Alós-Ferrer, Carlos; Buckenmaier, Johannes; Garagnani, Michele

Working Paper

Stochastic choice and preference reversals

Working Paper, No. 370

Provided in Cooperation with:
Department of Economics, University of Zurich

Suggested Citation: Alós-Ferrer, Carlos; Buckenmaier, Johannes; Garagnani, Michele (2020) : Stochastic choice and preference reversals, Working Paper, No. 370, University of Zurich, Department of Economics, Zurich, https://doi.org/10.5167/uzh-193661

This Version is available at:
https://hdl.handle.net/10419/228870

University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 370

Stochastic Choice and Preference Reversals

Carlos Alós-Ferrer, Johannes Buckenmaier and Michele Garagnani

December 2020

Stochastic Choice and Preference Reversals*

Carlos Alós-Ferrer†[1], Johannes Buckenmaier[1], and Michele Garagnani[1]

[1]Department of Economics, University of Zurich.

This Version: December 7th, 2020

Abstract

The preference reversal phenomenon is one of the most important, long-standing, and widespread anomalies contradicting economic models of decisions under risk. It describes the robust observation of frequent "standard reversals" where long-shot gambles are valued above moderate ones but then the latter are chosen, while the opposite "nonstandard reversals" happen rarely. This inconsistency casts severe doubts on commonly-used preference elicitation methods. Strikingly, alternative designs which should eliminate the phenomenon produce frequent nonstandard reversals instead, in a puzzling reversal of the phenomenon. We develop and test a model predicting when the phenomenon should occur, when its reversal should occur instead, and what determines the magnitude of the effects. The reversal of the phenomenon is predicted as a consequence of stochastic choice and risk aversion, without invoking any behavioral bias. The original phenomenon arises from stochastic choice and a systematic bias in monetary valuations, which is restricted to long-shot gambles. To validate the model, we conduct two experiments leading to the preference reversal phenomenon and its reversal, respectively. We employ a novel empirical approach allowing us to disentangle choice and valuation errors by relying on utilities estimated out of sample, and confirm that reversals are associated with errors in monetary valuations of long-shots, with the upward bias quantified at 293% in monetary terms. The data also confirms the model's novel predictions, showing that a larger bias strengthens the phenomenon and higher risk aversion strengthens its reversal. Surprisingly, our analysis implies that the magnitude of the original phenomenon has been underestimated so far.

JEL Classification: D01 · D81 · D91

Keywords: Preference reversals · Lottery choice · Stochastic choice · Overpricing · Risk aversion

*We are grateful to Miguel Ballester, Sudeep Bhatia, Ernst Fehr, Graham Loomes, Ganna Pogrebna, and Roberto Weber for helpful comments and discussions. Financial support from the German Research Foundation (DFG) through project Al-1169/4, part of the Research Unit "Psychoeconomics" (FOR 1882) is gratefully acknowledged.

†Corresponding author. Zurich Center for Neuroeconomics (ZNE), Department of Economics, University of Zurich (Switzerland). Blümlisalpstrasse 10, CH-8006 Zurich. E.mail: carlos.alos-ferrer@econ.uzh.ch

1 Introduction

The preference reversal phenomenon, where monetary valuations of gambles contradict risky choices, reveals a striking inconsistency between theoretically-equivalent preference elicitation methods (Lichtenstein and Slovic, 1971; Grether and Plott, 1979; Tversky and Thaler, 1990). This inconsistency is robust, systematic, and highly relevant for economic analysis, because individual and societal preferences are often estimated on the basis of monetary valuations or similar constructs (see, e.g., Bateman et al., 2002 for a detailed discussion). Thus, if such measurements contradict actual choices, welfare economics and most of normative economics would be on shaky grounds. Moreover, this puzzling phenomenon is fundamentally at odds not only with Expected Utility Theory, but with any preference-based theory of decisions under risk assuming that decision makers' preferences can be represented by a stable utility function, including Cumulative Prospect Theory and Rank-Dependent Utility. Accordingly, the phenomenon has received enormous attention in economics (e.g., Holt, 1986; Karni and Safra, 1987; Tversky et al., 1990; Cubitt et al., 2004; Schmidt and Hey, 2004; Butler and Loomes, 2007).[1]

The classic preference reversal paradigm involves pairs of lotteries, typically consisting of a relatively safe lottery, called the P-bet ("P" for "probability"), and a riskier lottery offering a larger prize (a long shot), called the $-bet. Individual preferences over such pairs are elicited both through pairwise choices and by eliciting valuations separately for each lottery, e.g. using stated minimal selling prices. The anomaly refers to the observation that decision makers often exhibit a preference for the comparatively-safe P-bet in the choice task, but explicitly reveal a larger monetary valuation for the $-bet compared to the P-bet, a pattern that we will call a "standard" reversal. This empirical pattern is extremely robust (in the words of Butler and Loomes, 2007, "easy to produce, but much harder to explain;" see Seidl, 2002 for a comprehensive survey). Standard reversals occur much more frequently than the opposite pattern, in which $-bets are chosen but P-bets receive a higher valuation ("nonstandard reversals"). This asymmetry has universally been taken as evidence that reversals cannot be due just to random errors arising from stochasticity in choices, elicited valuations, or both (e.g., Schmidt and Hey, 2004; Loomes, 2005). Hence, any account of preference reversals needs to explain the *asymmetries* in choices and valuations which underlie the phenomenon.

A large number of competing, partial explanations has been put forward over the years, including systematic violations of transitivity (Safra et al., 1990) or procedural invariance (Goldstein and Einhorn, 1987), among others (see Section 5 for a more detailed discussion of the broad literature on preference reversals and the relation to our results).

[1] The preference reversal phenomenon is not restricted to lottery choice. A closely-related phenomenon has been documented for health utility measurements (e.g. Stalmeier et al., 1997; Bleichrodt and Pinto Prades, 2009; Oliver, 2013; Attema and Brouwer, 2013). The phenomenon is representative of a wider class of inconsistencies between preference elicitation methods, which include reversals between pricing and rating (Schkade and Johnson, 1989) and between certainty and probability equivalents (Hershey and Schoemaker, 1985; Johnson and Schkade, 1989; Delquié, 1993; Collins and James, 2015). Preference reversals have also been found to occur under ambiguity (Maafi, 2011; Trautmann et al., 2011).

The prominence hypothesis (Tversky et al., 1988) generally attributes inconsistencies to choice errors, arguing that decision makers focus on a prominent attribute (e.g., the winning probability) and overweight it in choice tasks compared to evaluation tasks. The scale compatibility hypothesis (Tversky et al., 1990) attributes the phenomenon to errors in the evaluation method, arising because decision makers overweight attributes which naturally map onto the (monetary) evaluation scale. Accounts based on choice inconsistencies (Schmidt and Hey, 2004) argue that preference reversals occur because evaluation tasks are less natural and hence more noisy than choices, resulting in differences in error rates. However, starting with the seminal work of Tversky et al. (1990) it has been consistently shown that the overall phenomenon persists in experimental settings that control for these explanations (e.g. Pommerehne et al., 1982; Cubitt et al., 2004). While there is general agreement that several of the explanations given above influence preference reversals, no single account has been able to fully explain when and why the phenomenon should be expected, and which factors ameliorate or exacerbate it.

Even more striking than the preference reversal phenomenon is the fact that, if the monetary valuation task is replaced by an ordinal ranking task, the anomaly is reversed. That is, instead of resulting in similar rates of standard and nonstandard reversals, this alternative implementation results in a *reversal of the preference reversal phenomenon* (Casey, 1991; Bateman et al., 2007; Alós-Ferrer et al., 2016) where nonstandard reversal rates, which are rather low in the original design, now exceed the standard ones. This is striking, because ranking tasks are conceptually closer to binary choices, and hence should avoid systematic differences across elicitation methods as those posited by the prominence or the scale compatibility hypotheses. This puzzling reversal of the phenomenon cannot be accounted for by any of the explanations of preference reversals previously proposed in the literature. For instance, Bateman et al. (2007) argued that ranking methods introduced "distorting effects of their own."

We contend that a successful theoretical explanation of the preference reversal phenomenon must explain the phenomenon in itself, the determinants of its magnitude (what strengthens or weakens it), and also (crucially) under which conditions the opposite phenomenon becomes dominant. In the present work, we develop and test a stochastic choice model that answers all these questions. The phenomenon and its reversal are shown to hinge on the interaction between well-established properties of stochastic choice, risk aversion, *and* an overpricing bias leading to large and frequent evaluation errors for long shots. Our model further allows us to encompass and organize previous empirical findings in the literature which were hard to interpret up to now.

To test the model's predictions, and also to motivate and test the validity of the model's assumptions, we rely on a novel empirical approach combining preference reversal experiments with utility estimations out of sample. The design provides empirical evidence not previously available in preference reversal experiments. Our empirical tests then include the preference reversal phenomenon, the reversal of the phenomenon, and novel predictions on the causal effects of overpricing and risk aversion.

The paper is structured in two parts. In the first (Sections 2 and 3) we present the results of two experiments, PRICE and RANK, designed to elicit the preference reversal phenomenon and its reversal, respectively. Specifically, experiment PRICE relies on a standard design with a monetary valuation task as the ones where the preference reversal phenomenon has been observed, while experiment RANK makes use of a non-monetary, ranking-based evaluation instead, as previous experiments where the reversal of the phenomenon has been observed. Crucially, however, and differently to any previous preference-reversal experiments, the experiments included an additional phase designed to implement an out-of-sample estimation of individual preferences (using choices among lotteries designed for this purpose, and unrelated to the ones in other parts of the experiment). This allowed us to obtain individual-level estimates of preferences independently from the choices used to study preference reversals, which in turn provided an external criterion to identify choice and valuation errors. In the second part of the paper (Section 4) we develop a stochastic choice model incorporating received insights about the structure of errors from the stochastic choice literature, which we further empirically test in our two experiments. We then show how the model accounts for both the preference reversal phenomenon and its reversal, and then go beyond that account to derive novel predictions from the model, which we test empirically in our data. Here again, the estimated preferences become essential: On the one hand, they serve as a natural scale to study regularities in stochastic choice which underlie the model. On the other hand, they allow us to test novel predictions of the model, which link reversals to individual risk attitudes and to overpricing seen as a deviation from the own preferences.

We now explain the motivation and results of our two experiments, before providing a brief summary of our model and its predictions. The objective of our PRICE experiment was to provide a solution to a major difficulty in the literature: Even if one postulates that errors of different types might be the cause of the preference reversal phenomenon, when valuations and choices contradict each other, it is unclear whether errors in choices or errors in valuations are the source of the reversals. To address this difficulty, experiment PRICE includes three tasks. Two of them are as in standard preference reversal experiments, namely a choice task where participants make binary choices between P-bets and $-bets, and a monetary valuation task where participants provide their willingness-to-accept for the same lotteries used in the choice task. The new task, conducted before the other two, is an independent choice task based on a different set of 32 lottery pairs, which were chosen following optimal design theory (e.g., Silvey, 1980; Moffatt, 2015) in order to maximize the precision of an econometric estimation of individual utility functions, and hence preferences. By committing to these estimated utilities, we obtain an external criterion to determine whether individual behavior in the choice and valuation tasks in the subsequent preference reversal experiment is consistent with individual preferences, and which can be called an error. In this way, we obtain (and commit to) a criterion before reversals occur. To the best of our knowledge,

we are the first to provide a direct classification of errors in choices and evaluations as deviations from an independently-estimated preference relation.

The results of experiment PRICE show that valuation errors for \$-bets (but not for P-bets) are a major driver of classical preference reversals between choices and willingness-to-accept valuations. Valuation errors are driven by an overpricing bias that causes an asymmetry in error rates between the valuation and the choice task. Beyond this specific bias, we confirm that our utility estimation is accurate by showing that elicited monetary valuations of P-bets are captured extremely well by their estimated certainty equivalents; that is, willingness-to-accept valuations are not noisy in general. The high incidence of evaluation errors can be traced back to an overpricing bias that almost exclusively applies to \$-bets. This overpricing of \$-bets is systematic and has an economically relevant magnitude, which we estimate: on average, elicited individual monetary valuation of \$-bets exceed their preference-based value (certainty equivalent as derived from our estimated utilities) by a whopping 293%. Thus, our results for experiment PRICE identify valuation errors due to an overpricing bias for long shots as a major driver of preference reversals, and beyond that quantify the extent of the bias in monetary terms.

It is important to note that these results do *not* depend on the specifics of the estimation procedure. In particular, they do not obtain simply because preferences estimated from choices (although out-of-sample) reflect other choices more accurately than valuation decisions. In fact, the above results remain robust when alternative utility functions are estimated from imputed choices derived from the comparison of elicited valuations (stated prices) for each given choice pair, even within-sample, and errors (both in pricing and for choices) are classified with respect to those. The reason is simply that there is no systematic bias between choices and monetary valuations in general, but rather a bias in the valuation of \$-bets only.

The motivation for our experiment RANK is as follows. If asymmetric evaluation errors of the kind we expected and found were the only cause of preference reversals, the latter should disappear if biases in evaluations could be shut down. A number of previous contributions have attempted to do precisely this by using evaluation methods which are conceptually closer to choices. Bostic et al. (1990) find that choice-based elicitation methods lead to less overpricing (in their case, in the sense of bid prices exceeding expected values) than pricing-based methods and were accompanied by a reduced number of reversals.[2] However, if the objective is to shut down evaluation biases by entirely dispensing with the monetary scale, an obvious alternative is to rely on purely-ordinal rankings instead. Bateman et al. (2007) used ranking-based evaluation tasks which however included sure amounts and were used to derive certainty equivalents, hence indirectly incorporating a monetary scale. They found a reduction in standard reversals, although the overall phenomenon (more standard than nonstandard reversals) subsisted.

[2]Hershey and Schoemaker (1985) and Collins and James (2015) find preference reversals to be less frequent (but not to disappear) when valuations are elicited as probability equivalents instead of certainty equivalents. See Section 5.

Alós-Ferrer et al. (2016) used purely-ordinal ranking tasks void of any reference to a monetary scale, and observed a "reversal of the preference reversal phenomenon" (a term previously introduced by Casey, 1991) where the rate of nonstandard reversals exceeded the rate of standard ones. As commented above, no account of the preference reversal phenomenon can be complete without explaining this striking, additional phenomenon.

To provide causal evidence that asymmetries in error rates between choices and evaluations drive preference reversals, experiment RANK relied on a ranking-based evaluation task but also included a separate sample of pairwise lottery choices, allowing to estimate individual utility functions as in experiment PRICE. In this second experiment, as expected, the reversal of the preference reversal phenomenon obtains, that is, the asymmetry between standard and nonstandard reversal rates is reversed. If this reversal were caused by bias-driven errors one would expect choice errors to exceed evaluation errors, mirroring the pattern observed in the PRICE experiment. This is not the case. Relying on our estimated utilities, we show error rates in the choice and evaluation phases are very similar. Thus, asymmetries in error rates resulting from a bias of evaluations relative to choice alone cannot provide a full account of preference reversals, and in particular cannot explain the reversal of the preference reversal phenomenon.

In the remainder of the paper, we show that a unified account can explain both the preference reversal phenomenon and its reversal (and, in particular, that the latter is *not* due to any new bias or additional distorting effects), while also deriving new testable predictions. To this end, we develop a stochastic choice model for preference reversal experiments. We postulate a monotonic relation between error rates and "strength of preference," captured by differences in estimated certainty equivalents. Specifically, in the absence of a systematic bias error rates should be larger when differences in certainty equivalents (derived from the previously-estimated utilities) are small. This prediction, which is a standard assumption in random utility models (McFadden, 2001), arises from long-standing insights from psychophysics (Dashiell, 1937; Moyer and Landauer, 1967) and has been recently demonstrated in the domain of decisions under risk (Alós-Ferrer and Garagnani, 2018). This assumption is confirmed by our data: in both experiments, we find evidence for strength-of-preference effects both for actual choices and for imputed choices derived from the comparison of elicited valuations for each given choice pair.

The model predicts the standard preference reversal phenomenon as a consequence of stochastic choice and biased \$-bet valuations. The most important insight arising from the model, however, is that the overall proportion of choices and valuations in favor of P-bets is a crucial determinant of the rates of standard and nonstandard reversals. Since the expected value of both lotteries within a pair are usually very similar in preference reversal experiments (but \$-bets are riskier), P-bets tend to be chosen more frequently due to risk aversion. We show that a stronger bias in the monetary valuations of \$-bets exacerbates the preference reversal phenomenon, while a higher degree of risk aversion has the opposite effect. In particular, the implicit assumption in the literature that in the absence of an evaluation bias one should expect comparable rates of reversals is incorrect.

Since decision makers are typically risk averse, in the absence of a bias (as in our RANK experiment) the model predicts a lower rate of standard than nonstandard reversals, that is, the reversal of the preference reversal phenomenon is expected to occur. This prediction yields two additional implications. First, the reversal of the preference reversal phenomenon is *not* the result of another bias in evaluations but the direct consequence of stochastic choice and risk aversion. Second, the magnitude of the classic preference reversal phenomenon has actually been *underestimated*, since the default in the absence of any bias is the reverse asymmetry and not equal rates of reversals.

The key intuition for the results can be conveyed by the following stylized example. The design of preference reversal experiments relies on using P-bets and \$-bets with similar expected values, but \$-bets are riskier. Since decision makers are typically risk averse, P-bets typically have certainty equivalents above those of \$-bets. Thus, to fix ideas, consider a single lottery pair where this is the case, that is, $CE(P) - CE(\$) > 0$. Since stochastic choice prescribes that errors are decreasing as differences in certainty equivalents increase, in this case valuations where the \$-bet is ranked higher than the P-bet, which are errors, are less probable than valuations where the opposite happens. Standard reversal rates are computed as the percentage of pairs where the valuation of \$-bets is higher, *conditional on the P-bet being chosen.* Thus, such standard reversal rates should be relatively small under risk aversion. On the contrary, for the same pair with $CE(P) - CE(\$) > 0$, nonstandard reversal rates are computed as the percentage of pairs where the valuation of P-bets is higher, *conditional on the \$-bet being chosen.* In this case, however, the choice is an error, and valuations where the P-bet is ranked higher should be very frequent. Thus, nonstandard reversal rates should be relatively large. We conclude that the *reversal* of the preference reversal phenomenon is a consequence of stochastic choice and risk aversion, without any need to invoke a behavioral bias.

This intuition captures the results in experiment RANK well, for in this case our data shows that there is no systematic difference between the choice task and the ranking-based evaluation task. However, in regular preference-reversal experiments as our experiment PRICE, and as shown by our analysis, the monetary valuation bias embodies a large and persistent upward bias for the valuation of \$-bets. This shifts the stochastic imputed choices derived from the valuations in such a way that, for a large fraction of choices where $CE(P) - CE(\$) > 0$, the probability that the \$-bet is valued above the P-bet is larger than the probability of the opposite event. The logic above then flips entirely, and, even under risk aversion, the preference reversal phenomenon obtains. Thus, the interaction of stochastic choice, risk aversion, and a bias in the monetary valuations of \$-bets fully explains both the preference reversal phenomenon and its reversal.

The model readily delivers novel, testable predictions which further strengthen our conclusions. First, the difference between the rates of standard and nonstandard reversals is expected to be larger the more pronounced the bias in evaluations is. Since we have a quantitative measure of overpricing at the individual level (derived from stated willingness to accept and estimated certainty equivalents), we can directly test this hy-

pothesis in experiment PRICE by comparing subjects with a large overpricing bias to those with a small bias. We find that the difference in reversal rates is larger for subjects with a large bias, in line with the prediction of the model.

Second, the model also provides a testable prediction on the effect of risk aversion. In the absence of evaluation bias, the ratio of standard to nonstandard reversals is predicted to be decreasing in the degree of risk-aversion. Having individually estimated risk attitudes, we can test this prediction directly in experiment RANK, where the evaluation bias is absent. In line with the prediction, we find a significantly smaller ratio of standard to nonstandard reversals for subjects with a high degree of risk aversion compared to subjects with low risk aversion.

Last, our results allow us to encompass and organize previous empirical findings in the literature which were hard to interpret up to now. For instance, most of the available preference reversal experiments rely on evaluations elicited through willingness to accept (WTA), and only a handful rely on willingness to pay (WTP). It has been previously argued that the latter might be a less biased method than the former (Schmidt and Hey, 2004). However, empirical results using WTP have (puzzlingly) found both the preference reversal phenomenon and its reversal. Our results show that these findings might have been misinterpreted, because an unbiased evaluation method does *not* imply that the rates of standard and nonstandard reversals should be equal. In Section 5, we show that previous findings using WTP and other evaluation methods are explained by our results, which predict that, if an evaluation method exhibits no bias, whether the preference reversal phenomenon or its reversal is observed is determined by the exhibited aggregate risk aversion in the experiment, which in turn depends on the particular set of lotteries used, and possibly on other elements of the experimental design.

The paper is structured as follows. Section 2 describes our experimental design and the utility estimation procedure. Section 3 presents our results on valuation and choice errors in experiment PRICE including a robustness analysis, quantifies the magnitude of overpricing, and presents results from experiment RANK. Section 4 presents our stochastic choice model and provides a unified account of preference reversals. Section 5 discusses the related literature, and in particular how previous empirical findings are accounted for in light of our results. Section 6 concludes. The Appendix contains additional details and robustness analyses.

