Decisions Under Risk are Decisions Under Complexity *

Ryan Oprea†

August 3, 2024

Abstract

We provide evidence that classic lottery anomalies like probability weighting and loss aversion are not special phenomena of risk. They also arise (and often with equal strength) when subjects evaluate deterministic, positive monetary payments that have been disaggregated to resemble lotteries. Thus, we find, e.g., apparent probability weighting in settings without probabilities and loss aversion in settings without scope for loss. Across subjects, anomalies in these deterministic tasks strongly predicts the same anomalies in lotteries. These findings suggest that much of the behavior motivating our most important behavioral theories of risk derive from complexity-driven mistakes rather than true risk preferences.

Keywords: Complexity, fourfold pattern, probability weighting, loss aversion, prospect theory, bounded rationality, economics experiments

JEL codes: C91, D91 G0,

*This paper was previously circulated under the title "Simplicity Equivalents." I thank David Freeman for invaluable advice and assistance in designing the experiment and interpreting the results. I also thank Ned Augenblick, Andrew Caplin, Stefano DellaVigna, Benjamin Enke, Erik Eyster, Cary Frydman, Paul Glimcher, Thomas Graeber, Alex Imas, Chad Kendall, Kirby Nielsen, Matthew Rabin, Alex Rees-Jones, Dmitry Taubinsky, Ferdinand Vieider, Emanuel Vespa and Peter Wakker for valuable conversations. Dario Ochoa and Ravi Vora provided excellent research assistance. I am also grateful to audiences at Harvard University, the University of Chicago, UC Berkeley, NYU, Cornell, Carnegie Mellon, UC Davis, Bocconi, the University of Zurich, Queens University, USC, Columbia University, Wharton, Brown University, Northwestern University, University College London, Maison des sciences economiques, NYU Abu Dhabi, the Cognitive Noise Workshop (Harvard), BRIQ, CESifo, the Stanford-Ohio State Experimental Workshop, D-TEA Paris, SAET, HKBU-NTU-Osaka-Kyoto-Sinica Theory Seminar and the Society for Neuroeconomics. All mistakes are my own. This research was supported by the National Science Foundation under Grant SES-1949366 and was approved by UC Santa Barbara IRB.

†Economics Department, University of California, Santa Barbara, Santa Barbara, CA, 93106, roprea@gmail.com.

1 Introduction

In this paper we show that some of the most important lottery anomalies from the behavioral risk literature are not special phenomena of risk. They arise too (and often with equal strength) in the valuation of objects that are descriptively similar to lotteries but that are perfectly deterministic. We argue that this suggests that many important anomalies occur because lotteries are *complex* (costly or difficult to properly evaluate) rather than because they are risky. To the degree this is true, many anomalies that are commonly interpreted as expressions of risk preferences should instead be interpreted as systematic mistakes that are only indirectly related to risk.[1]

In each task in our experiment, we elicit subjects' dollar valuations for a set of 100 "boxes," each of which contains some dollar amount. For example, in one of our tasks (called G90) we ask subjects to value a set consisting of 90 boxes that each contain $25 and 10 boxes that each contain $0. Acquiring a set of boxes influences the subject's earnings in the experiment according to a *payoff rule*, and we compare how subjects value these sets under two contrasting payoff rules.

By opening one of the boxes from the set at random and paying the subject the amount inside, we turn the set into a *lottery* (i.e., G90 becomes a risky prospect of earning $25 with probability 0.9), and the dollar value the subject attaches to it becomes a *certainty equivalent*: the certain dollar amount the subject judges to be equivalently valuable to the risky lottery. Eliciting subjects' certainty equivalents (and related measures) for a standard set of such lotteries, we replicate the *fourfold pattern of risk* and *loss aversion*, two of the key anomalies in the lottery valuation literature. The fourfold pattern of risk is a tendency for risk postures to apparently change with the magnitude of probababilities and the gain/loss framing of lottery payments (typically attributed to what the literature calls "probability weighting"); loss aversion is a tendency for losses to receive more apparent weight than gains in valuations of mixed lotteries (typically attributed to "loss averse" preferences that are more strongly influenced by losses than gains). These anomalous behaviors have served as primary empirical motivation for many of our most important behavioral theories of risk preferences, including prospect theory (Kahneman & Tversky 1979, Tversky & Kahneman 1992). Because of their influence, we refer to these anomalies collectively as the *classical pattern.*

Our contribution is to compare these valuations to the valuations of what we call "deterministic mirrors" of the same lotteries. A deterministic mirror of a lottery consists of the same set of 100 boxes used to describe the lottery, but is characterized by a different payoff rule: instead of paying the dollar amount in one of the 100 boxes selected at random as a lottery does, a mirror pays the sum of the rewards in *all* of the boxes, weighted by the total number of boxes. Thus instead of paying

[1] Throughout the paper we will use the word "preferences" to refer to a decision maker's welfare relevant rank ordering of lotteries. This is a narrower way of using the term than some treatments in which "preference" refers simply to the decision maker's observed choice (i.e., the revealed preference of an agent). We will use the word "mistakes" to refer to failures to make choices that are consistent with the normatively correct ordering described by preferences. However, in this usage we do not preclude the possibility that mistakes are optimal – e.g., that choices maximize overall welfare, given constraints such as information processing costs.

$25 with probability 0.9 (as a lottery does), the mirror of G90 pays 0.9×$25=$22.50 with certainty. The dollar values subjects attach to these objects are no longer certainty equivalents, since there is no uncertainty in deterministic mirrors. Instead such valuations are *simplicity equivalents* – the simply-described dollar amounts subjects judge to be equivalently valuable to the relatively more complexly-described mirror.

Mirrors are descriptively identical to lotteries and valuing them requires similar information processing, but in contrast to lotteries mirrors contain no risk and therefore their valuations provide no scope for the rational expression of risk or loss preferences. In contrast to a lottery, valuing a mirror at anything other than its deterministic value (i.e., its corresponding lottery's expected value) is unambiguously a dominated, money-losing mistake. Nonetheless, we find that:

1. The fourfold pattern arises in the valuations of deterministic mirrors just as it does in lotteries, and with roughly the same strength. Importantly this means that we find strong evidence of what is usually called "probability weighting" in settings without probabilities.

2. Loss aversion arises in deterministic mirrors even though at the relevant margins they cannot actually produce losses. Thus we find strong evidence of what is usually called "loss aversion" in settings without risk of loss.

3. Across subjects, the severity of each of these anomalies in lotteries is strongly predicted by their severity in deterministic mirrors, suggesting that the behaviors in the two settings are strongly linked, deriving from a common behavioral mechanism (which, clearly, cannot be grounded in risk or risk preferences).

Additional treatments show that these results are robust to variation in the method of elicitation, the arithmetic difficulty of the valuation task, subject pool and stakes (Oprea 2024).

We interpret these results as evidence that lottery anomalies like probability weighting and loss aversion are not primarily rational expressions of non-standard risk preferences as is often believed (e.g., in some interpretations of prospect theory), but are instead in large part systematic mistakes induced by the complexity (the costs and difficulties) of properly valuing lotteries and similarly disaggregated objects.[2] We argue that this suggests that such anomalies tell us little about *tastes* for risk or loss and therefore should not be accommodated in welfare analysis or policy design. On the other hand we argue that the appearance of these anomalies as complexity responses in

[2]In using the term "complexity," we follow the standard definition from computer science: the cost of implementing the algorithm or procedure required to properly solve a problem. When we say a lottery is "complex," we mean only that its value is not transparent to the decision maker because the procedure required to optimally aggregate its disaggregated components into a value is costly or difficult. If these costs and difficulties are sufficiently severe – if a lottery is sufficiently complex – the decision maker may be induced to use a less optimal procedure instead (Simon 1955), producing mis-valuations. See Oprea (2020) for direct evidence that procedures produce complexity costs of this sort, Banovetz & Oprea (2022) for evidence that these costs drive people to use simpler-than-optimal choice procedures and Camara (2023) for a theoretical analysis of some consequences for decision-making.

deterministic domains suggests that the patterns of heuristic errors they describe likely distort choice in a far broader range of contexts than has so far been appreciated.

Our paper is organized as follows. In Section 2 we describe our experimental design and in Section 3 we report our results. In Section 4 we offer our interpretation of our findings and in Section 5 we discuss how our work connects to the prior literature. In Section 6 we conclude.

2 Experimental Design

In our experiment, we elicit valuations for a set of 12 lotteries known to produce some of the key valuation anomalies in the behavioral literature on risky choice (discussed in Section 2.1 below). We describe these lotteries in frequentist terms as a set of 100 boxes, each containing some amount of money, one of which will be opened at random to determine the subject's payment. For example, a lottery we call G75 consists of 75 boxes containing $25 and 25 boxes containing $0, meaning this is a lottery that pays out $25 with 75% chance.

Our contribution is to also ask the same subjects to value what we call *deterministic mirrors* of each of these 12 lotteries. Deterministic mirrors are described exactly as lotteries are, but with a change to the payoff rule that removes risk from the lottery. Instead of paying the contents of one box opened at random as in a lottery, in a mirror payoffs are determined by opening *all 100 boxes*, summing their values and weighting the total by the total number of boxes. In other words, the mirror pays the expected value of its corresponding lottery with certainty.

We designed the experiment to hold constant the information processing required to value a lottery and a mirror, varying only the presence of risk. To do this, we describe these two kinds of objects in identical, frequentist terms using virtually identical presentations, instructions and computer displays. We also ran the experiment using a within-subjects design: subjects valued a set of 12 lotteries (the Lottery treatment) and also a set of 12 mirrors (the Mirror treatment) of those same lotteries. Importantly, we randomize the order in which subject were assigned the 12 mirrors vs. the 12 lotteries. To further minimize scope for contagion across treatments, we were careful not to tell subjects first assigned lotteries that they would later be asked to value mirrors or vice versa. We also were careful to use examples and comprehension questions to make it very clear that lotteries and mirrors have very different payoff rules. Instructions are reproduced in Supplemental Appendix B.

In our main design, we elicit valuations of lotteries and mirrors using multiple price lists (MPLs), the most commonly used method in the literature. An MPL is simply a series of closely related binary choice problems (between options A and B) "stacked" on top of one another in a table. In MPLs designed to elicit lottery/mirror valuations, Option A in each row of the table simply repeats the lottery/mirror we are eliciting the value for, while Option B is a degenerate lottery (i.e. all 100 boxes contain the same dollar amount) whose value increases from row-to-row in $1 steps. The subject clicks either Option A or B on each row of the MPL to express her preference; if some row of the MPL is randomly selected for payment (see Section 2.2 below), her choice in that row is implemented to determine her earnings. By examining on which row the subject "switches" from preferring A to B we get an interval estimate of the subject's dollar valuation for the lottery or mirror. Following standard practice, throughout our data analysis we use the mid-point between rows of the MPL to measure values. Our software imposes a "single switching" rule that allows the subject to switch from Option A to Option B only once in the MPL. In Supplemental Appendix A.3 we report a robustness treatment in which we elicit values using the Becker-Degroot-Marschak (BDM) method (Becker et al. 1964) instead. Screenshots are provided in Supplemental Appendix A.8.2.

While the literature interprets the resulting valuations of lotteries as *certainty equivalents* – the certain dollar payments subjects value equivalently to risky lotteries – the same interpretation cannot be applied to mirrors which contain no uncertainty. Instead values for mirrors are *simplicity equivalents*: the simply-described payment amount subjects value equivalently to the more complexly described (but no less certain) mirror. Our question throughout the paper is whether simplicity equivalents have the same properties and suffer the same anomalies as certainty equivalents.

2.1 Lotteries

Eight of the twelve lotteries we ask subjects to value we call "fourfold lotteries," because they are designed to replicate what Tversky and Kahenman (1992) call the "fourfold pattern of risk" – a pattern that summarizes much of what we've learned about the certainty equivalents of lotteries in the last half century of empirically studying them. The pattern is typically measured using the certainty equivalents subjects assign to simple, two-state lotteries of the form $L(p; \$X)$ (i.e. lotteries that pay \$X with probability p and \$0 otherwise). Following the literature, we measure the pattern using valuations of two sets of such lotteries. Our "gains lotteries" pay $X = \$25$ with probabilities $p \in \{0.1, 0.25, 0.75, 0.9\}$ (yielding lotteries we call G10, G25, G75 and G90), and $Y = \$0$ otherwise. Our "loss lotteries" pay $X = -\$25$ with probabilities $p \in \{0.1, 0.25, 0.75, 0.9\}$ (yielding lotteries we call L10, L25, L75 and L90), and $Y = \$0$ otherwise. (In addition to the fourfold lotteries, we also include lotteries G50 and L50 which pay \$25 and -\$25 respectively with probability 0.5.)

Stated in terms of these lotteries, the fourfold pattern is a tendency for the certainty equivalent to (i) be lower than the lottery's expected value (revealing apparent risk aversion), for low probability prospects of losses (e.g., lotteries L10, L25) and high probability prospects of gains (e.g. lotteries G75, G90) and to (ii) be greater than the lottery's expected value (revealing apparent love for risk) for high probability prospects of losses (e.g., L75, L90) and low probability prospects of gains (e.g., G10 and G25). The fourfold pattern is therefore a series of apparent reversals in risk posture that occur when probabilities switch from low to high, and when lottery outcomes switch from losses to gains.

The final two tasks we assign to subjects we call "loss aversion" tasks because they are designed to measure the empirical regularity of loss aversion. Loss aversion is an apparent tendency to place excess weight on negative payments relative to positive payments when evaluating lotteries that mix gains and losses. Alternatively, loss aversion can be described as risk aversion towards mixed lotteries (lotteries that contain both strictly positive and negative payoffs). In our main design we measure loss aversion by eliciting *lottery equivalents* of a certain payoff (of \$0), instead of certainty equivalents of a lottery. Specifically, we ask subjects to choose between a sure payment of \$0 (repeated in each row of the MPL) and a menu of lotteries $(0.5; \$X, \$Y)$ in which $Y < 0$ is fixed at some negative value in each row and $\$X > 0$ increases from row-to-row of the MPL in \$2 steps. By looking for the value of \$X at which the subject switches from preferring the sure payment of \$0 to preferring the lottery, we get an implicit measure of the excess linear weight λ the subject places on negative relative to positive payoff events when valuing mixed lotteries. Lottery equivalents of this kind are popular for measuring loss aversion because they are believed to fix the reference point against which gains and losses are assessed at 0, allowing for cleaner measurement (Hershey & Schoemaker 1985, Sprenger 2015). We include tasks with \$Y equal to -10 and -15, giving us tasks $A10$ and $A15$, respectively.[3]

We present each subject with a version of each of these 12 tasks using lottery incentives (the Lottery treatment) and a version using deterministic mirror incentives (the Mirror treatment). In addition, in our main design we repeat two randomly selected lottery tasks and repeat the same two mirror tasks, allowing us to study the relationship between anomalies and inconsistencies across repeated choices (see Supplemental Appendix A.6). Thus, in total, subjects completed 14 tasks in each of the Lottery and Mirror treatments. The order of these two treatments (all Lottery tasks followed by all Mirror tasks or all Mirror tasks followed by all Lottery tasks) is randomized at the subject level.

2.2 Implementation

A total of 673 subjects participated in the experiment. We collected data from the main MPL treatment ($N = 184$) in April of 2023 on Prolific using custom Javascript programmed by the author and deployed via Qualtrics. Subjects were paid a \$6 base payment and, with 20% chance, were additionally paid the outcome from a randomly selected MPL and MPL row. They spent a median of 27.5 minutes in the experiment and earned an average of \$13.11 per hour. At the end of the experiment we included a short battery of three cognitive reflection tasks (Frederick 2005), a short demographic survey (focused on the subject's technical education) and a number of questions about subjects' strategies and beliefs during the experiment. Details are provided in Supplemental Appendix A.5 and the results are discussed in Section 3.3. In addition, we collected data from robustness treatments in which we (i) elicited values using the BDM mechanism instead of MPLs

[3] Because of possible losses subjects were given endowments of of \$5 for the gains lotteries, \$30 for Loss Lotteries \$15 for A10 and \$20 for A15

($N = 100$), (ii) attempted to lower the arithmetic difficulty of valuation by changing the numbers used in the design ($N = 90$) and (iii) used a university subject pool with more intensive training and higher incentives ($N = 113$).[4] Details on these robustness treatment are provided in Section 3.3 and Supplemental Appendix A.

2.3 Interpreting the Classical Pattern

The tasks we include in our design are characteristic of those typically used to measure the fourfold pattern of risk and loss aversion, two of the cardinal empirical patterns in the experimental literature on decision making under risk. Throughout the paper we will refer to these two patterns of anomalies jointly as the *classical pattern* and deviations from expected value that are characteristic of this pattern *pattern-consistent* deviations.

The fourfold pattern has been an important motivation for theories of decision making under risk in the behavioral economics literature and in psychology. Most importantly, it has typically been interpreted as evidence of *probability weighting*: a putative tendency for decision makers to value low probability prospects as if they are more likely and high probability prospects as if they are less likely than they really are. The empirical regularity of loss aversion has similarly been influential in inspiring behavioral theories of preferences: the apparent excess weight on losses it describes are widely interpreted as evidence that risk preferences are *reference dependent* and that losses (relative to a status-quo reference point of zero) have a greater impact on utility than gains.

Because these features of risk preferences (probability weighting, reference dependence, loss aversion) are not easily accommodated by standard expected utility theory (EUT), the classical pattern has served as the empirical foundation for a number of alternative theories of risk preferences. By far the most influential of these is *prospect theory* (Kahneman & Tversky 1979, Tversky & Kahneman 1992, Wakker 2010), which assembles both tendencies into a unified model of preferences. In particular, prospect theory describes the classical pattern as growing out of the joint influence of a probability weighting function that shapes utility as a function of probabilities (giving rise to probability weighting) and a reference dependent value function that computes value based on distance from the status quo rather than final wealth (giving rise to reference dependence), and is steeper in losses than gains (giving rise to loss aversion). The classical pattern is a key motivation for and is central to the empirical study of prospect theory: the fourfold pattern describes prospect theory's distinctive predictions concerning the evaluation of lotteries involving gains or losses (Tversky and Kahneman (1992) call it "the most distinctive implication of prospect theory"), and loss aversion its distinctive predictions concerning the evaluation of lotteries that mix gains and losses.

