX problems to do so. Chew and Waller is the paper that has a motivation closest to ours, emphasizing how connected CR-CC-MX problems can provide more information about risk preferences. However, both of these papers' experiments cover very few parameter configurations.

[9] Blavatskyy et al. (2022) and Blavatskyy et al. (2023) 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.

parameters. We use the following continuous measures as our primary outcome variables:[10]

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

In Panel A of Table 2, we conduct OLS regressions using these measures as dependent variables. We replicate the qualitative findings from Blavatskyy et al. (2022) and Blavatskyy et al. (2023) and additionally find that CRE and CCE are differentially sensitive to changes in p and r. This result suggests that what one would learn about each phenomenon and their connection is likely to vary based on the parameters examined.

In Panel B of Table 2, we divide prior CRE and CCE experiments into two categories. The first consists of those conducted at the canonical (p, r) values, while the second consists of those conducted at alternative (p, r) values. We find that a sizable CRE and CCE emerge for the experiments conducted at their respective canonical parameter values. However, we also find that both effect sizes are significantly smaller in magnitude across all experiments conducted at non-canonical values. Strikingly, an RCCE emerges on average for CC experiments conducted away from the canonical (p, r) configuration.

Finally, we observe that most prior studies have not focused on testing whether the CRE and CCE are robust to the choice of experimental parameters. Indeed, many of those prior studies did not have the CRE or CCE 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 across all studies 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 across all studies 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.

2.4 Prior Literature on Mixture Preferences

Separate from the literature on the CRE and the CCE, two distinct strands of work speak to mixture preferences. The first is a small literature on direct tests of the betweenness axiom. This literature is reviewed in two key papers: Camerer and Ho (1994) summarize and contribute to the early literature on betweenness and Blavatskyy (2006) follows up on this. Both papers conclude that individuals frequently violate betweenness, but the evidence is mixed on the direction of violation, with some studies concluding more mixture loving and others concluding more mixture aversion. These tests of betweenness are almost never in the context of connected CR-CC-MX problems.

[10] We use $\widehat{\Pr}(X|YZ)$ to denote the proportion of the population that chooses option X from the choice set YZ. Note that $CRE - RCRE$ is equivalent to the proportion choosing combination AD minus the proportion choosing BC in a CR problem, and $CCE - RCCE$ is equivalent to the proportion choosing combination AD minus the proportion choosing $B'C$ in a CC problem.

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	12.44	50.70
	(11.59)	(14.01)
Common Ratio: r	−61.14	−45.16
	(8.23)	(19.10)
Sample	Blavatskyy et al. (2023)	Blavatskyy et al. (2022)
Outcome Mean	19.25	8.64
Observations	143	81

Panel B. Canonical vs. Non-Canonical

	(3) Canonical	(4) Other	(5) Difference
	(i): KT Parameters		
CRE − RCRE	25.63	15.49	−10.71
	(16.09)	(24.80)	[−2.87]
Experiments	53	90	143
	(ii): Allais Parameters		
CCE − RCCE	21.37	−0.57	−21.94
	(18.18)	(22.09)	[−4.74]
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; $p = 0.8$, $r \in [0.2, 0.3]$) or Allais (1953) ($p = 0.9$ or 0.91, $r = 0.1$ or 0.11), respectively. Standard deviations are in parentheses, and t-statistics are in brackets.

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 present individuals with the same decision problem repeated multiple times, either mixed 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.

2.5 Choices versus Valuations

The prior literatures on the CRE, CCE, and MXE suffer from one additional shared limitation: They have focused almost exclusively on paired binary choice tasks. Comparing behavior across two binary choice tasks can lead to biased conclusions in the presence of noise. In McGranaghan et al. (2024), we build on prior work to demonstrate this problem within the context of the CRE, and we then show how one might avoid this problem by instead comparing behavior across analogous *valuation tasks*.

Motivated by the McGranaghan et al. (2024) development, our primary analysis in this paper

will focus on three valuation tasks:

***AB* Valuation Task**: state an $h_{AB} \gtrless M$ such that $(M, 1) \sim (h_{AB}, p)$

***AB′* Valuation Task**: state an $h_{AB'} \gtrless M$ such that $(M, 1) \sim (h_{AB'}, pr; M, 1-r)$

***CD* Valuation Task**: state an $h_{CD} \gtrless M$ such that $(M, r) \sim (h_{CD}, pr)$.

To see how such valuation tasks might 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 the person 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 will then be

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

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 empirical objects of interest Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX}.[11]

In contrast to comparing behavior across valuation tasks, comparing behavior across binary choice tasks can lead to biased inference. To illustrate, suppose noise operates in the same way for a choice task as it does for the analogous valuation task. Specifically, for a fixed (p, r, M), the person has effective indifference values h_{AB}, $h_{AB'}$, and h_{CD} that result from their underlying indifference values plus noise. In a binary choice task, the person chooses the safer option when the specific payment value H offered in the binary choice task is smaller than the relevant effective indifference value. For example, in an AB task, the person chooses A over B when $H < h_{AB}$, which can be rearranged to $-\varepsilon_{AB} < h^*_{AB} - H$. Note that $h^*_{AB} - H$ represents the *distance to indifference* between the underlying indifference value h^*_{AB} and the offered H. Intuitively, the person will choose the safer option when the relevant distance to indifference is sufficiently large relative to a realized shock.

Now consider the implications for comparing behavior across two binary choice tasks. For example, for a CR problem, this formulation implies $\Pr(A|AB) = \Pr(-\varepsilon_{AB} < h^*_{AB} - H)$ and $\Pr(C|CD) = \Pr(-\varepsilon_{CD} < h^*_{CD} - H)$.[12] Even if there is no underlying CRP and thus $h^*_{AB} = h^*_{CD}$, $E[\varepsilon_{AB}] = E[\varepsilon_{CD}] = 0$ is not enough to guarantee that $CRE - RCRE = 0$. For paired choice tasks to deliver an unbiased test of the CRE, the *distributions* of ε_{AB} and ε_{CD} must be the same. Suppose

[11] 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 Appendix C.1, we develop an analogous argument within the context of this paper, and we conduct both means and sign tests in our analysis in Section 4.

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

instead that there is differential mean-zero noise, and in particular that the noise has a bigger impact in the CD choice than in the AB choice. Then, even when $h^*_{AB} = h^*_{CD} \equiv h^*$, a positive distance to indifference ($h^* - H > 0$) implies we would observe a CRE ($CRE - RCRE > 0$), while a negative distance to indifference ($h^* - H < 0$) implies we would observe an RCRE ($CRE - RCRE < 0$). In other words, our conclusion would be biased, and moreover, the extent of the bias depends on the experimenter's choice of the experimental payment value H.

In McGranaghan et al. (2024), we formalize this argument in more detail in the context of CR problems and then run experiments that suggest there is indeed a differential noise problem in the direction described in the prior paragraph. In Appendix C.2, we provide the analogous formal argument in the context of the three tasks that constitute a connected CR-CC-MX problem, and then illustrate how the potential for misleading conclusions is even greater when attempting to identify preference patterns by comparing behavior across three binary choices.

To obtain robust inference, our primary analysis will focus on identifying patterns of Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX} using valuation tasks. However, our experiment will also conduct binary choice tasks that are linked to the valuation tasks. For these choice tasks, we intentionally select values for the payment value H based on pilot data to generate data with a roughly equal number of positive and negative distances to indifference and approximate indifference on average. We do so for two reasons. First, some researchers have argued that binary choice tasks are more reliable than valuations for eliciting preferences due to their simplicity and transparency (e.g., Brown and Healy, 2018 and Freeman et al., 2019). By collecting data from linked binary choice tasks, we can assess the relationship between valuations and choices, and whether the two types of data yield the same messages when appropriately accounting for the potential biases in choice-based inference. Second, these choice data will allow us to directly assess the direction and extent of any differential noise problem within the context of CC and MX problems, which the prior literature has not examined.

A final benefit of valuation tasks is that they provide magnitudes that allow for a direct interpretation. Specifically, with valuation tasks, we observe Δ_{CR}, Δ_{CC}, and Δ_{MX}. These represent direct measures of the magnitudes of their corresponding underlying preferences Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX}. In contrast, with binary choice tasks, we 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 sign of these quantities were unbiased, the magnitudes have no clear interpretation as they represent a combination of the magnitude of the underlying Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX} and the magnitude of the differential noise.

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.[13] We consider four possible values of $p \in \{0.3, 0.5, 0.8, 0.9\}$

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

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. The open circles in Panel B of Figure 1 plot our choice of (p, r) combinations.

Figure 2 provides an overview of the experiment timeline. Our experiment consists of two stages. In stage 1, we randomly assign each participant to four of our 20 (p, r) combinations, and they complete a total of 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 participant completes 24 binary choice tasks for the same four (p, r) combinations as in stage 1. Participants complete all 20 valuations before proceeding to the 24 binary choices, and we randomize the order of questions within each stage.