2 The Experiments

A total of 190 subjects (127 females, average age 23.43) participated in 6 experimental sessions and were randomly assigned to one of two experiments, PRICE and RANK (95 subjects each). That is, the experiments were formally conducted concurrently, as treatments within a larger experiment, to ensure full comparability, but we will discuss them separately. Participants were recruited from the student population of the University of Cologne using ORSEE (Greiner, 2015), excluding students majoring in psychology

and economics (who might have learned about the preference reversal phenomenon) and subjects who had previously participated in experiments involving lottery choice. The experiments were programmed in PsychoPy (Peirce, 2007).

Each experiment consisted of two parts. The first part was an estimation phase used to estimate individual preferences for each subject. The second part was the actual preference reversal experiment consisting of an evaluation phase and a choice phase. We decided to fix the order at evaluation then choice, since previous research indicates that the phenomenon occurs independently of the order (Alós-Ferrer et al., 2016). The choice phase was identical across experiments, but the evaluation phase differed as explained below.

2.1 Estimation Phase

The goal of the estimation phase was to obtain a measure of each subject's individual preference. Subjects faced 32 lottery pairs that were unrelated to the P-bet/$-bet pairs used in the second part of the experiment.[3] We used each subject's choices for these 32 lottery pairs to estimate an individual utility function (see Section 2.5 below). This was done out of sample in the sense that estimation relied exclusively on the choices in the first part, but was used as an external measure to classify choices and evaluations as errors or correct responses in the second part of each experiment (Section 3.2 shows that the results are robust when utilities are estimated out of monetary valuations instead).

2.2 Experiment PRICE

In addition to the estimation phase, experiment PRICE consisted of an evaluation phase and a choice phase. Taken together, these phases correspond to a standard preference reversal experiment, and hence we expected to obtain the preference reversal phenomenon. In the evaluation phase, we elicited subjects' valuations for 60 P-bets and 60 $-bets. Subjects stated their willingness-to-accept (WTA) valuations for each lottery. Specifically, subjects were asked to state their minimal selling price for each of the 120 lotteries.[4] Each lottery was presented on a separate screen. All lotteries were of the form $A = (p, x)$, that is, A pays an amount x with probability p and zero otherwise. Subjects' WTA evaluations were limited to the range $[0, x]$.

In the choice phase, subjects faced again the lotteries from the evaluation phase but now presented in 60 pairs, each consisting of a $-bet and a P-bet, and in a different,

[3]See Appendix E for the complete list of lotteries used in the experiment. This part also included four pairs with dominated choices as a consistency check, but across both experiments subjects made only 5 dominated choices (out of $190 \times 4 = 760$).

[4]We relied on WTA valuations because this is the most common choice in the literature. Preference reversals are also readily obtained if one uses willingness-to-pay (WTP) valuations or sequential elicitation methods instead (Butler and Loomes, 2007). There are, however, some differences, and it has been shown that preference reversals are somewhat less frequent when monetary valuations are elicited using WTP (Dubourg et al., 1994; Morrison, 1998).

randomized order. For each of the 60 pairs, subjects were asked to choose which lottery they would prefer to play.

2.3 Experiment RANK

Experiment RANK used the same estimation and choice phases as in PRICE, but employed a different, ranking-based elicitation procedure in the evaluation phase. The reason is that ranking-based elicitation methods have been shown to elicit the reversal of the preference reversal phenomenon (Alós-Ferrer et al., 2016). In the evaluation phase, the lotteries were presented in blocks of six, and subjects assigned ranks to them from their most (rank 1) to their least preferred option (rank 6) according to how much they would have liked to play each lottery. Each block contained three P-bets and three $-bets. To ensure comparability between treatments, the lotteries in PRICE were also presented in 20 "rounds," separated by screens announcing the next round. Each such round consisted of six lotteries presented sequentially, with the set of lotteries in a round corresponding to one block in RANK.

2.4 Procedures and Payment

Lotteries were presented in the form of colored pie charts, with colors (green and blue) counterbalanced across subjects. The screen position (left or right) of lotteries within pairs was also counterbalanced within subjects, with half of the pairs displaying a $-bet on the right. To control for order effects, each subject was randomly assigned to one of four different, pre-randomized sequences of lottery pairs.[5]

Before beginning the experiment, subjects were provided with written instructions and had to answer four control questions to ensure their understanding of the concept of a lottery and its pie chart representation. Detailed instructions for all parts were presented on-screen before the start of the respective task. At the end of the experiment, subjects were asked to complete a short questionnaire eliciting various demographics (gender, age, field of studies) and numerical literacy (Lipkus et al., 2001).

There was no feedback during the course of the experiment, that is, subjects did not receive any information regarding their earnings until the very end of the experiment. All decisions were made independently and at a subject's individual pace. Subjects never had to wait for the decisions of another subject. Only at the end of the experiment, when they had made all decisions, were they required to wait until everybody had completed the experiment before their payoff was computed and revealed.

Payment procedures were explained within the instructions and carried out truthfully. To determine a subject's payoff, one lottery from each phase was randomly selected and paid (Azrieli et al., 2018). For the estimation phase and the choice phase in the second part, one of the lottery pairs in the corresponding phase was randomly selected and the lottery chosen by the participant was played out. The evaluation phase used a

[5]We found no evidence for order effects on our main variables of interest.

variant of the (incentive-compatible) Ordinal Payment Method (Goldstein and Einhorn, 1987; Tversky et al., 1990; Cubitt et al., 2004). We opted for the more intuitive ordinal incentive scheme instead of the Becker-DeGroot-Marschak procedure because the latter is sometimes found to be noisier (Alós-Ferrer et al., 2016). The computer selected one round (for PRICE) or one block (for RANK) at random, and then randomly selected two of the six lotteries in the round/block. The one that the participant had priced or ranked higher was then played out. The total payoff from the experiment was the sum of the three amounts received from the estimation, evaluation, and choice phases. In addition subjects received a lab-mandated show-up fee of €4 for an average total remuneration of €19.76. Sessions lasted between 70 and 85 minutes including instructions and payment.

2.5 Description of the Estimation Procedure

The lottery choices in the estimation phase were only used to estimate subjects' individual preferences out-of-sample. The 32 lottery pairs used in this phase were constructed to maximize the precision of the estimated preferences. To achieve this we rely on optimal design theory (Silvey, 1980) in the context of non linear (binary) models (Ford et al., 1992; Atkinson, 1996), in agreement with the recommendations of Moffatt (2015).[6]

We assume that the structure of errors follows an additive random utility model (e.g., Thurstone, 1927; Luce, 1959; McFadden, 2001). However, all results throughout the paper remain qualitatively unchanged if we adopt a random preference model (Loomes and Sugden, 1998; Apesteguía and Ballester, 2018) instead (see Appendix C). Estimation of individual risk attitudes relies on a well-established maximum likelihood procedure (e.g., see Train, 2003; Moffatt, 2005; Bellemare et al., 2008). We refer the interested reader to Appendix A for a detailed description of the estimation procedure.

For the functional form of the utility function, we adopt the normalized constant absolute risk aversion (CARA) function as in Conte et al. (2011), which is given by

$$u(x) = \begin{cases} \frac{1-\exp(-rx)}{1-\exp(-rx_{\max})}, & \text{if } r \neq 0 \\ \frac{x}{x_{\max}}, & \text{if } r = 0, \end{cases}$$

where $x_{\max}$ is the upper bound of the outcome variable x. All our results remain qualitatively unchanged if we assume a CRRA utility function instead; we provide the corresponding analysis in Appendix B.

[6] We chose to estimate risk attitudes from a sequence of pairwise lottery choices over alternatives such as the multiple price list (MPL) method (Holt and Laury, 2002). The reason is that the latter imposes a strong correlation structure on the choice sequence, namely a unique switching point (see Andersen et al., 2006, for a discussion of the weaknesses of MPL methods). Moreover, Beauchamp et al. (2019) show that MPL methods are susceptible to the compromise effect, which may lead to biased results.

The average estimated risk propensity $\hat{r}$ is 0.152 (median 0.159, SD 0.103). Overall, 23 subjects (12.11%) are classified as risk-seeking and some participants are close to being risk neutral, but the majority of subjects is risk averse.[7]

3 Preference Reversals, Errors, and Overpricing

This section presents our results on evaluation and choice errors and the magnitude of overpricing. First, we analyze the data from experiment PRICE. In Section 3.1, we observe the standard asymmetry between reversals but, crucially, we show that it reflects errors in the elicitation of valuations and not choice errors. Section 3.2 presents a robustness analysis, which shows that the observed asymmetry between valuation and choice errors does not hinge on estimating utilities out of choices. On the contrary, it also obtains when utilities are estimated using elicited valuations instead. In Section 3.3, we show that evaluation errors affect almost exclusively \$-bets and are systematic, that is, a major driver of standard reversals is an overpricing bias caused by the elicitation procedure, which distorts subjects' evaluations of \$-bets upwards. In particular, the elicited monetary valuations are remarkably well-predicted by our independently-estimated preferences for P-bets, which further demonstrates the validity of our preference estimation procedure. For \$-bets, we explicitly provide a quantification of the overpricing bias in terms of its (large) economic magnitude. Last, in Section 3.4 we present the results of experiment RANK, where we use a ranking-based elicitation procedure (Bateman et al., 2007) that shuts down overpricing by design. The absence of overpricing, however, does not result in equal reversal rates but leads to the so-called reversal of the preference reversal phenomenon. In stark contrast to PRICE, we find similar rates of evaluation and choice errors, confirming that the bias was successfully shut down. It follows that the reversed asymmetry in RANK cannot be explained by a higher incidence of errors of either type indicating that bias-driven errors alone cannot fully account for preference reversals. In Section 4 we will return to this point and show that monotonicities in stochastic choice are the missing piece of the puzzle.

3.1 Valuation Errors (Experiment PRICE)

We first confirm the (standard) preference reversal phenomenon in this experiment. For each of the 95 subjects that participated in experiment PRICE, we collected WTA valuations for 120 lotteries (evaluation phase) and pairwise choices for the same lotteries (choice phase), presented in 60 pairs, each consisting of a P-bet and a \$-bet. Given a lottery pair $(\mathrm{P}, \$)$, a *standard preference reversal* occurs for subject i if this subject chooses P but states a higher price for \$, and a *nonstandard preference reversal* occurs

[7] An agent with a risk propensity equal to the average in our sample, $\hat{r} = 0.152$, would have a certainty equivalent of about \$3.25 when facing a lottery paying \$10 with 50% probability and zero otherwise.

if the opposite pattern is observed.[8] Reversals are extremely frequent, with an average individual reversal rate of 50.63% (not distinguishing types of reversals).[9] However, this number does not capture the asymmetry which is the essence of the preference reversal phenomenon. The rate of standard (resp. nonstandard) reversals is the number of standard (resp. nonstandard) reversals divided by the number of P-bet choices (resp. $-bet choices). Figure 1 (left panel) displays violin plots for the rates of standard and nonstandard reversals, revealing a clear asymmetry between the two types of reversals. Conditional on the P-bet being chosen the propensity to state an inconsistent price ordering is significantly higher (63.02%) than when the $-bet was chosen (3.66%) according to a Wilcoxon signed-rank (WSR) test ($N = 86$, $z = 8.008$, $p < 0.001$).[10]

Having established that the phenomenon is present in experiment PRICE, we now investigate the role of errors, classified according to the individual preferences estimated in the first part of the experiment as follows. Consider a pair of lotteries (A, B) with $A = (p, x)$ and $B = (q, y)$. That is, A pays x with probability p and 0 otherwise, and B pays y with probability q and 0 otherwise. Let u_i be the estimated utility function of subject i. We say that choosing A from the pair (A, B) is a *correct* choice for i if $EU_i(A) = pu_i(x) \geq qu_i(y) = EU_i(B)$, and we call it an *error* otherwise. Evaluation errors are defined analogously. That is, for a pair of lotteries (A, B), let $WTA_i(A)$ and $WTA_i(B)$ be the elicited prices for A and B, respectively. We say that evaluating A higher than B for a pair (A, B), i.e. $WTA_i(A) \geq WTA_i(B)$, is *correct* for i if $EU_i(A) \geq EU_i(B)$, and an *error* otherwise.

Figure 1 (center panel) displays the individual proportions of choice and evaluation errors. Choice errors are substantially less frequent (24.26%) than evaluation errors (58.02%; WSR test, $N = 95$, $z = -8.035$, $p < 0.001$). That is, the asymmetry in reversal rates is accompanied by an asymmetry in error rates. This result indicates that preferences elicited through WTA valuations are inherently more noisy than preferences revealed by choices. To the best of our knowledge, this is the first time that a direct test of this fact has been performed by relying on independently-estimated utilities and hence allowing for an *ex ante*, unambiguous definition of errors.

This striking result directly shows that the classic asymmetry in reversal rates can be traced back to an asymmetry between choice and evaluation errors. This claim can be further substantiated through a simple exercise. If the phenomenon is caused by errors in elicited WTA valuations, the asymmetry between the two types of reversals

[8] The literature often uses the names "predicted" and "unpredicted" for standard and nonstandard reversals, but they refer simply to the empirical fact that the first type of reversals is more frequent, and not to those being "predicted" by some underlying theory. Hence, we prefer the terminology "standard."

[9] Individual reversal rates were calculated excluding pairs where the P-bet and the $-bet were identically valued (6.74%, 384 out of a total of 5,700 comparisons). Our results are unchanged when pairs with identical valuations are included and classified as either non-reversals or reversals.

[10] Tests for differences in reversal rates can only include subjects for which both rates can be computed. For subjects with very few P-bet or $-bet choices, reversal rates tend to be on the extremes at 0% or 100%. Therefore, when calculating standard and nonstandard reversal rates we only include subjects with at least four choices of each type. We obtain qualitatively the same results when all subjects for which rates can be computed are used in the analysis.

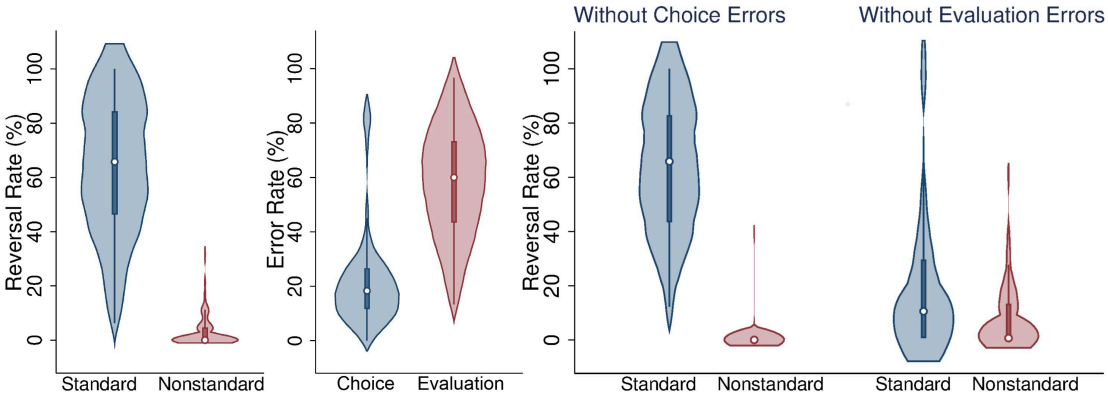


Figure 1: Experiment PRICE. Reversal rates, errors, and reversal rates excluding errors. *Notes:* Violin plots show the median, the interquartile range and the 95% confidence intervals as well as rotated kernel density plots on each side. Left: Distribution of individual rates of standard and nonstandard reversals. Center: Average of individual error rates, both for choices and evaluations (imputed choices). Right: Average of individual reversal rates (standard and nonstandard) when choice or evaluation errors are excluded.

should be smaller or even disappear if we restrict the analysis to pairs that do not involve an evaluation error (that is, if we exclude the choice and evaluations of pairs where the evaluations contradict elicited preferences). On the contrary, considering only pairs that do not involve a choice error should just remove noise and leave the asymmetry unaffected. Figure 1 (right panel) shows the reversal rates when choice or evaluation errors are excluded, respectively. As expected, excluding choice errors has little effect on reversal rates (average standard reversal rate 63.48%, nonstandard reversal rate 0.88%), and naturally the asymmetry remains (WSR test, $N = 50$, $z = 6.155$, $p < 0.001$). In contrast, when evaluation errors are excluded the rate of standard reversals drops drastically from 63.48% to 22.40% (WSR test, $N = 45$, $z = 5.842$, $p < 0.001$). Most importantly, the asymmetry between the two types of reversals is greatly reduced albeit the difference remains statistically significant (standard reversal rate: 22.40%; nonstandard reversal rate: 7.48%; WSR test, $N = 50$, $z = 2.555$, $p = 0.011$).[11] This corroborates that evaluation errors are indeed a major driver of the asymmetry behind the preference reversal phenomenon.

In summary, using out-of-sample estimation of individual preferences to classify choices and evaluations as errors or correct responses, we find that errors are significantly more frequent in the evaluation phase than in the choice phase. This is in partial agreement with earlier arguments stating that willingness-to-accept valuations are inherently noisier than choice (Schmidt and Hey, 2004). However, our analysis provides the first direct confirmation of this hypothesis, in the sense of defining errors by an ex-

[11] Additional robustness analyses yield similar results. Alternative estimation exercises that assume a CRRA utility function (Appendix B) or a random parameter model (Appendix C) instead reveal a comparable reduction of the asymmetry between the two types of reversals. In both cases the remaining difference is only marginally significant at the 10% level.

ternal criterion such as an independently estimated preference. Beyond that, we show that when evaluation errors are excluded the asymmetry in standard and nonstandard reversals is greatly reduced, whereas it is unaffected when choice errors are excluded instead. This finding supports the interpretation that errors in the evaluation phase are among the main drivers of the preference reversal phenomenon.

3.2 Robustness Check: Utilities Estimated from Monetary Valuations

So far we have relied on individual utility functions that were estimated on the basis of 32 binary lottery choices that were unrelated to the P-bets and $-bets used in the second part of the experiment. Although we are convinced that this approach is appropriate (and we will show below that the estimation is accurate), we acknowledge that one might argue that preferences estimated from choices (although out-of-sample) are likely to reflect other binary choices (like those in the choice phase) better than monetary valuations (like those in the evaluation phase) simply because the decision situations are more similar. This is *not* the case. Intuitively, this argument suggests that one should obtain the opposite result when utilities are estimated from monetary valuations instead. Although this is an intuitive line of thought, we can clearly refute this conjecture.

We now present an alternative estimation exercise which serves as a robustness check addressing this potential concern. Specifically, we repeat the estimation exercise described in Section 2.5 using the imputed choices derived from the actual willingness-to-accept valuations. That is, we take the $(\text{P}, \$)$-bet pairs used in the preference reversal experiment and consider P to be "chosen" by subject i if and only if $WTA_i(\text{P}) > WTA_i(\$)$.[12] We then estimate new utility functions u_i' using these imputed, valuation-based "choices." We remark that, in contrast to our original analysis, this estimation is made within sample and hence gives the new utilities an even better chance to correctly predict the relative order of valuations compared to an out-of-sample estimation. When we consider choice and evaluation errors defined relative to these new utility functions u_i', the asymmetry in error rates does not reverse; choice errors are still substantially less frequent (52.72%) than evaluation errors (69.11%; WSR test, $N = 95$, $z = 6.216$, $p < 0.001$). Repeating the exercise described above, we then consider only pairs that do not constitute choice or evaluation errors, respectively, with respect to the new utility functions u_i'. The results are qualitatively unchanged: When choice errors are excluded, standard reversals become even more frequent (70.96%), increasing the asymmetry. However, if evaluation errors are excluded, standard reversals decrease significantly (50.63%; WSR test, $N = 67$, $z = 5.621$, $p < 0.001$), reducing the asymmetry.

This robustness check confirms that the result that evaluation errors are a major driver of preference reversals does not hinge on estimating utilities out of choices but instead is qualitatively unchanged when utilities are estimated from WTA valuations. In particular, none of the comparisons "flips" when we use such pricing-based utility

[12] For 6.74% of all lottery pairs the stated valuation was the same for both the P-bet and the $-bet, that is, $WTA_i(\text{P}) = WTA_i(\$)$. These observations are not considered for the estimation.

estimates. The reason is simply that there is no systematic bias between choices and monetary valuations in general, but rather a bias in the valuation of $-bets only.

3.3 WTA Valuations vs. Certainty Equivalents

We have identified errors in WTA valuations as a main driver of preference reversals. It is natural to ask what causes such high error rates in monetary valuations. Section 3.2 shows that the difference in error rates across phases is not an artifact of the estimation method. That is, even when utilities are estimated from (biased) WTA valuations, evaluation errors are still more frequent. An alternative argument explaining the high error rates is that stating monetary valuations might be intrinsically cognitively more demanding than direct binary choices. Hence, one might speculate that error rates are higher for the evaluation phase simply due to these differences between tasks. If this argument were correct, however, one would have to expect overall higher error rates in the evaluation phase than in the choice phase, both for P-bets and for $-bets. We now show that this is not supported by the data.

Figure 2 (left panel) plots the stated WTA valuations against the estimated certainty equivalent (CE) for each of the 120 lotteries using the out-of-sample procedure, distinguishing P-bets and $-bets. For P-bets, the (WTA, CE) pairs are tightly clustered around the regression line, which is itself close to the diagonal. Stated valuations and the corresponding certainty equivalents show a strong and highly significant correlation for P-bets, with correlation coefficient close to unity (Spearman's $\rho = 0.930$, $N = 60$, $p < 0.001$). Thus, stated valuations for the P-bets are well-predicted by the out-of-sample estimation of individual utilities, which on the one hand further confirms the validity of our estimated utilities and on the other hand shows that there is no general, systematic difference across elicitation tasks (monetary valuation and choices).