[4] The remaining 186 subjects in our dataset participated in an earlier run of our main MPL design. A referee noticed a typo in one of the examples used in the instructions, leading us to re-run the experiment. Data from that earlier run of the main design is reported in Supplemental Appendix A.6: the results are nearly identical to those from our main design.

To the degree the classical pattern is indeed driven by risk preferences (i.e. tastes for risk that cause valuations to deviate from expected value), it should disappear when we remove risk from lotteries in our Mirror treatment. Because mirrors pay expected value with certainty, they effectively *induce* risk neutral EUT preferences in subjects, making any valuations that depart from expected value dominated mistakes under *any* rational theory of subjects' own native preferences. Thus, to the degree this distinctive pattern continues to arise in the absence of risk, we have evidence for an alternative interpretation of the classical pattern: that it is a pattern of systematic *mistakes*, arising not because lotteries are risky, per se, but rather because they are complex (costly or difficult to properly value).[5]

3 Main Results

In Figure 1, we plot raw data from the experiment. On the x-axis we plot the probability of the lottery's non-zero payment and for the fourfold lotteries (G10, G25, G75, G90, L10, L25, L75 and L90) we plot the deviation of the certainty/simplicity equivalent from expected value on the y-axis.[6] Positive values of this difference are conventionally interpreted as revelations of risk-loving preferences and negative values as revelations of risk-averse preferences. We plot means for each lottery/mirror and include error bars that span two standard errors in each direction.

In gray, we plot data from the Lottery treatment and find strong evidence of the characteristic reversals of the fourfold pattern. Subjects appear risk averse for high probability gains (G75, G90) but risk loving for low probability gains (G10, G25). These postures appear to flip in each case for losses: subjects instead appear risk loving for high probability losses (L75, L90) and risk averse for low probability losses (L10, L25). For every fourfold lottery we can reject the hypothesis of equality of valuations with expected value at the five percent level using Wilcoxon tests.

Our first main finding is that valuations are virtually identical in the Mirror treatment, plotted as hollow dots: subjects display each of the four components of the fourfold pattern of risk, even though mirrors contain no risk at all. In each case these deviations from expected value continue to be significant at the five percent level using Wilcoxon tests. In the left two panels of Figure 2, we re-plot this same data under the lens of probability weighting, following standard conventions from the literature (Tversky & Kahneman 1992). In these panels we plot (i) the probability

[5]Our method for identifying the role of complexity-derived mistakes in lottery anomalies is therefore to remove risk from lotteries and examine whether the anomalies persist. An alternative approach might be to attempt to remove complexity from lotteries to see if anomalies *disappear*. This alternative approach seems more difficult to implement and especially to verify: while we can directly remove risks from lotteries, it is more difficult to tell ex ante how to make valuing a lottery easy to evaluate or to verify that we were successful at doing so ex post. Nonetheless, convincing evidence gathered using such an alternative approach would be a powerful complement to the approach we take here.

[6]For legibility, we omit data from the G50 and L50 lotteries from these plots, since these are not involved in either the fourfold pattern or loss aversion. We plot data from these lotteries instead in Figure 2 and examine them in more detail in Supplemental Appendix A.7.

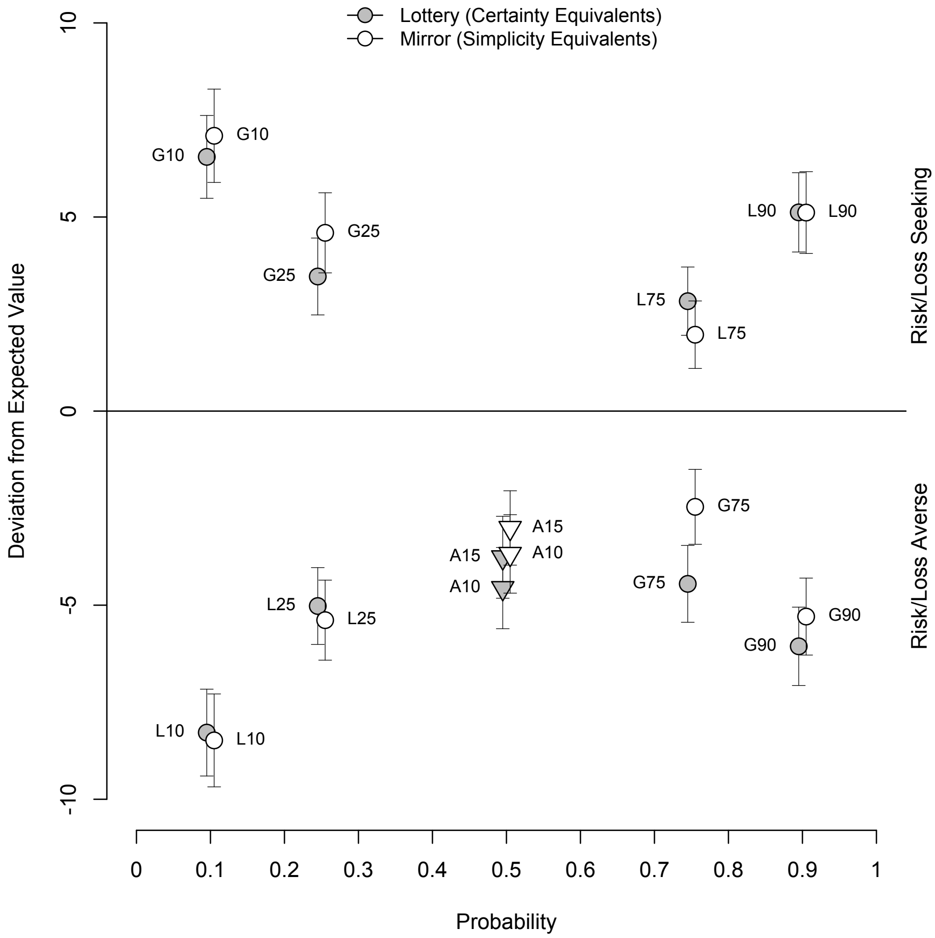


Figure 1: Mean deviations from expected value in lotteries (solid gray dots) and mirrors (hollow dots). *Notes: For fourfold lotteries, the y-axis measures the difference between subjects' certainty/simplicity equivalent and expected value, as stated in the axis label. The x-axis is the probability of the non-zero payoff. For loss aversion tasks, the y-axis measures instead the difference between the certain/simple payoff and the expected value of the mixed lottery/mirror. Two-standard-error bars are included for every task.*

of the non-zero payment on the x-axis and (ii) the ratio of the certainty/simplicity equivalent to the non-zero payment of the lottery on the y-axis. This provides a naive visual estimate of the "probability weighting function" for both gains (left panel) and losses (middle panel). In lotteries (plotted in gray) we find conventional evidence of probability weighting, with subjects acting as if they overweight low probability prospects and underweight high probability prospects relative to expected value (shown as a dashed 45-degree line). Our main finding is that we observe virtually identical probability weighting in mirrors (plotted as hollow dots), even though there are no probabilities in these tasks.

Result 1 *The fourfold pattern appears in deterministic mirrors, just as it does in lotteries. We*

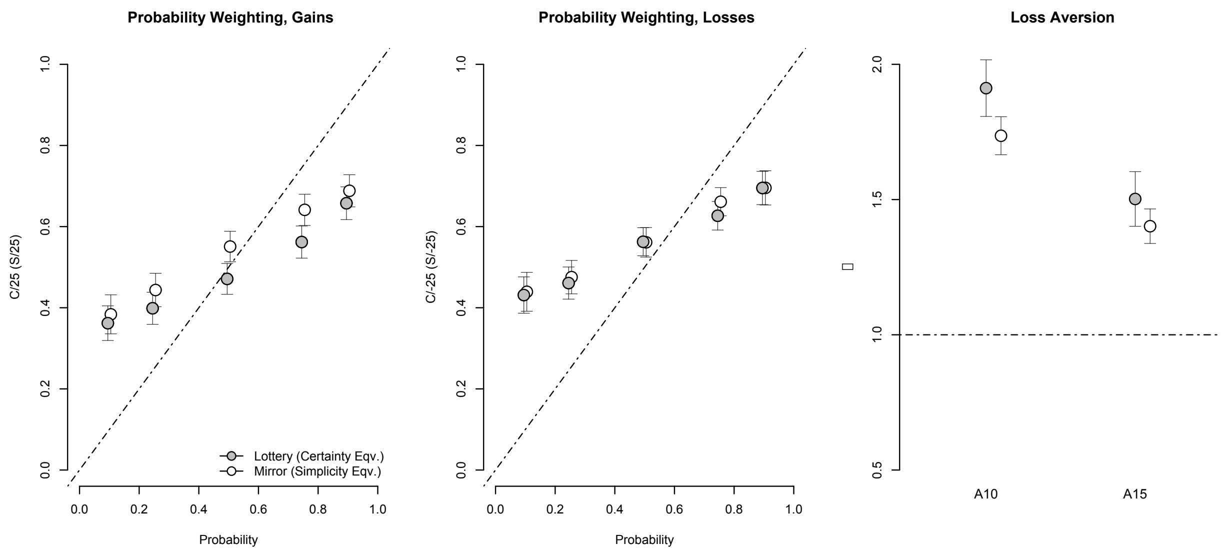


Figure 2: Naive visualization of the probability weighting functions (left two panels) and the loss aversion parameter, λ. *Notes: The first two panels plot a naive estimate of the probability weighting function (following Tversky & Kahneman (1992)) by plotting the ratio of the certainty/simplicity equivalent to the non-zero payment amount as a function of the probability of the non-zero payoff amount. The final panel plots a naive estimate of λ, the standard linear parameter of loss aversion, under the assumption of a reference point of zero.*

find strong evidence of "probability weighting" in settings without probabilities.

We also plot data from our loss aversion tasks (A10, A15) in Figure 1. For these tasks, we plot on the y-axis the difference between the certain payment of \$0 and the expected value of the lottery subjects judge to be equivalently valuable to 0; a negative value of this statistic is conventionally interpreted as evidence of loss aversion since it is evidence of subjects rejecting positive expected value lotteries that contain possible losses. To complement this raw analysis, in the right-most panel of Figure 2, we plot estimates of λ, the estimated linear weight placed on losses (relative to gains). A $\lambda > 1$ is conventionally interpreted as evidence of loss aversion.

For lotteries (plotted in gray), we find standard evidence of loss aversion. As Figure 1 shows, subjects deviate significantly from the loss-neutral benchmark in a loss-averse direction ($p < 0.01$ in both cases by Wilcoxon tests). Figure 2 plots estimates of λ ranging from 1.5 to nearly 2 suggesting that subjects place significant additional weight on losses relative to gains in their valuations. As with the fourfold pattern, our main finding is that behavior is very similar in mirrors. Valuations in Figure 1 significantly deviate from the optimal "loss neutral" benchmark in the same "loss averse" direction as lotteries ($p < 0.01$ in all cases by Wilcoxon tests). Loss aversion can also be described as risk aversion towards mixed lotteries. For this it is useful to contrast the expected earnings subjects are willing to sacrifice to avoid risk in *unmixed* 50/50 lotteries and mirrors (our L50 and G50 tasks, in which risk but not loss preferences can influence choice) to the expected earnings subjects are willing to sacrifice to avoid risk in our mixed 50/50 lotteries (tasks A10 and A15). The average subject is willing to sacrifice \$3.02 more in expected earnings to avoid risk in mixed than

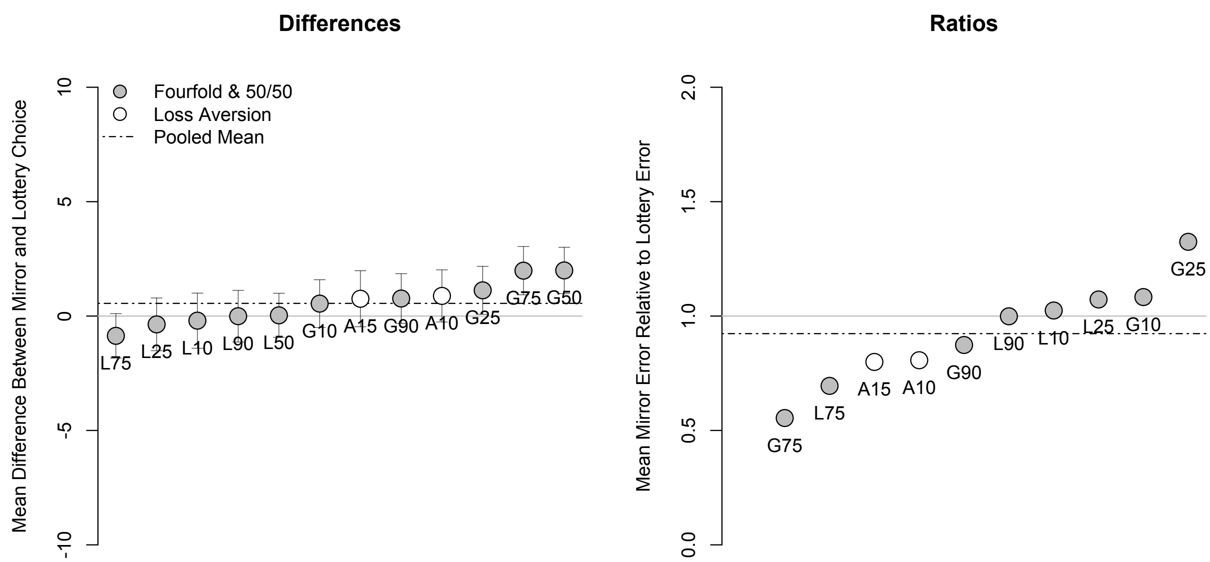


Figure 3: Comparisons between lottery and mirror tasks. *Notes: The left panel plots the mean difference between mirror and lottery net risk/loss seeking for each task, with two-standard error bars plotted in each direction. The right panel plots the ratio of total pattern-consistent errors in mirrors relative to lotteries for each task. Dashed lines plot the mean of each measure, pooling across tasks.*

unmixed 50/50 lotteries and similarly \$3.22 more to avoid "mixed" than "unmixed" 50/50 mirrors, suggesting that subjects are, on average, substantially more averse to mixed than umixed 50/50 lotteries and mirrors.

Importantly, conditional on making seemingly loss-neutral or loss-averse choices, there is *no actual scope for loss* in the mirrors of A10 and A15. We thus find strong evidence of apparent loss aversion in settings without actual risk of loss.

Result 2 *Apparent loss aversion appears in deterministic mirrors, just as it does in lotteries. We thus find strong evidence of "loss aversion" in settings without risk of loss.*

3.1 Relative Magnitudes

Visual inspection of Figures 1 and 2 suggests that the classical pattern not only arises in mirrors, it also arises with similar severity in mirrors as it does in lotteries. To study this more carefully we examine relative magnitudes of these anomalies in mirrors and lotteries in several ways. First, we examine the within-subjects difference in the lottery and mirror errors plotted in Figure 3 (effectively, measures of net risk/loss seeking) for the median and mean subject. For the median subject, this difference is zero in all 12 of our tasks, meaning that *in every task the median subject* *makes the same choices in lotteries and mirrors.* The left hand panel of Figure 3 plots the *mean* difference in errors for each of our fourfold and loss aversion tasks and for tasks G50 and L50 (with two standard error bars extending in each direction). We find some variation in the mean across tasks, with mirror errors sometimes larger and sometimes smaller than lottery errors. But these differences tend to be small and standard error bars overlap 0 for most of the tasks. Paired Wilcoxon tests do not allow us to reject the hypothesis that mirror and lottery choices are the same at the 5% level in nine out of our ten fourfold/loss aversion tasks (the exception is G75).

Second, we examine the ratio of mirror and lottery summed errors, normalized to be positive if they run in the direction of the classical pattern. This gives us a sense of the proportion of each anomaly in lotteries that also appears in mirrors in the aggregate. These proportions are plotted for each task in the right hand panel of Figure 3. Once again, we find variation across tasks, with anomalies sometimes more severe in lotteries and sometimes more severe in mirrors, but ratios are overall distributed roughly symmetrically around 1 (the benchmark for equal severity of the anomaly in mirrors and lotteries). Aggregating, for fourfold tasks the mean ratio is 97% across tasks suggesting that, overall, the fourfold pattern is about 97% as strong in mirrors as in lotteries. For our two loss aversion tasks, the mean ratio is 80%, suggesting that loss aversion may be moderately less severe in mirrors than lotteries by this metric.

Putting this evidence together, for the fourfold pattern there is little evidence that mirror behavior is systematically different from lottery behavior: the median subject makes no distinction between the two treatments, we generally can't statistically distinguish choices in the two treatments, and the relative magnitudes of pattern-consistent deviations is, on average, close to 1. For loss aversion, evidence is mixed. On the one hand, once again, median differences are zero, mean differences are close to zero and paired Wilcoxon tests fail to reject the hypothesis that the two behaviors are the same. However, in the aggregate, the mean magnitude of loss averse errors is somewhat smaller for mirrors than lotteries. Viewed in the context of the variation in the proportion of mirror/lottery errors in the datset overall (pictured in the right panel of Figure 3), it is difficult to tell whether this is due to sampling variation or is a real systematic difference in loss aversion relative to the fourfold pattern. However, a straight reading of the evidence suggests that loss aversion may be somewhat weaker on average (perhaps 80% as strong) in mirrors that it is in lotteries.[7]

Result 3 *The severity of the fourfold pattern is similar in lotteries and mirrors. By most measures the severity of loss aversion is likewise similar, though some evidence suggests that loss aversion may be moderately weaker in mirrors than in lotteries.*

[7]The left-hand panel of Figure 3 is also useful in that it allows us to study to what degree our loss aversion results are driven by the fact that they rely on 50/50 lotteries (which are plausibly arithmetically simpler than other lotteries in our design). In particular, it allows us to compare treatment differences in our mixed 50/50 lotteries to differences in our unmixed 50/50 lotteries, G50 and L50. There we find that treatment differences are small for mixed lotteries when compared to G50 but large when compared to L50. We therefore have little systematic evidence that the larger treatment gap we find in loss aversion tasks is driven by the fact that they rely on 50/50 lotteries.