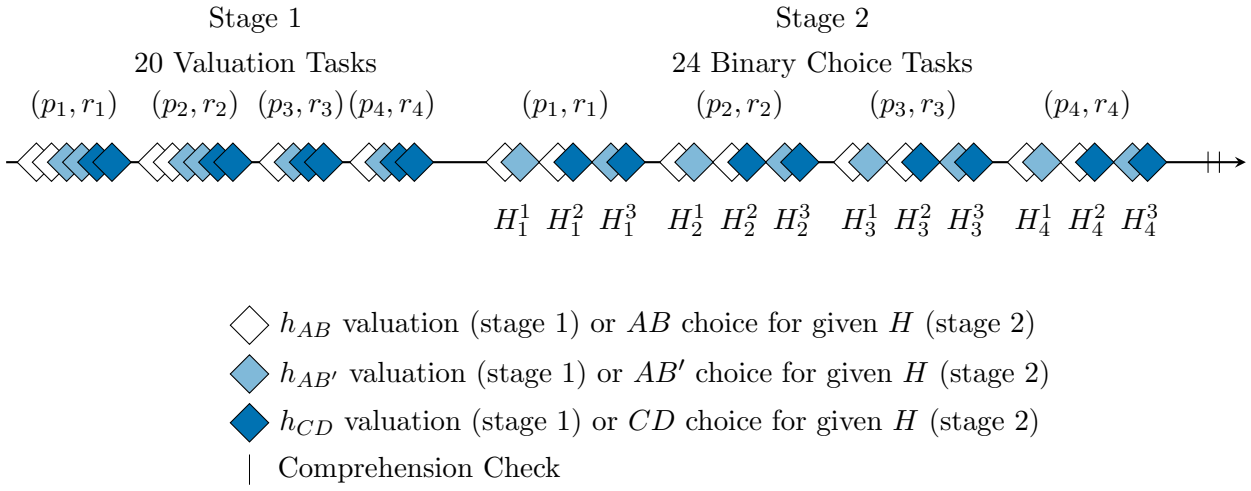


Figure 2: Experiment Timeline

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 each possible value of p in combination with at least one r value. Conversely, we draw r values with replacement so that a participant might face the same r across multiple values of p.[14]

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

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

For each (p, r) combination, we fix $M = \$(p \cdot 30)$ and elicit the relevant valuations using a

[14] We draw values of p *without* replacement to avoid additional instances of multiple elicitations of the same task since the AB task does not depend on r.

multiple price list.[15] The left-hand option in each price list remains fixed at either lottery $(M, 1)$ or lottery (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 $\$(p \cdot 30)$ to $\$(p \cdot 30 + 50)$. For instance, 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.[16]

For each price list, we enforce a unique switching point. When participants click a row in the left panel, it highlights the left-hand option in that row and all rows above. Analogously, when they click a row in the right panel, it highlights the right-hand option in that row and all the rows below. They can adjust their choices as much as they want before submitting their final choices for that valuation task. Appendix Figures F.1 to F.3 provide example screenshots of all three variants of the valuation task.

For two randomly drawn (p, r) combinations that a participant sees, we elicit each valuation twice, generating observations $\{h_{AB}, h'_{AB}, h_{AB'}, h'_{AB'}, h_{CD}, h'_{CD}\}$. For the remaining two (p, r) combinations, we only elicit the CD valuation twice, generating observations $\{h_{AB}, h_{AB'}, h_{CD}, h'_{CD}\}$.[17] 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 multiple elicitations of the same valuation are separated from each other by at least three other valuation tasks.

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 represent 12 paired choice tasks, three for each (p, r) combination: a CR paired choice task that involves an AB choice and a CD choice, a CC paired choice task that involves an AB' choice and a CD choice, and an MX paired choice task that involves an AB choice and an AB' choice. Appendix Figures F.4 to F.6 present examples of each type of binary choice task.

Each binary choice that a participant sees in stage 2 is akin to a single isolated "row" from one of the price lists they saw in stage 1. Specifically, for each (p, r) combination, we again fix $M = \$(p \cdot 30)$. Then, for each of the three types of paired choice tasks, we draw a random value of H from the relevant row in Table 3 without replacement. Hence, for a specific (p, r) combination,

[15] 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 should choose $H = \$30$ for all price lists.

[16] 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. To ensure that our valuations are as comparable as possible, we focus on h-valuation tasks in this paper.

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

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 pre-specified based on pilot data to be values where: i) the distance to indifference ranges from positive to negative across H values, and (ii) the distances to indifference are expected to be roughly balanced. Thus, the design permits exploration of inference problems of the paired-choice paradigm discussed in Section 2.5.

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 task.

It is interesting to note the scale of our stage 2 paired choice tasks relative to the prior CR and CC literature. We can think of 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 the two meta-studies discussed in Section 2.3, there are 143 CR experiments with 14,794 total observations and 81 CC experiments with 8,947 observations.

3.3 Additional Design Details

Before beginning stage 1, participants complete an unincentivized attention check and a quiz about the payment mechanism. After completing stage 2, participants complete two incentivized comprehension checks to gauge their understanding of the multiple price list format and the binary choice tasks. The first comprehension check tests whether individuals can correctly fill out a price list given a specified indifference value. The second comprehension check tests whether participants can correctly answer a binary choice question when given another person's responses to a multiple price list. Appendix Figures F.7 and F.8 provide example screenshots of these comprehension checks.[18] 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. Appendix Figure F.9 provides an example of one of these puzzles.

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

3.4 Recruiting

We recruited 2,102 participants through Prolific who had at least a high school education, were between the ages of 18 and 31, were living in the United States or Western Europe, had a minimum Prolific approval rating of 99 percent, were fluent in English, and had completed 50 to 1,000 previous Prolific submissions. We focus on 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. 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 based on the option they selected. 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 24 minutes and 27 seconds, and the average bonus payment for selected participants was $15.76.

4 Results

Our main analysis uses the AB, AB', and CD valuations collected in stage 1 of our experiment. As described in Section 3.1, we collect some valuations twice and randomly label them h_{XY} and h'_{XY} for $XY \in \{AB, AB', CD\}$.[19] We collect multiple elicitations of individual valuations for three main reasons.

First, we use the multiple elicitations for an initial assessment of our stage 1 data. Specifically, we examine the correlations between the two elicitations of the same valuation. For example, we compute the correlation between h_{CD} and h'_{CD} for a particular (p, r) combination. Appendix Table A.3 reports these correlations for each valuation by (p, r). All correlations are positive, ranging from 0.254 to 0.696. These strong correlations suggest that our valuation data capture meaningful information about the underlying preferences of interest.

Second, we need multiple elicitations of individual valuations to construct independent measures of our three main objects of study. 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}.

[19] 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.

If we had only a single elicitation for each valuation, then measurement error in the valuations would create mechanical correlations between Δ_{CR}, Δ_{CC}, and Δ_{MX}. For instance, if we use the same measure h_{CD} to construct both Δ_{CR} and Δ_{CC}, then measurement error in h_{CD} would create a mechanical positive correlation between Δ_{CR} and Δ_{CC}. Our use of multiple elicitations of individual valuations avoids this problem.

Third, we use the multiple elicitations of the same object to disentangle noise from preferences, which we describe in detail in Section 4.2.

4.1 Main Results: Δ_{CR}, Δ_{CC}, and Δ_{MX}

Our main results focus on Δ_{CR}, Δ_{CC}, and Δ_{MX}. We first analyze mean preferences across all participants. Figure 3 presents the mean values of Δ_{CR}, Δ_{CC}, and Δ_{MX} by p separately for each r. Recall that EU predicts all of these values to be zero, while many non-EU models predict a systematic $\Delta_{CR} > 0$ and $\Delta_{CC} > 0$. Panels A and B characterize the average CRP and CCP in our data, and illustrate that CRP and CCP are differentially sensitive to parameter changes and do not seem to be global phenomena. Values of Δ_{CR} are largely invariant to changes in p but are sharply decreasing in r, leading to $\Delta_{CR} \approx 0$ at higher r values. In contrast, values of Δ_{CC} are largely invariant to changes in r but are sharply increasing in p, with $\Delta_{CC} < 0$ at lower p values. Panel C reveals that our participants consistently prefer mixtures, sometimes quite strongly. The average Δ_{MX} is positive for most parameters and becomes substantially larger at lower values for p and r.

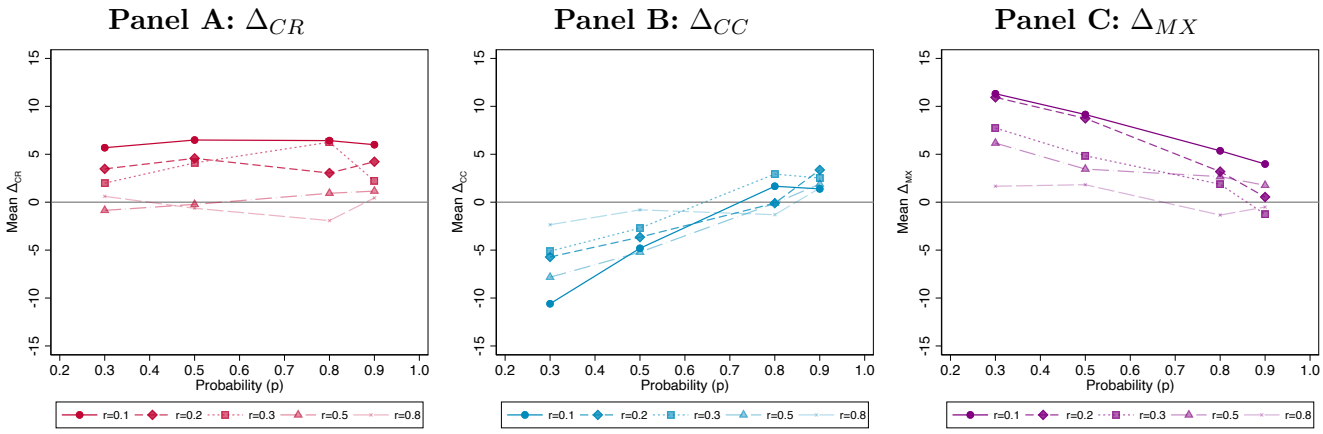
Looking across the three panels in Figure 3, these results yield variable patterns of preferences with two striking regularities. On one hand, near the canonical parameterizations for CR and CC problems (i.e., at $p = 0.8$ or 0.9 and $r = 0.1$, 0.2, or 0.3), we find both $\Delta_{CR} > 0$ and $\Delta_{CC} > 0$. Thus, CRP and CCP emerge at canonical parameter values, consistent with the findings from the paired-choice paradigm. In addition, at the canonical parameter values, Δ_{MX} tends towards zero such that the generally positive attitude towards mixtures observed in Panel C would be missed by focusing exclusively on the canonical parameterizations. On the other hand, for small p and small r, we find a robust pattern of $\Delta_{CR} > 0$, $\Delta_{CC} < 0$, and $\Delta_{MX} > 0$. This pattern contrasts sharply with the predictions of leading non-EU models in Table 1.