In particular, and contrary to the generalized impression in the literature, the evaluation method through stated WTA faithfully reflects certainty equivalents derived from independently-estimated expected utilities, in the case of P-bets. However, the same cannot be said for $-bets, for which the picture is much more dispersed and far away from the diagonal, and the correlation is much lower (Spearman's $\rho = 0.424$, $N = 60$, $p = 0.001$). That is, for $-bets the valuations of lotteries do not agree with the subjects' estimated certainty equivalents. Rather, they are noisy and systematically biased upwards. This analysis strongly suggests that excess noise in evaluation tasks (compared to choice tasks) is not a general, unqualified phenomenon, but rather it is associated with specific lotteries (long shots in the case of preference reversals).

This result reveals a systematic *overpricing bias* for $-bets. That is, elicitation procedures based on WTA valuations introduce a systematic bias that leads subjects to disproportionately overstate the valuations of $-bets. Thanks to our design, we can go beyond this observation and actually *quantify* the economic magnitude of this bias. For each lottery A and each subject i, we consider the difference between the stated WTA val-

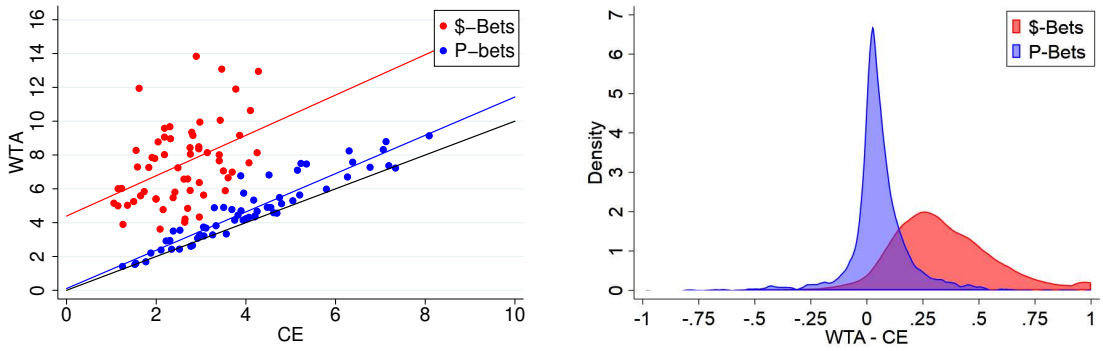


Figure 2: Accuracy of valuations in PRICE.

Notes: Left: Correlation between stated WTA and predicted certainty equivalent for P-bets and $-bets. Each point corresponds to one lottery representing the average WTA and the average CE across all subjects in the PRICE experiment. Right: Distribution of overpricing measure for P-bets and $-bets.

uation and the CE for that lottery, $WTA_i(A) - CE_i(A)$, where $CE_i(A) = u_i^{-1}(EU_i(A))$ is the certainty equivalent derived from subject i's utility function u_i estimated in the first part of the experiment. Figure 2 (right panel) displays the distribution of this variable (normalized by its largest value in the data set) separately for P-bets and $-bets. The two distributions are clearly different, which is confirmed by an equality of distributions test (Kolmogorov-Smirnov test, $D = 0.664$, $p < 0.001$). For P-bets, the distribution is concentrated around zero (mean 0.053, median 0.041, SD 0.152). On the contrary, for $-bets the distribution is more dispersed and shifted to the right, showing a sizable and systematic bias to overstate valuations (mean 0.339, median 0.309, SD 0.234).

The overpricing of $-bets is of a large and economically-relevant magnitude. To quantify this observation in monetary terms, we consider overpricing relative to the certainty equivalent, that is, $[WTA_i(A) - CE_i(A)]/CE_i(A)$. According to this measure, the average *monetary* increase in the valuations (overpricing) of the $-bets is 2.929 (median 2.596, SD 1.947), relative to the certainty equivalent. That is, on average, $-bets are overpriced by a whopping 293%. In contrast, for P-bets the average bias of the stated valuations is much smaller (around 17 times smaller than in the case of $-bets), amounting just to 0.175 relative to the certainty equivalent (median 0.108, SD 0.193).

The observed systematic overpricing of $-bets is in line with the long-standing Compatibility Hypothesis (Tversky et al., 1990), which states that in a pricing-based evaluation task attention is focused on the salient monetary dimension, hence there is a tendency to give a higher valuation to the $-bet. It is worth noticing that we also observe a tendency to overprice the P-bets, although of a comparatively much smaller magnitude. Salience theory (Bordalo et al., 2012, 2013), which links overweighting to payoff magnitudes, predicts overpricing for both the P-bet and the $-bet but the effect should be larger for the latter, which is in line our findings (in Appendix D we provide a more detailed analysis of overpricing in terms of a salience-based anchoring mechanism; see also the literature discussion in Section 5).

In summary, stated WTA valuations are well-captured by certainty equivalents derived from estimated individual utility functions for P-bets, but not for $-bets. Stated valuations show clear evidence for a systematic and large overpricing of $-bets, whereas P-bets are, in comparison, accurately evaluated. Most importantly, for $-bets the extent of this overpricing bias is of an economically-relevant magnitude.

3.4 Reversing the Reversal (Experiment RANK)

Our second experiment (RANK) used a ranking-based elicitation method in the evaluation phase (instead of the WTA valuations used in PRICE). The purpose of this experiment was to show that the asymmetry in errors observed in PRICE is exclusive of the pricing-based elicitation method used in that experiment. To this end, RANK used a different ranking-based evaluation method where subjects were asked to provide ordinal rankings for blocks of lotteries. We chose this evaluation method because a few studies have previously shown that ranking-based tasks can reduce standard reversals (Bateman et al., 2007; Alós-Ferrer et al., 2016). In contrast to those studies, however, our focus is on the differences in error rates between the evaluation methods. That is, if standard reversals in PRICE are mainly caused by errors in WTA valuations due to overpricing, a reduction in standard reversals in RANK is likely the result of (and should be accompanied by) a lower rate of ranking errors.

In RANK, subjects chose the P-bet 68.33% of the time, whereas they ranked it above the $-bet in 72.35% of all pairs. In contrast, in PRICE the P-bet was chosen 69.63% of the time but only 23.87% of the P-bets were priced higher than the corresponding $-bet. Behavior in the choice phase (percentage of P-bet choices at the individual level) in RANK was not significantly different compared to PRICE according to a Mann-Whitney-Wilcoxon (MWW) test ($N = 190$, $z = 0.738$, $p = 0.461$), but choices imputed through WTA or rank comparisons were (MWW test, $N = 190$, $z = -10.183$, $p < 0.001$). As expected, reversals were drastically reduced in experiment RANK, amounting only to 19.46% on average, compared to 50.63% in experiment PRICE (pooling both types of reversals). The difference is significant (MWW test, $N = 190$, $z = -9.869$, $p < 0.001$).

Figure 3 (left panel) displays the individual error rates in choices and WTA/ranking pairs separately for both experiments. The frequencies of choice errors are similar in both experiments, and indeed there is no difference in error rates for choices between PRICE (24.26%) and RANK (27.95%; MWW test, $N = 190$, $z = -1.194$, $p = 0.233$). Crucially, however, in RANK errors in the evaluation phase are not more frequent than choice errors. In fact, both types of errors are of a similar magnitude, with evaluation errors (mean 25.26%) being even slightly less frequent than choice errors (mean 27.95%; WSR test, $N = 95, z = -2.915, p = 0.002$). Thus, in contrast to PRICE we find no asymmetry in error rates in RANK.

We now turn to reversal rates. Figure 3 (left panel) displays violin plots for individual rates of standard and nonstandard reversals, for both experiments. If the asymmetry in

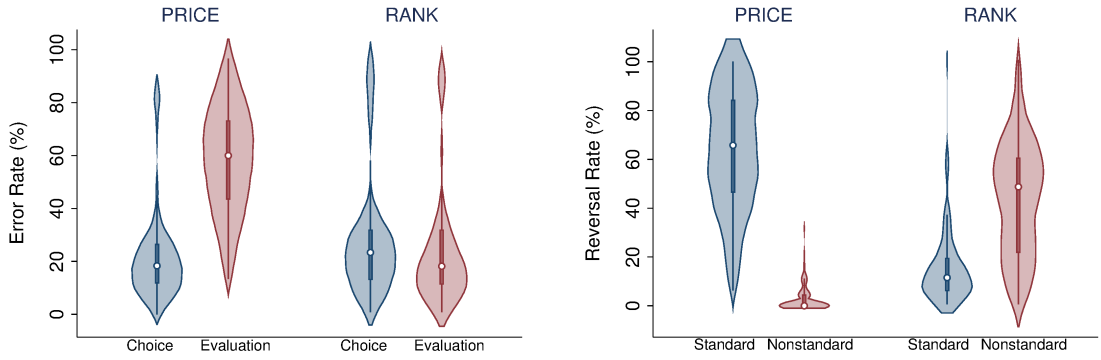


Figure 3: Treatment comparison.

Notes: Left: Individual error rates for choices and pairs of ranks, for both experiments. Error rates are comparable for choices, but are greatly reduced in ranking evaluations. Right: Individual reversal rates for standard and nonstandard reversals, for both experiments. Standard reversals are greatly reduced for experiment RANK.

reversal rates observed in PRICE were solely due to an asymmetry in error rates, then one would expect comparable rates of standard and nonstandard reversals in RANK as there the error rates are very similar. We will show in Section 4 that this intuition is actually incorrect. Here, we just show that it is not confirmed by the data. In RANK, we find that the rate of nonstandard reversals (44.41% of all \$-bet choices) is higher than the rate of standard reversals (15.73%; WSR test, $N = 86$, $z = -5.893$, $p < 0.001$). That is, using a ranking-based elicitation method leads to the so-called "reversal of the preference reversal phenomenon" (Casey, 1991; Alós-Ferrer et al., 2016), which so far has been considered a puzzle. In Section 4 below, we will return to it and provide an explanation (which our novel approach allows to test for).

In summary, the standard asymmetry between standard and nonstandard reversals can be shut down (and even reversed) when a ranking-based elicitation is used instead of a pricing-based one. This result lends further support to the conclusion that the asymmetry in errors between evaluations and choice, which can itself be traced back to a systematic overpricing of \$-bets, is a major cause of the asymmetry in reversals observed in PRICE and in a large number of previous preference-reversal experiments documented in the literature. However, the reversal of the preference reversal phenomenon observed in RANK cannot be explained by an asymmetry between errors in the evaluation phase and choice errors. This latter result suggests that bias-driven errors alone are insufficient to explain both phenomena. This puzzle will be resolved in the following section.

4 A Stochastic Choice Model for Preference Reversals

In this section we develop a stochastic choice model which explains the mechanisms behind both the preference reversal phenomenon and its reversal. The model allows us to further derive novel predictions on what determines the magnitude of these phenomena,

that we test and confirm with our data. The key insight is that considering the *sources* of errors in addition to those resulting directly from a bias (e.g. overpricing) fully explains the asymmetries in reversal rates. We remark that this is the first theoretical framework which is able to explain and predict both the preference reversal phenomenon and its reversal, which was previously considered puzzling.

Subsection 4.1 briefly discusses the empirical motivation of our model. Subsection 4.2 presents an intuition for the results, and especially for why the reversal of the preference reversal phenomenon, far from resulting from a bias, is merely a consequence of stochastic choice and risk aversion. Subsection 4.3 presents the actual model. Subsection 4.4 shows when the model predicts the reversal of the preference reversal phenomenon and derives a novel comparative static prediction (increased risk aversion should strengthen the phenomenon), which we then test. Subsection 4.5 shows that the bias on the evaluation of $-bets discussed in the previous section then might reverse the prediction, resulting in the original preference reversal phenomenon. This subsection also derives the novel prediction that a larger bias should strengthen the phenomenon, which we then also test. Last, Subsection 4.6 shows that the model predicts a sharp difference in reversal ratios depending on evaluation methods, which we again illustrate in the data, and indeed even allows us to predict those reversal ratios at the individual level (or a lower bound, depending on evaluation method) using only the choice phase.

4.1 Empirical motivation

It is well-known that human choices exhibit a certain degree of stochasticity, that is, subjects do not always give the same answer even when confronted with the same question multiple times (e.g., Davidson and Marschak, 1959; Tversky, 1969; Camerer, 1989; Hey and Orme, 1994; Ballinger and Wilcox, 1997; Alós-Ferrer et al., 2020), a fact that has motivated the literature on random utility models (RUMs; McFadden, 2001). Choice, however, is not purely random. On the contrary, error rates display well-known regularities, the most important of which is that error rates stand in a monotonic relation with choice difficulty (more errors for harder choices), and choices are harder when the alternatives are more similar. Overwhelming evidence for this "strength of preference" effect has been provided in the cognitive sciences (e.g., Dashiell, 1937; Moyer and Landauer, 1967; Laming, 1985; Wichmann and Hill, 2001).

In the domain of economic decisions under risk, a classical study by Mosteller and Nogee (1951) provided preliminary evidence toward an analogous effect. Recently, Alós-Ferrer and Garagnani (2018) has demonstrated that, in decisions under risk, error rates are monotonically decreasing in the difference between utilities of the alternatives. Since we have obtained estimated utilities for all our subjects, we can use them to shed light on the role of strength of preference (and, hence, stochastic choice) in explaining preference reversals. The key observation is that, by relying on our estimated individual utilities, we

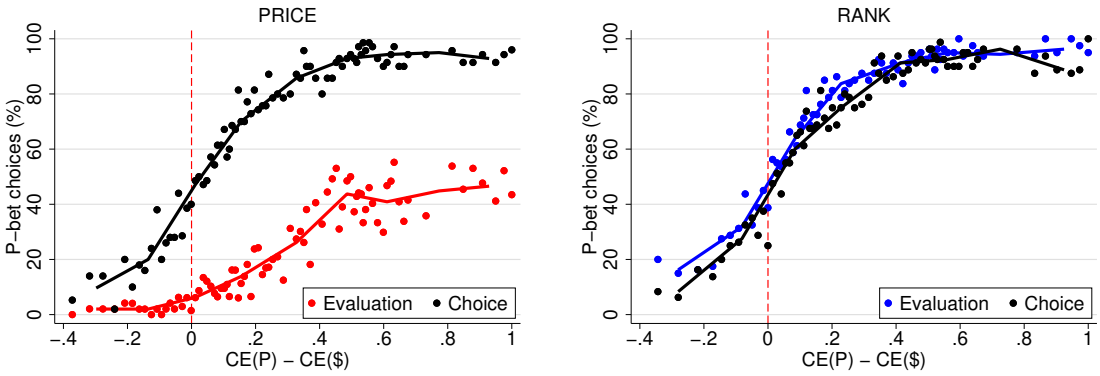


Figure 4: Treatment comparison.

Notes: Proportion of P-bet choices (or choices imputed through WTA valuation or ranking pairs) in the choice and evaluation phase of each experiment, as a function of $CE(\text{P}) - CE(\$)$. Each point represents the average choice frequency over all observations (subjects and lotteries) within a given bin of differences in CEs. The regression line is constructed using a median-band procedure which divides the x-axis in eight parts and then linearly connects the median values of the y-variable.

are able to identify a monotonic relationship between differences in certainty equivalents and the probability that the P-bet is chosen over or evaluated higher than the \$-bet.

Consider a lottery pair $(\text{P}_k, \$_k)$ in a preference reversal experiment, where P_k and $\$_k$ represent the P-bet and the \$-bet in the pair, respectively. Fix a decision maker i, and denote by $\rho_c(\text{P}_k, \$_k)$ the probability that i chooses P_k from the pair $(\text{P}_k, \$_k)$ in the choice phase of the experiment. This defines a *stochastic choice function*. Analogously, consider the evaluation phase, which might consist of WTA valuations as in our experiment PRICE or of rankings as in our experiment RANK. Denote by $\rho_v(\text{P}_k, \$_k)$ the probability that P_k is evaluated (via valuation or ranking) higher than $\$_k$. Note that this *stochastic evaluation function* ρ_v provides only imputed choices derived from evaluations and not the evaluations themselves.

For a given lottery pair $(\text{P}_k, \$_k)$ let

$$\Delta^i(\text{P}_k, \$_k) = CE_i(\text{P}_k) - CE_i(\$_k)$$

denote the difference in estimated certainty equivalents for subject i. Figure 4 (left panel) plots the proportion of choices and evaluations in PRICE that are in favor of the P-bet against the individual CE differences Δ^i. For choices, we find a monotonic sigmoidal relation between the propensity to choose the P-bet and the difference in certainty equivalents. That is, the P-bet is chosen more often for larger CE differences, showing that the latter can be used to capture the effects of strength of preference in our context. In PRICE, this empirical stochastic choice function is roughly symmetric around zero and takes the value one half for CE differences close to zero. For evaluations, we also find a monotonically increasing relation between the propensity to evaluate the P-bet above the \$-bet and Δ^i. However, the empirical stochastic evaluation function is clearly

shifted downwards taking a value well below one half around zero. Even for relatively large differences in CE, the propensity to evaluate the P-bet higher than the $-bet barely reaches 50%. This observation is compatible with the overpricing bias identified in the previous section, which leads to a strong downwards shift of the empirical stochastic evaluation function relative to the empirical stochastic choice function in PRICE.

Figure 4 (right panel) plots the analogous graph for experiment RANK. The empirical stochastic choice function again takes a sigmoidal shape exhibiting a monotonically increasing relation. The second curve plots the proportion of pairs for which the P-bet was ranked higher within the pair. In stark contrast to the PRICE experiment, the empirical stochastic evaluation function for RANK exhibits no systematic shift relative to the stochastic choice function. That is, for RANK the stochastic evaluation function exhibits no bias relative to the stochastic choice function.

In summary, we conclude that both choices and evaluations involved in preference reversals display a monotonic "strength of preference" effect relative to the subjective difference between options. The latter seems to be well-captured by the difference in estimated certainty equivalents. The empirical stochastic choice functions are very similar in both experiments, PRICE and RANK. However, and most importantly for our purposes, when we consider the imputed choices inferred from rankings or WTA valuations, we observe a clear difference. For RANK, the stochastic evaluation function is essentially indistinguishable from the stochastic choice function, whereas for PRICE the stochastic evaluation function is clearly shifted downwards. The latter reflects overpricing and indicates that WTA valuations of $-bets react differently (compared to choices) to differences in estimated certainty equivalents.

4.2 Intuition for the results

We now give the intuition of how a stochastic choice framework that incorporates the monotonicities observed above can explain both the preference reversal phenomenon and its reversal. The key insight is that the comparison of standard and nonstandard reversal rates hinges upon conditioning on the actual choice of either a P-bet or a $-bet. Fix a lottery pair $(P_k, \$_k)$ with CE difference Δ_k. Then, conditional on a P-bet choice the likelihood to observe a standard reversal for $(P_k, \$_k)$ is simply the likelihood that $\$_k$ is evaluated higher than P_k, which is $1 - \rho_v(P_k, \$_k)$. Analogously, conditional on a $-bet choice the likelihood to observe a nonstandard reversal is $\rho_v(P_k, \$_k)$. Consider the example illustrated in the left panel of Figure 5. The stochastic valuation function $\rho_v(\Delta)$ is monotonically increasing and exhibits no bias, that is, $\rho_v(0)$ is exactly one half. Then for any lottery pair with a certainty equivalent difference of zero, the likelihood to evaluate the P-bet above the $-bet is exactly 50%, and consequently we should expect similar rates of standard and nonstandard reversals. However, if the certainty equivalent difference Δ is strictly positive, then the probability that the P-bet is evaluated above

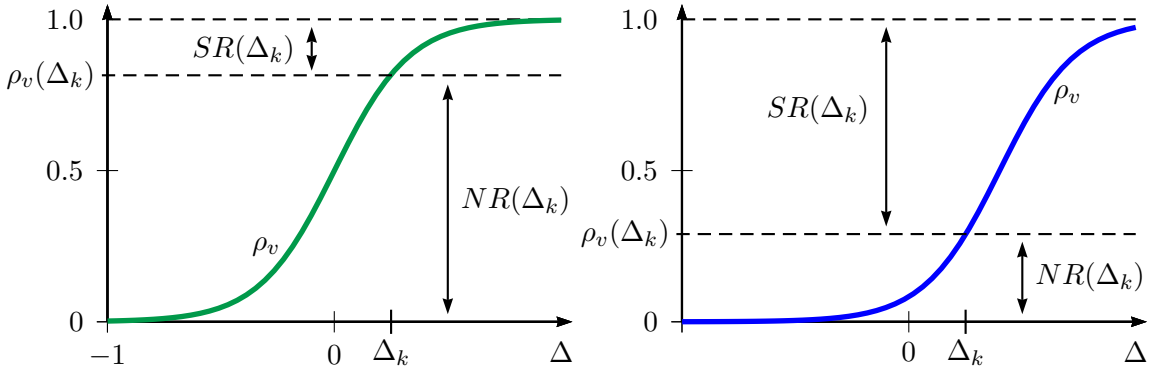


Figure 5: Stochastic evaluation functions and reversal rates.

the \$-bet increases. Consequently, one should expect more nonstandard than standard reversals for pairs with $\Delta_k > 0$ as the left-hand side of the figure illustrates.