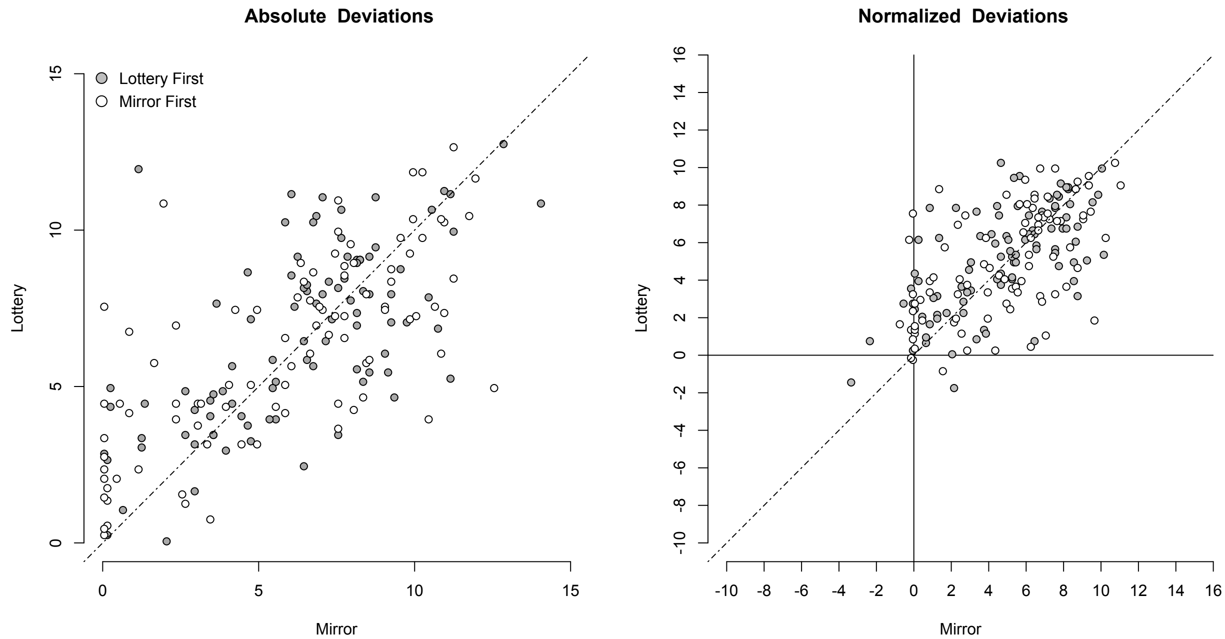


Figure 4: Deviations from expected value maximizing choices in mirrors (x-axis) versus lotteries (y-axis), by subject. *Notes: Each dot represents a separate subject. On the x-axes we plot the subject's data from the Mirror treatment and on the y-axes the same subject's data from the Lottery treatment. The left panel plots the mean absolute deviation from expected value. The right panel plots the mean deviation, normalized to be positive if it runs in the direction of the classical pattern. Gray dots are subjects who were assigned the Lottery treatment first, hollow dots subjects who were assigned the Mirror treatment first.*

3.2 Correlation and Heterogeneity

To what degree do these anomalies (the fourfold pattern and loss aversion) occur in risky lotteries and riskless mirrors for the same reason, driven by the same behavioral mechanism? To study this, we make use of our within-subjects design and examine the statistical relationship between anomalous behavior in mirrors and lotteries across subjects. Since there is no risk in mirrors, to the extent that evidence of anomalies is strongly *correlated* in lotteries and mirrors, we have evidence that they are likely both driven by the complexity of evaluation (the property lotteries and mirrors share) rather than by risk or risk preferences (a property absent from mirrors).

The left panel of Figure 4 plots a separate dot for each subject, with the x-axis plotting that subject's mean absolute deviation from the expected value maximizing choice in mirrors and the y-axis plotting the same deviation in lotteries. The plot shows a great deal of heterogeneity in the magnitude of errors across subjects, but a strikingly strong correlation between lottery and mirror errors of 0.68 ($p < 0.001$). The right panels instead examines mean deviations normalized to be positive if they run in the direction of the classical pattern (i.e., the fourfold pattern or loss aversion), again plotting the mean value for mirrors on the x-axis and for lotteries on the y-axis. We make three observations. First, virtually all deviations are concentrated in the northeast quadrant, suggesting that subjects make highly asymmetric errors on net in the distinctive direction of the classical pattern in both lotteries and mirrors. Second, although the severity of these deviations is highly heterogeneous, there is again a very strong correlation (0.62, $p < 0.001$) between mirror and lottery deviations, suggesting the two tendencies likely derive from a related behavioral mechanism. Finally, the correlation is virtually identical when subjects began in the Mirror treatment and move on to the Lottery treatment and vice versa (see Supplemental Appendix A.1 for details).

Result 4 *The severity of the classical pattern is strongly correlated, across subjects, in lotteries and mirrors.*

We make three additional observations about these correlations. First in Supplemental Appendix A.8.1, we show that these correlations are similar for the fourfold pattern and loss aversion when calculated separately. Second, our measures are unavoidably noisy and their correlations are therefore likely to be under-estimated due to attenuation bias (Gillen et al. 2019). Thus, as high as these correlations are, they should be viewed as lower bound estimates of the relationship between lottery and mirror behaviors. Third, at the end of the experiment, we asked subjects to report whether they used completely/mostly different strategies in lotteries and mirrors or identical/mostly similar strategies. 75% of subjects reported using identical or mostly similar strategies in the two treatments, strongly matching our behavioral findings and reinforcing the idea that the pattern is driven in each case by the same behavioral mechanism.

Finally, a key prediction of standard risk preference-based interpretations of the classical pattern (e.g., prospect theory) is that the pattern should only arise in the presence of risk. What proportion of subjects actually fit this description? To study this, we classify a subject as "complexity sensitive" if she deviates from expected value in the direction of the classical pattern by at least one price list row in her average mirror valuation, and "risk sensitive" if she deviates by at least one row *more* in the average lottery than in the average mirror. Standard risk preference-based theories predict that subjects' tendencies to exhibit the pattern will be risk sensitive but complexity *insensitive.* We find that only 14% of subjects can be classified this way, exhibiting the pattern in lotteries but not mirrors. Most subjects therefore deviate from the most basic prediction of risk preference-based theories. By contrast the vast majority of subjects (82%) can be classified as complexity sensitive, showing clear evidence of the classical pattern even in the absence of risk. Indeed, the modal subject (59% of subjects overall) is complexity sensitive but *not* risk sensitive, showing no more evidence of the pattern in lotteries than in mirrors. These results seem to suggest a dominant role for complexity in driving the classical pattern.

Result 5 *Most subjects make systematic mistakes matching the classical pattern in the absence of risk. Most subjects display the pattern no more strongly in the presence of risk than in its absence.*

An additional 22% of subjects are both complexity and risk sensitive by this classification, exhibiting the pattern in mirrors but even more strongly in lotteries. These complexity- and risk-sensitive subjects are either (i) displaying the pattern more strongly in lotteries than mirrors due to the additional influence of risk preferences on top of the complexity-driven mistakes they already display in mirrors or (ii) are simply more vulnerable to mistakes in the presence of risk than in its absence. As we discuss in Section 5, the latter interpretation is consistent with Martinez-Marquina et al. (2019) who provide evidence that already complex problems become more complex in the presence of risk.

3.3 Robustness and Additional Evidence

In the Supplemental Appendix we report results from several robustness treatments (using a total of 489 additional subjects) and several robustness exercises that aid in the interpretation of these results. First, in Supplemental Appendix A.1 we show that these results are not driven by order effects or contagion between treatments. Removing the second-assigned treatment and conducting an entirely between-subjects comparison between subjects first-assigned the Lottery vs. Mirror treatments, we find virtually identical results.

Second, in Supplemental Appendix A.3, we show that these results are not a special outgrowth of our use of multiple price lists (MPLs), but are instead a broader phenomenon of valuation. There we report a variation on our main design using the BDM mechanism (Becker et al. 1964) rather than multiple price lists ($N = 100$), and find very similar results: both the fourfold pattern and loss aversion arise similarly in lotteries and mirrors. However, as in previous work (discussed in Section 5, below), we find that the "shape" of these anomalies change with the method of elicitation: the fourfold pattern is more severe at high probabilities and loss aversion is weaker in BDM than in our MPL elicitations. Remarkably we find the exact same "elicitation effects" in mirrors (i.e. switching from MPL to BDM changes behavior nearly identically in lotteries and mirrors), suggesting that these elicitation effects themselves have little to do with risk. Indeed, we find that varying the method of elicitation (BDM vs. MPL) has a far larger effect on behavior than does varying the objective function itself (lottery vs. mirror): switching from MPL to BDM has a roughly four times larger effect on valuations than does removing risk from the lottery altogether. This strongly reinforces the interpretation offered below that valuations of objects like lotteries and mirrors do not transparently reveal preferences, but instead derive from subjects' use of relatively shallow heuristics that are highly sensitive to superficial details of the choice environment. In a follow up experiment to ours, Vieider (2023) reports evidence that simple binary choices over mirrors are nearly identical to binary choices over lotteries, suggesting that our findings may extend to risky decision-making more generally.

Third, in Supplemental Appendix A.4 we provide evidence that these results are not due to simple math errors or an aversion to mathematical difficult. There we report a version of our main design in which we attempt to lower the arithmetic difficulty of calculating, e.g., expected value, by (i) describing frequencies using "4 boxes" rather than "100 boxes" (e.g., a 75% chance of earning \$25 is described as 3 boxes containing \$25 rather than 75 boxes containing \$25) and (ii) in some exercises using rewards of \$20 rather than \$25 (making the arithmetic particularly simple).

We find these interventions have little effect on our results, suggesting that our findings do not derive from mathematical mistakes or even an aversion to difficult arithmetic. Subjects seem to be valuing lotteries heuristically not because the math is too difficult, but rather in order to avoid the "fixed costs" of properly setting up the valuation problem and of deploying the cognitive resources required to carefully evaluate disaggregated objects like lotteries and mirrors.

Fourth, in Supplemental Appendix A.2, we show that these results are slightly weakened but not eliminated with higher stakes and a more sophisticated subject pool. We report a version of our main design (i) run on undergraduate students ($N = 113$) instead of an online sample and (ii) using significantly higher (quintupled) incentives. We find broadly similar results: the fourfold pattern and loss aversion continue to arise strongly in both lotteries and mirrors. However, the fourfold pattern and especially loss aversion are somewhat weaker in mirrors relative to lotteries than in our other treatments (the fourfold pattern shrinks to 82% as strong and loss aversion 54% as strong). Perhaps the most important finding from this treatment is that the fourfold pattern is substantially smaller in *both lotteries and mirrors* in this treatment than in our main treatment. The weakening of the pattern in both mirrors and lotteries (and in mirrors relative to lotteries) when we use stronger incentives and a more mathematically sophisticated subject pool is consistent with a boundedly rational interpretation of these anomalies, since it suggests that higher incentives to optimize (and plausibly lower costs of optimizing among students), weaken these effects to some degree.

Finally, we collected a number of additional pieces of data in our main experiment that we correlate with the severity of the classical pattern in lotteries and mirrors (see Supplemental Appendix A.5 for details), giving us some insight into the behaviors that drive the classical pattern. For instance, we find that (i) fast decision-making, (ii) noisy, inconsistent choices in repeated instances of the same task and (iii) poor performance on cognitive reflection tasks administered post-experiment are all positively correlated with the severity of the classical pattern. We also asked subjects after the experiment (iv) how likely they believed it was that they made suboptimal choices (measuring "cognitive uncertainty," a'la Enke & Graeber (2023)), (v) how imprecise they thought their decision-making process was (on a 100-point Likert scale) and (vi) how little attention subjects believe they themselves paid to payoffs and proportions in the descriptions of mirrors (again, using a 100-point Likert scale), and found that all of these were significantly correlated with the pattern too. These results therefore link the classical pattern in both lotteries and mirrors to hasty, noisy, imprecise and inattentive decision-making and suggest that subjects were largely aware that they were making imperfect decisions in these valuations (i.e. in important respects they *know* they are heuristically valuing these objects). Importantly, this is virtually identically true in lotteries and mirrors: we find highly consistent correlations between the classical pattern and all of these measures in the two settings, reinforcing our conclusion that the pattern is driven by the same behavioral mechanism in lotteries and mirrors.

Putting these strands of evidence together, the twin appearance of the classical pattern in lotteries and mirrors suggests that it represents a response not to risk but rather to the complexity of valuation. Perhaps surprisingly, this complexity does not seem to be primarily rooted in the arithmetic required in valuation, but in other cognitively taxing aspects of the task. For instance simply thinking through how one's preferences connect to the primitives of lotteries and mirrors and articulating the implications for behavior plausibly requires significant mental effort, even if one has little difficulty with the math once the problem is "set up." We speculate that subjects make a kind of "extensive margin" choice when deciding how to approach valuation tasks like these, deciding first whether to (i) do a precise, careful job of evaluation, or instead to (ii) casually or informally approximate value using heuristic methods. Following approach (i) requires more mental effort, strain and time than approach (ii), leading many subjects to pursue approach (ii) instead. Auxiliary evidence from Supplemental Appendix A.5 seems consistent with this account, since this evidence shows that features of behavior that we would expect to accompany casual or informal valuation procedures (e.g., hasty, inconsistent, imprecise inattentive and error-prone choices) are highly predictive of the severity of the classical pattern.

4 Interpretation

We interpret these findings as evidence that preferences for even the simplest seeming lotteries are not transparent to decision-makers (as is often implicitly assumed in the literature), and that lottery valuations therefore do not reliably reveal subjects' risk preferences. Instead lotteries are *complex* in the sense that their values are costly or difficult for subjects to properly assess, and lottery valuations therefore often reveal the consequences of systematic heuristic mistakes instead of true preferences for risk. We show this by inducing risk-neutral preferences in standard lottery valuation tasks (i.e., "deterministic mirrors") and showing that subjects systematically fail to reveal those induced preferences in their valuations. Instead, subjects make systematic valuation errors that take the distinctive shape of the classic fourfold pattern of risk and loss aversion, two key empirical regularities in the literature that have inspired a number of behavioral theories of risk preferences. These systematic mistakes in mirrors strongly predict the same distinctive behaviors in lotteries, suggesting that the key empirical regularities typically used to measure putative components of preferences like probability weighting, reference dependence and loss aversion in lotteries are likely to a great extent driven by heuristic mistakes as well.

In offering this interpretation, it is important to emphasize that we do not make several seemingly related claims. We do not claim, for instance, on the basis of these data that risk preferences or even loss preferences do not exist, but only that they are unlikely to be reliably revealed in lottery valuations. Indeed, our finding of potentially stronger loss aversion in lotteries than mirrors might even be evidence that true loss averse preferences act as a secondary driver of loss aversion in lotteries on top of the mistakes that exclusively drive the same phenomenon in mirrors (though see Section 5 below for a caution concerning this interpretation). Likewise, our data suggests that the classical pattern sometimes appears as mistakes in deterministic settings (revealing that the pattern is not a special phenomenon of risk), but we do not have any basis to claim that this is a universal phenomenon. It is possible that some alternative framings of deterministic mirrors might make it less difficult or costly to infer the mirror's true value, attenuating or eliminating this effect. On the other hand, we would not be surprised (on the basis of evidence from Martinez-Marquina et al. (2019), discussed in the next section) if this were less true of lotteries: risk itself may make inferring true value difficult, regardless of the framing. As a result, there may well be some settings in which there is a larger wedge between lottery and mirror behavior than in our experiment.[8] However, this possibility has little impact on our main conclusion: that the classical pattern is in large part a description of the heuristic mistakes people make when the values of disaggregated objects (like lotteries or their mirrors) are difficult or costly for them to properly assess.[9]

Importantly, the literature has, in recent years, offered a number of descriptions of heuristic behaviors that are capable of generating the classical pattern without appeal to risk or risk preferences. A growing literature interprets lottery anomalies as growing out of *imprecise valuation strategies* and the tactics decision makers use to compensate for noise in their internal representations of numbers or calculations of aggregates. The "noisy coding" literature (Woodford 2020, Glimcher 2022), for instance, shows that if decision makers shade noisy evaluations of lotteries towards prior beliefs in a Bayesian manner, this can produce the fourfold pattern (Steiner & Stewart 2016, Khaw et al. 2022, Vieider 2023, Frydman & Jin 2023) and even loss aversion (Woodford 2012*a*) under some assumptions. Enke & Graeber (2023) show, similarly, that uncertainty about the quality of value calculations, combined with cognitive defaults, can produce the fourfold pattern. Blavatskyy (2007) shows that if decision makers do nothing more in response to valuation noise than ensure that their valuations do not exceed the bounds of the lottery's support, the fourfold pattern will emerge as a result. Closely related is the literature on "decision by sampling," which roots the classical pattern in heuristics built on imprecise comparisons between past and present circumstances (Friedman 1989, Stewart et al. 2006). Another literature shows that the pattern can arise instead from *inattentive valuation strategies*: Bordalo et al. (2012) show that if decision makers put excess weight on salient components of lotteries when valuing them, this can generate the fourfold

[8]To give an extreme example, if we were to directly tell subjects in our experiment the expected value, we would expect the classical pattern to shrink much more in mirrors than in lotteries. This is because knowing the expected value directly removes the difficulties of valuation in mirrors (since the value of mirrors *just is* the expected value), but it doesn't in any clear way remove those same difficulties in lotteries. In order to value lotteries, the subjects still must face the difficult task of accessing their preferences for risk and linking those preferences to the properties of the lottery, plausibly preserving the temptation to avoid these difficulties by valuing lotteries instead using the heuristics responsible for the classical pattern. Indeed, there is little evidence that simply showing subjects statistics like expected value meaningfully affects their evaluations of lotteries (Lichtenstein et al. 1969, Montgomery & Adelbratt 1982, Beauchamp et al. 2020).