Tables 4 and 5 provide information on the statistical significance of the patterns in Figure 3. Table 4 presents the magnitudes of all 60 points in Figure 3 along with whether each is statistically different from zero using both mean and sign tests.[20,21] The table highlights that quantities in

[20] Appendix Table A.4 provides all means and sign tests with supporting details.

[21] 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 which we can directly compare 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

Figure 3: 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 all 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 420 observations. Expected utility predicts $\Delta = 0$ in all three panels.

excess of roughly 1.5 in absolute value in Figure 3 are statistically different from zero. Table 5 regresses each measure (Δ_{CR}, Δ_{CC}, and Δ_{MX}) on p and r, and confirms the patterns in Figure 3. Specifically, Δ_{CR} is larger for smaller r and is roughly independent of p, Δ_{CC} is larger for larger p and is roughly independent of r, and Δ_{MX} is larger for both smaller r and smaller p.[22]

Whereas Figure 3 and Tables 4 and 5 characterize mean responses, it is also important to investigate preferences at the individual level. To relate our data to the directional model predictions in Table 1, we focus on participants' directional responses; for example, whether a person exhibits $\Delta_{CR} > 0$, $\Delta_{CR} = 0$ (denoted $\oslash$CR), or $\Delta_{CR} < 0$ for CR problems. With three possible directional responses for each of the three measures of interest, there are 27 combinations. Figure 4 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 4 is both across participants and within participants across (p, r).

Figure 4 reveals that the mean preferences in Figure 3 and Tables 4 and 5 mask substantial heterogeneity. Response patterns that correspond to the standard behavioral hypothesis of people having both $\Delta_{CR} > 0$ and $\Delta_{CC} > 0$ are highlighted in green. Such observations constitute only a small fraction of the observed patterns (21.0 percent) and are no more frequent than those that exhibit the combination of $\Delta_{CR} > 0$ and $\Delta_{CC} < 0$ (21.3 percent). Indeed, as foreshadowed by Figure 3, the most frequently observed single response pattern is the combination of $\Delta_{CR} > 0$,

parameter choices, especially the choice of r. Relative to McGranaghan et al. (2024), the set of parameters considered in this paper yields slightly more cases in which preferences are in the direction of CRP, particularly for cases in which $r = 0.1$. This underscores the role low values of r play in generating evidence of CRP, which many non-EU models implicitly assume is a global phenomenon.

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

Table 4: Mean Δ_{CR}, Δ_{CC}, and Δ_{MX} by p and r

Panel A: Mean Δ_{CR}

	$p = 0.3$	$p = 0.5$	$p = 0.8$	$p = 0.9$
$r = 0.1$	$5.68^{*,\dagger}$	$6.49^{*,\dagger}$	$6.42^{*,\dagger}$	$6.00^{*,\dagger}$
$r = 0.2$	$3.48^{*,\dagger}$	$4.57^{*,\dagger}$	$3.04^{*,\dagger}$	$4.22^{*,\dagger}$
$r = 0.3$	$1.99^{*,\dagger}$	$4.10^{*,\dagger}$	$6.26^{*,\dagger}$	$2.23^{*,\dagger}$
$r = 0.5$	-0.85	-0.23	0.93	1.16
$r = 0.8$	0.61	-0.63	$-1.92^{*,\dagger}$	0.45

Panel B: Mean Δ_{CC}

	$p = 0.3$	$p = 0.5$	$p = 0.8$	$p = 0.9$
$r = 0.1$	$-10.60^{*,\dagger}$	$-4.81^{*,\dagger}$	1.66	1.39
$r = 0.2$	$-5.72^{*,\dagger}$	$-3.65^{*,\dagger}$	-0.10	$3.36^{*,\dagger}$
$r = 0.3$	$-5.11^{*,\dagger}$	$-2.70^{*,\dagger}$	2.93^{*}	2.52^{*}
$r = 0.5$	$-7.83^{*,\dagger}$	$-5.22^{*,\dagger}$	-0.11	1.89^{*}
$r = 0.8$	$-2.35^{*,\dagger}$	$-0.80^{\dagger}$	$-1.31^{\dagger}$	1.46^{*}

Panel C: Mean $\Delta_{CR} - \Delta_{CC}$

	$p = 0.3$	$p = 0.5$	$p = 0.8$	$p = 0.9$
$r = 0.1$	$16.28^{*,\dagger}$	$11.30^{*,\dagger}$	$4.76^{*,\dagger}$	$4.61^{*,\dagger}$
$r = 0.2$	$9.19^{*,\dagger}$	$8.22^{*,\dagger}$	$3.14^{*,\dagger}$	$0.86^{\dagger}$
$r = 0.3$	$7.10^{*,\dagger}$	$6.80^{*,\dagger}$	$3.32^{*,\dagger}$	$-0.29^{\ddagger}$
$r = 0.5$	$6.98^{*,\dagger}$	$4.99^{*,\dagger}$	1.04	-0.73
$r = 0.8$	$2.96^{*,\dagger}$	0.17	-0.62	-1.01

Panel D: Mean Δ_{MX}

	$p = 0.3$	$p = 0.5$	$p = 0.8$	$p = 0.9$
$r = 0.1$	$11.32^{*,\dagger}$	$9.15^{*,\dagger}$	$5.36^{*,\dagger}$	$3.98^{*,\dagger}$
$r = 0.2$	$10.94^{*,\dagger}$	$8.74^{*,\dagger}$	$3.19^{*,\dagger}$	0.54
$r = 0.3$	$7.74^{*,\dagger}$	$4.85^{*,\dagger}$	$1.87^{*,\dagger}$	-1.26
$r = 0.5$	$6.16^{*,\dagger}$	$3.45^{*,\dagger}$	$2.66^{*,\dagger}$	$1.76^{\dagger}$
$r = 0.8$	$1.67^{*,\dagger}$	$1.82^{*,\dagger}$	-1.34	-0.50

Notes: Table presents mean values along with corresponding hypothesis tests for $\Delta_{CR} = h_{AB} - h_{DE}$, $\Delta_{CC} = h_{AB'} - h'_{DE}$, $\Delta_{CR} - \Delta_{CC}$, and $\Delta_{MX} = h'_{AB} - h'_{AB'}$. Panels A, B, and C aggregate data across all 8,408 observations, with each entry corresponding to roughly 420 observations. Panel D aggregates across the 4,204 observations for which we elicit h'_{AB} and $h'_{AB'}$, with each entry corresponding to roughly 210 observations. Expected utility null hypothesis corresponds to zero mean or zero sign difference. * denotes that the value is significantly different from zero at the 5 percent level using a means test. † denotes a significant deviation in the direction of the reported sign at the 5 percent level using a sign test.

Table 5: Predicting the Prevalence of CR, CC, and MX by p and r

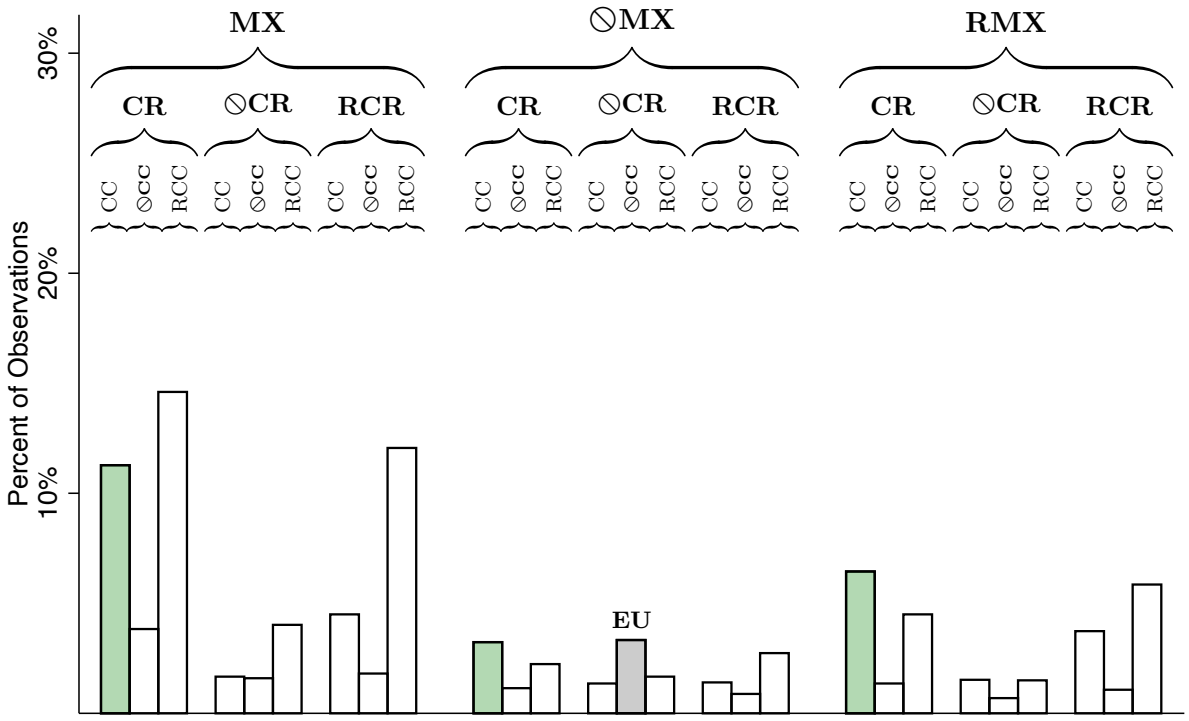
	(1)	(2)	(3)	(4)
Outcome:	Δ_{CR}	Δ_{CC}	$\Delta_{CR} - \Delta_{CC}$	Δ_{MX}
Probability (p)	1.00	13.96	−12.96	−10.95
	(0.67)	(0.71)	(0.92)	(0.95)
Common Ratio (r)	−9.16	1.82	−10.97	−9.24
	(0.66)	(0.74)	(0.84)	(0.86)
Outcome Mean	2.74	−1.72	4.46	4.07
Observations	8,408	8,408	8,408	4,204

Notes: Table presents ordinary least squares regressions of $\Delta_{CR} = h_{AB} - h_{DE}$, $\Delta_{CC} = h_{AB'} - h'_{DE}$, $\Delta_{CR} - \Delta_{CC}$, and $\Delta_{MX} = h'_{AB} - h'_{AB'}$ on experimental parameters (p, r). Columns (1)-(3) use all 8,408 observations from 2,102 participants, while column (4) uses 4,204 observations from 2,102 participants. Specification also includes a constant that is not reported. Standard errors clustered at individual level in parentheses.