In typical preference reversal experiments, lottery pairs $(\mathrm{P}_k, \$_k)$ are constructed such that the expected values of $\$_k$ and P_k are similar, but $\$_k$ is riskier. Since decision makers tend to be risk averse, it follows that for the majority of lottery pairs CE differences are positive. Hence, in the absence of a bias in evaluations, risk aversion leads to the reversal of the preference reversal phenomenon with more nonstandard than standard reversals. Additionally, a novel, comparative-statics prediction arises: higher risk aversion (or, equivalently, larger Δ for a given lottery pair) will exacerbate the difference in reversal rates and hence strengthen the reversal of the preference reversal phenomenon.

Now consider the second example illustrated in the right panel of Figure 5. Here, the stochastic valuation function $\rho_v(\Delta)$ is biased toward the \$-bets, that is, $\rho_v(0)$ is below one half. In that case, also for some pairs with positive differences in certainty equivalents, the likelihood to evaluate the P-bet above the \$-bet is smaller than 50%. As a result, we should expect more standard than nonstandard reversals even in the presence of (moderate) risk-aversion. Hence, a strong-enough bias toward \$-bets in evaluations (e.g. due to overpricing) leads to the preference reversal phenomenon. Additionally, a further, novel comparative-statics prediction obtains, namely that a stronger bias in evaluations should strengthen the phenomenon.

4.3 The model

Consider a preference-reversal experiment D with K binary lottery choices,

$$D = \{(\mathrm{P}_1, \$_1), \ldots, (\mathrm{P}_K, \$_K)\},$$

where P_k and $\$_k$ represent the P-bet and the \$-bet in the kth pair, respectively. Suppose that a given decision maker is asked to express her preferences using two different elicitation procedures to which we refer as "choice" and "evaluation" for simplicity. We assume that the decision maker can be characterized by the *stochastic choice function* $\rho_c(\mathrm{P}_k, \$_k)$

and the *stochastic evaluation function* $\rho_v(\mathrm{P}_k, \$_k)$ described in Subsection 4.1. That is, for the pair $(\mathrm{P}_k, \$_k)$, $\rho_c(\mathrm{P}_k, \$_k)$ is the probability that P_k is chosen, and $\rho_v(\mathrm{P}_k, \$_k)$ is the probability that P_k is evaluated (e.g. via WTA valuations or rankings) higher than $\$_k$.

For such a decision maker the likelihood to observe a standard reversal for a pair $(\mathrm{P}_k, \$_k)$ is $\rho_c(\mathrm{P}_k, \$_k)(1 - \rho_v(\mathrm{P}_k, \$_k))$, whereas the likelihood to observe a nonstandard reversal is $(1 - \rho_c(\mathrm{P}_k, \$_k))\rho_v(\mathrm{P}_k, \$_k)$. Consequently, the expected (average) standard reversal rate in experiment D is

$$SR(D, \rho_c, \rho_v) = \sum_{k=1}^{K} \frac{\rho_c(\mathrm{P}_k, \$_k)}{\sum_{l=1}^{K} \rho_c(\mathrm{P}_l, \$_l)} (1 - \rho_v(\mathrm{P}_k, \$_k)) \tag{1}$$

and the expected (average) nonstandard reversal rate is

$$NR(D, \rho_c, \rho_v) = \sum_{k=1}^{K} \frac{(1 - \rho_c(\mathrm{P}_k, \$_k))}{\sum_{l=1}^{K} (1 - \rho_c(\mathrm{P}_l, \$_l))} \rho_v(\mathrm{P}_k, \$_k). \tag{2}$$

Our stochastic choice model postulates that the functions ρ_c and ρ_v are monotonic in the difference in certainty equivalents within a lottery pair, as empirically observed in Subsection 4.1 above. That is, we assume that, as observed in our data, the propensity to choose or evaluate the P-bet over the \$-bet is increasing in the certainty equivalent difference, $\Delta^i(\mathrm{P}, \$) = CE_i(\mathrm{P}) - CE_i(\$)$.

Formally, given a preference reversal experiment $D = \{(\mathrm{P}_1, \$_1), \ldots, (\mathrm{P}_K, \$_K)\}$ we denote the corresponding CE differences for subject i by $\Delta^i = \{\Delta^i_1, \ldots, \Delta^i_K\}$ where $\Delta^i_k = \Delta^i(\mathrm{P}_k, \$_k)$. We say that a decision maker (DM) exhibits *strength-of-preference* (SoP) effects if ρ_c and ρ_v can be written as increasing functions of Δ^i_k. To simplify notation we henceforth drop the obvious dependencies on i and simply write Δ_k instead of Δ^i_k. For a DM exhibiting SoP effects, the expected (average) reversal rates in (1) and (2) are then functions of the CE differences Δ given by

$$SR(\Delta, \rho_c, \rho_v) = \sum_{k=1}^{K} \frac{\rho_c(\Delta_k)}{\sum_{l=1}^{K} \rho_c(\Delta_l)} (1 - \rho_v(\Delta_k)) \tag{3}$$

and

$$NR(\Delta, \rho_c, \rho_v) = \sum_{k=1}^{K} \frac{(1 - \rho_c(\Delta_k))}{\sum_{l=1}^{K} (1 - \rho_c(\Delta_l))} \rho_v(\Delta_k), \tag{4}$$

respectively.

4.4 Risk aversion and the reversal of the preference reversal phenomenon

Although historically the observation of the preference reversal phenomenon preceded the observation of its reversal, conceptually it is easier to explain the latter first. We now proceed to show that this phenomenon is simply a consequence of stochastic choice

(strength of preference effects) and risk aversion. In particular, no behavioral bias of any kind is needed to explain and predict it.

In standard preference reversal experiments, lottery pairs $(\mathrm{P}_k, \$_k)$ are constructed so that the expected values of $\$_k$ and P_k are similar, but $\$_k$ is riskier. Since DMs tend to be risk averse, one typically observes more P-bet choices than \$-bet choices. That is, if we denote the total proportion of P-choices by $\pi_c = \frac{1}{K}\sum_{k=1}^{K}\rho_c(\Delta_k)$, we will typically observe that $\pi_c > \frac{1}{2}$. In contrast, if π_c is one half, then the decision maker shows no risk aversion (on the aggregate level).

Analogously, we denote the total proportion of valuations in favor of P by $\pi_v = \frac{1}{K}\sum_{k=1}^{K}\rho_v(\Delta_k)$. In the absence of any bias across evaluation methods (as suggested by Figure 4 (right) for experiment RANK), we should expect $\pi_c = \pi_v$, but of course inconsistencies of both types will still occur due to the fact that both choices and evaluations are stochastic. For ease of reference, say that *evaluations are unbiased* if $\pi_c = \pi_v$. On the contrary, if there is a systematic difference between choices and evaluations, we will generally observe $\pi_c \neq \pi_v$. We say that a decision maker (Δ, ρ_c, ρ_v) *exhibits reduced risk-aversion in evaluations* (relative to choice) if $\pi_c > \pi_v$, that is, choices imputed through evaluations favor the riskier \$-bets more often on the aggregate, compared to direct binary choices (as suggested by Figure 4 (left) for experiment PRICE).

It is now easy to show that there is a direct link between the ratio of the standard and nonstandard reversal rates and the quantities discussed above. Specifically, using (1) and (2) we obtain that

$$\begin{aligned}\frac{SR(\Delta,\rho_c,\rho_v)}{NR(\Delta,\rho_c,\rho_v)} &= \frac{\left(\sum_{l=1}^{K}(1-\rho_c(\Delta_l))\right)\cdot\sum_{k=1}^{K}\rho_c(\Delta_k)(1-\rho_v(\Delta_k))}{\left(\sum_{l=1}^{K}\rho_c(\Delta_l)\right)\cdot\sum_{k=1}^{K}(1-\rho_c(\Delta_k))\rho_v(\Delta_k)} \\ &= \frac{(1-\pi_c)}{\pi_c}\left[\frac{\sum_{k=1}^{K}\rho_c(\Delta_k)-\rho_c(\Delta_k)\rho_v(\Delta_k))}{\sum_{k=1}^{K}\rho_v(\Delta_k)-\rho_c(\Delta_k)\rho_v(\Delta_k)}\right]\end{aligned}$$

which simplifies to

$$\frac{SR(\Delta,\rho_c,\rho_v)}{NR(\Delta,\rho_c,\rho_v)} = \frac{1-\pi_c}{\pi_c}\left[\frac{\pi_c-\frac{1}{K}\sum_{k=1}^{K}\rho_c(\Delta_k)\rho_v(\Delta_k)}{\pi_v-\frac{1}{K}\sum_{k=1}^{K}\rho_c(\Delta_k)\rho_v(\Delta_k)}\right] \tag{5}$$

The preference reversal phenomenon describes the consistent empirical finding that standard reversals are more frequent than nonstandard ones, when evaluations are monetary. The implicit (and sometimes also explicit) assumption in the literature is that, if any potential bias in evaluations relative to choice (e.g. overpricing) could be removed, the remaining standard and nonstandard preference reversals should be due to simple behavioral noise and result in identical reversal rates (standard and nonstandard). This assumption is incorrect. Corollary 1 below shows that, as a direct consequence of equation (5), in the absence of a systematic difference between evaluations and choices (as in our experiment RANK), behavioral noise does *not* lead to equal reversal rates. On the

contrary, our model then predicts the reversal of the preference reversal phenomenon, where nonstandard reversal rates are larger than standard ones. The proof of the following result is immediate given equation (5).

Corollary 1. *If evaluations are unbiased ($\pi_c = \pi_v$), aggregate risk aversion in choices ($\pi_c > \frac{1}{2}$) leads to the reversal of the preference reversal phenomenon, that is, $SR(\Delta, \rho_c, \rho_v) < NR(\Delta, \rho_c, \rho_v)$.*

This corollary fully explains the reversal of the preference reversal phenomenon, as observed e.g. in our experiment RANK. As we have seen, there is little empirical difference between the stochastic choices given by ρ_c and the choices imputed from stochastic evaluations given by ρ_v. In this case, the empirical observation of larger rates for nonstandard than for standard reversals is simply a consequence of stochastic choice and risk aversion.[13] Since decision makers are typically risk averse, the default (in the absence of a systematic difference between evaluations and choices) is a lower rate of standard than nonstandard reversals.

We can take this observation further and derive novel predictions from our model. Suppose that there is no systematic difference between choices and evaluations, in the sense that $\pi_v = \pi_c$. Then equation (5) simplifies dramatically to

$$\frac{SR}{NR} = \frac{1 - \pi_c}{\pi_c}.$$

This allows us to derive a comparative statics prediction. Say that ρ_c^1 *exhibits more risk aversion* than ρ_c^2 if $\pi_c^1 > \pi_c^2$. The next (straightforward) result shows that more risk aversion leads to a stronger reversal of the preference reversal phenomenon (in relative terms) if evaluations are unbiased (relative to choice).

Proposition 1. *Suppose two decision makers are characterized by (ρ_c^1, ρ_v^1) and (ρ_c^2, ρ_v^2) such that $\rho_c^1 = \rho_v^1$ and $\rho_c^2 = \rho_v^2$. If ρ_c^1 exhibits more risk aversion than ρ_c^2, then*

$$\frac{SR(\Delta, \rho_c^1, \rho_v^1)}{NR(\Delta, \rho_c^1, \rho_v^1)} < \frac{SR(\Delta, \rho_c^2, \rho_v^2)}{NR(\Delta, \rho_c^2, \rho_v^2)}.$$

Proposition 1 provides a testable implication on how differences in risk aversion affect the relation between the rates of standard and nonstandard reversals. The ratio of standard to nonstandard reversals should be decreasing as risk aversion increases. To test this prediction, we conducted a median split of subjects in experiment RANK according to their individually-estimated risk attitudes. The left panel of Figure 6 shows the average proportion of P-choices for the two groups. Indeed, in the choice phase subjects with high risk aversion (estimated in the estimation phase) choose the P-bet more often

[13] If decision makers are risk seeking, it follows analogously to Corollary 1 that the default is the preference reversal phenomenon. Indeed, considering only the 13 risk-seeking subjects in RANK we find more standard (28.3%) than nonstandard reversals (17.1%), hence reversing the reversal of the preference reversal phenomenon. However, due to the low number of observations this difference is not significant at any conventional level.

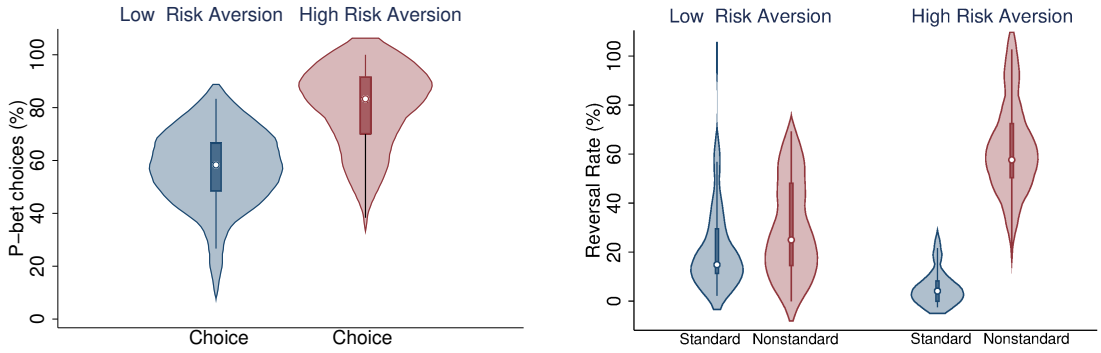


Figure 6: Left: Proportion of P-choices for high and low risk aversion in RANK. Right: Asymmetry in reversal rates for high and low risk aversion (median split) in RANK.

than subjects with low risk aversion (80.10% versus 56.31%; MWW test, $N = 95$, $z = 6.022$, $p < 0.001$). That is, the former group exhibits more risk aversion (as defined above) than the latter group. Thus Proposition 1 predicts a smaller ratio of standard to nonstandard reversals for high risk aversion subjects than for low risk aversion subjects. For the latter group we find that the rate of standard and nonstandard reversals is 22.21% and 29.87%, respectively. In contrast, for the high risk aversion group we have average rates of standard and nonstandard reversals of 8.03% and 61.92%, respectively. A comparison of the two groups reveals that the ratio of standard to nonstandard reversals for high risk aversion is significantly smaller than for low risk aversion (MWW test, $N = 89$, $z = -6.309$, $p < 0.001$), in line with Proposition 1. That is, risk aversion exacerbates the reversal of the preference reversal phenomenon in experiment RANK.

Summarizing, we have obtained two important and novel insights. First, with risk averse subjects, we should *expect* the reversal of the preference reversal phenomenon to obtain in any experiment (as RANK) where there is no systematic difference between choices and evaluations. Explaining this phenomenon does not require any bias, and its strength is monotonically related to the degree of risk aversion.

The second insight is more subtle. Ever since the preference reversal phenomenon was first discovered (Slovic and Lichtenstein, 1968; Lichtenstein and Slovic, 1971; Lindman, 1971; Grether and Plott, 1979), dozens of contributions have reported the effect to be both robust and large, with considerably-higher rates of standard reversals compared to the rates of nonstandard reversals. Ironically, it seems that this wide consensus has hidden a misunderstanding: the effect has actually been *underestimated*. The reason is that the literature has implicitly assumed that the "default" situation in the absence of whatever causal determinants were behind the phenomenon should have been an equality in reversal rates. Corollary 1 shows that this is incorrect. In the absence of those determinants (and, specifically, overpricing of \$-bets), the default situation is one where the rates of nonstandard reversals are *higher*, and hence the fact that they become lower in preference-reversal experiments with monetary valuations shows

that the underlying determinants are stronger than implicitly assumed. This is fully aligned with our previous observation that overpricing of $-bets is of an extremely high magnitude in monetary terms (Section 3.3).

4.5 Overpricing and the preference reversal phenomenon

We now turn to the traditional preference reversal phenomenon. As seen in the previous subsection, this phenomenon cannot be explained in the absence of an additional bias, since the actual prediction when evaluations and choices do not systematically differ is the reversal of the phenomenon. In Section 3, however, we have documented an upward, strong bias in WTA valuations in experiment PRICE. In this subsection, we show that this bias counteracts the effects discussed above.

The essence of the effect can be seen directly in equation (5). Suppose, as a thought experiment, that decision makers were risk neutral on the aggregate, in the sense that $\pi_c = \frac{1}{2}$. It is then an immediate implication of equation (5) is that the preference reversal phenomenon is predicted if evaluations favor the $-bets compared to choices, i.e. $\pi_c > \pi_v$. Formally, we obtain the following corollary.

Corollary 2. *In the absence of aggregate risk aversion in choices ($\pi_c = \frac{1}{2}$), reduced risk-aversion in evaluations ($\pi_c > \pi_v$) leads to the preference reversal phenomenon, that is, $SR(\Delta, \rho_c, \rho_v) > NR(\Delta, \rho_c, \rho_v)$.*

However, decision makers are typically risk averse. More generally, fix the proportion of P-bet choices π_c, and assume aggregate risk aversion in choices, $\pi_c > \frac{1}{2}$. Suppose evaluations are biased and DMs exhibit reduced risk-aversion in evaluations, $\pi_c > \pi_v$. Consider increasingly-biased DMs, corresponding to smaller and smaller values of π_v. While the left-hand fraction in (5) is smaller than one by risk-aversion, the right-hand fraction is larger than one and becomes larger as π_v becomes smaller, eventually leading to a larger rate of standard than nonstandard reversals. That is, for moderate levels of risk aversion, a sufficiently reduced risk aversion in evaluations compared to choices will result in the preference reversal phenomenon. As shown in Section 3, the bias in favor of $-bets for monetary evaluations is very large, and hence should be expected to offset the effect of standard levels of risk aversion.

We now quantify this latter intuition while working directly with the stochastic choice and evaluation functions ρ_c and ρ_v. A DM exhibiting SoP effects *exhibits a $-bias* in evaluations if $\rho_v(\Delta) < \rho_c(\Delta)$ for all Δ. That is, evaluations are biased away from the P-bet and toward the $-bet relative to choices. The left panel of Figure 7 shows examples of stochastic choice and evaluation functions for a DM exhibiting a $-bias in evaluations. Intuitively, if there is a $\delta > 0$ such that $\rho_v(\delta) = \frac{1}{2}$, then δ captures the extent of the bias of evaluations toward the $-bets, that is δ is the premium that makes the decision maker (stochastically) indifferent between P and $\$ + \delta$. Note that a DM exhibiting a $-bias in evaluations will also exhibit reduced risk-aversion in evaluations, i.e. $\pi_c < \pi_v$ in the terms of the previous subsection.

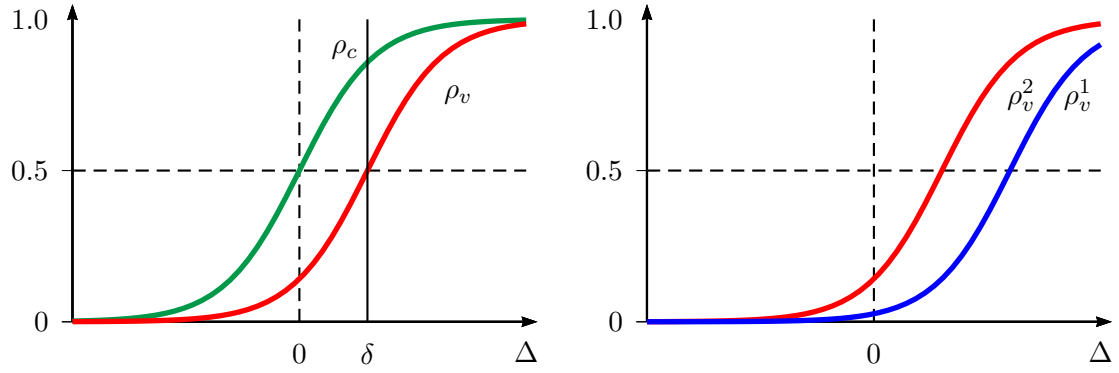


Figure 7: Illustration of overpricing as a result of stochastic valuations.

To show how a \$-bias in evaluations affects the asymmetry between reversal rates we need to be able to compare the extent of the bias across decision makers. Consider two decision makers characterized by (ρ_c^1, ρ_v^1) and (ρ_c^2, ρ_v^2) that exhibit SoP effects. We say that ρ_v^1 exhibits a *larger \$-bias* than ρ_v^2 if $\rho_v^1(\Delta_k) < \rho_v^2(\Delta_k)$ for all k. Intuitively, a larger \$-bias means that a DM is more likely to evaluate the \$-bet higher than the P-bet. Figure 7 (right) illustrates this graphically. The stochastic evaluation function ρ_v^1 is shifted to the right relative to ρ_v^2 resulting in a higher propensity to evaluate the \$-bet above the P-bet for any Δ.

Our next result shows that *ceteris paribus* a larger \$-bias in evaluations exacerbates the difference between the rates of standard and nonstandard reversals.