[9]To show that the classical pattern is a pattern of mistakes, it seems sufficient to show that there are *some* settings that superficially resemble lotteries in which the classical pattern appears but is not rationalizable by preferences. To show that the same is likely true of lotteries, it seems sufficient to show that the pattern is strongly predicted by clear instances of such mistakes. These conclusions seem unaffected by the possible existence of settings in which the same mistakes are easier to avoid.

pattern; given wide-spread evidence in psychology on the greater salience of negative relative to positive information (Baumeister et al. 2001, Rozin & Royzman 2001) it is easy to see how a similar mechanism might produce apparent loss aversion as well (e.g., Bhatia & Golman 2019).

What these styles of explanations have in common is that, unlike theories of behavioral risk preferences, they formally apply equally to lotteries and mirrors, and therefore can explain why these patterns occur both with and without risk. As we highlight in Section 3.3 (and Supplemental Appendix A.5), the severity of the classical pattern in our data is strongly correlated with measures of behavioral noise (choice inconsistency) and cognitive uncertainty (expressed uncertainty about the optimality of one's own actions), which seems especially suggestive of explanations rooted in the use of imprecise valuation strategies – a tentative conclusion that is consistent with several recent papers (Enke & Graeber 2023, Frydman & Jin 2023, Vieider 2023, Khaw et al. 2022) that provide more direct evidence linking lottery anomalies to the predictions of noisy cognition models. However, we emphasize that multiple tactics for avoiding complexity may coexist in the heuristics subjects use in lieu of rational valuation (see, e.g., Ba et al. (2023) for evidence that both inattentive and imprecise tactics are used by subjects in inference tasks). In light of our results, understanding in more depth what heuristic behaviors drive these anomalies seems like an important future task for the literature.

5 Connections to the Literature

Methodologically, the closest paper to ours is Martinez-Marquina et al. (2019) who, like us, compare behavior in risky and deterministic versions of similar tasks. In their main exercise they study simple bidding tasks in which subjects fail to properly contingently reason about objects of uncertain value, leading them to overbid on those objects. Their main finding is that such overbidding occurs also in deterministic versions of these tasks (in which objects are worth their expected value with certainty), but that the rate of overbidding is 20 percentage points lower. In another task, Martinez-Marquina et al. (2019) ask subjects to bet on which of two stochastic states will occur by allocating lottery tickets across states, and vary whether subjects are paid stochastically based on the realized state or based on the expected value of their bet. They find that "probability matching" (a mistake in which subjects bet on each state in proportion to the state's likelihood of occurring instead of rationally betting everything on the more likely state) is about 8-percentage points more common in the risky version of the task than in the deterministic version.

These findings suggest that problems involving risk are often more complex than isomorphic deterministic problems, generating more mistakes. This provides a useful caution in interpreting our results because it suggests that the complexity of lotteries may be greater than that of mirrors in at least some choice settings. To the degree this is true, we should view the decomposition afforded by our approach as providing a conservative, *lower bound* estimate of the role complexity plays in driving lottery anomalies – in at least some cases, we should expect lottery errors to be more severe than mirror errors, even if errors in each are produced by complexity alone. Indeed, by some measures we find about a twenty percentage point difference in loss aversion in lotteries relative to mirrors, which may be a result of true loss aversion compounding the effects of complexity in lotteries, but might instead be an instance of the same sort of increase in complexity Martinez-Marquina et al. (2019) find in the presence of risk. On the other hand, we find only a few percentage point difference in mistakes rates between lotteries and mirrors in our fourfold lotteries. One lesson from Martinez-Marquina et al. (2019) is that the size of the effect risk has on task complexity can vary dramatically across problems (e.g., it is much smaller in absolute terms in their probability matching task than in their bidding task) and it may be that the fourfold lotteries are easier to reason about than mixed lotteries, reducing scope for risk to amplify complexity. Understanding in greater depth when and how risk itself influences complexity seems an important topic for future investigation.

Topically, our paper connects to a vast literature on risky choice in economics and psychology. One particularly relevant strand of this literature suggests that lottery choices are heavily influenced by complexity and cognitive errors, mirroring a key conclusion of our paper. The literature has amassed significant evidence linking departures from expected value (including the ones described by the classical pattern) to cognitive ability (Benjamin et al. 2013, Choi et al. 2021), cognitive load (Benjamin et al. 2013, Deck & Jahedi 2015, Gerhardt et al. 2016), choice inconsistency (Khaw et al. 2021, Enke & Graeber 2023) and inattention (Pachur et al. 2018), all of which suggest that many such departures may be driven by judgement errors. The literature has also produced evidence that making lotteries more complex (typically by adding outcomes to otherwise similar lotteries) produces stronger departures from expected value (Huck & Weizsacker 1999, Bernheim & Sprenger 2020, Puri 2023); Enke & Shubatt (2023) identify a number of lottery characteristics that predict subjects' failures to correctly infer the expected value of lotteries and shows that these same characteristics strongly predict departures from expected value maximization in lottery choice.

Another strand of the literature casts doubt on the assumption that lottery valuations reveal stable preferences, echoing another theme of our paper. A growing literature shows that people's preferences for risk, including aspects of preferences related to the classical pattern, change dramatically when the method of elicitation is changed (Friedman et al. 2017, 2022, Beauchamp et al. 2020, Bauermeister et al. 2018, Harbaugh et al. 2010, Holzmeister & Stefan 2021), casting some doubt on the idea that we are directly measuring preferences in these tasks. A large literature (e.g., Hertwig et al. 2004) shows that "decisions from experience" (decisions made between lotteries whose properties are discovered by sampling them) produce very different behavior than conventional "decisions from description," including a reversal of probability weighting. Another literature shows that anomalous phenomena often attributed to preferences including small stakes risk aversion (Ert & Haruvy 2017, Charness et al. 2023), probability weighting (Van de Kuilen 2009) and the Allais paradox (Van de Kuilen & Wakker 2006) are transient, declining or even disappearing with experience, calling into question the idea that these phenomena measure welfare-relevant preferences at all. Two recent papers provide direct evidence that people prefer the normative axioms of EUT, but fail to reveal those preferences in their lottery choices due, apparently, to complexity-derived mistakes (Nielsen & Rehbeck 2022, Benjamin et al. 2023).

Because of its intimate connections to the classical pattern, our paper connects to a long literature on prospect theory, by far the most influential "behavioral" theory of decision under risk (e.g., Kahneman & Tversky 1979, Tversky & Kahneman 1992, Barberis 2013, Wakker 2010), which was built to explain the classical pattern and related phenomena. Prospect theory describes the classical pattern as growing out of risk preferences, but the literature has long been ambivalent about the interpretation of these preferences and in particular whether prospect theory describes decision makers' welfare-relevent tastes for risk and loss or whether it instead describes judgement errors. We view our results as strong support for the latter interpretation.

Finally, our paper is connected to a growing literature in economics on how complexity shapes human decision making (e.g., Oprea 2020, Camara 2023). A strand of this literature that is particularly relevant to our paper focuses on a first order implication of complexity: that it causes decision-makers to be inensitive to features of decision problems that matter for optimal choice. The classical pattern can be interpeted, in large part, as an outgrowth of just this sort of insensitivity, an observation that goes back at least to Tversky & Kahneman (1992). Enke & Graeber (2023) use overt complexity manipulations and direct measures of "cognitive uncertainty" about the optimality of subjects' own choices to show that phenomena as seemingly-distinct as probability weighting, errors in forming expectations and failures of Bayesian updating arise in large part due to this kind of complexity-derived insensitivity. Ba et al. (2023) show that well-known anomalies in belief updating can similarly be interpreted as resulting from insensitivities to primitives (in conjunction with incomplete allocations of attention). Abeler & Jäger (2015) show that increasing the complexity of taxes results in insensitivities to marginal tax rates. Enke et al. (2023) use methods like ours ("atemporal mirrors" of intertemporal choice problems) in conjunction with measures of cognitive uncertainty and complexity manipulations to show that hyperbolic discounting in intertemporal decision-making is also primarily a phenomenon of complexity-driven insensitivity. These kinds of results underscore and expand upon our interpretation of our results by suggesting that the patterns of insensitivity that describe the classical pattern may be generic to the evaluation of complex things – a possibility that may unify a great number of anomalies in behavioral economics.

6 Conclusion

We provide evidence that some of the central lottery anomalies in behavioral economics (those used to measure phenomena like probability weighting, reference dependence and loss aversion), are not special phenomena of risk and therefore are unlikely to reflect decision makers' risk preferences. Instead, they are to a great extent patterns of heuristic mistakes that occur because lotteries are complex to properly evaluate, i.e., because their values are not transparent to decision makers, but are instead costly or difficult to infer. There are two implications of this. First, theories of risk preferences designed to explain these anomalies (e.g., prospect theory) are unlikely to contain much normative content and therefore should not be accomodated in the inference of welfare or the design of policy. Second, our finding of systematic departures from neoclassical benchmarks in perfectly deterministic settings suggests that many of our descriptive theories of preferences for risk are really descriptive theories of the way people evaluate complex things. Because of this, many of the phenomena that have animated the rich behavioral literature on decision making under risk likely have a much broader scope of application than has been so far appreciated.

References

Abeler, J. & Jäger, S. (2015), 'Complex Tax Incentives', *American Economic Journal: Economic Policy* **7**(3), 1–28.

Ba, C., Bohren, J. A. & Imas, A. (2023), 'Over- and underreaction to information', *Unpublished manuscript* .

Banovetz, J. & Oprea, R. (2022), 'Complexity and procedural choice', *Unpublished Manuscript* .

Barberis, N. C. (2013), 'Thirty years of prospect theory in economics: A review and assessment', *Journal of Economic Perspectives* **27**(1), 173–196.

Bauermeister, G.-F., Hermann, D. & Musshof, O. (2018), 'Consistency of determined risk attitudes and probability weightings across different elictation methods', *Theory and Decision* **84**, 627–644.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C. & Vohs, K. D. (2001), 'Bad is stronger than good', *Review of General Pscychology* **5**(4), 323–370.

Beauchamp, J., Benjamin, D., Laibson, D. & Chabris, C. (2020), 'Measuring and controlling for the compromise effect when estimating risk preference parameters', *Experimental Economics* **23**, 1069–1099.

Becker, G. M., DeGroot, M. H. & Marschak, J. (1964), 'Measuring utility by a single-response sequential method', *Behavioral Science* **9**, 226–232.

Benjamin, D., Brown, S. & Shapiro, J. (2013), 'Who is 'behavioral'? cognitive ability and anomalous preferences.', *Journal of the European Economic Association* **11**(6), 1231–1255.

Benjamin, D. J., Fontana, M. A. & Kimball, M. S. (2023), 'Reconsidering risk aversion', *Unpublished manuscript* .

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

Bhatia, S. & Golman, R. (2019), 'Attention and reference dependence', *Decision* **6**(2), 145–170.

Blavatskyy, P. (2007), 'Stochastic expected utility theory', *Journal of Risk and Uncertainty* **34**, 259–286.

Bordalo, P., Gennaioli, N. & Shleifer, A. (2012), 'Salience theory of choice under risk', *Quarterly Journal of Economics* **127**(3), 1243–1285.

Camara, M. (2023), 'Computationally tractable choice', *Unpublished Manuscript* .

Chapman, J., Snowberg, E., Wang, S. W. & Camerer, C. (2022), 'Looming large or seeming small? attitudes towards losses in a representative sample', *Unpublished Manuscript* .

Charness, G., Chemaya, N. & Trujano-Ochoa, D. (2023), 'Learning your own risk preferences', *Unpublished manuscript* .

Choi, S., Kim, J., Lee, E. & Lee, J. (2021), 'Probability weighting and cognitive ability', *Management Science* **forthcoming**.

Deck, C. & Jahedi, S. (2015), 'The effect of cognitive load on economic decision making: A survey and new experiments', *European Economic Review* **78**, 97–119.

Enke, B. & Graeber, T. (2023), 'Cognitive uncertainty', *Quarterly Journal of Economics* **forthcoming**.

Enke, B., Graeber, T. & Oprea, R. (2023), 'Complexity and time', *Unpublished Manuscript* .

Enke, B. & Shubatt, C. (2023), 'Quantifying lottery choice complexity', *Unpublished manuscript* .

Ert, E. & Haruvy, E. (2017), 'Revisiting risk aversion: Can risk preferences change with experience', *Economic Letters* pp. 91–95.

Frederick, S. (2005), 'Cognitive reflection and decision making', *Journal of Economic Perspectives* **19**(4), 25–42.

Friedman, D. (1989), 'The s-shaped value function as a constrained optimum', *American Economic Review* **79**(5), 1243–1248.

Friedman, D., Habib, S., James, D. & Williams, B. (2022), 'Varieties of risk preference elicitation', *Games and Economic Behavior* **forthcoming**.

Friedman, D., Isaac, R. M., James, D. & Sunder, S. (2017), *Risky curves: on the empirical failures of expected utility*, Routledge.

Frydman, C. & Jin, L. J. (2023), 'On the source and instability of probability weighting', *Unpublished Manuscript* .

Fudenberg, D. & Puri, I. (2022), 'Evaluating and extending theories of choice under risk', (manuscript).

Gerhardt, H., Biele, G. P., Heekeren, H. R. & Uhlig, H. (2016), 'Cognitive load increases risk aversion', *Unpublished manuscript* .

Gillen, B., Snowberg, E. & Yariv, L. (2019), 'Experimenting with measurement error: techniques with applications to the caltech cohort study', *Journal of Political Economy* **127**(4), 1826–1863.

Glimcher, P. (2022), 'Efficiently irrational: illuminating the riddle of human choice', *Trends in Cognitive Science* **26**(8), 669–687.

Harbaugh, W. T., Krause, K. & Vesterlund, L. (2010), 'The fourfold pattern of risk attitudes in choice and pricing tasks', *Economic Journal* **120**(545), 595–611.

Hershey, J. & Schoemaker, P. (1985), 'Probability versus certainty equivalence methods in utility measurement: Are they equivalent?', *Management Science* **31**(10), 1213–1231.

Hertwig, R., Barron, G., Weber, E. U. & Erev, I. (2004), 'Decisions from experience and the effect of rare events in risky choice', *Psychological Science* **15**(8), 534–539.

Holzmeister, F. & Stefan, M. (2021), 'The risk elicitation puzzle revisted: across methods (in)consistency?', *Experimental Economics* **24**, 593–616.

Huck, S. & Weizsacker, G. (1999), 'Risk, complexity and deviations from extpected-value maximizagion: Results of a lottery choice experiment', *Journal of Economic Psychology* **149**(9), 1644–1683.

Kahneman, D. & Tversky, A. (1979), 'Prospect theory: An analysis of decision under risk', *Econometrica* **47**(2), 263–291.

Khaw, M. W., Li, Z. & Woodford, M. (2021), 'Cognitive imprecision and small-stakes risk aversion', *Review of Economic Studies* **88**(4), 1979–2013.

Khaw, M. W., Li, Z. & Woodford, M. (2022), 'Cognitive imprecision and stake-dependent risk attitudes', *Unpublished manuscript* .

Lichtenstein, S., Slovic, P. & Zink, D. (1969), 'Effect of instruction in expected value on optimality of gambling decisions', *Journal of Experimental Psychology* **79**(2), 236–240.

Martinez-Marquina, A., Niederle, M. & Vespa, E. (2019), 'Failures in contingent reasoning: the role of uncertainty', *American Economic Review* **109**(10), 3437–3474.

Montgomery, H. & Adelbratt, T. (1982), 'Gambling decisions and information about expected value', *Organizational Behavior and Human Performance* **29**(1), 39–57.

Nielsen, K. & Rehbeck, J. (2022), 'When choices are mistakes', *American Economic Review* .

Oprea, R. (2024), 'Data and code for: Decisions under risk are decisions under uncertainty'.

Oprea, R. D. (2020), 'What Makes a Rule Complex', *American Economic Review* **110**(12), 3913–3951.

Pachur, T., Schulte-Mecklenbeck, M., Murphy, R. O. & Hertwig, R. (2018), 'Prospect theory reflects selective allocation of attention', *Journal of Experimental Psychology: General* **147**(2), 147–169.

Puri, I. (2023), 'Simplicity and risk', *Unpublished Manuscript* .

Rozin, P. & Royzman, E. B. (2001), 'Negativity bias, negativity dominance, and contagion', *Personality and Social Psychology Review* **5**(4), 296–320.

Simon, H. A. (1955), 'A Behavioral Model of Rational Choice', *The Quarterly Journal of Economics* **69**(1), 99–118.

Sprenger, C. (2015), 'An endowment effect for risk: Experimental tests of stochastic reference points', *Journal of Political Economy* **123**(6), 1456–1499.

Steiner, J. & Stewart, C. (2016), 'Perceiving prospects properly', *American Economic Review* **106**(7), 1601–1631.

Stewart, N., Chater, N. & Brown, G. D. (2006), 'Decision by sampling', *Cognitive Psychology* **53**, 1–26.

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

Van de Kuilen, G. (2009), 'Subjective probability weighting and the discovered preference hypothesis', *Theory and Decision* **67**, 1–22.

Van de Kuilen, G. & Wakker, P. (2006), 'Learning in the allais paradox', *Theory and Decision* **33**, 155–164.

Vieider, F. (2023), 'Decisions under uncertainty as bayesian inference on choice options', *Unpublished Manuscript* .

Wakker, P. & Deneffe, D. (1996), 'Eliciting von neumann-morgenstern utilities when probabilities are distorted or unknown', *Management Science* **42**(8), 1131–1150.

Wakker, P. P. (2010), *Prospect theory: For risk and ambiguity*, Cambridge university press.

Woodford, M. (2012*a*), 'Prospect theory as efficient perceptual distortion', *American Economic Review Papers and Proceedings* **102**, 41–46.

Woodford, M. (2012*b*), 'Prospect theory as efficient perceptual distortion', *American Economic Review Papers and Proceedings* **102**, 41–46.