$\Delta_{CC} < 0$, and $\Delta_{MX} > 0$, though even this modal pattern accounts for only 14.6 percent of observations.[23]

[23] Appendix Figures A.1 to A.3 depict the distributions of responses for various (p, r) subsets. Near the canonical

Figure 4: Histogram of Response Patterns



Notes: Figure presents histogram of $(sign(\Delta_{CR}), sign(\Delta_{CC}), sign(\Delta_{MX}))$ combinations, where $\Delta_{CR} = h_{AB} - h_{CD}$, $\Delta_{CC} = h_{AB'} - h'_{DE}$, 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 $\Delta_{CR} > 0$ and $\Delta_{CC} > 0$, a frequent prediction of leading behavioral models (see Table 1).

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

4.2 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. We then use these estimates for two purposes. First, we assess how much of the variability in our data is due to heterogeneity in preferences versus noise. Second, we derive what the distribution of response patterns from Figure 4 would look like in the absence of decision noise.

To disentangle noise from preferences, we leverage the fact that we collect multiple elicitations

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 13.8 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 now constitutes 21.4 percent. However, there is substantial variability at all problem parameters.

for each valuation. Intuitively, if the noise is independent across the two elicitations, then the covariance between them is solely determined by the heterogeneity in preferences. We provide a brief overview of our approach here, while 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 we define 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 two elicited 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 if we had only a single elicitation for the XY valuation, then we would be unable to decompose its variance into the separate preference and noise components. The second equation illustrates our central intuition: If the noise draws for the two elicitations of the XY valuation are independent, then the covariance between those two elicitations identifies the variance of the preference heterogeneity.

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 representation 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. Appendix Table A.5 reports our estimates of these 12 terms for each (p, r) combination.[24]

Given these estimates, we can assess how much of the variability in our data is due to hetero-

[24] Appendix D.5 describes a more sophisticated estimation using MLE. Because that approach requires additional distributional and implementation assumptions, we prefer the approach described in the text. However, the MLE approach yields virtually identical conclusions.

geneity in preferences versus noise. Consider first the variability in the indifference values h_{AB}, $h_{AB'}$, and h_{CD}. The last three columns of Appendix Table A.5 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}. Appendix Table A.6 uses the estimates in Appendix Table A.5 to calculate these six variance terms for each (p, r) combination.[25] The last three columns of Appendix Table A.6 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}, 31 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 4 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 Appendix Table A.5 to generate 100,000 draws from a joint normal distribution for $(h^*_{AB}, h^*_{AB'}, h^*_{CD})$. We then convert each draw into a Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX}.[26] This approach allows us to isolate the preference patterns underlying the response patterns depicted in Figure 4.

Figure 5 reproduces the distribution of response patterns in Figure 4 but adds black dots to denote the distribution of preference patterns that we simulate using the approach described above. Note that any pattern that implies an intransitivity in the underlying preferences $(h^*_{AB}, h^*_{AB'}, h^*_{CD})$ cannot emerge from this simulation.[27] Of the 27 possible response patterns, 14 of them imply an intransitivity. Neither our decomposed preferences nor any transitive theory of preferences would be able to accommodate such patterns, which are marked in gray in Figure 5.[28]

[25] When using this approach, nothing guarantees that the calculated $var(\Delta^*_Z) > 0$, and indeed there is one instance where this problem arises (for Δ_{MX} when $(p, r) = (0.3, 0.5)$). We ignore this case and focus on the other 59 cases.

[26] Specifically, we convert each h^* draw into the midpoint of its two closest integers (e.g., any draw strictly between \$2 and \$3 is converted to \$2.50). We then use these adjusted h^* terms to generate the Δ^* terms. Note that this approach permits simulations of the Δ^* terms to be zero. See Appendix D.3 for complete details.

[27] For example, pattern RCRP-CCP-MXP would require $h^*_{CD} > h^*_{AB}$, $h^*_{AB'} > h^*_{CD}$, and $h^*_{AB} > h^*_{AB'}$, and this combination could never emerge from a single realization of $(h^*_{AB}, h^*_{AB'}, h^*_{CD})$.

[28] 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 may give an indication of the prevalence of noise relative to preferences in our data. 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.

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 exercise permits both strict and weak patterns, and indeed, all 13 patterns have a positive share of simulated preferences. Figure 5 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 preference patterns, but only 44 percent of response patterns. Our decomposition and simulation exercise highlights 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 preferences and 11 percent of observations. According to our decomposition, these patterns would be slightly less frequent in the absence of decision noise. Together, the seven marked patterns account for 81 percent of preferences and 55 percent of responses, representing the majority of both. As we will see, these marked patterns, many of which are inconsistent with the prominent non-EU theories in Table 1, emerge as the seven patterns predicted by our model of upside potential presented in Section 5.

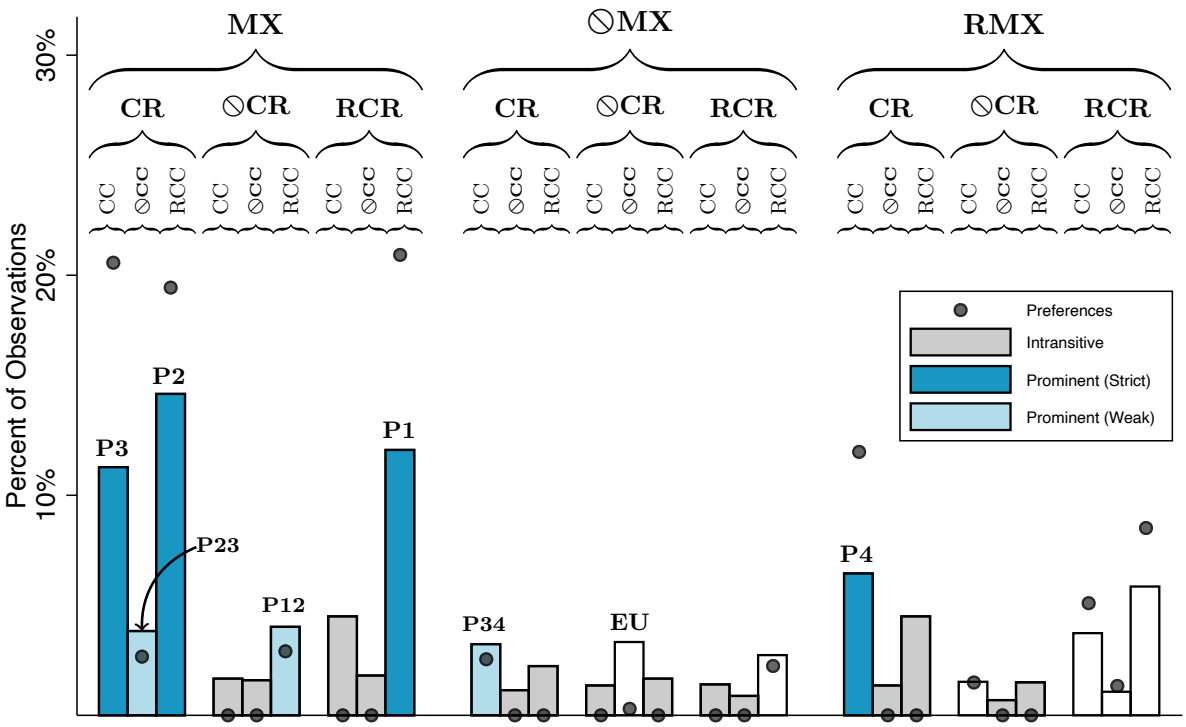
4.3 Choices versus Valuations

Some researchers have argued that binary choice tasks are more reliable than valuations for eliciting preferences due to their simplicity and transparency (e.g., Brown and Healy, 2018 and Freeman et al., 2019). In this section, we use our stage 2 binary choice data to address this critique, and we further show that, in fact, our stage 2 choice data yield much the same conclusions about average preferences as our stage 1 valuations data.

In our design, we can directly link each binary choice that a participant sees in stage 2 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. In other words, each stage 2 binary choice corresponds to a specific row of a stage 1 price list.