Proposition 2. *Let $D = \{(P_1, \$_1), \ldots, (P_K, \$_K)\}$ be a preference-reversal experiment. Suppose two decision makers are characterized by (ρ_c, ρ_v^1) and (ρ_c, ρ_v^2) and both exhibit SoP effects. If ρ_v^1 has a larger \$-bias than ρ_v^2, then $SR(\Delta, \rho_c, \rho_v^1) - NR(\Delta, \rho_c, \rho_v^1) > SR(\Delta, \rho_c, \rho_v^2) - NR(\Delta, \rho_c, \rho_v^2)$ for any stochastic choice function ρ_c.*

Proof. Consider an arbitrary stochastic choice function ρ_c. We have

$$\begin{aligned} SR(\Delta, \rho_c, \rho_v^1) - NR(\Delta, \rho_c, \rho_v^1) &= \frac{\sum_{k=1}^K \rho_c(\Delta_k)(1-\rho_v^1(\Delta_k))}{\sum_{l=1}^K \rho_c(\Delta_l)} - \frac{\sum_{k=1}^K (1-\rho_c(\Delta_k))\rho_v^1(\Delta_k)}{\sum_{l=1}^K (1-\rho_c(\Delta_l))} \\ &= \sum_{k=1}^K \frac{\rho_c(\Delta_k)}{\sum_{l=1}^K \rho_c(\Delta_l)} - \left(\frac{\rho_c(\Delta_k)}{\sum_{l=1}^K \rho_c(\Delta_l)} + \frac{(1-\rho_c(\Delta_k))}{\sum_{l=1}^K (1-\rho_c(\Delta_l))} \right) \rho_v^1(\Delta_k) \\ &> \sum_{k=1}^K \frac{\rho_c(\Delta_k)}{\sum_{l=1}^K \rho_c(\Delta_l)} - \left(\frac{\rho_c(\Delta_k)}{\sum_{l=1}^K \rho_c(\Delta_l)} + \frac{(1-\rho_c(\Delta_k))}{\sum_{l=1}^K (1-\rho_c(\Delta_l))} \right) \rho_v^2(\Delta_k) \\ &= SR(\Delta, \rho_c, \rho_v^2) - NR(\Delta, \rho_c, \rho_v^2). \end{aligned}$$

Since ρ_v^1 exhibits a larger \$-bias than ρ_v^2, it follows that $\rho_v^1(\Delta_k) < \rho_v^2(\Delta_k)$ for all $\Delta_k \in \Delta$, which implies the inequality. This completes the proof. □

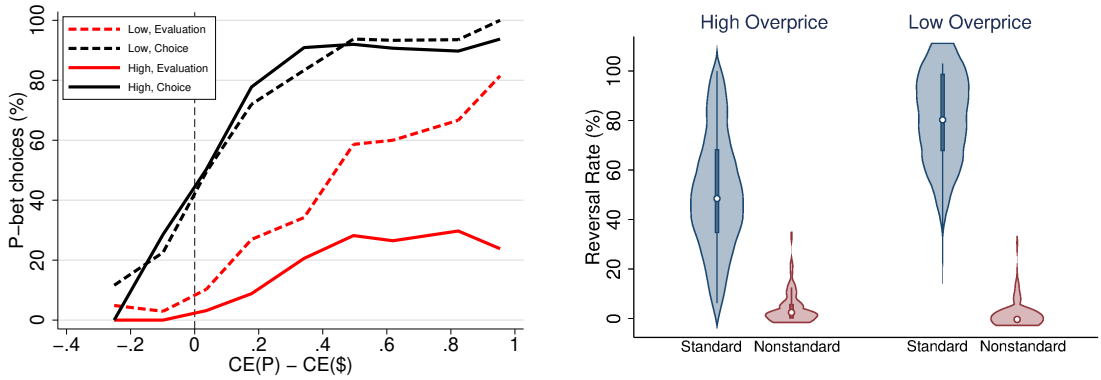


Figure 8: High versus low overpricing.

Notes: Left: Proportion of P-bet choices (or choices imputed through WTA valuation or ranking pairs) in the choice and evaluation phase as a function of $CE(\text{P}) - CE(\$)$, separately for for high and low overpricing (median split) in PRICE. Right: Asymmetry in reversal rates for high and low overpricing (median split) in PRICE.

As a consequence, for given risk aversion, a DM is more likely to exhibit the preference reversal phenomenon the larger his or her \$-bias is. Given the large and systematic bias documented in Section 3, it is not surprising that the preference reversal phenomenon is observed empirically for most decision makers when monetary valuations are used.

This result also delivers a further novel, testable implication of the effects of a \$-bias on the asymmetry in reversal rates. For experiment PRICE, Proposition 2 predicts that more overpricing, that is, a larger \$-bias, increases the difference between standard and nonstandard reversals. That is, if the bias due to overpricing is strong enough the preference reversal phenomenon is expected to occur, and the asymmetry is expected to be larger the more pronounced the bias.

To test this prediction we divide subjects into two groups based on the estimated average overpricing $(WTA_i(A) - CE_i(A)$, averaged over all lotteries A, for each i; recall Section 3.3) using a median split. Figure 8 (left) shows the empirical (average) stochastic choice and evaluation functions separately for subjects with high and low overpricing. For the former, the stochastic evaluation function is shifted downwards compared to the latter, that is, the group with high estimated overpricing indeed exhibits a larger \$-bias in evaluations. Figure 8 (right) shows violin plots of the reversal rates for both groups. We find that, in line with the prediction of Proposition 2, overpricing exacerbates the asymmetry between standard (low 51.04%, high 77.45%) and nonstandard reversals (low 4.67%, high 2.44%). This difference between reversal rates is statistically significant (MWW test, $N = 86$, $z = -4.664$, $p < 0.001$). Summarizing, we find that the asymmetry between the rates of standard and nonstandard reversals is larger for subjects that exhibit more overpricing, confirming the prediction of Proposition 2.

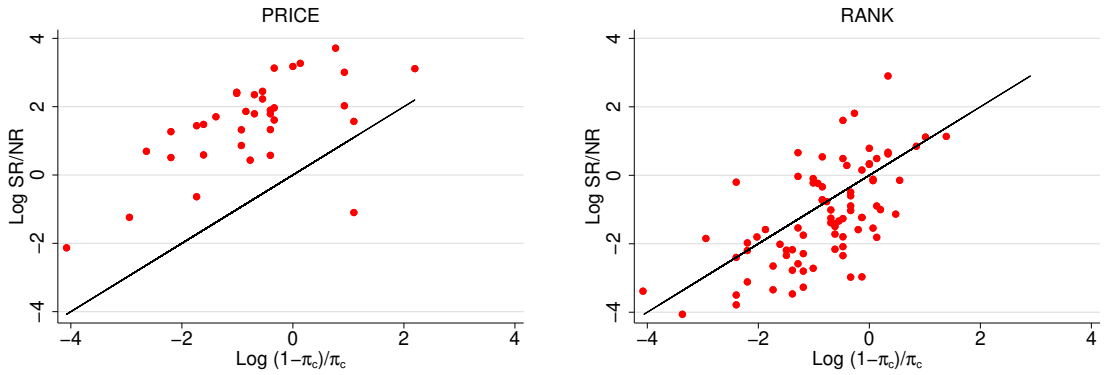


Figure 9: Logarithm of of the individual reversal ratio $\frac{SR}{NR}$ against logarithm of the individual choice odds $\frac{1-\pi_c}{\pi_c}$, separately for PRICE (left) and RANK (right).

4.6 The relationship between reversal ratios and choice ratios

Equation (5) yields additional implications related to the comparison of evaluation tasks, which we can illustrate with our data. For PRICE, a \$-bias in evaluations due to overpricing leads to a reduced risk aversion for valuations, $\pi_v < \pi_c$, which implies $\frac{SR}{NR} > \frac{1-\pi_c}{\pi_c}$. Similarly, for RANK we have $\pi_v \sim \pi_c$, which implies $\frac{SR}{NR} \sim \frac{1-\pi_c}{\pi_c}$. That is, we can use Equation (5) to predict individual reversal ratios from individual choice odds (which do not involve evaluations). Figure 9 plots the ratio of standard to nonstandard reversals against the ratio of the proportion of \$-bet choices to P-bet choices (both on a log scale) for PRICE (left) and RANK (right). In PRICE, for all individuals but one the reversal ratio is larger than the corresponding choice odds. In contrast, for RANK the points are clustered around the diagonal, that is, for many individuals the reversal ratio is predicted very well by the choice odds.

5 Related Literature

In this section, we first briefly discuss the relation of our work to the existing (vast) literature on the preference reversal phenomenon, and specifically, how our work differs from previous, key contributions. Then, we proceed to show how our results allow us to organize and explain previous empirical findings, including several which were previously seen as inconsistent.

5.1 Differences with respect to the previous literature

Our results are in line with the natural and long-standing hypothesis that elicited monetary valuations might be inherently noisier than choices (Lichtenstein and Slovic, 1971). In a repeated-choice experiment where subjects made choices and stated monetary valuations for the same pairs of lotteries several times, Schmidt and Hey (2004) found that excluding pairs with inconsistent valuations reduced reversals, whereas excluding pairs

with inconsistent choices did not affect reversals. Under the assumption that inconsistencies are associated with more errors, their results align with our finding that evaluation errors a major driver of preference reversals. In contrast to this work, our approach is based on an out-of-sample estimation providing a direct classification of errors for choices and evaluations. Further, our results show that excess noise in valuations is not a general phenomenon, but instead is mostly associated with \$-bets.

Both our theoretical analysis and our experimental results highlight the importance of overpricing to explain the preference reversal phenomenon, while showing at the same time that overpricing alone cannot provide a full account of preference reversals. In their seminal work, Tversky et al. (1990) proposed the scale compatibility hypothesis, which singled out overpricing as a likely cause of preference reversals, but as the authors state themselves their method only allowed to "identify the sign of the discrepancy between pricing and choice; the labels do not imply that the bias resides in the pricing." More recently, Loomes and Pogrebna (2017) find valuations elicited via a BDM procedure to be systematically higher compared to stochastic indifference points inferred from repeated binary choices. They find valuations to be biased upwards compared to stochastic indifference points for both P-bets and \$-bets, but, in line with overpricing, the difference for the latter is of a larger magnitude. Relative to these papers, our empirical contributions are that we are able to directly attribute the bias to the evaluation phase, to show that the source of the asymmetry in errors is actually the systematic and large overpricing of \$-bets, and beyond that, to even quantify the extent of the bias with respect to certainty equivalents. Crucially, our stochastic choice model shows that overpricing combined with monotonicities in stochastic choice and risk aversion provides a unified account of preference reversals, their reversal, and the magnitude of both.

An important difficulty with the scale compatibility hypothesis is that it attributes the phenomenon, and the overpricing of \$-bets in particular, exclusively to the psychological effects of activating a monetary scale through the request to generate a price. This would imply that the phenomenon should disappear or even be reversed if valuations were elicited through probability equivalents, i.e. if decision makers are asked to report the probability p leaving them indifferent between a given lottery and obtaining a fixed amount X with probability p. Experiments using probability equivalents were carried out by Hershey and Schoemaker (1985) and Collins and James (2015), but the expected result was not found. Although preference reversals were indeed less frequent, they did not disappear or revert (however, see the next subsection). A possible explanation is that it is not just the activation of the monetary scale that creates the overpricing bias, but simply the salience of the high monetary outcome of the \$-bet, which becomes evident when a monetary scale is used but also for comparisons with a lottery (p, X) for a large X. This interpretation is in line with salience theory (Bordalo et al., 2012, 2013), which assumes that decision makers overweight salient states with salience being linked to payoff magnitudes (or differences thereof). Bordalo et al. (2012) argue that overpricing is particularly pronounced for valuations since in this case the lottery is evaluated

against the alternative of "not having it," which renders the upside more salient (in contrast, any ranking task must provide other lotteries as alternatives). Hence, stated valuations might be anchored on the largest monetary outcome of a lottery. Since $-bets are constructed using large outcomes, those become salient, exacerbating anchoring. In Appendix D we confirm this intuition by showing that the empirically-observed overpricing is compatible with a salience-based anchoring mechanism.

Another influential strand of the literature has examined whether preference imprecision (leading to behavioral noise) might be among the causes of reversals (Butler and Loomes, 1988, 2007, 2011; Dubourg et al., 1994; Morrison, 1998; Butler et al., 2014; Cubitt et al., 2015). Elicitation methods in these studies typically ask subjects to explicitly express imprecision in their preferences, e.g. by using interval measures. This approach has been fruitfully applied to the study of various decision-theoretic anomalies, and several authors (e.g., Butler and Loomes, 2007) have suggested that preference imprecision might contribute to the preference reversal phenomenon. Our design is fundamentally different and we cannot examine preference imprecision in this sense. There is, however, a close link between preference imprecision and behavioral noise, as pointed out by, e.g., Loomes (2005) and Cubitt et al. (2015, p. 2). Our approach linking error rates and "strength of preference" is related but differs in that we explicitly model stochastic choice and evaluations as a functions of (estimated) differences in certainty equivalents. In that sense, our stochastic choice model incorporates behavioral noise but cannot be encompassed simply by appealing to errors in the absence of overpricing of $-bets.

5.2 Explaining previous results from the literature

We now proceed to reexamine existing preference reversal experiments from the literature in light of our results, following an approach analogous to the one in Subsection 4.6. Figure 10 plots the ratio of the average proportion of $-bet choices to P-bet choices (choice odds) against the ratio of the reversal rates (reversal ratio) for 16 previous experiments. Points in the upper part show the preference reversal phenomenon, whereas points below zero show the reversal of the phenomenon. Points left of the vertical line at zero indicate (average) risk-aversion ($\pi_c > 0.5$), whereas points to the right indicate (average) risk-seeking behavior.

The majority of the experiments (11 out of 16) find the classic asymmetry with more standard than nonstandard reversals, two experiments find basically identical rates of reversals of both types, and three experiments find a reversal of the preference reversal phenomenon. The majority (again 11 out of 16) of the experiments exhibit (aggregate) risk aversion with an average proportion of P-bet choices exceeding the average proportion of $-bet choices and only five experiments exhibit (aggregate) risk-seeking behavior. Comparing choice odds and reversal ratios across experiments allows us to obtain further insights on the extent of the bias for different evaluation methods. As we argue below, those insights in combination with differences in observed risk aversion across experi-

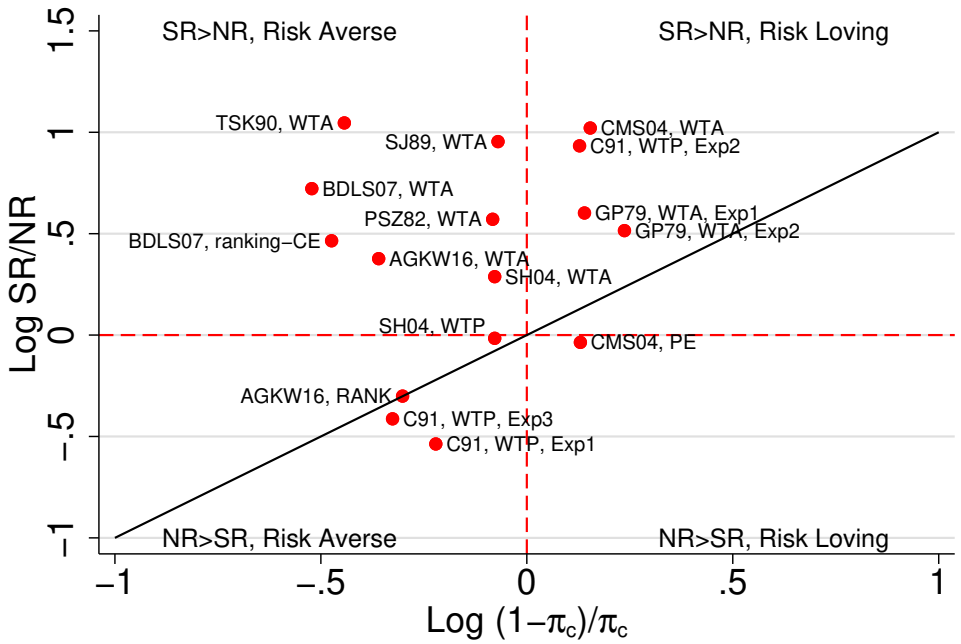


Figure 10: Logarithm of the aggregate reversal ratio $\frac{SR}{NR}$ against logarithm of the aggregate choice odds $\frac{1-\pi_c}{\pi_c}$ for different experiments in the literature.

ments can account for most of the differences in reversal rates found in the literature and can shed light on earlier findings which were viewed as inconsistent.

Previous contributions that elicited evaluations using WTA-valuations consistently find evidence for the preference reversal phenomenon with a higher rate of standard than nonstandard reversals. In the figure, those correspond to the points labeled with "WTA:" GP79 (Grether and Plott, 1979), PSZ82 (Pommerehne et al., 1982), SJ89 (Schkade and Johnson, 1989), TSK90 (Tversky et al., 1990), CMS04 (Cubitt et al., 2004), SH04 (Schmidt and Hey, 2004), BDLS07 (Bateman et al., 2007), and AGKW16 (Alós-Ferrer et al., 2016). All of those are above the diagonal line, as predicted by equation (5).

We now discuss experiments deviating from the standard WTA design. In addition to the WTA-based experiment listed above, Schmidt and Hey (2004) ran a second experiment (SH04, WTP) where evaluations were elicited via WTP-valuations (Willingness To Pay). The lotteries in both experiments were such that on average subjects exhibited almost no risk aversion ($\pi_c = 0.54$), with no difference across experiments in this respect. Hence, the odds ratio is close to one and one should expect no asymmetry between reversal rates if there were no bias in the evaluation task. While for WTA the standard asymmetry was observed, in agreement with our results suggesting a strong bias for this method, for WTP reversal rates were indistinguishable, which suggests that WTP-valuations exhibit no bias, or at least a smaller \$-bias than those based on WTA.

Evaluations through WTP were also used in a series of (non-incentivized) experiments (C91) by Casey (1991). If WTP-valuations exhibit a small or no bias, equation

(5) would suggest that whether the preference reversal phenomenon or its reversal is observed should mainly depend on aggregate risk aversion in the experiment. Specifically, for risk-seeking behavior we should expect the preference reversal phenomenon, and for risk aversion we should expect the reversal of the phenomenon. These are essentially the results in Casey (1991), which were then viewed as inconsistent. In Experiments 1 and 3 (with large, hypothetical stakes) the average proportion of P-choices was 68% and 62%, respectively, and the reversal of the phenomenon was observed, as we would expect. In contrast, in Experiment 2 (with small stakes), the P-bet was chosen only 43% of the time on average, and the preference reversal phenomenon was observed. Thus, if the \$-bias in WTP-valuations is negligible, the apparently inconsistent findings of Casey (1991) are explained by Proposition 1, which predicts a smaller reversal ratio for large stakes, where subjects exhibited significantly more risk aversion, compared to small stakes.

This begs the question of why the literature has not noticed that the preference reversal phenomenon crucially depends on using WTA-valuations instead of WTP ones, even though it has been previously suggested that WTP valuations might exhibit a smaller bias (Schmidt and Hey, 2004). We believe that the reason is that the literature has overlooked that the absence of a bias does *not* imply an equality between the rates of standard and nonstandard reversals. As shown by our results, what the absence of a bias actually suggests is that whether one observes more reversals of one or the other kind is mainly determined by the aggregate risk aversion captured in the experiment, which itself depends on the particular set of lotteries used, and possibly on other elements of the design. Our analysis suggests that the evidence arising from experiments using WTP valuations is internally consistent and reflect the insights derived from our work.

Cubitt et al. (2004) compare two evaluation methods, one based on WTA valuations (CMS04, WTA) and another using probability equivalents (CMS04, PE). In both experiments subjects exhibited a small degree of risk-seeking behavior ($\pi_c = 0.42$). WTA-valuations led to the preference reversal phenomenon, in agreement with our results suggesting that these valuations entail a strong bias. However, there was no difference between reversal rates for PE-valuations. This could be seen as surprising because, in light of the scale compatibility hypothesis, shifting the focus from outcomes to probabilities should induce a P-bias, as opposed to the standard \$-bias. In other words, the fact that valuations through probability equivalents did not "flip" the phenomenon casts doubts on the compatibility hypothesis. However, our results suggest that the results of Cubitt et al. (2004) could be interpreted exactly as originally intended. The key is that, again, the absence of a difference between reversal rates is *not* diagnostic. Figure 10 shows that experiment (CMS04, PE) is actually located below the diagonal, which in view of our results is compatible with an overpricing bias for P-bets.

Bateman et al. (2007) indirectly inferred monetary valuations using a task that asked subjects to rank P-bets and \$-bets separately but together with sure amounts (BDLS07, ranking-CE). They found a smaller asymmetry in reversal rates compared to a second experiment using WTA-valuations (BDLS07, WTA). Observed aggregate risk aversion

was similar across treatments ($\pi_c^{\text{rank}} = 0.75$ and $\pi_c^{\text{wta}} = 0.77$). Thus, our results suggest that the \$-bias in their ranking-based task was smaller than when WTA valuations were used instead. Going one step further, Alós-Ferrer et al. (2016) used a pure ranking task (AGKW16, RANK), which led to the reversal of the preference reversal phenomenon, whereas the preference reversal phenomenon was observed in another treatment that used WTA-valuations (AGKW16, WTA). Notably, (AGKW16, RANK) sits squarely on the diagonal, in full agreement with our conclusions and equation (5).

In summary, our framework accommodates previous findings from the literature and provides explanations for several observations which were considered inconsistent or hard to explain. This is made possible by taking into account the combination of differences in the bias in evaluations and experiment-specific differences in observed risk aversion.

6 Conclusions

Among all the empirical anomalies contradicting basic microeconomic principles, the preference reversal phenomenon is one of the most robust and worrying ones. With a history going back just over 50 years (Slovic and Lichtenstein, 1968), its impact on the study of economic decisions has been remarkable. Its implications are wide-ranging, since it uncovers a basic inconsistency between choices and preference elicitation methods based on monetary or other cardinal valuations, casting pervasive doubts on most measurement methods underlying welfare economics and economic policy design. Accordingly, it is important to uncover and quantify the sources of the phenomenon.