Woodford, M. (2020), 'Modeling imprecision perception, valuation and choice', *Annual Review of Economics* **12**.

Supplemental Appendix

Decisions Under Risk are Decisions Under Complexity

Ryan Oprea

A Additional Results

A.1 Between-Subjects Comparison

A natural concern about our results is that they may be a consequence of contagion between mirrors and lotteries resulting from our use of a within-subjects design. Perhaps subjects re-use heuristics they first employ in lotteries in their later mirror decisions or vice versa, causing behavior in the two treatments to be similar on average for reasons artificial to our design?

We can evaluate this alternative interpretation simply by restricting attention to *subjects facing the first of the two treatments they are assigned*, transforming our within-subjects design (with potential contamination) into a between-subjects design (without scope for contamination). This transformation is credible because subjects facing their first treatment (Mirror or Lottery) were not aware that they would later be facing the other treatment (Lottery or Mirror), removing scope for even prospective contamination. Figure 5 reconstructs Figure 1 using only this subset of the data and produces nearly identical qualitative results, suggesting that these results are not an artifact of cross-treatment contamination. Subjects continue to display very similar evidence of the pattern in mirrors and lotteries even when they have not yet experienced (or even learned of the existence of) the other treatment. Valuations continue to deviate significantly (at at least the 5% level via Wilcoxon tests) from expected value in the direction of the pattern in both lotteries and mirrors for all valuations.

A related concern is that the correlations between lotteries and mirrors visualized in Figure 4 are driven by subjects carrying over their behavior from the first treatment into the second, rather than by a deep connection in behavioral mechanism between the two treatments. A reason to doubt this interpretation is that (as just discussed) we find nearly identical initial behavior across the two treatments before subjects know the other treatment exists. What's more, the correlations between the two treatments in Figure 4 are also nearly identical regardless of the order of treatments.[10] Since contagion doesn't seem to be a first order driver of behavior and the correlations between treatments are not affected by order, the correlations are instead likely to be driven by subjects using similar valuation strategies in the two different treatments in the first place.

Thus our evidence suggests that our results are not an artifact of order effects or cross-treatment contagion.

Result 6 *The classical pattern continues to arise in both mirrors and lotteries in between-subjects comparisons. There is little evidence of contagion or order effects in the data.*

A subtler version of the same concern is that, for reasons that have little do with contagion, subjects might be drawn to heuristics usually reserved for interpreting (or valuing) probabilities

[10] For absolute deviations (the left panel of Figure 4) the correlation is 0.71 when mirrors come first and 0.65 when lotteries come first; for normalized deviations (the right panel) it is 0.61 when mirrors come first and 0.64 when lotteries come first.

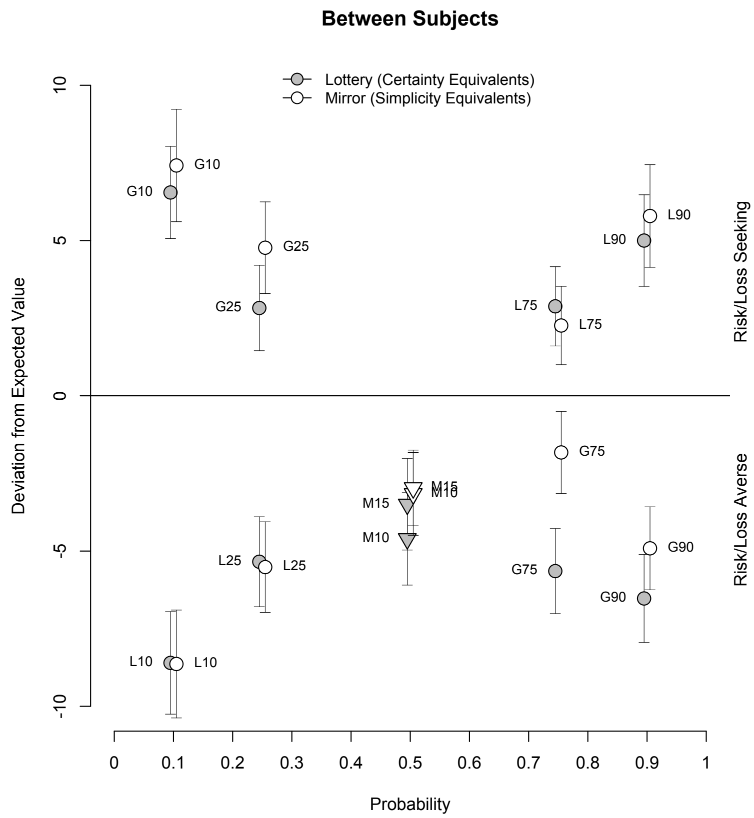


Figure 5: Between-subjects mean deviations from expected value in lotteries (gray dots) and mirrors (hollow dots) from the main (MPL) treatment. *Notes: For fourfold lotteries, the y-axis measures the difference between subjects' certainty/simplicity equivalent and expected value, as stated in the axis label. The x-axis is the probability of the non-zero payoff. For loss aversion tasks, the y-axis measures instead the difference between the certain/simple payoff and the expected value of the mixed lottery/mirror. Two-standard-error bars are included for every task.*

when valuing riskless mirrors. For instance, it may be that subjects apply risk preferences or distort probabilities in mirrors simply because they contain probabilities and subjects are acccustomed to responding to probabilities in a distorted way whenever they see them. However, it is important to emphasize that we deliberately attempted to rule this out in our design by framing the entire exercise in frequentist terms. Mirrors were described entirely as a "box opening" exercise in order to allow us to completely avoid mention of probabilities, likelihoods or randomness in our framing and instructions of this treatment. Consequently, subjects who were initially assigned mirrors (and who, recall, were not told that they would later be assigned lotteries) had no basis for importing lottery-like responses to the deterministic weights we assigned in these valuation tasks. The fact that (as Figure 5 shows) these subjects continue to display the pattern strongly suggests that such "mis-importation" of loettry behavior is unlikely to account for our results. If subjects apply probabilistic reasoning to these frequentist problems, arguably we should equally expect them to do so in virtually any other deterministic valuation task as well.

A.2 Student Sample

An important question is whether our results are a consequence of implementation choices such as (i) our use of an online subject pool rather than a conventional student pool, (ii) limitations in training of subjects due to our online implementation, or (iii) the scale of incentives we used in our design (recall we only pay subjects based on their choices with 20% chance). Perhaps our results are artifacts of unsophisticated subjects, insufficient trading or weak incentives – any of which could plausibly exaggerate noisy and biased behavior.

We ran a nearly identical version of our main design using 113 undergraduate students at UC Santa Barbara in a manner that removes (or at least reduces) these concerns by using more intensive training and stronger incentives. First, this experiment used undergraduate students at a selective university rather than an online subject pool. Experiments were run on Zoom in conventional, fixed experimental sessions monitored by the experimenter, allowing subjects to ask the experimenter clarifying questions in real time before and during the experiment. Second, this experiment featured more intensive training than in our main design. Specifically, we quadrupled the number of comprehension questions subjects were asked immediately before each of the treatments (Mirror and Lottery). These questions were designed to highlight for subjects the differences between the incentives of lotteries vs. mirrors in order to remove the possibility that subjects mistook one payoff rule for the other. Finally, this experiment quintupled the incentives in the main experiment by paying subjects based on a random lottery with certainty (rather than with 20% chance).

The UCSB sessions were conducted in February and March 2021 using 113 subjects from the subject pool of the Laboratory for the Integration of Theory and Experiments (LITE) at UC Santa Barbara. Because of the Covid-19 pandemic, the physical laboratory was closed at this time so the five sessions of data collection were held remotely on Zoom. In each session no more than 25 subjects from the undergraduate population at UC Santa Barbara were invited by email to log into our Zoom account at a pre-specified time. They were then given a link to the experimental software and were allowed to ask the experimenter questions throughout the session.

Relative to the main sessions run on Prolific, the UC Santa Barbara sessions differed in three major respects:

- The main sessions conducted on Prolific were more demographically diverse, drawing subjects from throughout the United States and included largely non-student subjects. By contrast, the UCSB sessions included only students from the University of California, Santa Barbara, a selective public university.
- As the instructions in Supplemental Appendix B discuss, we gave subjects four identical quiz questions concerning the nature of payments immediately prior to the Lottery treatment and again prior to the Mirror treatments in the Prolific sessions. Because the answers to these questions differed across the two treatments, these questions allowed us to make payoff differences across treatments salient to subjects. In the UCSB sessions we quadrupled the number of questions, adding additional questions in both the gains and loss domain. Thus these sessions intensified subjects' training.
- In the Prolific sessions we gave subjects a $6 fixed payment for participation and paid 20% of subjects (randomly selected, ex post) a bonus based on their decision in a random price list and row. By contrast, in the UCSB sessions we paid subjects a $5 fixed payment and, in addition, paid *all* subjects a bonus based on their decision in a random price list and row. Incentives were therefore substantially larger in the UCSB sessions.

Additionally, the sessions differed in two respects that are less likely to have influenced the results reported in the paper:

- In the Prolific sessions, we asked subjects a number of unincentivized questions at the end of the experiment about their decision-making (reviewed in Supplemental Appendix A.5). In the UCSB sessions, we included only the cognitive reflection test and a single cognitive uncertainty measure.
- The UCSB sessions included four additional price lists not included in the Prolific sessions. These were rather more complex lotteries designed to gather non-parametric measures of prospect-theoretic value function curvature using methods suggested by Wakker & Deneffe (1996). These lists, intriguingly, produced evidence of similar degree of value function curvature in Mirrors and Lotteries, but the results were extremely noisy and sensitive to specification. For this reason (and because these results are only of secondary importance to our main motivating questions), we did not use these lists in our main Prolific sessions.

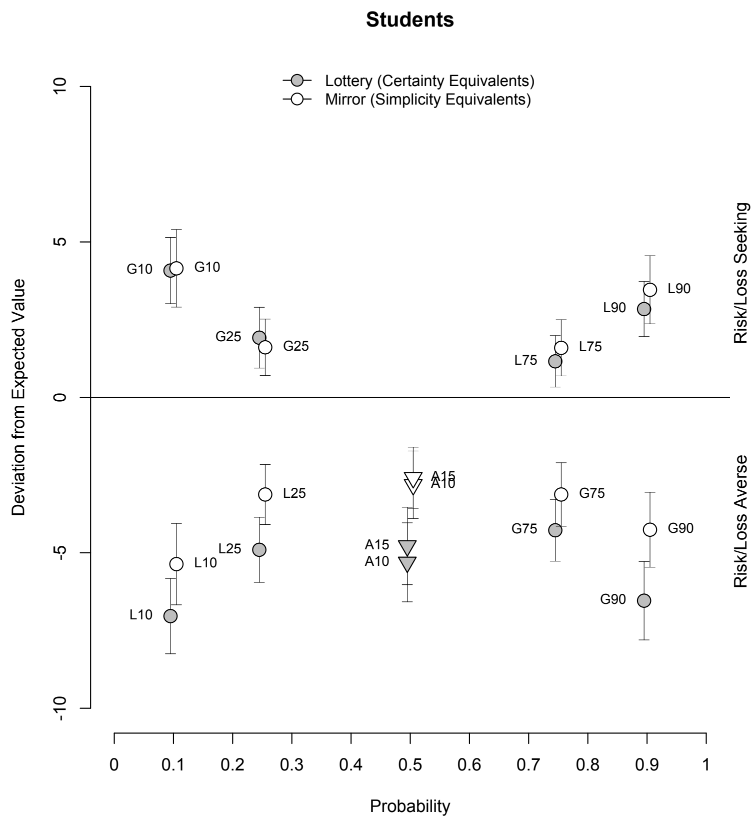


Figure 6: Student sample deviations from expected value in lotteries (gray dots) and mirrors (hollow dots). *Notes: For fourfold lotteries, the y-axis measures the difference between subjects' certainty/simplicity equivalent and expected value, as stated in the axis label. The x-axis is the probability of the non-zero payoff. For loss aversion tasks, the y-axis measures instead the difference between the certain/simple payoff and the expected value of the mixed lottery/mirror. Two-standard-error bars are included for every task.*

In all other respects, including instructions, software and decision tasks the UCSB sessions were identical to the main sessions.

Figure 6 plots the results from these sessions and they strongly suggest that these features of the implementation are not driving our results. The plot shows continued evidence that the full classical pattern appears in mirrors and to a similar degree as in lotteries; we can reject the hypothesis that subjects choose expected value in every list for both lotteries and mirrors (at the 1% level by Wilcoxon tests). We also continue to find a similarly strong correlation between the pattern in the two cases ($\rho = 0.64$ for absolute deviations and $\rho = 0.5$ for deviations normalized in the direction of the pattern). The main difference in this sample is that by some metrics there is a somewhat larger "gap" between the strength of the pattern in mirrors and lotteries: summed errors in the direction of the fourfold pattern are overall 82% as large and loss aversion 54% as large in Mirrors as Lotteries.

Result 7 *A robustness sample of university students with increased training and quintupled incen-*

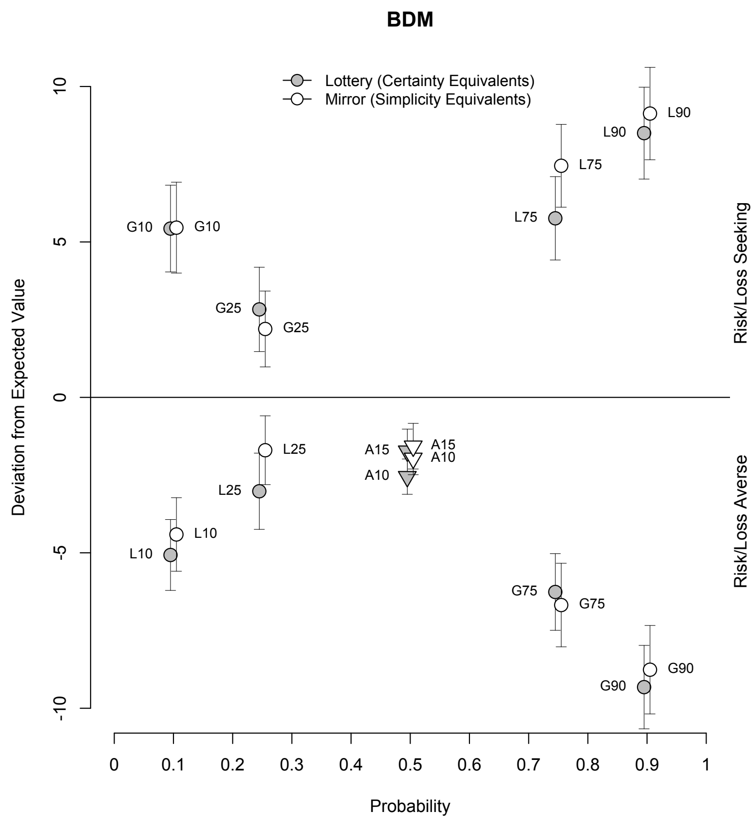


Figure 7: Mean deviations from expected value in BDM lotteries (gray dots) and mirrors (hollow dots). *Notes: For fourfold lotteries, the y-axis measures the difference between subjects' certainty/simplicity equivalent and expected value, as stated in the axis label. The x-axis is the probability of the non-zero payoff. For loss aversion tasks, the y-axis measures instead the difference between the certain/simple payoff and the expected value of the mixed lottery/mirror. Two-standard-error bars are included for every task.*

tives produces results similar to those in the main dataset.

A second interesting difference is, as is clear from Figure 6, we find a somewhat weaker fourfold pattern in this data than in the main sample: valuations are closer to expected value. But this is true in both lotteries and mirrors, meaning whatever mechanism drives these sample effects is linked to complexity (shared by lotteries and mirrors) rather than risk or risk preferences. Conversely, we find an intensification of loss aversion in lotteries in our student sample, but no such intensification in mirrors, driving the increased gap in loss aversion. This increase in loss aversion in student samples has been reported in recent work (Chapman et al. 2022) and may suggest that the "gap" in loss aversion between lotteries and mirrors is driven by true loss averse preferences operating as a secondary driver of loss averse valuation.

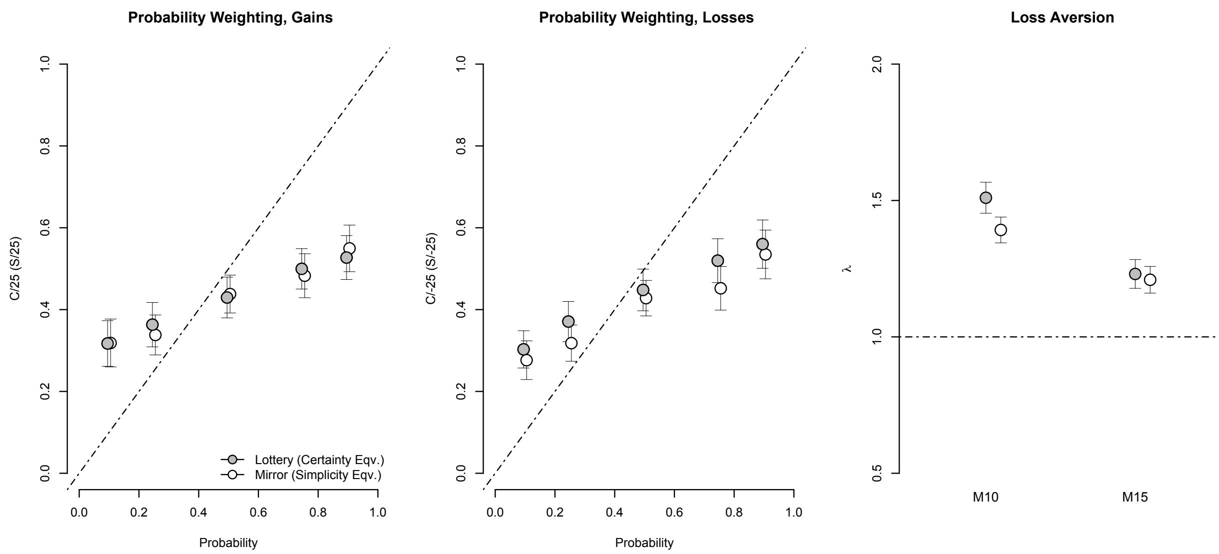


Figure 8: Naive visualization of the probability weighting functions (left two panels) and the loss aversion parameter, λ in the BDM treatment. *Notes: The first two panels plot a naive estimate of the probability weighting function (following Tversky & Kahneman (1992)) by plotting the ratio of the certainty/simplicity equivalent to the non-zero payment amount as a function of the probability of the non-zero payoff amount. The final panel plots a naive estimate of λ, the standard linear parameter of loss aversion, under the assumption of a reference point of zero.*

A.3 BDM Treatment

Another natural question about our results is whether they are a special outgrowth of our use of mulitiple price lists (MPLs), or if they are a more general phenomenon of valuation. To answer this question we ran the the BDM treatment ($N = 100$, collected on Prolific in April 2023) in which we replicated our main MPL design but elicited certainty/simplicity equivalents using the Becker-Degroot-Marschak or “BDM” mechanism (Becker et al. 1964). In our BDM tasks, subjects are shown the lottery being evaluated (e.g., G10) and asked to express their willingness to pay either to acquire (WTP-to-acquire) or to avoid (WTP-to-avoid) this lottery. Specifically, subjects were asked to enter a certainty/simplicity equivalent C in the lottery’s support in a text box (see Figure 18 for a screen shot from a WTP-to-acquire task and Figure 19 for a screen shot from a WTP-to-avoid task). The subject was informed that the computer would later draw a random price P in the support. For WTP-to-acquire tasks, if C<P the subject is not assigned a payment based on the lottery in question; if C≥P, the subject acquires the lottery and pays price P. For WTP-to-avoid tasks, if C<P the subject is assigned the lottery; if C≥P, the subject is not assigned the lottery and pays price P.