Panels A-C of Figure 6 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 collected into percentile bins, and the vertical axis then reflects the average stage 2 $CRE-RCRE$ within that bin for the corresponding stage 2 binary choice tasks. Panels B and C are analogous for the Δ_{CC} and Δ_{MX} measures. Across Panels A-C, we see that preferences measured using stage 1 valuations 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} measures would predict, less likely to exhibit the CCE pattern in choices than their Δ_{CC} measures would predict, and more likely to exhibit the MXE pattern in choices than their Δ_{MX} measures would predict. Nonetheless, clearly choices and valuations are both capturing meaningful

Figure 5: Preference Patterns vs. Data Patterns



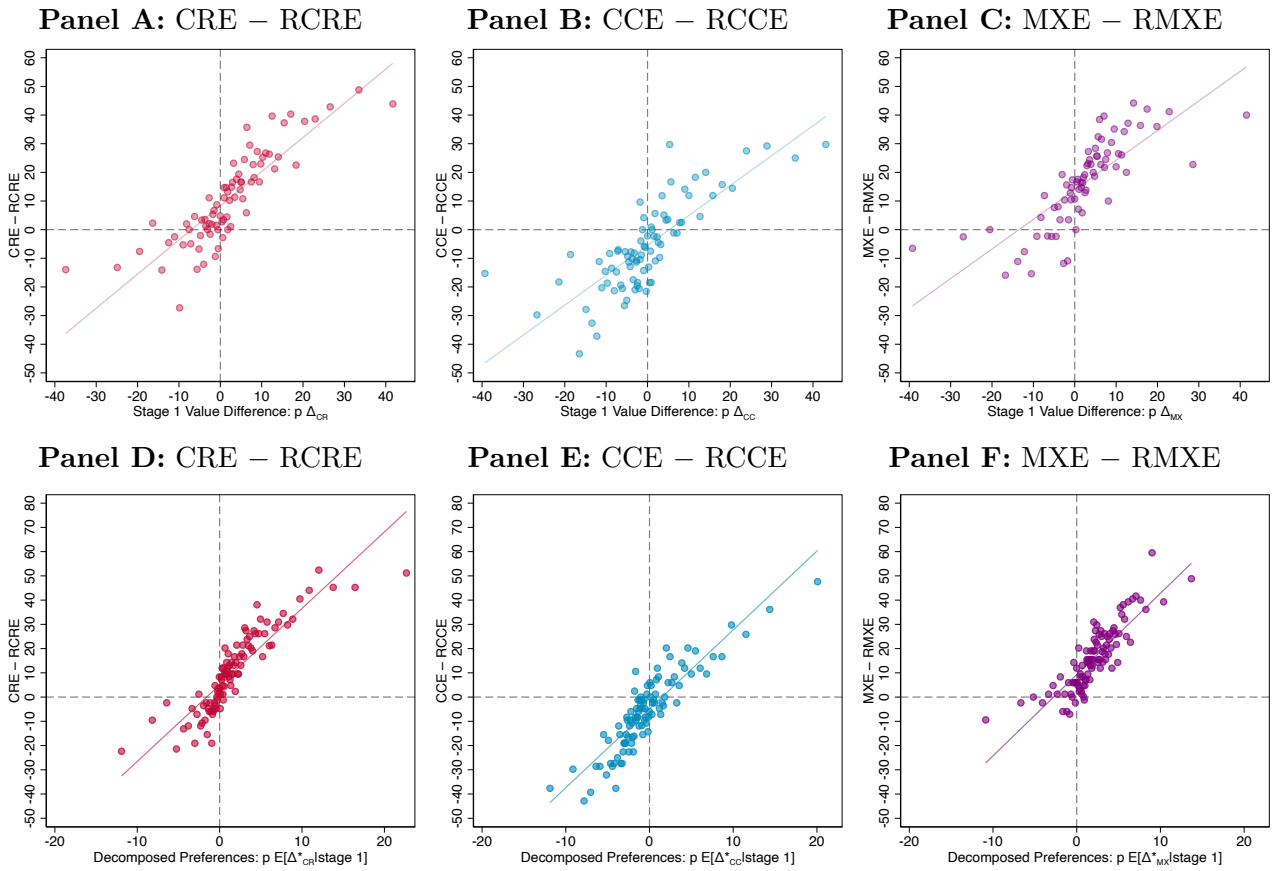
Note: Figure presents histograms of $(sign(\Delta_{CR}), sign(\Delta_{CC}), sign(\Delta_{MX}))$ combinations, where $\Delta_{CR} = h_{AB} - h_{CD}$, $\Delta_{CC} = h_{AB'} - h'_{DE}$, and $\Delta_{MX} = h'_{AB} - h'_{AB'}$, for the 4,204 observations for which we elicit h'_{AB} and $h'_{AB'}$. Each variable 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 ⦸CR to $\Delta_{CR} = 0$). Bars represent the share of raw responses and are identical to those in Figure 4. Black dots represent the share of simulated preferences following the decomposition and simulation described in Section 4.2. 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.

information about the underlying preferences of interest.[29]

Panels D-F of Figure 6 highlight how the link between stage 1 valuations and stage 2 choices becomes even tighter when we use our decomposition from Section 4.2 to reduce the noise in participants' stage 1 valuations. Specifically, we combine our estimated population distribution of preferences with a participant's elicited stage 1 valuations to generate posterior expectations (from the perspective of the analyst) for that participant's $(h^*_{AB}, h^*_{AB'}, h^*_{CD})$ (see Appendix D.4 for details). We then convert those valuations into posterior expectations for that participant's $(\Delta^*_{CR}, \Delta^*_{CC}, \Delta^*_{MX})$, and use these in Panels D-F. Panels D-F further demonstrate how choices and valuations both capture meaningful information about the underlying preferences of interest.

[29] 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 6: Predicting Stage 2 Results using Stage 1 Valuations

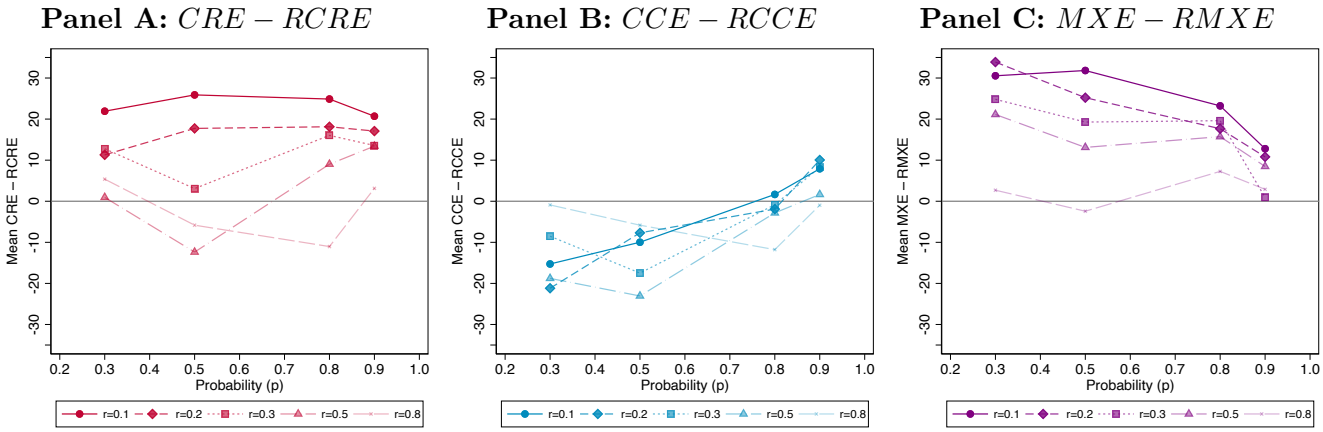


Notes: Figure relates individual measures of stage 1 Δ_{CR}, Δ_{CC}, and Δ_{MX} to stage 2 measures of $CRE - RCRE$, $CCE - RCCE$, and $MXE - RMXE$, respectively. Panels A-C use raw stage 1 responses. Panels D-F use the estimated population distribution of preferences from the decomposition in Section 4.2 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 82, 78, 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 Appendix C.3 for details).

In terms of the aggregate results, our stage 1 valuations and stage 2 choices yield the same qualitative conclusions regarding our three main preferences of interest. Figure 7 is the analog for Figure 3 when we use choices instead of valuations, where the measures on the vertical axes are now $CRE - RCRE$, $CCE - RCCE$, and $MXE - RMXE$. For each (p, r) combination, we combine choice data across all payment values for H. While Figure 7 is not as stark as Figure 3, we still see our main conclusions regarding average preferences: The CR pattern is largely invariant to p, with a CRE emerging for small r but not for large r; the CC pattern is largely invariant to r, where there is a mild CCE for large p but a strong $RCCE$ for small p; and the MX pattern depends on

both p and r, where an MXE emerges quite broadly, and it is substantially stronger for smaller p and smaller r. Appendix Table A.7 is analogous to Table 5 and confirms the statistical significance of these conclusions.

Figure 7: Mean Effects by p for each r



Notes: Figure depicts average values of $CRE-RCRE$, $CCE-RCCE$, and $MXE-RMXE$. Each panel aggregates over all 8,408 observations. Each point in each panel aggregates over the six payment values for H noted in Table A.2 and corresponds to approximately 420 observations. The line for zero choice difference is indicated in all three panels but only serves as a reference (zero does not correspond to the expected utility null hypothesis unless distance to indifference is zero on average across the two choice tasks in each problem).

Although our stage 2 choices yield the same qualitative message as our stage 1 valuations, Section 2.5 highlights a major drawback of using binary choice tasks: When comparing behavior across binary choice tasks, conclusions can be biased if noise has a differential impact across those binary choice tasks. In McGranaghan et al. (2024), we demonstrated 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 Appendix C.3, we perform a similar analysis using the data from the current experiment. We find evidence of differential noise on all three dimensions. For CR problems, we again find that noise has a larger impact on the CD decision than on the AB decision. For CC problems, we find that noise has a larger impact on the AB' decision than on the CD decision. For the MX problems, we find that noise has a larger impact on the AB' decision than on the AB decision.[30] Given that we find evidence of differential noise for all three problems, it is perhaps surprising that our choice data yield the same conclusions as our valuations data. This tight connection likely stems from our deliberate choice to balance parameters for stage 2; specifically, it likely results from our choice of a balanced set of possible payment values for H to use for each (p, r, M).

Furthermore, our results from stage 2 highlight an additional drawback of binary choice data:

[30]Interestingly, these conclusions differ from the predictions of EU with additive i.i.d utility noise, which would imply 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 Appendix C.3 for more details.

We can only make statements on qualitative patterns for behavior. Converting these into statements about the magnitudes of preferences is difficult because the differential noise influences them. Hence, while choices provide a valuable method to validate our main results, valuations are more robust and informative in studying risk preference patterns across tasks.

5 Proposed Rationalization: A Preference for Upside Potential

The modal pattern of CRP-RCCP-MXP and the more general prevalence of MXP violate both EU and leading non-EU models in Table 1. To rationalize our data, one might naturally turn to prospect theory, and indeed there is some scope to broaden the predictions of prospect theory beyond those in Table 1 by considering more flexible functional forms. However, we show in Section 5.3 that such an approach struggles to explain the patterns in our data. In the absence of a rationalizing theory, we speculate on a possible model that might be able to explain those patterns.