In this work, we provide the first unified account explaining when the phenomenon occurs, when it should be reversed, and what determines the extent of one or the other. The discrepancies between reversal rates reflecting inconsistencies between choices and evaluations arise due to the interaction of stochastic choice, risk aversion, and an overpricing bias which is essentially restricted to monetary valuations of long shot gambles. At the same time, we are the first to embed a measurement of preference reversals within a larger empirical framework which allows for an independent estimation of decision makers' preferences and, crucially, allows to unambiguously classify both choices and evaluations as correct or erroneous. This innovation allows us to conduct new empirical tests to validate the assumptions and confirm the predictions of the model, including novel hypotheses on the determinants of the magnitude of the phenomenon.

Our first empirical results serve to validate the assumptions of the model. Since we can compute the difference in certainty equivalents for each choice pair, we can confirm the presence of strength-of-preference effects in our experimental data, leading to higher error rates whenever this difference is small. When evaluations are monetary (experiment PRICE), we observe a discrete shift (comparing choices and evaluations) in the functional relations linking certainty equivalent differences and (imputed) choice frequencies, which reflects an overpricing of long shot gambles (\$-bets). This shift is absent when evaluations are carried out through an ordinal task (experiment RANK).

Our results also confirm the model's predictions. In the absence of a systematic difference (bias) between choices and evaluations, as is empirically the case with ranking-based evaluations, the model predicts the *reversal of the preference reversal phenomenon*, which was previously observed but considered puzzling. This prediction obtains without any need to invoke a behavioral bias. The fact that choice is stochastic and errors reflect strength of preference, coupled with risk aversion, is sufficient to explain larger (nonstandard) reversal rates when long shots are chosen than when they are not. Our experiment RANK documents this phenomenon. The model also predicts that the reversal of the preference reversal phenomenon should be stronger for more risk averse individuals, a prediction which we confirm by relying on the independently-estimated risk attitudes.

The model also shows that an upward bias in evaluation of $-bets, as empirically observed for monetary valuations, suffices to offset the effects of risk aversion and explain the preference reversal phenomenon, where larger (standard) reversal rates obtain when moderate lotteries are chosen over long shots than when long shots are chosen. Our experiment PRICE documents this standard phenomenon. A further, novel prediction is that the effect will be stronger the stronger the bias is. Again, our design allows us to test and confirm this new prediction by examining the bias at the individual level.

Crucially, our independently-estimated preferences allow us to compare errors across elicitation methods (choices and evaluations), while the previous literature was limited to pointing out inconsistencies among them. This innovation allows us to show that errors are significantly more frequent in monetary evaluations than in choices, but they are of similar magnitudes for ranking evaluations and choices. Further, we are able to pinpoint those errors and show that the observation that monetary valuations are more noisy than choices is confined to one type of lotteries, namely $-bets (long shots). That is, contrary to generalized beliefs, willingness-to-accept valuations are not particularly noisy in general as an elicitation method. Stated valuations of "regular" lotteries (P-bets) closely track those derived from estimated individual utility functions. There is a clear discrepancy for long shots, which is reflected both in increased noise and in a systematic upward bias where such lotteries are overpriced. We are further able to quantify the extent of the overpricing phenomenon, and we find it to be of a very-large economic magnitude. On average, the individual valuations of long shot lotteries were overestimated by 293% relative to their estimated certainty equivalents.

Our model and data provide a consistent, unified, systematic account of the preference reversal phenomenon which explains long-standing puzzles. We are able to explain the origins of the phenomenon and the determinants of its magnitude within the same framework that explains when the opposite effect dominates and what the determinants of the magnitude of the latter are. We thus paint a complete picture of the phenomenon which delivers several surprises. For instance, the reversal of the phenomenon does not result from a behavioral bias, but merely from risk aversion. As a consequence, the literature has *underestimated* the extent of the preference reversal phenomenon for half a century, by comparing it to an incorrect default (the equality of reversal rates).

Our results can be argued to paint an optimistic picture for applied economics. The preference reversal phenomenon is not a fundamental, blanket difficulty revealing unavoidable inconsistencies across preference revelation methods, but rather a specific, concrete anomaly arising because of a systematic bias affecting a particular subset of (maybe extreme) alternatives. Further, the reversal of the phenomenon is not an anomaly at all. This is, of course, not to say that the preference reversal phenomenon can be ignored, since alternatives of the long shot type are prone to occur in applied contexts. However, knowing that the phenomenon mostly originates in an upward bias in the valuations of long shot alternatives, and not in, say, a general bias caused by an evaluation method or the stochastic nature of choice in itself, helps narrow down the scope of its implications and potentially avoid them entirely for specific applications.

References

Alós-Ferrer, C., E. Fehr, and N. Netzer (2020). Time Will Tell: Recovering Preferences when Choices are Noisy. *Journal of Political Economy forthcoming*.

Alós-Ferrer, C. and M. Garagnani (2018). Strength of Preference and Decisions Under Risk. Working Paper, University of Zurich.

Alós-Ferrer, C., D.-G. Granić, J. Kern, and A. K. Wagner (2016). Preference Reversals: Time and Again. *Journal of Risk and Uncertainty 52*(1), 65–97.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2006). Elicitation Using Multiple Price List Formats. *Experimental Economics 9*(4), 383–405.

Apesteguía, J. and M. A. Ballester (2018). Monotone Stochastic Choice Models: The Case of Risk and Time Preferences. *Journal of Political Economy 126*(1), 74–106.

Atkinson, A. C. (1996). The Usefulness of Optimum Experimental Designs. *Journal of the Royal Statistical Society 51*(1), 59–76.

Attema, A. E. and W. B. Brouwer (2013). In Search of a Preferred Preference Elicitation Method: A Test of the Internal Consistency of Choice and Matching Tasks. *Journal of Economic Psychology 39*, 126–140.

Azrieli, Y., C. P. Chambers, and P. J. Healy (2018). Incentives in Experiments: A Theoretical Analysis. *Journal of Political Economy 126*(4), 1472–1503.

Ballinger, T. P. and N. T. Wilcox (1997). Decisions, Error and Heterogeneity. *Economic Journal 107*(443), 1090–1105.

Bateman, I., B. Day, G. Loomes, and R. Sugden (2007). Can Ranking Techniques Elicit Robust Values? *Journal of Risk and Uncertainty 34*(1), 49–66.

Bateman, I. J., R. T. Carson, B. Day, M. Hanemann, N. Hanley, T. Hett, M. J. Lee, G. Loomes, S. Mourato, E. Ozdemiroglu, D. W. Pearce, R. Sugden, and J. Swanson (2002). *Economic Valuation with Stated Preference Techniques: A Manual*. Cheltenham, United Kingdom: Edward Elgar.

Beauchamp, J. P., D. J. Benjamin, D. I. Laibson, and C. F. Chabris (2019). Measuring and Controlling for the Compromise Effect when Estimating Risk Preference Parameters. *Experimental Economics forthcoming*.

Bellemare, C., S. Kröger, and A. van Soest (2008). Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities. *Econometrica 76*(4), 815–839.

Bleichrodt, H. and J. L. Pinto Prades (2009). New Evidence of Preference Reversals in Health Utility Measurement. *Health Economics 18*(6), 713–726.

Bordalo, P., N. Gennaioli, and A. Shleifer (2012). Salience Theory of Choice under Risk. *Quarterly Journal of Economics 127*(3), 1243–1285.

Bordalo, P., N. Gennaioli, and A. Shleifer (2013). Salience and Consumer Choice. *Journal of Political Economy 121*(5), 803–843.

Bostic, R., R. J. Herrnstein, and R. D. Luce (1990). The Effect on the Preference-Reversal Phenomenon of Using Choice Indifferences. *Journal of Economic Behavior and Organization 13*, 193–212.

Bruner, D. M. (2017). Does Decision Error Decrease with Risk Aversion? *Experimental Economics 20*(1), 259–273.

Butler, D., A. Isoni, G. Loomes, and K. Tsutsui (2014). Beyond Choice: Investigating the Sensitivity and Validity of Measures of Strength of Preference. *Experimental Economics 17*(4), 537–563.

Butler, D. and G. Loomes (1988). Decision Difficulty and Imprecise Preferences. *Acta Psychologica 68*(1-3), 183–196.

Butler, D. J. and G. Loomes (2007). Imprecision as an Account of the Preference Reversal Phenomenon. *American Economic Review 97*(1), 277–297.

Butler, D. J. and G. Loomes (2011). Imprecision as an Account of Violations of Independence and Betweenness. *Journal of Economic Behavior and Organization 80*, 511–522.

Camerer, C. F. (1989). Does the Basketball Market Believe in the 'Hot Hand'. *American Economic Review 79*, 1257–1261.

Cappellari, L. and S. P. Jenkins (2003). Multivariate Probit Regression Using Simulated Maximum Likelihood. *The Stata Journal 3*(3), 278–294.

Casey, J. T. (1991). Reversal of the Preference Reversal Phenomenon. *Organizational Behavior and Human Decision Processes 48*(2), 224–251.

Collins, S. M. and D. James (2015). Response Mode and Stochastic Choice Together Explain Preference Reversals. *Quantitative Economics 6*(3), 825–856.

Conte, A., J. D. Hey, and P. G. Moffatt (2011). Mixture Models of Choice Under Risk. *Journal of Econometrics 162*(1), 79–88.

Cubitt, R. P., A. Munro, and C. Starmer (2004). Testing Explanations of Preference Reversal. *Economic Journal 114*(497), 709–726.

Cubitt, R. P., D. Navarro-Martinez, and C. Starmer (2015). On Preference Imprecision. *Journal of Risk and Uncertainty 50*(1), 1–34.

Dashiell, J. F. (1937). Affective Value-Distances as a Determinant of Aesthetic Judgment-Times. *American Journal of Psychology 50*, 57–67.

Davidson, D. and J. Marschak (1959). Experimental Tests of a Stochastic Decision Theory. In W. Churchman and P. Ratoosh (Eds.), *Measurement: Definitions and Theories*, Volume I, Part I, pp. 233–269. New York: Wiley.

Delquié, P. (1993). Inconsistent Trade-Offs Between Attributes: New Evidence in Preference Assessment Biases. *Management Science 39*(11), 1382–1395.

Dubourg, W. R., M. W. Jones-Lee, and G. Loomes (1994). Imprecise Preferences and the WTP-WTA Disparity. *Journal of Risk and Uncertainty 9*(2), 115–133.

Ford, I., B. Torsney, and C. J. Wu (1992). The Use of a Canonical Form in the Construction of Locally Optimal Designs for Non-Linear Problems. *Journal of the Royal Statistical Society 54*(2), 569–583.

Goldstein, W. M. and H. J. Einhorn (1987). Expression Theory and the Preference Reversal Phenomena. *Psychological Review 94*(2), 236–254.

Greiner, B. (2015). Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *Journal of the Economic Science Association 1*, 114–125.

Grether, D. M. and C. R. Plott (1979). Theory of Choice and the Preference Reversal Phenomenon. *American Economic Review 69*(4), 623–638.

Halton, J. H. (1960). On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals. *Numerische Mathematik 2*(1), 84–90.

Hershey, J. C. and P. J. H. Schoemaker (1985). Probability versus Certainty Equivalence Methods in Utility Measurement: Are They Equivalent? *Management Science 31*(10), 1213–1231.

Hey, J. D. and C. Orme (1994). Investigating Generalizations of Expected Utility Theory Using Experimental Data. *Econometrica 62*(6), 1291–1326.

Holt, C. A. (1986). Preference Reversals and the Independence Axiom. *American Economic Review 76*(3), 508–515.

Holt, C. A. and S. K. Laury (2002). Risk Aversion and Incentive Effects. *American Economic Review 92*(5), 1644–1655.

Johnson, E. J. and D. A. Schkade (1989). Bias in Utility Assessments: Further Evidence and Explanations. *Management Science 35*(4), 406–424.

Karni, E. and Z. Safra (1987). 'Preference Reversal' and the Observability of Preferences by Experimental Methods. *Econometrica 55*(3), 675–685.

Laming, D. (1985). Some Principles of Sensory Analysis. *Psychological Review 92*(4), 462–485.

Lichtenstein, S. and P. Slovic (1971). Reversals of Preference Between Bids and Choices in Gambling Decisions. *Journal of Experimental Psychology 89*(1), 46–55.

Lindman, H. R. (1971). Inconsistent Preferences Among Gambles. *Journal of Experimental Psychology 89*(2), 390–397.

Lipkus, I. M., G. Samsa, and B. K. Rimer (2001). General Performance on a Numeracy Scale Among Highly Educated Samples. *Medical Decision Making 21*(1), 37–44.

Loomes, G. (2005). Modelling the Stochastic Component of Behaviour in Experiments: Some Issues for the Interpretation of Data. *Experimental Economics 8*(4), 301–323.

Loomes, G. and G. Pogrebna (2017). Do Preference Reversals Disappear When We Allow for Probabilistic Choice? *Management Science 63*(1), 166–184.

Loomes, G. and R. Sugden (1995). Incorporating a Stochastic Element into Decision Theories. *European Economic Review 39*(3–4), 641–648.

Loomes, G. and R. Sugden (1998). Testing Different Stochastic Specifications of Risky Choice. *Economica 65*(260), 581–598.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.

Maafi, H. (2011). Preference Reversals Under Ambiguity. *Management Science 57*(11), 2054–2066.

McFadden, D. L. (2001). Economic Choices. *American Economic Review 91*(3), 351–378.

Moffatt, P. G. (2005). Stochastic Choice and the Allocation of Cognitive Effort. *Experimental Economics 8*(4), 369–388.

Moffatt, P. G. (2015). *Experimetrics: Econometrics for Experimental Economics*. London: Palgrave Macmillan.

Morrison, G. C. (1998). Understanding the Disparity between WTP and WTA: Endowment Effect, Substitutability, or Imprecise Preferences? *Economics Letters 59*(2), 189–194.

Mosteller, F. and P. Nogee (1951). An Experimental Measurement of Utility. *Journal of Political Economy 59*, 371–404.

Moyer, R. S. and T. K. Landauer (1967). Time Required for Judgements of Numerical Inequality. *Nature 215*(5109), 1519–1520.

Oliver, A. (2013). Testing Procedural Invariance in the Context of Health. *Health Economics 22*(3), 272–288.

Peirce, J. W. (2007). PsychoPy – Psychophysics Software in Python. *Journal of Neuroscience Methods 162*(1), 8–13.

Pommerehne, W. W., F. Schneider, and P. Zweifel (1982). Economic Theory of Choice and the Preference Reversal Phenomenon: A Reexamination. *American Economic Review 72*(3), 569–574.

Safra, Z., U. Segal, and A. Spivak (1990). Preference Reversal and Nonexpected Utility Behavior. *American Economic Review 80*(4), 922–930.

Schkade, D. A. and E. J. Johnson (1989). Cognitive processes in preference reversals. *Organizational Behavior and Human Decision Processes 44*(2), 203–231.

Schmidt, U. and J. D. Hey (2004). Are Preference Reversals Errors? An Experimental Investigation. *Journal of Risk and Uncertainty 29*(3), 207–218.

Seidl, C. (2002). Preference Reversal. *Journal of Economic Surveys 16*(5), 621–655.

Silvey, S. D. (1980). *Optimal Design: An Introduction to the Theory for Parameter Estimation*, Volume 1. New York: Chapman and Hall.

Slovic, P. and S. Lichtenstein (1968). Relative Importance of Probabilities and Payoffs in Risk Taking. *Journal of Experimental Psychology Monograph 78*(3, Part 2), 1–18.

Stalmeier, P. F. M., P. P. Wakker, and T. G. G. Bezembinder (1997). Preference Reversals: Violations of Unidimensional Procedure Invariance. *Journal of Experimental Psychology: Human Perception and Performance 23*(4), 1196–1205.

Thurstone, L. L. (1927). A Law of Comparative Judgement. *Psychological Review 34*, 273–286.

Train, K. E. (2003). *Discrete Choice Methods with Simulation*. New York: Cambridge University Press.

Trautmann, S. T., F. M. Vieider, and P. P. Wakker (2011). Preference Reversals for Ambiguity Aversion. *Management Science 57*(7), 1320–1333.

Tversky, A. (1969). Intransitivity of Preferences. *Psychological Review 76*, 31–48.

Tversky, A., S. Sattath, and P. Slovic (1988). Contingent Weighting in Judgment and Choice. *Psychological Review 95*(3), 371–384.

Tversky, A., P. Slovic, and D. Kahneman (1990). The Causes of Preference Reversal. *American Economic Review 80*(1), 204–217.

Tversky, A. and R. H. Thaler (1990). Anomalies: Preference Reversals. *Journal of Economic Perspectives 4*(2), 201–211.

Vieider, F. M. (2018). Violence and Risk Preference: Experimental Evidence from Afghanistan, Comment. *American Economic Review 108*(8), 2366–2382.

Wichmann, A. F. and N. J. Hill (2001). The Psychometric Function: I. Fitting, Sampling, and Goodness of Fit. *Attention, Perception, & Psychophysics 63*(8), 1293–1313.

Wilcox, N. T. (2008). Stochastic Models for Binary Discrete Choice Under Risk: A Critical Primer and Econometric Comparison. In J. C. Cox and G. W. Harrison (Eds.), *Risk Aversion in Experiments*, Volume 12 of *Research in Experimental Economics*, pp. 197–292. Bingley, UK: Emerald.

Wilcox, N. T. (2011). Stochastically More Risk Averse: A Contextual Theory of Stochastic Discrete Choice Under Risk. *Journal of Econometrics 162*(1), 89–104.

Appendix A Description of RUM Estimation

To estimate individual utility functions from the binary lottery choices in part one of the experiment we follow the approach described in Moffatt (2015, Chapter 13). All $T = 32$ trials used for the utility estimation involved binary choices between lotteries of the form $A = (p, x)$ and $B = (q, y)$, where A pays x with probability p and B pays y with probability q, and 0 otherwise. We index the trials in the experiment by $t = 1, ..., 32$, that is, at trial t subjects face the choice between $A_t = (p_t, x_t)$ and $B_t = (q_t, y_t)$. Further, we index the $N = 190$ subjects by $i = 1, ..., N$. In the main analysis we assume a normalized constant absolute risk aversion (CARA) function as in Conte et al. (2011), which is given by

$$u(x \mid r) = \begin{cases} \frac{1-e^{-rx}}{1-e^{-rx_{\max}}}, & \text{if } r \neq 0 \\ \frac{x}{x_{\max}}, & \text{if } r = 0, \end{cases} \tag{6}$$

where $x_{\max} = \max\{x_1, \ldots, x_T, y_1, \ldots, y_T\}$ is the maximum outcome across all T lottery pairs (trials). The normalization ensures that $u(x \mid r)$ is increasing also for negative values of r (indicating risk-lovingness). However, the results are qualitatively unchanged when we assume utility function with constant relative risk aversion (CRRA) instead (see Appendix B). Under the assumption of Expected Utility maximization, subject i with utility function $u(x \mid r_i)$ chooses A_t over B_t if the difference in expected utilities is positive, that is,

$$\nabla_t(r_i) := p_t u(x_t \mid r_i) - q_t u(y_t \mid r_i) = \frac{p_t(1 - e^{-r_i x_t}) - q_t(1 - e^{-r_i y_t})}{1 - e^{-r_i x_{\max}}} > 0. \tag{7}$$

In order to be able to estimate the parameters of the model, we now add noise to the model. There are two standard approaches in the literature: The Fechner or Random Utility Model (RUM) and the Random Preference Model (RPM). RUM assumes that each subject is characterized by a risk parameter r_i that is fixed across trials, whereas RPM assumes that a subject's risk parameter varies randomly between trials but is drawn from a certain distribution. Since our goal is to classify choices across multiple trials as errors or correct responses, the main analysis reported in the paper use a RUM-based estimation.[14]

Following the RUM approach, we add an error term $\varepsilon_{it} \sim N(0, \sigma^2)$ with $\sigma^2 > 0$ to (7). That is, the lottery A_t is chosen if the following condition holds

$$\nabla_t(r_i) + \varepsilon_{it} > 0 \tag{8}$$

[14] In Appendix C we carry out an analogue estimation using the RPM approach and report the corresponding results. We find that all results are qualitatively unchanged.

Define the binary choice indicator for trial t

$$\gamma_{it} = \begin{cases} 1 & \text{if } A_t \text{ chosen by subject } i \\ -1 & \text{if } B_t \text{ chosen by subject } i. \end{cases}$$

Then the probability of a choice conditional on the risk-parameter r_i is given by

$$P(\gamma_{it} \mid r_i) = P\left(\gamma_{it}\nabla_t(r_i) > \gamma_{it}(-\varepsilon_{it})\right) = P\left(\gamma_{it}\frac{\nabla_t(r_i)}{\sigma} > \gamma_{it}\frac{-\varepsilon_{it}}{\sigma}\right) = \Phi\left(\gamma_{it}\frac{\nabla_t(r_i)}{\sigma}\right) \tag{9}$$

where Φ is the standard normal cumulative distribution function.

The conditional probabilities above were derived conditional on a subject's risk parameter r_i. In other words, estimating this model over the entire population would imply homogeneity in risk attitude across subjects. In order to allow for between-subject heterogeneity, we let the risk attitudes vary across the population. In particular, we assume that the individual risk attitudes in the population are distributed normally in our subject pool according to

$$r \sim N(\mu, \eta^2).$$

Hence, the log-likelihood of a sample given by the matrix $\Gamma = (\gamma_{it})$ consisting of T trials and N subjects is

$$\log L = \sum_{i=1}^{N} \ln \int_{-\infty}^{\infty} \prod_{t=1}^{T} \Phi\left(\gamma_{it}\frac{\nabla_t(r)}{\sigma}\right) f(r \mid \mu, \eta) dr \tag{10}$$

where $f(r \mid \mu, \eta) = \frac{1}{\sqrt{2\rho\eta^2}} e^{-\frac{1}{2}\left(\frac{r-\mu}{\eta}\right)^2}$ is the density function of the risk parameter r.