The fourfold lotteries used in this design are identical to those used in the main MPL treatment. However, because it is difficult to measure quantities other than certainty/simplicity equivalents using the BDM we are unable to use lottery equivalents (i.e. tasks A10 and A15) to measure loss aversion. We instead measure loss aversion by eliciting certainty/simplicity equivalents for

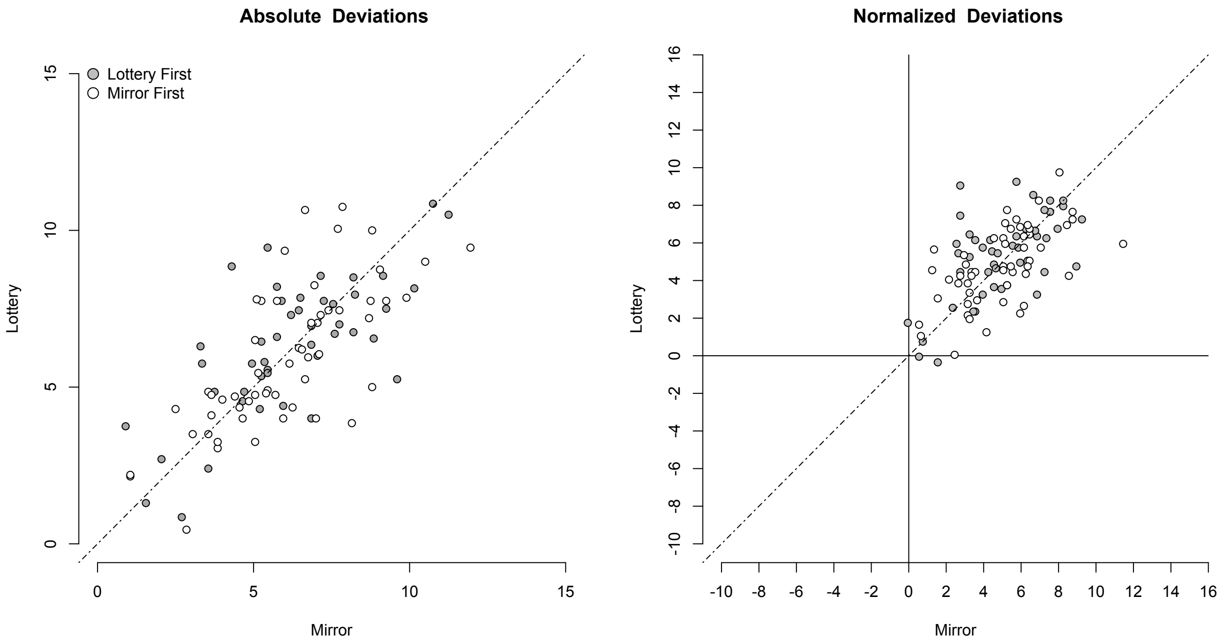


Figure 9: Deviations from expected value maximizing choices in mirrors (x-axis) versus lotteries (y-axis) in the BDM treatment, by subject. *Notes: Each dot represents a separate subject. On the x-axes we plot the subject's data from the Mirror treatment and on the y-axes the same subject's data from the Lottery treatment. The left panel plots the mean absolute deviation from expected value. The right panel plots the mean deviation, normalized to be positive if it runs in the direction of the classical pattern. Gray dots are subjects who were assigned the Lottery treatment first, hollow dots subjects who were assigned the Mirror treatment first.*

(i.e. willingness to pay to avoid) the mixed lotteries/mirrors $M10 = (0.5; 5, -10)$ and $M15 = (0.5; 5, -15)$. We chose these negative expected value lotteries to create scope for subjects to reveal not only loss averse but also loss seeking valuations in simple elicitations of subjects' willingness-to-pay-to-avoid. We thus subjects' WTP-to-acquire lotteries G10, G25, G50, G75 and G90 and their WTP-to-avoid lotteries L10, L25, L50, L75, L90, M10 and M15 in these experiments. Subjects were paid at the end based on their choice in one randomly selected task and one randomly selected price, P.

We present our findings in Figures 7, 8 and 9, designed to mirrors Figures 1, 2 and 4, respectively, from the body of the paper. As Figure 7 shows, from the perspective of our main motivating questions, the BDM results are very similar to those for the main MPL treatment. As with MPL, we find evidence of the fourfold pattern in both lotteries and mirrors. Summing up pattern-consistent choices, we find that the fourfold pattern is 99% as strong in mirrors as in lotteries. We also find evidence of loss aversion in both lotteries and mirrors, with loss aversion 83% as strong in the latter as in the former. In each of our 12 tasks we find (as in MPL) that for the median subject the difference in valuations between lotteries and mirrors is 0. As Figure 9 shows, mirror and lottery deviations are strongly correlated: we find a correlation of 0.72 for absolute deviations and 0.6 for normalized deviations, closely matching results from the MPL treatment.

We conclude that our main results are not driven by our use of multiple price lists, but instead reflect a more general phenomenon of valuation.

Result 8 *Results from the BDM treatment are similar to those from our main MPL treatment.*

Visual comparison of Figures 7 and 1 (from the body of the paper) reveals significant effects of the method of elicitation (MPL vs. BDM) on estimates of the severity of the fourfold pattern in lotteries. In particular, deviations from expected value are somewhat smaller at low probabilities and much larger at high probabilities in BDM than in MPL. As we discuss in Section 5 these kinds of "elicitation effects" are standard in the literature – measured risk preferences (including prospect theoretic parameters) tend to vary (often within-subject) across choice contexts. What is new here is that we find virtually identical elicitation effects in mirrors, with the shape of the fourfold pattern changing in identical ways in the two contexts. As a result, mirror behavior tracks lottery behavior across elicitation methods, suggesting that elicitation effects themselves have little to do with risk or risk preferences.

This is important for the interpretation of our results, because it strongly reinforces our finding that lottery valuations fail to reveal risk preferences. As we show in the paper, changing the objective function itself – the preferences these elicitations are generally deployed to measure – by inducing risk neutral preferences using mirrors has surprisingly small effects on lottery valuations. By contrast, changing seemingly superficial details of the elicitation method has large, first-order effects *that are identical across objective functions* (i.e. across lotteries and mirrors). Comparing the change in valuations due (i) to changes in the objective function (mirrors vs. lotteries) and (ii) to changes in the method of elicitation (MPL vs. BDM), we find that the latter effect is at least twice as large for every one of our tasks and typically much larger. Pooling across all of our tasks we find that changes in the method of elicitation are, on average, four times larger in absolute value than changes in the underlying objective function (2.96 vs. 0.737). This differential strongly suggests that valuation is dominated by heuristic behaviors that respond to details of the choice environment but that have little connection to underlying preferences.

It is also apparent when comparing Figures 7 and 1 that loss aversion estimates differ between MPL and BDM, a finding that may be notable for a different reason. A very conventional explanation for the large differences in loss aversion in these two cases is that while A10/A15 (studied in MPL) likely firmly establishes a reference point of zero, M10/M15 (studied in BDM) plausibly establishes a reference point of less than zero. As a result λ, as we've calculated it, plausibly underestimates loss aversion in M10/M15 relative to A10/A15 due to a change in the reference point across the two cases. Perhaps surprisingly, we find an identical "reference point effect" in mirrors, with an almost identical weakening of loss averse behavior in M10/M15 relative to A10/A15. Of course, this effect (like the effects observed in the fourfold pattern) might be due, in both lotteries and mirrors, to the change in the elicitation mechanism. However, to the degree we interpret this weaking of measured loss aversion as a reference point effect, the results suggest that not only does loss aversion and probability weighting survive the removal of risk, so does reference dependence

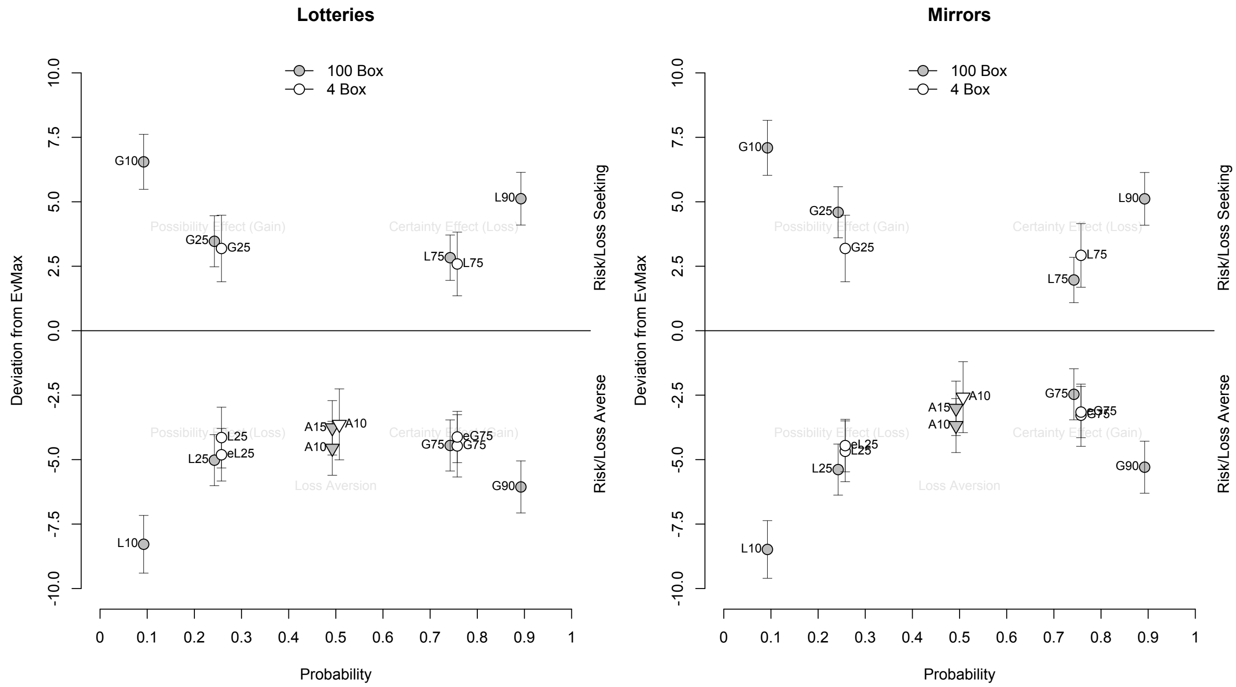


Figure 10: Results from the Easier treatment (4 box), overlaid on on results from the main sample (100 box). *Notes: Panels are included for Lotteries (left) and Mirrors (right). For fourfold lotteries, the y-axis measures the difference between subjects' certainty/simplicity equivalent and expected value (as stated in the axis label). The x-axis is the probability of the non-zero payoff. For loss aversion tasks, the y-axis measures instead the difference between the certain payoff and the expected value of the mixed lottery. Two-standard-error bars are included for every lottery. .*

and its sensitivity to manipulations of the reference point.

Result 9 *Variation in the method of elicitation influences the anomalies of the classical pattern in identical ways in lotteries and mirrors. Variation in the method of elicitation has a substantially larger effect on these anomalies than does variation in the objective function itself.*

A.4 4-Box Treatment

A further question about our main results is whether they are a consequence of arithmetic difficulties that arise due to the numbers used in the main design. For instance, we describe lotteries/mirrors using 100 outcomes (i.e., 100 boxes) and perhaps it is difficult to perform calculations involving this many outcomes. Likewise, the non-zero payment in our fourfold lotteries was $25 which does not produce whole-number expected values in any of our lotteries – perhaps this makes it unnecessarily difficult to calculate true value in mirrors and the expected value in lotteries. Perhaps subjects are constrained in their ability to perform arithmetic, causing them to make errors that show up as the classical pattern.

A strong ex ante reason to doubt this interpretation is that all of our treatments were run online, meaning all of our subjects had ready access to powerful calculators that make the arithmetic trivial. This means that subjects were not constrained in their ability to (with some minor effort) precisely calculate expected value. To reinforce this point, we report a robustness treatment we call "4-Box" in which we reduce the difficulty of the arithmetic required to calculate expected value. First, in our main dataset, likelihoods are described using 100 boxes, each of which contains a dollar amount, and non-zero payments are described as appearing in 10, 25, 50, 75 or 90 of the boxes. In the 4-Box treatment we shrink the outcome space from 100 boxes to 4 boxes *without changing the underlying probabilities*. Doing this allows us to express payoffs occurring with 0.25, 0.5 and 0.75 probabilities as dollar amounts contained in 1,2 or 3 of the boxes instead of 25, 50 or 75 of the boxes, plausibly making the problem easier to reason about and mathematical calculations easier to conduct. Thus, in the 4-Box treatment we repeat the G25, G50, G75, L25, L50, L75 and A10 lotteries but describe them using 4 boxes instead of 100.[11]

The 4-Box treatment was conducted in May of 2022 using 90 subjects on Prolific and MPL elicitation. The treatment repeated Lotteries G25, G50, G75, L25, L50, L75 and A10. The reason we did not include G10, G90, L10 and L90 is because the main idea of the treatment is to describe probabilities in frequentist terms using four outcomes (four "boxes") instead of 100. While 25%, 50% and 75% odds can be described using this coarse of a state space, clearly 10% and 90% cannot. The treatment also included (i) a repetition of L50 and G50 and (ii) treatments sG75 and sL25 which replaced the non-zero payment of $25 in G75 and L25 with $20. The instructions, implementation and payoff rules from the 4-Box treatment were identical to those in the main treatment except for the descriptions of frequencies. Instead of describing 25%, 50%, 75% and 100% as payouts contained in 25 out of 100, 50 out of 100, 75 out of 100 and 100 out of 100 boxes (as in the rest of the dataset), we described them as being contained in 1 out of 4, 2 out of 4, 3 out of 4 and 4 out of 4 boxes.

Figure 10 plots the results. It includes one panel for lotteries and another for mirrors and in these panels repeats the data pictured in Figure 1 using solid dots (100 box data), for reference. On each of these panels we overlay, using hollow dots, data from 4-box versions G25, G50, G75, L25, L50, L75 and A10 lotteries from the 4-Box treatment. We make two observations. First, the pattern continues to arise (for both lotteries and mirrors) under this simplified framing – Wilcoxon tests continue to allow us to reject the hypothesis of valuation at expected value for both Lotteries and Mirrors ($p < 0.01$ throughout). Second, valuations change little in either lotteries or mirrors when we move from 100-box to 4-box frames – Wilcoxon tests allow us to reject the hypothesis of identical valuation in 100-box and 4-box lotteries for only one of the ten comparisons (G25 mirrors). We conclude that the number of outcomes has at most a secondary effect on the appearance and

[11] It is important to highlight that this treatment does not make lotteries/mirrors any less disaggregated (the lottery's support continues to contain two elements) and therefore it does not make it any less *complex* in the sense of Bernheim & Sprenger (2020), Puri (2023) and Fudenberg & Puri (2022). This treatment holds the amount of information that has to be processed (the number of elements that must be aggregated) constant but attempts to reduce the mathematical difficulty of that processing.

severity of the pattern.

A second potential source of arithmetic difficulties in the main dataset is the use of a non-zero payoff of \$25 in the fourfold lotteries, which may be more difficult to reason about than a rounder number that is more easily multiplied by the relevant probabilities/weights in the task. To examine this we added to the 4-Box treatment a repetition of lotteries L25 and G75 but with a payoff of \$20 instead of \$25. We ran this also with the 4-box (rather than 100-box) design, making intuitive calculations of expected value particularly easy (\$20 in 2 or 3 boxes is easily seen to imply expected values of \$10 or \$15 through simple whole-number division). We call these lotteries sL25 and sG75 and plot valuations from these lotteries in Figure 10. We find no overall reduction in the severity of the pattern. Again, this suggests that mere arithmetic difficulty has little power to explain our results.

Together, these treatment interventions (combined with our already maximally simple 2-outcome setting, featuring a zero-outcome in one of the two outcomes) produce perhaps the arithmetically simplest possible lotteries in which the pattern can be measured. Our \$20 lists ask subjects to value lotteries that have the minimal possible number of outcomes (for a true lottery), one of these outcomes pays nothing and can be ignored in computation, the numbers describing the likelihoods are small and the non-zero payoff is calibrated to allow for whole-number computations of expected value by simple division. Nonetheless, we continue to find strong evidence of the pattern both in lotteries and their mirrors even in these maximally arithmetically simple valuation tasks.