We develop a model based on introspection about why people might exhibit an MXP while also exhibiting a CRP and RCCP. Encouragingly, this model is also able to explain some of the other key patterns in our data at both the aggregate and individual levels. While this model is post-hoc and preliminary, we think our analysis in this section can provide insight for building 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 average valuations are $h_{AB} = 38$, $h_{AB'} = 29$ and $h_{CD} = 33$. These valuations imply:

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

Notice that the AB indifference implies risk aversion: Relative to lottery A, choosing lottery B yields additional expected value at the cost of additional 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 sounds a lot like the usual EU intuition from insurance and finance settings for a person deciding how much risk to take on. 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. Put differently, this preference suggests that indifference curves are convex in the Marschak-Machina triangle rather than linear as in EU.

We propose an alternative psychology: Rather than (or perhaps in addition to) caring about risk in the usual EU way, people decide how much *upside potential* they want by trading off (i)

the total probability of winning *something* versus (ii) expected winnings conditional on winning something. Applied here, the attraction of B' relative to B is that it increases the total probability of winning something. At the same time, the attraction of B' relative to A is that it increases expected winnings conditional on winning something.

To formalize this intuition, we assume that a person evaluates lottery $(X, q_X; Y, q_Y)$ with $X, Y > 0$ as

$$U = \underbrace{[q_X u(X) + q_Y u(Y)]}_{EU} \quad + \quad \underbrace{f(q_X + q_Y)}_{\substack{\text{weighted} \\ \text{upside probability}}} \quad \underbrace{\left[\frac{q_X \kappa(X) + q_Y \kappa(Y)}{q_X + q_Y}\right]}_{\substack{\text{conditional expected} \\ \text{valuation of upside}}}. \tag{2}$$

The first term is a standard expected utility component, whereas the second term captures preferences for upside potential. The latter is the product of the weighted total probability of winning something and the expected valuation of those winnings conditional on winning something. The function $f(q)$ permits nonlinearities in how the person cares about the probability of winning something. The function $\kappa(x)$ captures the person's relative valuation of different winning amounts.[31]

Before proceeding, we clarify two points about this model. First, while these preferences perhaps look like gain-loving preferences (i.e., the opposite of loss aversion), they are distinct because all winning probabilities are combined and then transformed via f. We further discuss the distinctions between our model and prospect theory in Section 5.3. Second, an important issue for this formulation is what outcomes are included in the set of winning outcomes. For our experiment, a natural assumption is that all positive outcomes are included in the set of winning outcomes, and we explore here the implications of this assumption.[32]

In equation (2), if $f(q) = q$ then the model effectively reduces to EU.[33] To generate an MXP, we need f to be convex so that, when considering whether to secure larger expected winnings conditional on winning something, giving up some probability of winning is not too costly (so one is willing to move from A to B'), while giving up too much probability of winning is too costly (so one is not willing to move from B' to B).

To generate a simple, tractable version of the model with a convex f, we consider the model predictions when $f(q) = q^2$ and $u(x) = x$, in which case

$$U = [q_X X + q_Y Y] + (q_X + q_Y)\,[q_X \kappa(X) + q_Y \kappa(Y)]\,. \tag{3}$$

For this parameterized version of the model, the function κ is the free dimension that we can adapt to explain variation in behavior across individuals or across (p, r) combinations.

[31] This formulation assumes $\kappa(0) = 0$; it permits $\kappa(x) = u(x)$ but also allows for other possibilities.

[32] Future analyses that apply our intuition of preferences for upside potential to broader environments will need to address this issue, which is analogous to the issue of what reference point to use in prospect theory and other reference-dependent models. As in that literature, one might assume some exogenous rule, or that a person's assessment of which outcomes are winning outcomes depends on features of the choice set in the spirit of Kőszegi and Rabin, 2007.

[33] In particular, equation (2) becomes $U = q_H[u(H) + \kappa(H)] + q_M[u(M) + \kappa(M)]$, which is just EU with Bernoulli utility function $u(x) + \kappa(x)$.

To ensure that preferences are continuous, we apply equation (3) even when q_H or q_M goes to zero or one. Hence, for a binary lottery (X, q) with $X > 0$ (as in our lotteries B, C, and D),

$$U = qX + q^2\kappa(X),$$

and for a certain lottery $(X, 1)$ (as in our lottery A),

$$U = X + \kappa(X).$$

The former generates effects reminiscent of Allais' original intuition regarding the non-substitutability of probability units and the source of certainty effects. Allais (1952) argues that attitudes towards a 10% chance of winning a prize might depend on whether one already "owns" the other 90%, implying important complementarities between the first 90% and the last 10% of winning.[34] Such complementarities emerge naturally from our formulation.

Applying the model to our experiment, the triplet $(h^*_{AB}, h^*_{AB'}, h^*_{CD})$ solves

$$\begin{aligned} M + \kappa(M) &= ph^*_{AB} + p^2\kappa(h^*_{AB}) && (4)\\ M + \kappa(M) &= prh^*_{AB'} + (1-r)M + (pr + 1 - r)\left[pr\kappa(h^*_{AB'}) + (1-r)\kappa(M)\right] && (5)\\ rM + r^2\kappa(M) &= prh^*_{CD} + (pr)^2\kappa(h^*_{CD}). && (6) \end{aligned}$$

In 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 (4), (5), and (6). For any $(p, r) \in (0, 1)^2$ and $\kappa(x)$ that is strictly increasing in x:

(1) A person's Δ^*_{CR}, Δ^*_{CC}, and Δ^*_{MX} satisfy:

(a) $\Delta^*_{CR} > 0$ if and only if $\kappa(M) > p^2\kappa(h^*_{AB}) > p^2\kappa(h^*_{CD})$;
$\Delta^*_{CR} < 0$ if and only if $\kappa(M) < p^2\kappa(h^*_{AB}) < p^2\kappa(h^*_{CD})$; and
$\Delta^*_{CR} = 0$ if and only if $\kappa(M) = p^2\kappa(h^*_{AB}) = p^2\kappa(h^*_{CD})$.

(b) $\Delta^*_{CC} > 0$ if and only if $\kappa(M) > \left(\frac{p}{2-p}\right)\kappa(h^*_{AB'}) > \left(\frac{p}{2-p}\right)\kappa(h^*_{CD})$;
$\Delta^*_{CC} < 0$ if and only if $\kappa(M) < \left(\frac{p}{2-p}\right)\kappa(h^*_{AB'}) < \left(\frac{p}{2-p}\right)\kappa(h^*_{CD})$; and
$\Delta^*_{CC} = 0$ if and only if $\kappa(M) = \left(\frac{p}{2-p}\right)\kappa(h^*_{AB'}) = \left(\frac{p}{2-p}\right)\kappa(h^*_{CD})$.

(c) $\Delta^*_{MX} > 0$ if and only if $\kappa(M) < p\kappa(h^*_{AB'}) < p\kappa(h^*_{AB})$;
$\Delta^*_{MX} < 0$ if and only if $\kappa(M) > p\kappa(h^*_{AB'}) > p\kappa(h^*_{AB})$; and
$\Delta^*_{MX} = 0$ if and only if $\kappa(M) = p\kappa(h^*_{AB'}) = p\kappa(h^*_{AB})$.

[34] Translated from its original formulation in French, Allais' quote reads: "It seems impossible to me to think that, if I am cautious, the psychological weight given to a one-in-ten chance to win 100 million should be the same whether or not I already own the remaining nine-in-ten chance. *There is in reality an extremely pronounced complementarity effect that must be taken into account. This effect is not in itself irrational.*" (p.39, emphasis is Allais'.)

(2) $\Delta^*_{CR} \leqslant 0$ implies $\Delta^*_{CC} < 0$ and $\Delta^*_{MX} > 0$, and $\Delta^*_{CC} \leqslant 0$ implies $\Delta^*_{MX} > 0$. (Equivalently, $\Delta^*_{MX} \leqslant 0$ implies $\Delta^*_{CR} > 0$ and $\Delta^*_{CC} > 0$, and $\Delta^*_{CC} \geqslant 0$ implies $\Delta^*_{CR} > 0$.)

(3) The person must exhibit one of the following seven patterns of behavior:[35]

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).

Part 1 of Proposition 1 characterizes conditions for when a person has a CRP, a CCP, and an MXP. These conditions are derived directly from combinations of equations (4) through (6). While these conditions are expressed in terms of the endogenous variables h^*_{AB} and $h^*_{AB'}$, part 2 uses these conditions to restrict the set of combinations of preferences that can arise from the model. For instance, one can easily see that the conditions in part 1 for $\Delta^*_{CR} \leqslant 0$ imply that we must have $\Delta^*_{CC} < 0$ and $\Delta^*_{MX} > 0$ given that $p^2 < p$ and $p^2 < p/(2-p)$ for any p. Finally, Part 3 lists the seven patterns of behavior that are consistent with part 2.

It is instructive to highlight the predictions of our model when $\kappa(x) = \phi x$ for some $\phi > 0$. Proposition A2 in Appendix E.1 establishes for this case that a person must exhibit P2, P23, or P3. This includes our modal pattern CRP-RCCP-MXP, and more generally requires that a person have an MXP. Hence, an MXP arises in this framework not because of specific assumptions on the shape of $\kappa(\cdot)$, but rather due to the way that probabilities enter the upside-potential preferences (in particular, due to f being convex).

That said, a linear κ function cannot explain all patterns in our data, and cannot even explain all instances where our modal pattern emerges. For instance, in our motivating case where $p = 0.5$, $r = 0.2$, and the average valuations are $h_{AB} = 38$, $h_{AB'} = 29$ and $h_{CD} = 33$, part 1 of Proposition 1 implies the $\kappa(\cdot)$ function must satisfy:

$$\frac{1}{2}\kappa(29) > \frac{1}{3}\kappa(29) > \kappa(15) > \frac{1}{4}\kappa(38).$$

These conditions can be satisfied only if κ is initially convex and then concave.[36] We explore this point in more detail in Section 5.2.3.

[35] 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.