In order to evaluate the integral in (10) we use the method of maximum simulated likelihood (MSL) (see Train, 2003, for details). Specifically, we will approximate this integral by the following average

$$\frac{1}{H}\sum_{h=1}^{H}\left(\prod_{t=1}^{T}\Phi\left(\gamma_{it}\frac{\nabla_t(r_{ih})}{\sigma}\right)\right) \tag{11}$$

using a sequence of H (transformed) Halton draws $(r_{i1}, \dots, r_{iH})$ from $N(\mu, \eta^2)$ for each subject i (fixed over trials t). For the estimation, we use the Stata implementation "mdraws" of this procedure (Cappellari and Jenkins, 2003). Halton draws, a by-now-standard procedure, simulate random draws that ensure even coverage of the parameter space (e.g. avoiding clustering) using Halton sequences (Halton, 1960; Moffatt, 2015). Specifically, a Halton sequence is defined for a given prime number p, for example $p = 2$, is $(\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \frac{1}{16}, \frac{9}{16}, \dots)$. Such a sequence $(h_1, h_2, \dots)$ provide pseudo-random draws from the uniform distribution $U(0, 1)$. To obtain draws from $N(\mu, \eta^2)$ we apply

the following transformation $r_{ij} = \mu + \eta\Phi^{-1}(h_j)$ where Φ^{-1} is the inverse of the normal cumulative distribution function.

The MSL approach amounts to replacing the integral in (10) by (11) and then maximize the resulting function

$$\log \hat{L} = \sum_{i=1}^{N} \ln \frac{1}{H} \sum_{h=1}^{H} \left(\prod_{t=1}^{T} \Phi \left(\gamma_{it} \frac{\nabla_t(r_{ih})}{\sigma} \right) \right). \tag{12}$$

Maximization of (12) is carried out using standard MLE routines in Stata to obtain the estimates $(\hat{\mu}, \hat{\eta}, \hat{\sigma})$. Given those estimates we obtain the posterior expectation of each subject's risk attitude $\hat{r}_i$ conditional on their T choices applying Bayes' rule as follows

$$\hat{r}_i = E(r_i|\gamma_{i1}, ..., \gamma_{iT}) \approx \frac{\frac{1}{H}\sum_{h=1}^{H} r_{ih} \left(\prod_{t=1}^{T} \Phi \left(\gamma_{it} \frac{\nabla_t(r_{ih})}{\hat{\sigma}} \right) \right)}{\frac{1}{H}\sum_{h=1}^{H} \left(\prod_{t=1}^{T} \Phi \left(\gamma_{it} \frac{\nabla_t(r_{ih})}{\hat{\sigma}} \right) \right)}$$

for a sequence of Halton draws $(r_{i1}, \ldots, r_{iH})$ from $N(\hat{\mu}, \hat{\eta}^2)$.

Given the estimated individual mean risk parameter $\hat{r}_i$, we obtain

$$\hat{u}_i(x) = \frac{1 - e^{-\hat{r}_i x}}{1 - e^{-\hat{r}_i x_{\max}}} \text{ for } \hat{r}_i \neq 0$$

as the estimated utility function of subject i.

Appendix B Robustness Analysis: CRRA

As a further robustness check we repeat the RUM-based estimation exercise described in Appendix A using the following constant relative risk aversion (CRRA) utility function

$$u(x \mid r) = \begin{cases} \frac{x^{1-r}}{1-r}, & \text{if } r \neq 1 \\ \ln(x), & \text{if } r = 1 \end{cases}$$

instead of the CARA function used there. We then reproduce the results reported in Section 3.1 and 3.3. As we detail below, this robustness check confirms that the results reported in the main text do not hinge on the CARA specification of the utility function, but remain robust when the above CRRA specification is used instead.

The average of the estimated individual risk propensities using this alternative utility function is 0.514 (median 0.556, SD 0.256). 8 subjects (4.21%) are classified as risk loving, some participants have a risk parameter close to zero indicating risk neutrality, but the majority of subjects is moderately risk averse.

Figure B.1 (left panel) displays the proportions of choice and evaluation errors, classified according to the individually estimated CRRA utility functions. Choice errors are substantially less frequent (20.18%) than evaluation errors (56.00%; WSR test, $N = 95$, $z = -7.936$, $p < 0.001$). Figure B.1 (right panel) shows the reversal rates when choice

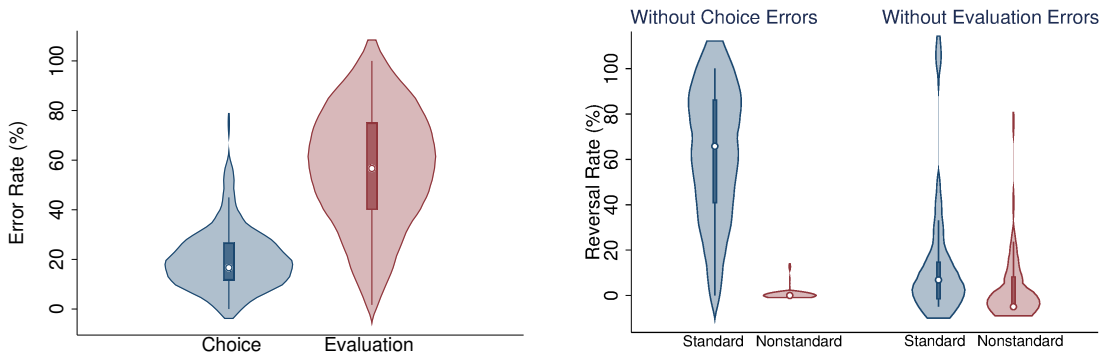


Figure B.1: Errors and reversal rates excluding errors in PRICE with CRRA utilities.

Notes: Left: Average of individual error rates, both for choices and evaluations. Right: Average of individual reversal rates (standard and nonstandard) when choice or evaluation errors are excluded.

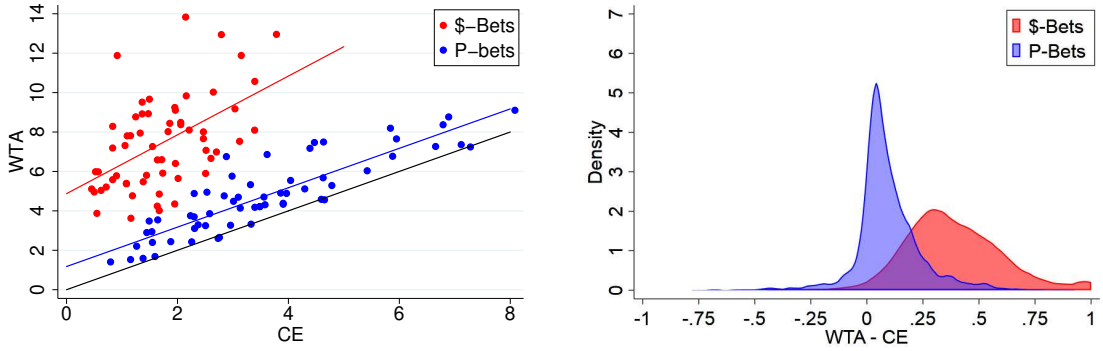


Figure B.2: Accuracy of valuations in PRICE with CRRA utilities.

Notes: Left: Correlation between stated WTA and predicted certainty equivalent for P-bets and $-bets. Each point corresponds to one lottery representing the average WTA and the average CE across all subjects in the PRICE experiment. Right: Distribution of overpricing measure for P-bets and $-bets.

errors or evaluation errors are excluded, respectively. We observe that, also in this case, excluding choice errors has little effect on reversal rates (average standard reversal rate 62.44%, nonstandard reversal rate 0.78%), and hence the asymmetry between both types of reversals persists (WSR test, $N = 46$, $z = 5.901$, $p < 0.001$). This is in stark contrast with the results when evaluation errors are excluded. In this case, the rate of standard reversals drops drastically from 62.44% to 17.12% (WSR test, $N = 36$, $z = 4.989$, $p < 0.001$). Most importantly, the asymmetry between the two types of reversals is greatly reduced (standard: 17.12%, nonstandard: 13.08%), however, the difference remains marginally significant (WSR test, $N = 82$, $z = 1.772$, $p = 0.076$).

Figure B.2 (left panel) plots stated WTA valuations against the estimated certainty equivalent for each of the 120 lotteries, distinguishing P-bets and $-bets. For P-bets, monetary valuations and estimated CEs show a strong and highly significant correlation (Spearman's $\rho = 0.866$, $N = 60$, $p < 0.001$). However, the same cannot be said for

$-bets, for which the picture is much more dispersed and far away from the diagonal (Spearman's $\rho = 0.501$, $N = 60$, $p < 0.001$).

Figure B.2 (right panel) displays the distribution of the difference between stated WTA valuation and CE, $WTA_i(A) - CE_i(A)$, separately for P-bets and $-bets. The two distributions are clearly different, which is confirmed by an equality of distributions test (Kolmogorov-Smirnov test, $D = 0.659$, $p < 0.001$). For P-bets, the distribution is concentrated around zero (mean 0.094, median 0.070, SD 0.153). On the contrary, for $-bets the distribution is clearly more dispersed and shifted to the right, showing a systematic bias to overstate valuations (mean 0.389, median 0.363, SD 0.220). The average *monetary* increase in the valuation (overpricing) of the $-bets is 2.749 (median 2.527, SD 1.116), relative to the certainty equivalent. That is, on average, $-bets are overpriced by a whopping 275%. In contrast, for P-bets the average bias of the stated valuations is much smaller, amounting just to 0.460 relative to the certainty equivalent of the P-bet (median 0.346, SD 0.367).

Appendix C Robustness Analysis: RPM Estimation

Recently, random utility models have been criticized (see Wilcox, 2008, 2011; Bruner, 2017; Apesteguía and Ballester, 2018; Vieider, 2018) and some of these author's have suggested Random Preference Models (RPM) as an alternative (e.g., Loomes and Sugden, 1995), since it is immune to some of these critiques. The RPM approach is based on the idea that a subject's risk parameter is not fixed but varies randomly. The drawback is that this also renders it arguably less appropriate in our context since our goal is to obtain a (fixed) measure of individual risk attitudes. Nevertheless, we think that it is useful as a robustness analysis to ensure that the obtained results do not hinge on the specifics of the estimation procedure. Consequently, in this section we estimate a RPM as a further robustness check and then reproduce the analogous analysis reported in Section 3.1 and 3.3 using the results of this estimation.

Appendix C.1 Description of RPM procedure

For the RPM estimation, we use the same setup with $N = 190$ subjects, $T = 32$ trials, and the CARA utility function given by (6). Additionally we will assume that A_t is the safer of the two lotteries, that is, $p > q$. In contrast to the RUM approach, the RPM assumes that a subject's risk parameter is not fixed across trials but varies randomly between trials. Specifically, we assume that subject i's risk parameter in trial t is distributed according to $r_{it} \sim N(m_i, \sigma^2)$ where m_i is subject i's mean risk attitude. Assuming Expected Utility maximization, in this setup subject i with utility function u_i chooses A_t over B_t if and only if

$$\Delta_t(r_{it}) = \frac{p_t(1 - e^{-r_{it}x_t}) - q_t(1 - e^{-r_{it}y_t})}{1 - e^{-r_{it}x_{\max}}} > 0.$$

Let r_t^* be the risk parameter that would make a subject exactly indifferent between the two lotteries in task t, that is, $\Delta_t(r_t^*) = 0$. Since A_t is always the safer lottery, we obtain the following equivalence

$$\Delta_t(r_{it}) > 0 \quad \Leftrightarrow \quad r_{it} > r_t^*.$$

Again using $\gamma_{it} \in \{1, -1\}$ as a binary indicator that A_t is chosen by subject i in trial t, the probability of a choice conditional on a subject's mean risk attitude m_i is given by

$$P(\gamma_{it}|m_i) = P(\gamma_{it} r_{it} > \gamma_{it} r_t^* | m_i) = P\left(\gamma_{it}\frac{r_{it} - m_i}{\sigma} > \gamma_{it}\frac{r_t^* - m_i}{\sigma}\right) = \Phi\left(\gamma_{it}\frac{m_i - r_t^*}{\sigma}\right)$$

where Φ is the standard normal cumulative distribution function. Next, in order to introduce between-subject heterogeneity we let the individual mean risk attitude vary across the population. In particular, we assume that

$$m_i \sim N(\mu, \eta^2).$$

Hence, the log-likelihood for a sample consisting of T trials and N subjects given by the matrix $\Gamma = (\gamma_{it})$ is

$$\log L = \sum_{i=1}^{N} \ln \int_{-\infty}^{\infty} \prod_{t=1}^{T} \Phi\left(\gamma_{it}\frac{m - r_t^*}{\sigma}\right) f(m \mid \mu, \eta) dm \tag{13}$$

where $f(m \mid \mu, \eta)$ is the density function of the mean risk attitude m. Using the MSL approach we replace the integral in (13) by the following approximation

$$\frac{1}{H}\sum_{h=1}^{H}\left(\prod_{t=1}^{T}\Phi\left(\gamma_{it}\frac{m_{ih} - r_t^*}{\sigma}\right)\right) \tag{14}$$

using a sequence of H (transformed) Halton draws $(m_{i1}, \ldots, m_{iH})$ from $N(\mu, \eta^2)$ for each subject i (fixed over trials t). We then maximize the resulting function

$$\log \hat{L} = \sum_{i=1}^{N} \ln \frac{1}{H}\sum_{h=1}^{H}\left(\prod_{t=1}^{T}\Phi\left(\gamma_{it}\frac{m_{ih} - r_t^*}{\sigma}\right)\right). \tag{15}$$

using standard MLE routines in STATA to obtain the parameter estimates $(\hat{\mu}, \hat{\eta}, \hat{\sigma})$. Given those estimates we can then compute the posterior expectation of a subject's mean risk attitude $\hat{m}_i$ conditional on the observed T choices using Bayes' rule as follows

$$\hat{m}_i = E(m_i|\gamma_{i1}, ..., \gamma_{iT}) \approx \frac{\frac{1}{H}\sum_{h=1}^{H} m_{ih}\left(\prod_{t=1}^{T}\Phi\left(\gamma_{it}\frac{m_{ih} - r_t^*}{\sigma}\right)\right)}{\frac{1}{H}\sum_{h=1}^{H}\left(\prod_{t=1}^{T}\Phi\left(\gamma_{it}\frac{m_{ih} - r_t^*}{\sigma}\right)\right)}$$

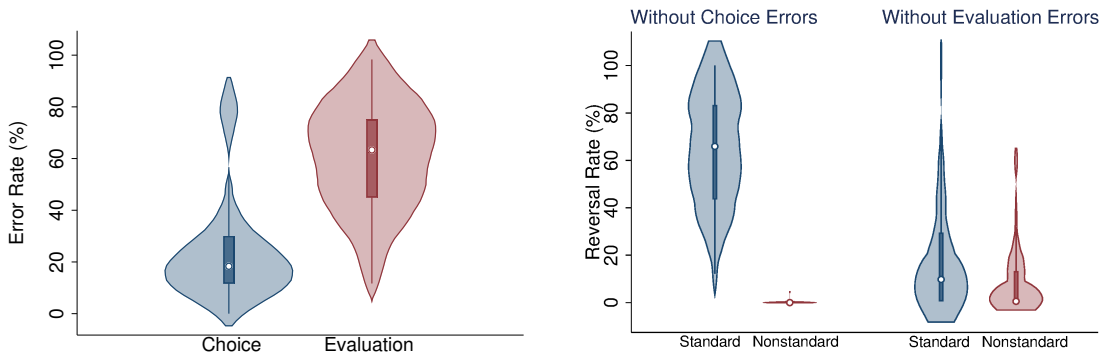


Figure C.1: Errors and reversals excluding errors in PRICE with RPM utilities.
Notes: Left: Average of individual error rates, both for choices and evaluations. Right: Average of individual reversal rates (standard and nonstandard) when choice or evaluation errors are excluded.

for a sequence of Halton draws $(m_{i1}, \ldots, m_{iH})$ from $N(\hat{\mu}, \hat{\eta}^2)$.

Given the estimated individual mean risk parameter $\hat{m}_i$, we obtain

$$\hat{u}_i(x) = \frac{1 - e^{-\hat{m}_i x}}{1 - e^{-\hat{m}_i x_{\max}}} \text{ for } \hat{m}_i \neq 0$$

as the estimated utility function of subject i.

Appendix C.2 Results based on RPM estimation

The average estimated risk propensity using this alternative procedure is $\hat{r} = 0.162$ (median 0.170, SD 0.122). 13.16% of subjects are classified as risk loving, some participants have a risk parameter close to zero indicating risk neutrality, but the majority of subjects is moderately risk averse.

Figure C.1 (left panel) displays the proportions of choice and evaluation errors, classified according to the individually estimated RPM utility functions. Choice errors are substantially less frequent (25.25%) than evaluation errors (60.21%; WSR test, $N = 95$, $z = -8.155$, $p < 0.001$). This confirms the results obtained using RUM-based estimations, that preferences elicited through monetary valuations are inherently more noisy than choices.

Figure C.1 (right panel) shows the reversal rates when choice errors or evaluation errors are excluded, respectively. Again, excluding choice errors has little effect on reversal rates (average standard reversal rate 64.07%, nonstandard reversal rate 0.01%), and hence the asymmetry between both types of reversals persists (WSR test, $N = 43$, $z = 5.712$, $p < 0.001$). This is in stark contrast with the results when evaluation errors are excluded. In this case, the rate of standard reversals drops drastically to 18.71% (WSR test, $N = 37$, $z = 5.303$, $p < 0.001$). Most importantly, the asymmetry between the two types of reversals is greatly reduced and indeed the difference is only marginally significant (nonstandard reversal rate: 8.80%; WSR test, $N = 45$, $z = 1.650$, $p = 0.099$).

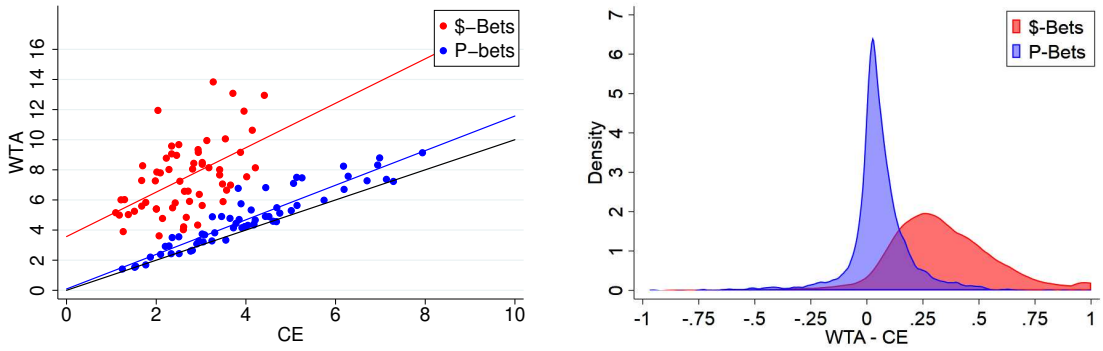


Figure C.2: Accuracy of valuations in PRICE with RPM utilities.

Notes: Left: Correlation between stated WTA and predicted certainty equivalent for P-bets and $-bets. Each point corresponds to one lottery representing the average WTA and the average CE across all subjects in the PRICE experiment. Right: Distribution of overpricing measure for P-bets and $-bets.

Figure C.2 (left panel) plots stated WTA valuations against the estimated certainty equivalents for each of the 120 lotteries, distinguishing P-bets and $-bets. For P-bets, stated monetary valuations and the estimated CEs show a strong and highly significant correlation (Spearman's $\rho = 0.927$, $N = 60$, $p < 0.001$). However, the same cannot be said for $-bets, for which the picture is much more dispersed and far away from the diagonal (Spearman's $\rho = 0.502$, $N = 60$, $p = 0.001$).

Figure C.2 (right panel) displays the distribution of the difference between the stated WTA valuation and the CE, $WTA_i(A) - CE_i(A)$, separately for P-bets and $-bets. The two distributions are clearly different, which is confirmed by an equality of distributions test (Kolmogorov-Smirnov test, $D = 0.650$, $p < 0.001$). For P-bets, the distribution is concentrated around zero (mean 0.054, median 0.043, SD 0.159). On the contrary, for $-bets the distribution is clearly more dispersed and shifted to the right, showing a systematic bias to overstate valuations (mean 0.327, median 0.310, SD 0.264). The average *monetary* increase in the valuation (overpricing) of the $-bets is 3.172 (median 2.830, SD 2.114), relative to the certainty equivalent. That is, on average, $-bets are overpriced significantly by a whopping 317%. In contrast, for P-bets the average bias of the stated valuations is much smaller, amounting just to 0.193 relative to the certainty equivalent of the P-bet (median 0.118, SD 0.208).

Appendix D Overpricing as Anchoring

In this section, we comment on the likely sources of overpricing, that is, the $-bias in elicited WTA valuations. We have established that overpricing of $-bets (and not a general bias in elicitation methods) is a major driver behind the preference reversal phenomenon. The compatibility hypothesis (Tversky et al., 1988, 1990) as well as salience theory (Bordalo et al., 2012) point toward some form of anchoring as the poten-

Table D.1: Random effects panel regressions on anchoring to the highest outcome.