Result 10 *Making valuation tasks arithmetically easier has only minor effects on the severity of the classical pattern in mirrors or lotteries.*

A.5 Correlates of the Pattern

Strong correlations between behavior in lotteries and mirrors suggest they are driven by a common mechanism and, because there is no risk in mirrors, suggest that this mechanism is rooted in the complexity the two types of tasks share. In order to gather some clues as to the common mechanisms that drive the pattern in both lotteries and mirrors, we collected a number of auxiliary measures and here we study to what degree these measures predict the severity of classical anomalies in both cases. We gathered three types of measures and we conduct a primarily exploratory analysis of how they relate to the incidence and severity of anomalous behavior in our main treatment. For this analysis, we restrict attention to our main MPL treatment where we have the most data and therefore the most reliable estimates.

First, we gathered several behavioral measures. Most importantly, we *repeated* two random valuation tasks in both lotteries and mirrors, allowing us to measure re-test consistency of choices in identical problems. The mean absolute difference in valuation between identical problems gives us a direct measure of **noise** in subjects' decision making. Next, we measured the average response **time** for each subject's choices – a commonly used measure of effort. Finally, after the main experiment

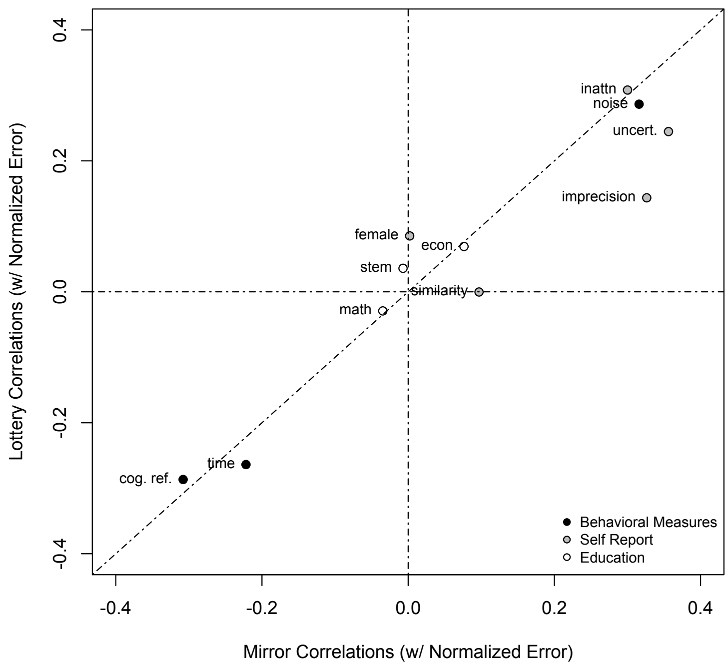


Figure 11: Correlates predictors (labeled dots) with pattern-consistent bias in mirrors (x-axis) and lotteries (y-axis). *Notes: Each dot is a different predictor and the x- and y-axes show the correlation of each predictor with pattern-consistent bias in mirrors and lotteries, respectively.*

we administered a three-question **cognitive reflection** test (Frederick (2005)), commonly used to measure how strongly subjects lean on intuitive vs. careful decision making.

Second, we administered several post-experiment questions that asked subjects to reflect on their choices. For instance, we we asked subjects how confident they were (in percentage terms) that they made the optimal choice in both lotteries and mirrors. Measures of **cognitive uncertainty** like this have proved predictive of the fourfold pattern and other anomalies in recent work (e.g., Enke & Graeber (2023), Enke et al. (2023)). We also asked subjects (separately for lotteries and mirrors) to report on a 100-point Likert scale how much attention they paid (0 for little attention, 100 for a lot of attention) to the number of boxes (i.e., to the probabilities) and to the dollar amounts (i.e., payoffs) when evaluating lotteries/mirrors. This gives us measures of **inattention** for each subject for both lotteries and mirrors. Likewise we asked subjects to use a 100-point Likert scale to estimate the degree to which they "guessed" (0) versus "made a precise (exact) decision" (100) in their valuations, again for both lotteries and mirrors. This gives us a self-reported measure of **imprecision** of decisions.

Third, we gathered several demographic measures, focused on measures that proxy for mathematical ability. We asked subjects to report their highest level of **math** education, coding subjects as 1 (relatively advanced mathematical training) if they had taken any college-level math and 0 otherwise. We asked a similar question about whether subjects had any college-level economics training. We also asked subjects their college major, coding them as **STEM** if they reported majoring in Science, Mathematics or Business. Finally, we asked for the subject's gender which is of interest because of debates in the literature about whether risk preferences are related to gender.

In Figure 11, we estimate the Pearson correlation between each measure and the mean error (normalized to be positive if in the direction of the classical pattern) in mirrors and lotteries, plotting the correlation coefficient ρ for (i) mirrors on the x-axis and (ii) lotteries on the y-axis.[12] We make several observations.

First and perhaps most importantly, there is a strikingly strong relationship between the correlates of the pattern in mirrors and lotteries. Correlation coefficients hover around the 45 degree line and there is a ρ =0.94 correlation between correlation estimates across the two valuation problems This relationship strongly reinforces our conclusion that the two types of behavior are driven by the same underlying behavioral mechanisms and that the driver of the pattern in lotteries is therefore likely rooted in the way people respond to disaggregation.

Result 11 *There is a strong similarity in the predictors of the classical pattern in lotteries and mirrors.*

Second, the strongest correlations are for variables that relate to the types of simpler-than-optimal decision procedures the literature has proposed as potential proximal mechanisms for the classical pattern. We find that (i) self-reported inattention is strongly positively correlated and (ii) correct responses in cognitive reflection test questions are strongly negatively correlated with pattern-consistent errors. This is potentially evidence in favor of the hypothesis that the classical pattern occurs because subjects use inattentive procedures, like those described by Bordalo et al. (2012). We also find that (i) noise in decision making and (ii) self-reported cognitive uncertainty are both strongly positively correlated with pattern-consistent errors. This is highly consistent with the hypothesis that the classical pattern occurs because subjects use imprecise strategies that produce cognitive noise (e.g., Blavatskyy 2007, Woodford 2012*a*, Steiner & Stewart 2016, Woodford 2020, Enke & Graeber 2023, Woodford 2012*b*, Khaw et al. 2022, Vieider 2023, Frydman & Jin 2023). Both types of accounts seem consistent with our finding that decision time is strongly negatively correlated with the classical pattern – potential evidence that such behavior is especially strong for subjects who expend less effort on the valuation task.

Third, by contrast, we find much weaker evidence that our other variables are very predictive of the classical pattern. Perhaps most importantly, we find little evidence linking the classical pattern to mathematical preparation. Prior mathematical or economic training and reporting majoring in a STEM training have, at best, weak predictive power in either lotteries or mirrors. This seems consistent with our finding that varying the arithmetic difficulty of calculating expected value has

[12] To reduce risks of attenuation, we pooled several of these measures. In particular, we averaged our post-experiment cognitive uncertainty, inattention and imprecision measures at the subject level and used a single pooled noise measure.

little effect on the classical pattern in our sample. This does not mean that there isn't a strong cognitive dimension to these findings, but rather that prior training in arithmetic calculation doesn't seem to be a major modulator of the effect.

We summarize the results of this correlational analysis as a further result:

Result 12 *The classical pattern is especially pronounced in subjects who (i) invest less time in valuation, (ii) report paying less attention to valuation, (iii) make mistakes on cognitive reflection test, (iv) make noisy or inconsistent decisions and (v) report cognitive uncertainty about the quality of their valuations.*

Together, these results seem to suggestthat subjects consciously (and perhaps deliberately) use hasty, casual, inattentive and imprecise strategies to value disaggregated objects like lotteries and mirrors and that this choice is an important driver of pattern-consistent errors.

A.6 Additional MPL data

We originally ran our main MPL treatment on Prolific in May of 2022 with 186 subjects. During the reviewing process, a helpful referee discovered a typo in one of the examples used to explain price lists in this treatment. In particular, the "Choosing a Set of Boxes" page of the instructions (reproduced in Section B.2.1 of this Appendix), in the second sentence of the second bullet point read "Set B has 50 boxes containint $10 and 50 boxes containing $0" instead of "Set B has 40 boxes containing $10 and 60 boxes containing $0" as it should have (and as the current instructions does). The referee raised legitimate concerns that this might have confused subjects and caused or intensified some of our findings.

For this reason, we re-ran the main MPL treatment (as described in the body of the paper) with this typo fixed, and report the results from the original run of the experiment in this Appendix. The original run of the experiment was nearly identical to the experiment that has taken its place in the main body of the paper. The only difference is that in the original version of the experiment we repeated tasks G50 and L50 for each subject, while in the revised the experiments we randomly selected two tasks to be repeated.

We present data from the original run of the experiment in Figures 12, 13 and 14, which mirror Figures 1, 2 and 4 for MPL in the main text. The results are virtually identical. As with the main dataset, we find in Figure We present data from the original run of the experiment in Figures 12 evidence of the fourfold pattern in both lotteries and mirrors. Summing up pattern-consistent choices, we find that the fourfold pattern is 97% as strong in mirrors as in lotteries. We also find evidence of loss aversion in both lotteries and mirrors, with loss aversion 64% as strong in the latter as in the former. This is somewhat smaller than in our main MPL sample and in our BDM treatment. Nonetheless, in each of our 12 tasks we find (as in the main dataset) that for the median subject the difference in valuations between lotteries and mirrors is 0. As Figure 14 shows, mirror

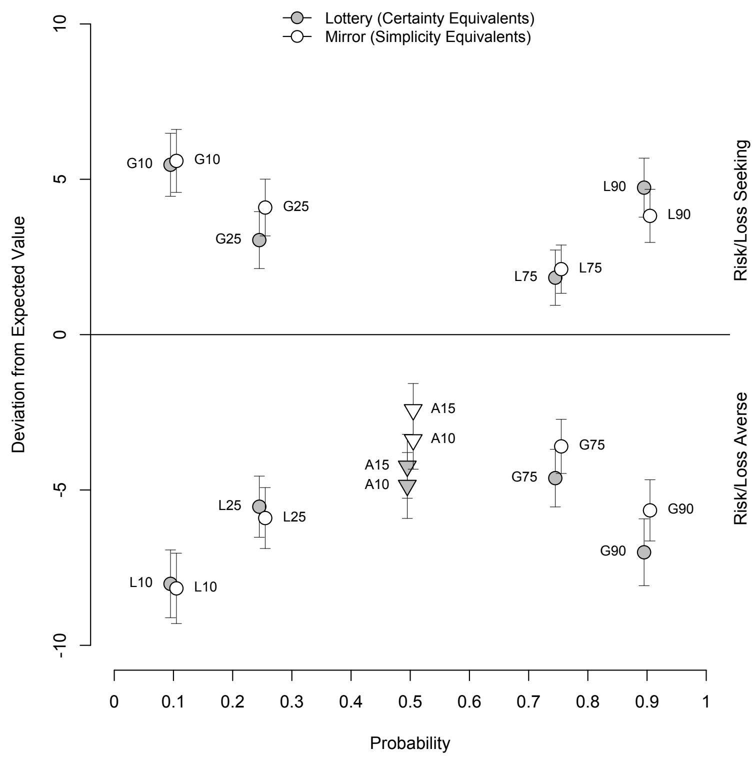


Figure 12: Mean deviations from expected value in lotteries (gray dots) and mirrors (hollow dots) in the original run of the main MPL treatment. *Notes: For fourfold lotteries, the y-axis measures the difference between subjects' certainty/simplicity equivalent and expected value, as stated in the axis label. The x-axis is the probability of the non-zero payoff. For loss aversion tasks, the y-axis measures instead the difference between the certain/simple payoff and the expected value of the mixed lottery/mirror. Two-standard-error bars are included for every task.*

and lottery deviations are strongly correlated. We find a correlation of 0.65 for absolute deviations and 0.59 for normalized deviations, closely matching results from the main MPL sample.

A.7 Analysis of G50 and L50

In Figure 15 we plot deviations from expected value for the 50/50 lotteries (G50 and L50) we included in all of our treatments, mirroring Figure 1 (recall, for space reasons we omitted these lotteries from the main Figure). These lotteries are potentially interesting because they seem particularly arithmetically easy to evaluation – they require math no more difficult than simple averaging (i.e. to calculate expected value). Evaluating behavior in these tasks provides another opportunity (alongside the analysis in Supplemental Appendix A.4) to examine to what degree our main results are rooted in the arithmetic difficulty of lottery/mirror valuation. Under the hypothesis that it is arithmetic difficulty that drives the pattern, we might expect errors to diminish in mirrors relative to lotteries in these easier problems, allowing lotteries to reveal true risk preferences.

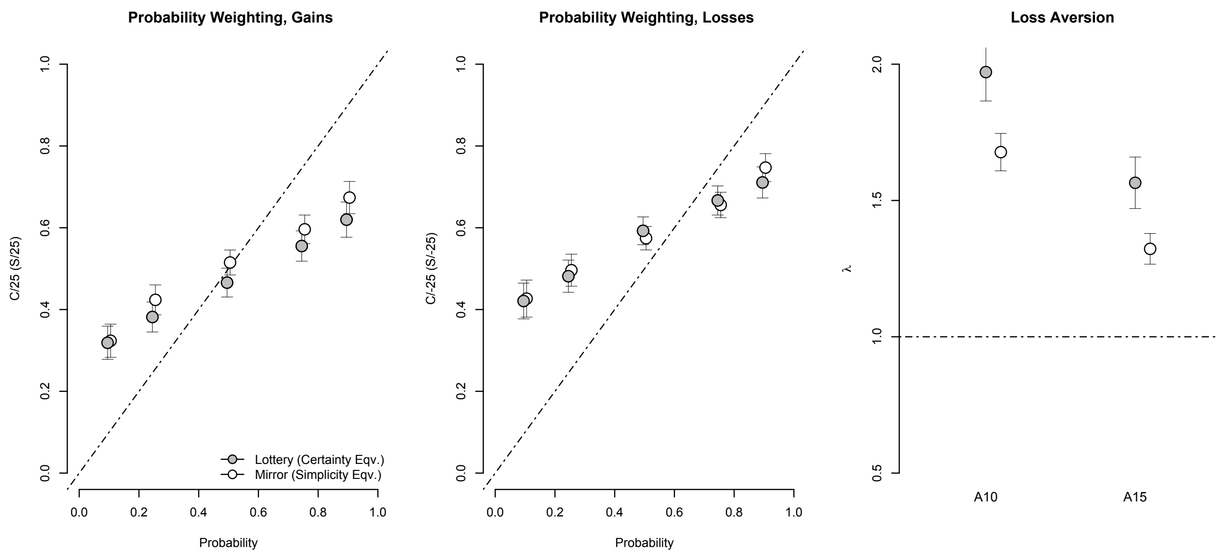


Figure 13: Naive visualization of the probability weighting functions (left two panels) and the loss aversion parameter, λ in the original run of the main MPL treatment. *Notes: The first two panels plot a naive estimate of the probability weighting function (following Tversky & Kahneman (1992)) by plotting the ratio of the certainty/simplicity equivalent to the non-zero payment amount as a function of the probability of the non-zero payoff amount. The final panel plots a naive estimate of* λ*, the standard linear parameter of loss aversion, under the assumption of a reference point of zero.*

However, we find little evidence of this. In most cases, lottery and mirror valuations are identical, and deviations are never systematically closer to expected value in mirrors than lotteries in any cases.[13]

A.8 Additional Tables and Figures

A.8.1 Correlations By Anomaly

Figure 16 repeats the analysis reported in the right-hand panels of Figure 4 separately for the fourfold pattern and loss aversion. In particular, the left hand panels plots mean bias measured in "fourfold lotteries" (G10, G25, G75, G90, L10, L25, L75, L90) for mirrors and lotteries (each dot, again, is an individual subject). In the right hand panel we do the same for biases from "loss aversion lists" (A10 and A15 or M10 and M15). In MPL, for fourfold lists (left hand panel) we measure a lottery-mirror correlation of $\rho = 0.63$ ($p < 0.001$) and for loss aversion (right hand panel) we measure $\rho = 0.50$ ($p < 0.001$). In BDM, for fourfold lists (left hand panel) we measure a lottery-mirror correlation of $\rho = 0.58$ ($p < 0.001$) and for loss aversion (right hand panel) we

[13] The lotteries G50 and L50 produce deviations in opposite directions in lotteries and mirrors in our main sample and student sample, which is the biggest difference we find. Clearly this isn't a very robust pattern as it doesn't show up in losses, BDM or 4-Box framings of the problem. What's more, in our main sample the deviations are more severe in mirrors than lotteries.