[36] Specifically, one can combine the second and third inequalities to derive that $(\kappa(29)-\kappa(15))/14 > (\kappa(15)-\kappa(0))/15$ and $(\kappa(29)-\kappa(15))/14 > (\kappa(38)-\kappa(29))/9$. Thus, κ must be on average convex between $x = 0$ and $x = 29$, and on average concave between $x = 15$ and $x = 38$.

5.2 Exploring Additional Model Predictions

While we developed our model to explain why people might exhibit an MXP while also exhibiting a CRP and RCCP, we next assess the extent to which our model can account for other patterns in the data at both the aggregate and individual levels

5.2.1 Distribution of Behavior

Figure 5 in Section 4.2 provides a histogram of observations across the 27 possible preference patterns that might emerge, along with a histogram of decomposed preferences across the 13 possible transitive preference patterns that might emerge. As we discuss in Section 4, 55 percent of raw response patterns and 81 percent of decomposed preference patterns reflect just seven distinct patterns of behavior. These patterns are denoted in Figure 5 with the labels $P1$, $P12$, $P2$, $P23$, $P3$, $P34$, and $P4$ because they correspond to the patterns indicated in part 3 of Proposition 1. Thus our model is able to rationalize not only the modal pattern of CRP-RCCP-MXP for which it was created, but also a substantial component of raw response and decomposed preference combinations that arise across individuals and across (p, r) combinations.

Importantly, the model does not permit every combination. In addition to those patterns that are ruled out by any theory (i.e., those that violate transitivity, indicated in gray in Figure 5), our model also rules out six more preference patterns (unshaded patterns in Figure 5). Intransitive patterns ruled out by any noise-free theory of choice account for 26 percent of raw response patterns (and 0 percent of decomposed preferences). Patterns additionally ruled out by our theory account for 18 percent of raw response patterns and 19 percent of decomposed preference patterns. Assuming the extent of intransitive choice patterns is a reasonable benchmark for the extent of noise-based observations in the data, it is encouraging to see that patterns ruled out by our model occur at a rate even lower than that generated by noise.

5.2.2 Link to Risk Aversion

An unexpected prediction of our model is that the condition for CRP from part 1 of Proposition 1 coincides with the condition for risk aversion in the AB task. That is, an individual will be risk averse in the AB task (i.e., will have h^*_{AB} such that $ph^*_{AB} > M$) if and only if $\kappa(M) > p^2\kappa(h^*_{AB})$ (see Proposition A3 in Appendix E.1). Though at the aggregate level we do not see mean risk seeking at any (p, r) combination, we can assess whether this prediction is supported at the individual level. Specifically, we use the approach described in Section 4.3 to relate the sign of risk aversion, $E[ph^*_{AB} - M|\text{stage 1}]$, to the sign of CRP, $E[\Delta^*_{CR}|\text{stage 1}]$. Roughly 13 percent of observations exhibit risk tolerance, $E[ph^*_{AB} - M|\text{stage 1}] < 0$; of these, the majority (54%) exhibit RCRP. The remaining 87 percent exhibit risk aversion; of these, the majority (69%) exhibit CRP. Thus, risk aversion is significantly correlated with common ratio preferences as our model of upside potential predicts (Fisher's exact test $p < 0.001$).

5.2.3 Variation in Behavior Across (p, r) Combinations

We next assess the model's ability to explain the variation in behavior across different (p, r) combinations. To do so, we use our data to estimate the κ function, which is the only free "parameter" in equation (3). We then assess the fit of the estimated model by comparing the model-predicted h^* and Δ^* values to the actual h and Δ in the data.

To keep this analysis tractable, we do not focus on behavior at the individual level.[37] Rather, as "data" for the estimation we use the mean responses for h_{AB}, $h_{AB'}$, and h_{CD} for each of the 20 (p, r) combinations, thus yielding 60 observations. Equations (E.2)-(E.4) implicitly define h^*_{AB}, $h^*_{AB'}$, and h^*_{CD} as a function of the experimental parameters (p, r, M) and the parameters of the function $\kappa(z)$. Our general approach specifies a function $\kappa(z;\theta)$ where θ is a vector of parameters to be estimated. We estimate our model on the 60 observations assuming each mean response $h_{XY} = h^*_{XY}(p, r, M, \theta) + \epsilon$ using non-linear least squares to provide an estimate of $\kappa(z;\theta)$. In Appendix E.2, we describe the estimation technique in detail and provide corresponding parameter estimates.

In the absence of an a priori sense of the shape of κ, we begin with a flexible functional form. 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. Hence, we specify a five-parameter piecewise-linear functional form for κ with kinks at 9, 15, 24, and 27.[38] The resulting estimate of $\kappa(z;\theta)$ shown in Panel A of Figure 8 exhibits an S-shape, featuring a convex segment for low values of x followed by concavity at 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 an $R^2 = 0.76$ and an $MSE = 3.53$.[39] The correlation between predicted and actual h's is an impressive 0.91, and the correlation between predicted and actual Δ's is 0.90.[40]

Motivated by the S-shape in panel A of Figure 8, we propose the following more parsimonious and analytically tractable three-parameter sigmoid function:

$$\kappa(z;\theta) = \theta_1 * \left[\frac{1}{1 + exp(\theta_2(z - \theta_3))}\right] - \theta_1 * \left[\frac{1}{1 + exp(\theta_2(0 - \theta_3))}\right].$$

In this formulation, the first bracketed term is a classic two-parameter sigmoid function (with parameters θ_2 and θ_3) that goes from zero (as $x \to -\infty$) to one (as $x \to \infty$). The third parameter (θ_1) is a multiplier on the bracketed term that makes the first term instead go from zero to θ_1.

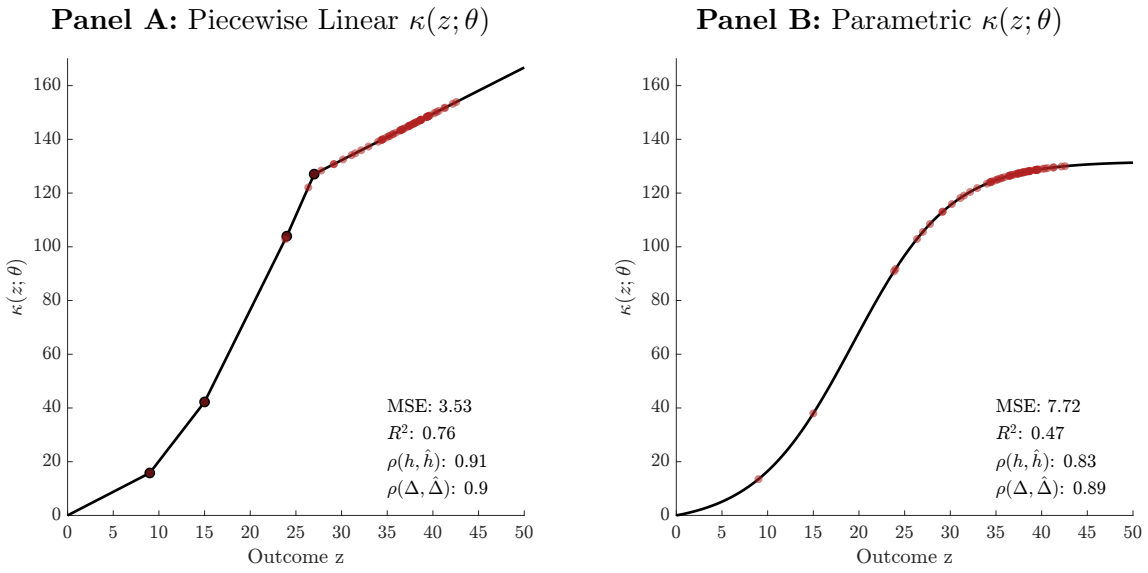
[37] In light of the variation in Figure 5, estimating our model at the individual level would require permitting heterogeneity in the κ function, which in turn would require making assumptions about the nature of that heterogeneity and would also substantially increase the computational burden.

[38] In Appendix E.2, we also report results when we add an additional kink at 36. Because that model performs roughly the same, the five-parameter model discussed in the text is our preferred model.

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

[40] Panels B and C of Appendix Figure E.2 show that these pairs are clustered relatively close to the 45-degree line.

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



Notes: Figure presents estimates for $\kappa(z;\theta)$ under our upside-potential 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 Appendix E.2.2 for details.

Finally, the second term subtracts off the value of the first term when it is evaluated at $x = 0$; including this term imposes that $\kappa(0) = 0$. Panel B of Figure 8 depicts the estimated κ given this specification. This more parsimonious version retains substantial predictive accuracy with $R^2 = 0.47$, $MSE = 7.72$, a correlation between predicted and actual h values of 0.83, and a correlation between predicted and actual Δ values of 0.89.

Overall, our model captures important variation in aggregate behavior—both the levels and the differences—across different (p, r) combinations, even in its relatively parsimonious form with only three degrees of freedom.

5.3 Can We Explain Our Data with Prospect Theory?

While our model of upside potential is effective at capturing the main data patterns, it is important to consider how its performance compares to other models. In particular, because there is some scope to broaden the predictions of prospect theory beyond those in Table 1 by considering more flexible functional forms for the probability weighting function $\pi(q)$, one might wonder whether we

can explain our data with prospect theory.