Anchoring	Model 1	Model 2
P-Bet		-0.231^{***}
		(0.016)
Constant	0.255^{***}	0.370^{***}
	(0.049)	(0.050)
N	11,400	11,400

Notes: Robust standard errors in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

tial source of overpricing. Intuitively, the high monetary outcome of the \$-bet is more salient compared to the relatively low monetary upside of the P-bet.

Again, thanks to our particular design, we can provide direct evidence for this hypothesis. Suppose that the stated willingness to accept for a lottery $A = (p, x)$ by a subject i, $WTA_i(A)$, deviates from the true certainty equivalent $CE_i(A)$ due to anchoring on the largest monetary outcome, x. That is,

$$WTA_i(A) = (1 - \mu_i) \cdot CE_i(A) + \mu_i \cdot x$$

where $\mu_i \in [0, 1]$ is the individual anchoring propensity, and anchoring occurs if $\mu_i > 0$. Assuming μ_i to be lottery-independent for simplicity, we obtain:

$$\frac{WTA_i(A) - CE_i(A)}{x - CE_i(A)} = \mu_i. \tag{16}$$

This latter equation can be estimated in our dataset (experiment PRICE). The willingness to accept $WTA_i(A)$ is directly elicited in the evaluation phase, the largest outcome x is known for each lottery, and, unlike in previous experiments, we can compute the certainty equivalent $CE_i(A)$ using the utilities obtained in the estimation phase. Hence, the left-hand fraction in equation (16) can be directly computed from our data. We assume overpricing to be a noisy phenomenon and, due to the salience of \$-bets, we allow for a difference in anchoring between the latter and P-bets. Let δ_P be a dummy taking the value one if A is a P-bet, and let μ be the population average anchoring propensity and write $\mu_i = \mu + \eta_i$. We obtain the simple model

$$\frac{WTA_i(A) - CE_i(A)}{x - CE_i(A)} = \mu + \mu_P \delta_P + \eta_i + \epsilon.$$

which corresponds, under standard assumptions on the error terms, to a simple random-effects panel estimation.

Table D.1 reports the corresponding regressions. Model 1 excludes the P-bet dummy and tests just for the overall anchoring effect. The constant (μ) is significantly larger than zero, which is consistent with anchoring. Model 2 corresponds to the equation described above. The constant is again significantly different from zero, confirming the

presence of anchoring, and the P-bet dummy has a significantly negative coefficient (μ_P), revealing that anchoring is of a smaller magnitude for P-bets, as expected. However, a post hoc linear combination test shows that anchoring for P-bets, although smaller, is also significantly different from zero ($\mu + \mu_P = 0.138$, $p = 0.005$). This is consistent with the small overpricing effect that we previously found for P-bets, and the slight shift observed in Figure 2.

One comment is in order. Model 2 above essentially states that, for \$-bets, average anchoring is around 37%, in the sense that the estimated (average) anchoring propensity, as defined by equation (16), is $\mu = 0.370$. This is a rather large magnitude. In Section 3.3, we computed the average overpricing factor, as given by $[WTA_i(A) - CE_i(A)]\,/CE_i(A)$, to be 2.929. Both observations are fully compatible (and mechanically related to each other). Putting equation (16) together with the average overpricing estimate implies $(x - CE_i(A))/CE_i(A) = 2.929/0.370 \backsimeq 7.92$, or $x \backsimeq 8.92 CE_i(A)$. Direct examination of the data reveals that, on average in our dataset, the largest outcome of a \$-bet is around 8.96 times its certainty equivalent, as derived from the estimated utilities.

In conclusion, our data is in line with the idea that overpricing is related to anchoring on the highest outcome of a lottery, which itself is a natural derivation from the compatibility hypothesis of Tversky et al. (1990). This also explains why the effect is much more accused for \$-bets, since the latter exhibit a much larger outcome than P-bets. Salience theory provides another explanation why subjects tend to anchor their monetary valuations on the upside of a lottery. It also implies that, when a lottery is judged within a pair (as in the choice phase) and not in isolation, the salience advantage of the \$-bets upside should be reduced. Indeed, Bordalo et al. (2012) find that when monetary valuations where elicited within a choice context, the order of elicited average valuations is consistent with choices on the aggregate level, but preference reversals on the individual level still occur.

Appendix E List of Lottery Pairs

Table E.1 shows the 32 lottery pairs used for the utility estimation (part one). Table E.2 shows the 60 lottery pairs used in the preference reversal experiment (part two).

Table E.1: Lottery pairs (A, B) used for the utility estimation in part one.

Lottery	Lottery A			Lottery B		
Pair	p	x	EV	q	y	EV
1	0.05	12	0.6	0.8	3	2.4
2	0.2	22	4.4	0.8	5	4
3	0.25	17	4.25	0.75	6	4.5
4	0.35	20	7	0.6	8	4.8
5	0.35	17	5.95	0.7	4	2.8
6	0.4	12	4.8	0.7	6	4.2
7	0.4	14	5.6	0.65	6	3.9
8	0.4	14	5.6	0.8	3	2.4
9	0.5	11	5.5	0.7	7	4.9
10	0.5	15	7.5	0.65	7	4.55
11	0.5	20	10	0.7	5	3.5
12	0.55	5	2.75	0.35	18	6.3
13	0.55	4	2.2	0.2	15	3
14	0.55	4	2.2	0.4	15	6
15	0.55	4	2.2	0.45	21	9.45
16	0.6	6	3.6	0.35	11	3.85
17	0.6	5	3	0.3	22	6.6
18	0.6	8	4.8	0.5	13	6.5
19	0.6	14	8.4	0.7	4	2.8
20	0.6	4	2.4	0.55	6	3.3
21	0.6	3	1.8	0.5	13	6.5
22	0.65	3	1.95	0.15	18	2.7
23	0.65	17	11.05	0.75	7	5.25
24	0.7	4	2.8	0.1	16	1.6
25	0.7	7	4.9	0.6	20	12
26	0.7	11	7.7	0.8	6	4.8
27	0.7	18	12.6	0.85	5	4.25
28	0.75	6	4.5	0.3	15	4.5
29	0.75	6	4.5	0.4	15	6
30	0.75	4	3	0.35	12	4.2
31	0.75	15	11.25	0.8	5	4
32	0.8	3	2.4	0.4	17	6.8

Table E.2: Lottery pairs $(\mathrm{P}, \$)$ used in the preference reversal experiment.

Lottery	P-Bets			$-Bets		
Pair	p	x	EV	q	y	EV
1	0.95	3	2.85	0.37	10	3.70
2	0.57	5	2.85	0.46	10	4.60
3	0.90	6	5.40	0.30	11	3.30
4	0.80	6	4.80	0.30	11	3.30
5	0.72	7	5.04	0.23	11	2.53
6	0.79	2	1.58	0.21	11	2.31
7	0.8	2	1.60	0.4	11	4.40
8	0.64	8	5.12	0.24	12	2.88
9	0.84	6	5.04	0.48	12	5.76
10	0.75	3	2.25	0.17	12	2.04
11	0.94	3	2.82	0.49	12	5.88
12	0.92	4	3.68	0.53	12	6.36
13	0.82	3	2.46	0.34	12	4.08
14	0.74	6	4.44	0.15	13	1.95
15	0.89	5	4.45	0.39	13	5.07
16	0.87	6	5.22	0.36	13	4.68
17	0.9	2	1.80	0.35	13	4.55
18	0.66	2	1.32	0.24	13	3.12
19	0.6	5	3.00	0.45	13	5.85
20	0.9	7	6.30	0.51	14	7.14
21	0.86	5	4.30	0.16	15	2.40
22	0.70	10	7.00	0.31	15	4.65
23	0.85	5	4.25	0.41	15	6.15
24	0.63	7	4.41	0.41	15	6.15
25	0.75	6	4.50	0.15	15	2.25
26	0.76	11	8.36	0.37	16	5.92
27	0.63	4	2.52	0.33	16	5.28
28	0.96	5	4.80	0.19	17	3.23
29	0.96	8	7.68	0.43	17	7.31
30	0.84	9	7.56	0.25	18	4.50
31	0.83	6	4.98	0.31	18	5.58
32	0.95	5	4.75	0.22	18	3.96
33	0.86	5	4.30	0.33	18	5.94
34	0.79	4	3.16	0.33	18	5.94
35	0.60	11	6.60	0.22	19	4.18
36	0.56	10	5.60	0.43	19	8.17
37	0.79	7	5.53	0.20	20	4.00
38	0.7	5	3.50	0.17	20	3.40
39	0.85	10	8.50	0.3	20	6.00
40	0.65	4	2.60	0.25	20	5.00
41	0.92	8	7.36	0.23	21	4.83
42	0.88	11	9.68	0.35	21	7.35
43	0.72	6	4.32	0.29	21	6.09
44	0.68	3	2.04	0.23	21	4.83
45	0.73	9	6.57	0.21	22	4.62
46	0.6	7	4.20	0.3	22	6.60
47	0.68	11	7.48	0.23	23	5.29
48	0.88	8	7.04	0.4	24	9.60
49	0.84	7	5.88	0.35	25	8.75
50	0.95	8	7.60	0.31	27	8.37
51	0.82	11	9.02	0.24	31	7.44
52	0.87	5	4.35	0.13	32	4.16
53	0.86	4	3.44	0.55	6	3.30
54	0.8	4	3.20	0.45	6	2.70
55	0.87	3	2.61	0.5	7	3.50
56	0.75	5	3.75	0.55	7	3.85
57	0.82	5	4.10	0.47	8	3.76
58	0.71	5	3.55	0.22	9	1.98
59	0.89	5	4.45	0.55	9	4.95
60	0.82	4	3.28	0.36	9	3.24

Appendix F Translated Instructions

[*These are the written instructions given to subjects before the experiment. The original instructions were in German. Text in brackets [...] was not displayed to subjects.]*

General Instructions

Welcome! In this experiment you will be asked to make a series of decisions that will determine your earnings at the end of the experiment. The total duration of the experiment is about 1 hour and 30 minutes.

If you have a question, please raise your hand and remain seated. An experimenter will come and answer your question.

It is important, that you read the instructions carefully before you make your decisions.

During the experiment you are not allowed to talk or communicate in any other way with the other participants. If you violate this rule, you might be excluded from the experiment.

We now explain the general course of the experiment: The experiment consists of three parts. In each part you have to make multiple decisions. At the end of the experiment you will be asked to answer a questionnaire.

In each part, you can earn money. How much money you earn will depend on your decisions in that part and chance. Your earnings in one part of the experiment are independent of your earnings and decisions in the other parts. Your earnings in each part will be added up and you will be paid the total amount anonymously and in cash at the end of the experiment. In addition to this amount you will receive €4 for your participation in the experiment.

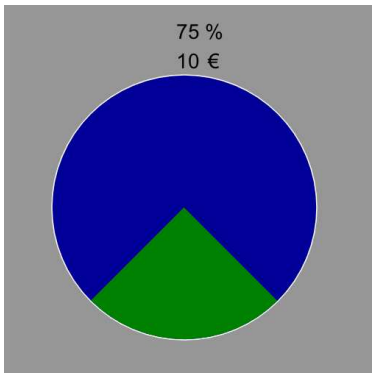
Below you will find further general information for the experiment. The specific instructions for each part will be shown on screen directly before the beginning of that part.

Instructions: Lotteries

In the three parts of the experiment you will be asked to make decisions about lotteries. Hence, we will now explain in detail what a lottery is:

A lottery consists of two potential outcomes, each of which will occur with a given probability. One of the two outcomes is always €0 (zero). The other outcome will differ from lottery to lottery. If a lottery is played out, this means that you will receive exactly one of the two possible outcomes (in Euro).

In the experiment lotteries will be represented by pie charts as in the example below. The colored areas of the pie chart illustrate the probabilities for the two corresponding outcomes.



Example:

The pie chart depicted above is an example of how we present a lottery. In this example, the lottery pays €10 with a probability of 75%, which is represented by the blue area. Additionally, this information is also shown numerically above the pie chart. Accordingly, the lottery pays €0 with a probability of 25%, which is represented by the green area. The second outcome is always €0 and occurs with the remaining probability. Please note that this information is not repeated numerically on screen.

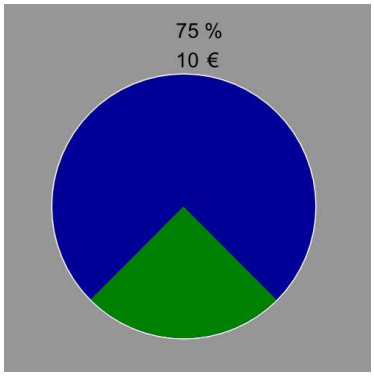
If a lottery is played out, this means that it will pay exactly one of the two outcomes. In the example above, the lottery pays €10 with a probability of 75% and €0 with the remaining probability of 25%.

Please note that the lottery shown above is only an example. The lotteries in the experiment will have different outcomes and probabilities.

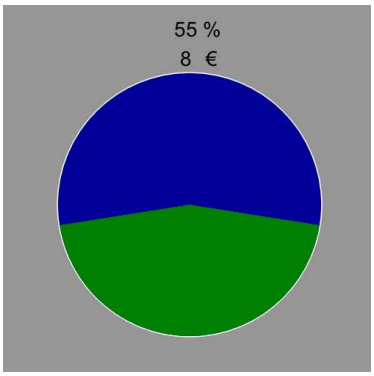
If you have a question, please raise your hand. If you have no further questions, you may proceed to the comprehension questions on the next page.

Comprehension questions: Lotteries

Below you see examples of two lotteries, similar to the ones you will face later on in the experiment. Please note that these lotteries are only examples.



Lottery A



Lottery B

Please answer the following comprehension questions:

1. What is the probability that Lottery A pays €10?

2. What is the probability that Lottery B pays €0?

3. Which amount does Lottery A pay with a probability of 25%?

4. Which amount does Lottery B pay with a probability of 55%?

Once you have answered all comprehension questions, please raise your hand. An experimenter will then check your answers.

Translated onscreen instructions

[These are the instructions for each part, which were presented separately on screen, at the beginning of each part. The original instructions were in German. Text in brackets [...] was not displayed to subjects.]

Welcome to this economic experiment. Thank you for supporting our research.
Please note the following rules:

1. During the experiment you are not allowed to communicate with each other.

2. If you have questions, please raise your hand.

3. Please refrain from using any features of the computer that are not part of the experiment.

Instructions for part 1

Your decisions: In this part of the experiment you will be presented with a series of lottery pairs. Your task is to choose one of the two lotteries from each pair.

On the screen you will see a lottery pair (consisting of two lotteries) represented by two pie charts. One of the lotteries will be shown on the left and the other will be shown on the right. You choose one of the lotteries by pressing the left or right arrow key on your keyboard. These keys are marked with a yellow sticker. To choose the lottery on the left, press the left arrow key "←." To choose the lottery on the right, press the right arrow key "→." Please note that your decisions will affect your earnings at the end of the experiment (a detailed description of how your earnings are determined will follow below).

There are no wrong or correct decisions. When you choose one of the lotteries, this simply shows that you prefer to play this lottery over the other lottery.

After you have made your decision, you will see the next lottery pair. In part 1 you will be presented with a total of 36 lottery pairs. After you have made a decision for each of the pairs, this part ends and we will start with the next part of the experiment.

Your earnings for part 1

After you have made a decision for each of the lottery pairs, the computer will randomly select one of the 36 lottery pairs. The computer then checks which of the two lotteries you have chosen for this randomly selected pair. The lottery you have chosen will be played out. The outcome of the lottery determines your earnings for part 1 of the experiment.

The lottery will be played out at the end of the experiment, that is, after you have completed all three parts of the experiment. Please note: Although your earnings for this part will be determined at the end of the experiment, they will only depend on your decisions in this part of the experiment and chance.

If you have any further questions, please raise your hand and remain seated.

Instructions for part 2

[Instructions for experiment PRICE]

Your decisions: In this part of the experiment you will be presented with a series of lotteries. When a lottery is presented to you on screen, you may simply assume, that you own that lottery and are asked to sell it.

Your task is to state the lowest price at which you are still willing to sell the presented lottery instead of keeping the lottery and playing it out.

There is no wrong or correct answer when stating the lowest price at which you are still willing to sell the lottery. When you enter your selling price for the lottery, simply ask yourself "Is this really the lowest price at which I am still willing to sell the lottery instead of playing the lottery?". Please note that your decisions will affect your earnings at the end of the experiment (a detailed description of how your earnings are determined will follow below).

Please enter the lowest price at which you are still willing to sell the lottery in the form "EURO.CENTS." Please note that you cannot enter a selling price that is larger than the highest outcome of the lottery.

After you have entered your selling price, the next lottery will be presented. In this part of the experiment you will see a total of 120 lotteries, presented in 20 rounds of 6 lotteries each. All rounds are independent. Once you have entered a selling price for each lottery in a round, the next round will start. Once you are done with all 20 rounds, you can continue with the next part of the experiment.

Your earnings for part 2

[Instructions for experiment PRICE]

After you have entered your lowest selling price for each of the lotteries, the computer will randomly draw one of the 20 rounds. From this round, the computer will then randomly select two of the six lotteries. The computer then checks for which of the two

lotteries you have entered the higher selling price (in case both prices are the same, the computer will randomly select one of the two lotteries with equal probability). This lottery will be played out and the outcome of that lottery determines your earnings for part 2 of the experiment.

The lottery will be played out at the end of the experiment, that is, after you have completed all three parts of the experiment. Please note: Although your earnings for this part will be determined at the end of the experiment, they will only depend on your decisions in this part of the experiment and chance.

If you have any further questions, please raise your hand and remain seated.

Instructions for part 2

[Instructions for experiment RANK]

Your decisions: In this part of the experiment you will be presented with a series of lotteries. When a lottery is presented to you on screen, you may simply assume, that you own that lottery and may play that lottery.

Your task is to order different lotteries according to your preference, that is, according to how much you would like to play them.

In each round you will see six different lotteries on screen. Please order the lotteries as follows:

- First, choose your first-ranked lottery, that is, the one of the six lotteries that you would like to play the most.
- Second, choose your second-ranked lottery, that is the one that you would like to play out the second most.
- Third, choose your third-ranked lottery, that is the one that you would like to play out the third most.
- Fourth, choose your fourth-ranked lottery, that is the one that you would like to play out the fourth most.
- Fifth, choose your fifth-ranked lottery, that is the one that you would like to play out the fifth most.
- Sixth, choose your sixth-ranked lottery, that is the one that you would like to play out the least.

To select a lottery simply click on the button below the lottery that you want to select. As soon as you assign a rank to a lottery, the corresponding rank (from 1 to 6) will be shown below that lottery.

In case you want to change the rank of the lotteries, please press the "Reset" button. This resets the ranking. After you have ranked the lotteries from rank 1 to rank 6, please press the "Continue" button to confirm your ranking and proceed to the next round.

Please note that there is no wrong or correct ranking. When ranking the lotteries, simply ask yourself which lottery you would like to play out the most, the second-most and so on. Please note that your decisions will affect your earnings at the end of the experiment (a detailed description of how your earnings are determined will follow below).

In this part of the experiment you will see a total of 120 lotteries, presented in 20 rounds of 6 lotteries each. All rounds are independent, that is, you will have to submit 20 rankings of 6 lotteries by assigning ranks from 1 to 6. Once you are done with all 20 rounds, you can continue with the next part of the experiment.

Your earnings for part 2

[Instructions for experiment RANK]

After you have ranked all lotteries, the computer will randomly draw one of the 20 rounds. From this round, the computer will then randomly select two of the six lotteries. The computer then checks which of the two lotteries you have ranked higher (that is, which one you want to play out more). This lottery will be played out and the outcome of that lottery determines your earnings for part 2 of the experiment.

The lottery will be played out at the end of the experiment, that is, after you have completed all three parts of the experiment. Please note: Although your earnings for this part will be determined at the end of the experiment, they will only depend on your decisions in this part of the experiment and chance.

If you have any further questions, please raise your hand and remain seated.

Instructions for part 3

Your decisions: In this part of the experiment you will be presented with a series of lottery pairs. Similar to part 1, your task is to choose one of the two lotteries from each pair. Please note that the lottery pairs are different from part 1.

On the screen you will see a lottery pair (consisting of two lotteries) represented by two pie charts. One of the lotteries will be shown on the left and the other will be shown on the right. You can choose one of the lotteries pressing the left or right arrow key on your keyboard. These keys are marked with a yellow sticker. To choose the lottery on the left, press the left arrow key "←." To choose the lottery on the right, press the right arrow key "→." Please note that your decisions will affect your earnings at the end of the experiment (a detailed description of how your earnings are determined will follow below).

There are no wrong or correct decisions. When you choose one of the lotteries, this simply shows that you prefer to play this lottery over the other lottery.

After you have made your decision, you will see the next lottery pair. In part 3 you will be presented with a total of 60 lottery pairs. After you have made a decision for each of the pairs, this part ends and you can start the questionnaire.

Your earnings for part 3

After you have made a decision for each of the lottery pairs, the computer will randomly select one of the 60 lottery pairs. The computer then checks which of the two lotteries you have chosen for this randomly selected pair. The lottery you have chosen will be played out. The outcome of the lottery determines your earnings for part 3 of the experiment.

The lottery will be played out at the end of the experiment, that is, after you have completed all three parts of the experiment. Please note: Although your earnings for this part will be determined at the end of the experiment, they will only depend on your decisions in this part of the experiment and chance.

If you have any further questions, please raise your hand and remain seated.