In Appendix E.2.3, we use the same approach as in Section 5.2.3 to estimate several variants of prospect theory. We use both original prospect theory (OPT) and cumulative prospect theory (CPT). For both approaches, we assume the value function takes the form $v(x) = x^{\alpha}$ and estimate the parameter α in addition to the parameters of the probability weighting function. We first confirm that commonly used functional forms for $\pi(q)$ perform poorly, where we consider both the one-parameter version from Tversky and Kahneman (1992) and the two-parameter version from Lattimore et al. (1992). The best-performing model was CPT with the two-parameter functional form. Panel A of Figure 9 depicts the estimated $\pi(q)$ function for that model. The in-sample fit is indeed poor, with a negative in-sample R^2 of -0.23, $MSE = 18.02$, a correlation between predicted and actual h's of 0.55, and a correlation between predicted and actual Δ's of 0.70.[41]

We next considered a flexible six-part piece-wise linear functional form that permits (but does not require) discontinuities at $q = 0$ and $q = 1$. For both OPT and CPT, we put the kinks at the values of q where $\pi(q)$ is frequently evaluated while also trying to have similar numbers of instances within each segment. The best-performing model was the CPT version, and the estimated $\pi(q)$ function for that model is depicted in panel B of Figure 9. We find the classical pattern of overweighting of low probabilities and underweighting of high probabilities, although substantially more overweighting appears than generally estimated. Despite its flexibility and additional degrees of freedom, even this prospect theory model substantially underperforms both estimated models of upside potential from Figure 8, with an $R^2 = 0.25$, $MSE = 11.02$, a correlation between predicted and actual h values of 0.71, and a correlation between predicted and actual Δ values of 0.72.[42]

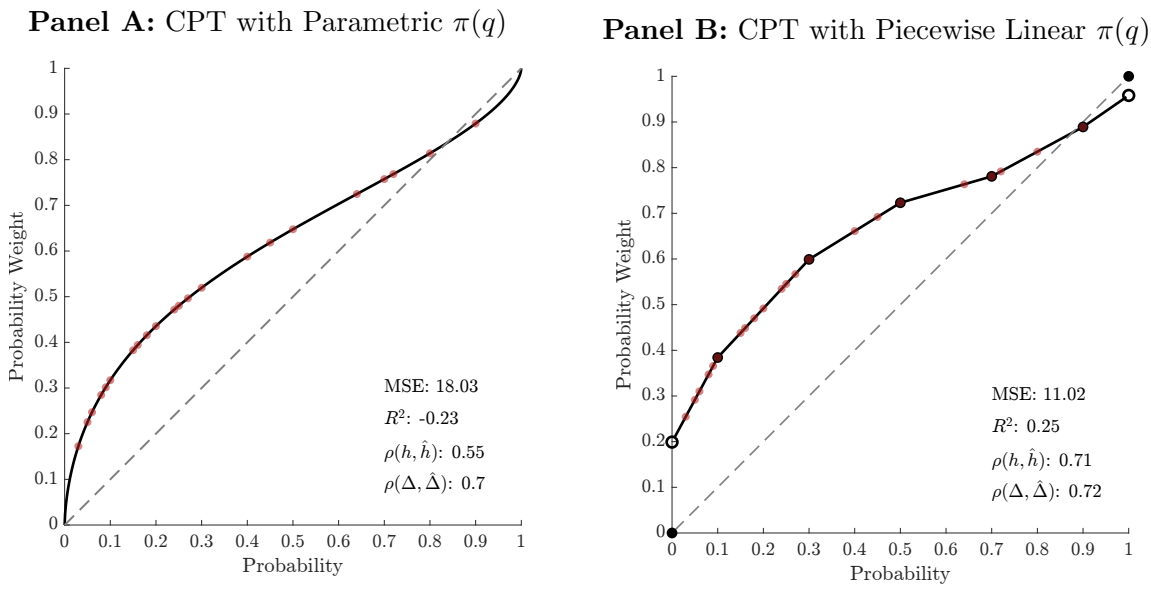
To understand why prospect theory models perform substantially worse than our model, in Appendix E.3 we explore how upside potential conceptually differs from both CPT and OPT. Within both CPT and OPT, the decision weight applied to each outcome depends only on that outcome's probability (or cumulative probability in the case of CPT). Under upside potential, the decision weight applied to each winning outcome depends on both that outcome's probability and the total probability of winning. For binary lotteries with one winning outcome, such as lotteries B, C, and D, the distinction is limited as the winning outcome's probability is also the total probability of winning. However, in trinary lotteries such as lottery B', the distinction is a critical source of difference. In other words, it is for decisions involving trinary lotteries, such as in our AB' task, where our upside-potential model becomes distinct from CPT and OPT.

Finally, while it is not relevant to our analysis in this paper, we highlight one further distinction between our upside-potential model and CPT. Under CPT, the weights attached to outcomes depend on their relative ranks. The weights applied to winning outcomes under our model are symmetric: each weight depends on that outcome's probability and the total probability of winning and doesn't asymmetrically shift with relative rank. Hence, our model is consistent with the growing

[41] Note that the negative R^2 term implies that one would be more accurate predicting the sample mean for every observation.

[42] See Appendix E.2.3 for details, including the estimates for all six prospect theory models that we consider.

Figure 9: Estimates for $\pi(q)$ under Prospect Theory



Notes: Figure presents estimates for probability weighting function $\pi(q)$ under CPT. 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 $\pi(q)$ when using the two-parameter probability weighting function from Lattimore et al. (1992), and Panel B presents estimates for $\pi(q)$ when using a six-part piecewise linear formulation that permits (but does not require) discontinuities at $q = 0$ and $q = 1$. For both panels, the estimation also involves estimating a value function $v(x) = x^{\alpha}$. Light red points in each panel correspond to locations where the function is evaluated in the estimation. In panel B, black dots denote kinks in the piecewise linear formulation; in estimation, the function is evaluated multiple 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 Appendix E.2.3 for details.

body of evidence suggesting rank-independence (Bernheim and Sprenger, 2020; Bernheim et al., 2022).

6 Discussion

In this paper, we study connected CR-CC-MX problems across a broad range of experimental parameters using valuations. 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 patterns of preferences. Because these empirical results are inconsistent with EU and leading non-EU models, we also posit a post-

hoc theoretical model designed to explain our modal mean preference. Encouragingly, this model successfully predicts some of the more nuanced data patterns we observe, suggesting it might hold insight for future theoretical work. We conclude by discussing some broader messages to take away from our analysis.

Our analysis highlights the danger of inferring broad preference features from isolated examples that cover a limited part of the parameter space. In particular, our results show that one must be careful when motivating global assumptions based on such examples when developing models. This caution applies beyond the domain of risk: Many prominent non-standard models of preferences—from the domains of ambiguity to intertemporal choice and beyond—are based on data from few canonical examples, without understanding empirically how behavior might be sensitive to relevant features of the decision problem. Moreover, the prior literature has often studied non-standard choice examples independently rather than together. A more comprehensive empirical foundation for theorizing emerges by simultaneously studying multiple preference features across a broad range of parameters. We also emphasize an important follow-up point: As we gain more insight into how patterns change across parameters and individuals, we may need to develop models that can accommodate sensitivities to parameters and individual differences, rather than predicting global, uniform effects. This doesn't necessarily require complex models with many degrees of freedom. In fact, our proposed model, based on a simple intuition, outperforms the leading non-EU model with fewer parameters.

Our analysis also reinforces a key methodological point from McGranaghan et al. (2024) that cautions against using binary choice data when making comparisons across decisions and instead encourages using valuation data for such comparisons. Many behavioral "effects" are based on evidence from comparing a pair of binary choice tasks. In our data, the CR, CC, and MX problems all show hallmarks of the confounds induced by differential noise, indicating that evidence from binary choice data could present a biased picture of the underlying preferences. While our binary choice data ultimately generate the same qualitative conclusions as our valuations data, they do so mainly due to our deliberate selection of balanced parameters—a selection made possible by the fact that we had pilot data on valuations. Hence, we hope future studies will use valuations rather than binary choices (or in addition to binary choices) when studying comparisons across decisions.

Our analysis relates to a nascent literature that emphasizes a direct preference for purposeful 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 often 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. Deliberate randomization between two successive presentations of an AB choice is clearly consistent with exhibiting an MXP in our context, though a limitation of such designs is that they cannot reveal an RMXP. Future work could explore whether there is an empirical connection between these behaviors, for example, 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.[43] 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, our model of upside potential permits conditions under which individuals will be averse to mixtures, rather than exhibiting solely a systematic preference for them.

On the surface, our empirical finding of an MXP and our post-hoc model's rationalization of that MXP seem to contradict recent work suggesting that lotteries with greater numbers of outcomes are penalized (Bernheim and Sprenger, 2020; Fudenberg and Puri, 2022; Puri, 2024). 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, our model can in fact simultaneously 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 Appendix E.1 for details). Of course, it would be valuable for future work to explore these connections further.

Finally, and most importantly, our results and analysis call for the development of new models of risk preferences alongside the development of a more complete empirical foundation for risky choice. Our empirical findings are at odds with 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 the CR-CC-MX structure is a critical step to take. We hope the psychology of upside potential, and its novel perspective on the source of risk attitudes, will prove useful for guiding such empirical explorations and the theories that will follow.

References

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

[43] 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.

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. (1952). Discussion of une axiomatisation de comportement raisonnable face a l'incertitude. In *XL Econometrie - Colloque de Mai 1952*, Colloques Internationaux du Centre National de la Recherche Scientifique, pages 29–40, Paris. Editions du Centre National de la Recherche Scientifique.

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., Royer, R., and Sprenger, C. (2022). Robustness of rank independence in risky choice. *AEA Papers and Proceedings*, 112:415–420.

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.

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.

Carrera, M., Royer, H., Stehr, M., Sydnor, J., and Taubinsky, D. (2022). Who chooses commitment? evidence and welfare implications. *Review of Economic Studies*, 89(3):1205–1244.

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

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

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.

Ingersoll, J. (2008). Non-monotonicity of the tversky-kahneman probability-weighting function: A cautionary note. *European Financial Management*, 14(3):385–390.

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

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

Lattimore, P. K., Baker, J. R., and Witte, A. D. (1992). The influence of probability on risky choice: A parametric examination. *Journal of Economic Behavior and Organization*, 17(3):377–400.

Loomes, G. and Sugden, R. (1986). Disappointment and dynamic consistency in choice under uncertainty. *The Review of Economic Studies*, 53(2):271–282.

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

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York.

McFadden, D. (1981). Econometric models of probabilistic choice behavior. In Manski, C. F. and McFadden, D., editors, *Structural Analysis of Discrete Data and Econometric Applications*, chapter 5, pages 198–272. MIT Press, Cambridge, MA.

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.

Puri, I. (2024). Simplicity and risk. *Journal of Finance*.

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