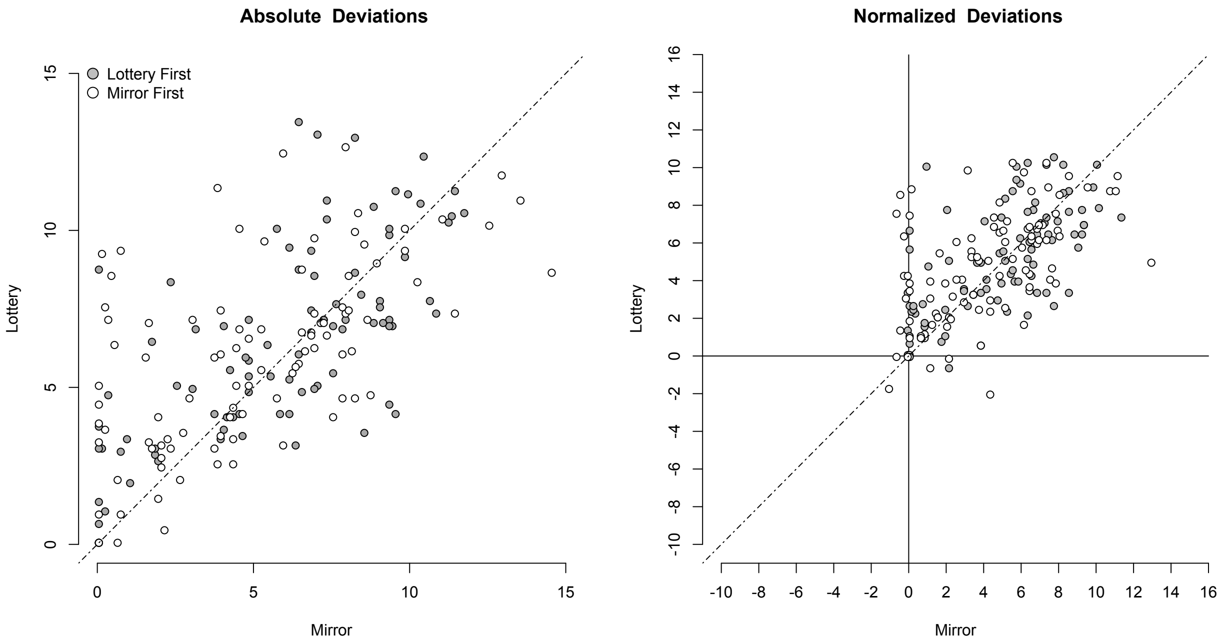


Figure 14: Deviations from expected value maximizing choices in mirrors (x-axis) versus lotteries (y-axis) in the original run of the main MPL treatment, by subject. *Notes: Each dot represents a separate subject. On the x-axes we plot the subject's data from the Mirror treatment and on the y-axes the same subject's data from the Lottery treatment. The left panel plots the mean absolute deviation from expected value. The right panel plots the mean deviation, normalized to be positive if it runs in the direction of the classical pattern. Gray dots are subjects who were assigned the Lottery treatment first, hollow dots subjects who were assigned the Mirror treatment first.*

measure $\rho = 0.616$ ($p < 0.001$).

A.8.2 Screenshots

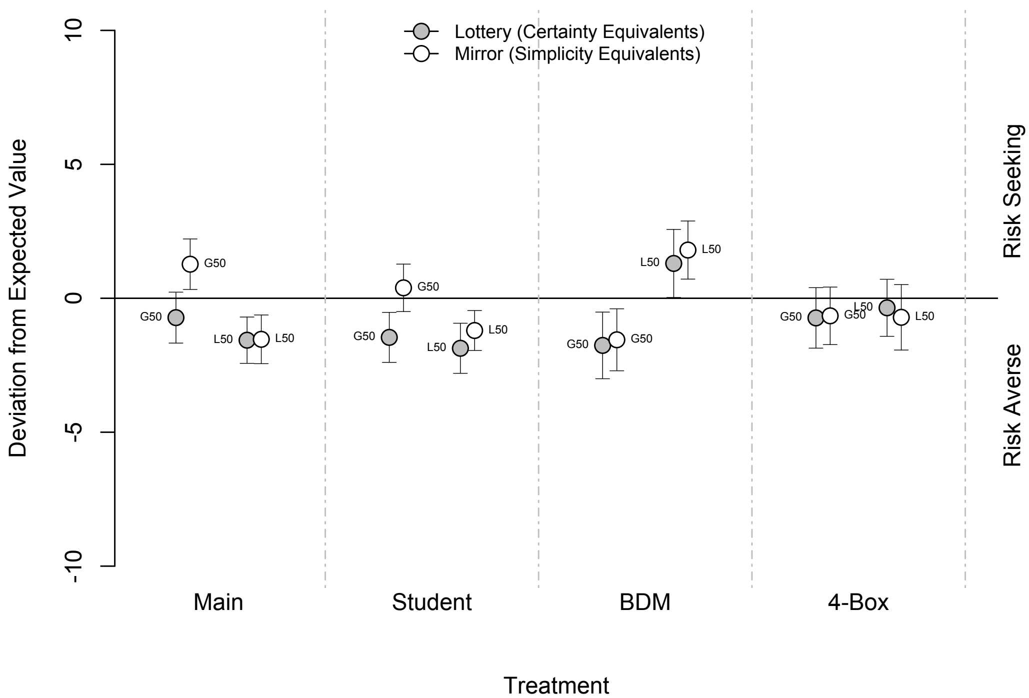


Figure 15: Mean deviations from expected value in 50/50 unmixed lotteries (gray dots) and mirrors (hollow dots). *Notes: The y-axis measures the difference between subjects' certainty/simplicity equivalent and expected value (as stated in the axis label). On the x-axis is experimental treatment.*

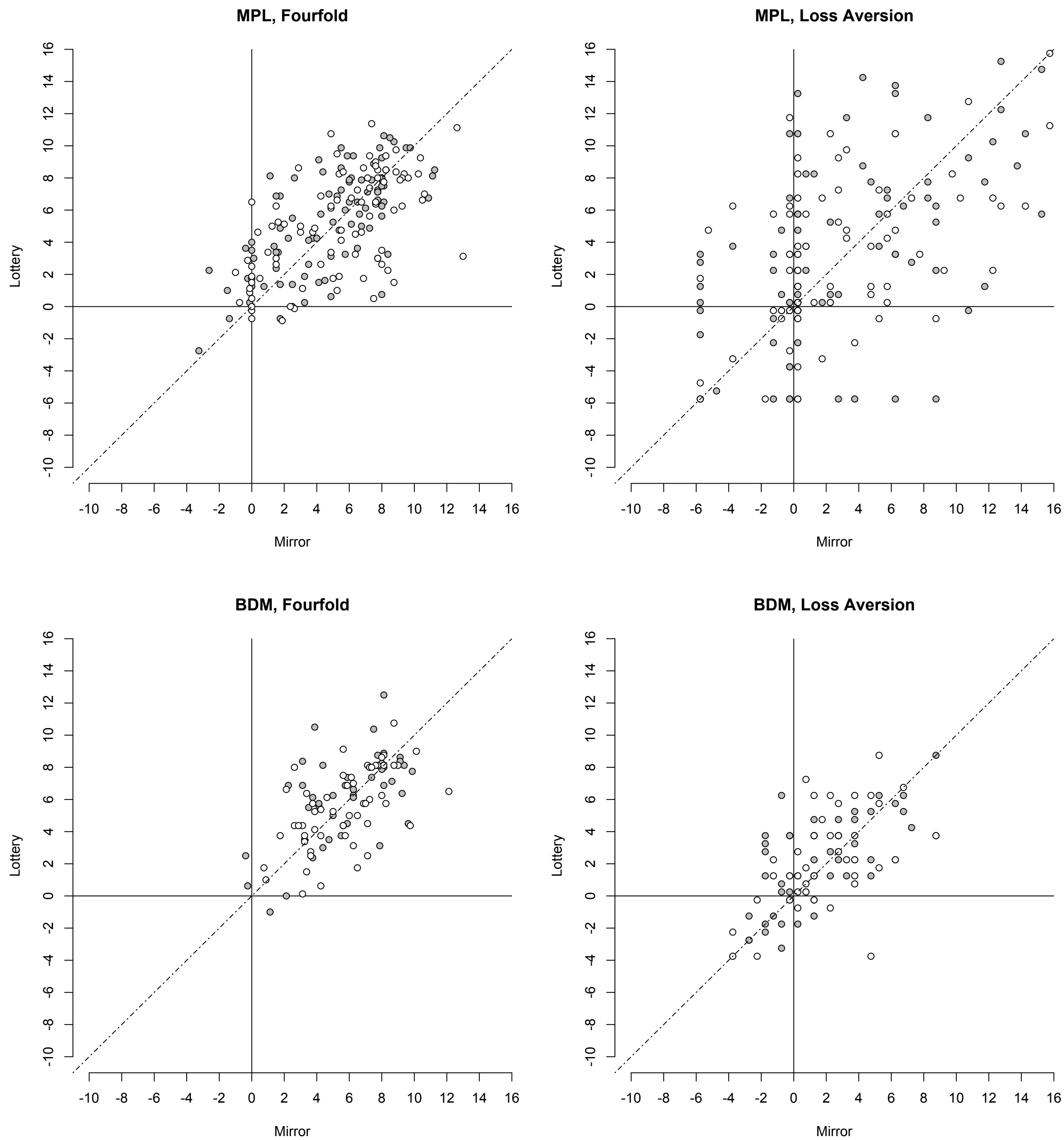


Figure 16: Deviations from expected value maximizing choices in mirrors (x-axis) versus lotteries (y-axis), by subject and pattern: the fourfold pattern (left panel) and loss aversion (right panel).
Notes: Each dot represents a subject. "Lottery First" designates subjects who were initially assigned lotteries (Mirror First is the reverse). The left panel plots mean "bias" (mean deviations normalized to be positive if they are in the direction of the classical pattern) for "fourfold lists" (G10, G25, G75, G90, L10, L25, L75, L90) while the right hand panel plots the same for "loss aversion lists" (A10 and A15 or M10 and M15).

Initial Money: $5.00

- Please select which Set (A or B) you'd prefer for each row of the table (each **version** of the problem) and click the Submit button.
- If this task is selected for payment, the computer will **randomly** select one row (one version) and use your choice in this row to determine your earnings.
- You will be paid $5 **plus** the value of **all** of the boxes from the Set you selected, **added up** and divided by 100.

	Set A	Set B	
Version	100 Boxes	10 Boxes	90 Boxes
1	$25.00	$25.00	$0.00
2	$24.00	$25.00	$0.00
3	$23.00	$25.00	$0.00
4	$22.00	$25.00	$0.00
5	$21.00	$25.00	$0.00
6	$20.00	$25.00	$0.00
7	$19.00	$25.00	$0.00
8	$18.00	$25.00	$0.00
9	$17.00	$25.00	$0.00
10	$16.00	$25.00	$0.00
11	$15.00	$25.00	$0.00
12	$14.00	$25.00	$0.00
13	$13.00	$25.00	$0.00
14	$12.00	$25.00	$0.00
15	$11.00	$25.00	$0.00
16	$10.00	$25.00	$0.00
17	$9.00	$25.00	$0.00
18	$8.00	$25.00	$0.00
19	$7.00	$25.00	$0.00
20	$6.00	$25.00	$0.00

Figure 17: Screenshot from a mirror task (list G10) under MPL. *Notes: In lottery tasks, the screen is identical except for the text in green which instead reads "...plus the value of one of the boxes from the Set you selected, randomly chosen by the computer."*

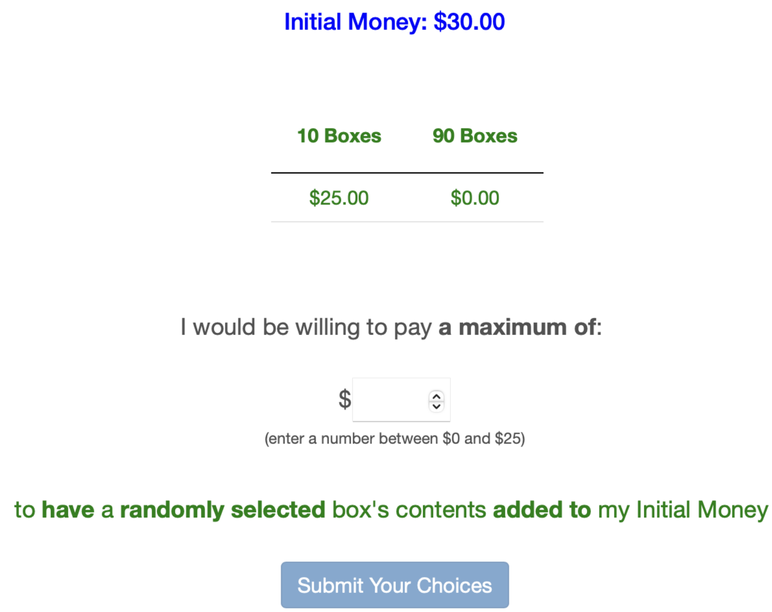


Figure 18: Screenshot from a lottery task (task G10) under BDM. *Notes: In mirror tasks, the screen is identical except for the text at the bottom which instead reads "...to have the average of these boxes' contents added to my Initial money.'*

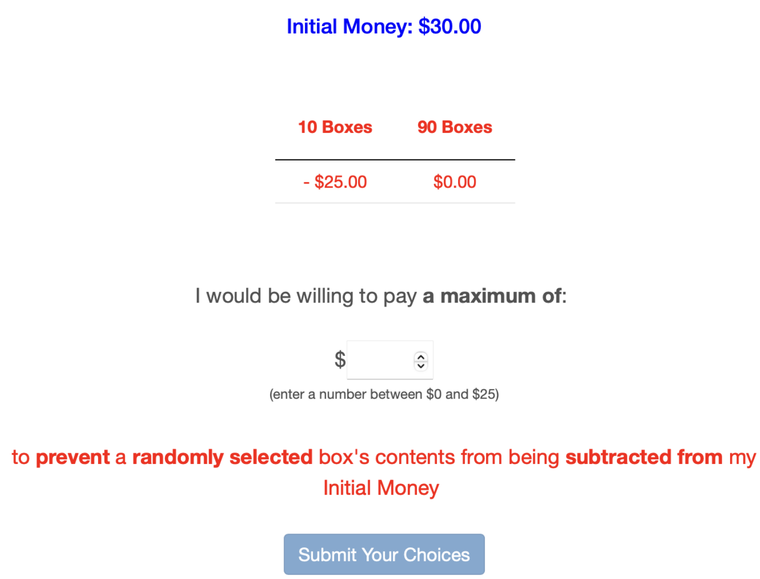


Figure 19: Screenshot from a lottery task (task L10) under BDM. *Notes: In mirror tasks, the screen is identical except for the text in green which instead reads "...to prevent the average of these boxes' contents from being subtracted from my Initial Money."*

B Instructions to Subjects

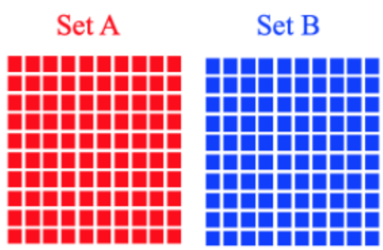
B.1 MPL Treatment

B.1.1 Beginning of Instructions

The first part of the instructions are given at the beginning of the session, regardless of whether subjects are assigned ,mirrors or lotteries first.

Boxes With Money

- In each of several tasks, we will give you an **INITIAL** sum of money.
- You will then choose which **set of BOXES** -- Set A (consisting of 100 boxes) or Set B (also consisting of 100 boxes) -- you would like the computer to open.



- Each box contains either a **POSITIVE** or **NEGATIVE** amount of money (or nothing). When the computer opens one or more boxes from your chosen set, the amount of money in the opened boxes will be added to (or subtracted from) your **INITIAL** money to determine your **FINAL EARNINGS**.

The Decision Table

- The two sets of boxes will be described in a **TABLE** like the one below. For each set, one or more counts of boxes (for instance 75, 25 or 100 boxes) are listed at the top, and the positive or negative amount of money in that number of boxes (for instance $20, $0, $7) is shown in the row of the Table.

Set A		Set B
75 Boxes	**25 Boxes**	**100 Boxes**
$20.00	$0.00	$7.00

- In the example above, Set A consists of **75** boxes each containing $20 and **25** boxes each containing $0. Set B consists of **100** boxes ALL of which contain $7.

- In the example below, Set A consists of **25** boxes with -$12 (negative $12) in each box and **75** boxes with $0 in each box. Set B consists of **100** boxes ALL of which contain -$3 (negative $3).

Set A		Set B
25 Boxes	**75 Boxes**	**100 Boxes**
- $12.00	$0.00	- $3.00

- Your job will be to click on the Table to decide which set of boxes (A or B) you would like the computer to pay you based on. Clicking on the Table will turn one of the sets yellow. Whichever set is highlighted in yellow will be selected by the computer to determine your **FINAL EARNINGS**.

Set A		Set B
75 Boxes	**25 Boxes**	**100 Boxes**
$20.00	$0.00	$7.00

- In the example above you have highlighted Set B and so will be paid based on that set.

B.1.2 Treatment Instructions

Next, one of the following two pages of instructions is given, depending on whether subjects are assigned mirrors or lotteries first. After subjects have completed making choices in the first treatment (Mirror or Lottery), they are given the *other* page from the Treatment Instructions, below.

A Random Box

- In the upcoming set of Tasks, the computer will **RANDOMLY** select one of the 100 boxes from whichever Set you've chosen (each box in the Set you chose is **EQUALLY** likely to be selected by the computer). If the amount in the box is positive, it will be **ADDED** to your initial money. If the amount is negative, it will be **SUBTRACTED** from your initial money.

- Example: In the example below, there are **100** boxes in each Set. For Set A, **50** boxes contain $16.00 and **50** of them contain $0.00. If you choose Set 1, there is therefore **50% chance** $16 will be added to your initial amount of money and a **50% chance** $0 will be added. For Set B all **100** boxes contain $4.00 so if you choose this Set, you have a **100% chance** of having $4 added to your initial money.

Set A		Set B
50 Boxes	50 Boxes	100 Boxes
$16.00	$0.00	$4.00

- Example: In the example below, there are also **100** boxes in each Set. For Set A, **50** boxes contain -$8.00 and **50** of them contain $0.00. If you choose Set 1, there is therefore a **50% chance** you will have $8 subtracted from your initial amount of money (you lose $8) and a **50% chance** you have $0 subtracted. For Set B all **100** boxes contain -$6.00 so if you choose this Set, you have a **100% chance** you will have $6 substracted from your initial money.

Set A		Set B
50 Boxes	50 Boxes	100 Boxes
- $8.00	$0.00	- $6.00

The Average Box

- In the upcoming tasks, the computer will pay you by calculating the **AVERAGE** amount of money across all 100 boxes for **whichever set you've chosen**. That is, it will add up the amount of money from each of the 100 boxes and divide that sum by 100. If the amount is positive, that amount will be **ADDED** to your initial money. If the amount is negative, it will be **SUBTRACTED** from your initial money.

- Example: In the example below, there are **100** boxes in each set. For Set A, **50** boxes contain $16.00 and **50** of them contain $0.00. If you choose Set A, the computer will therefore add (50×$16 + 50×$0)/100 = $8 to your initial amount of money. For Set B, all **100** boxes contain $4.00 so if you choose this set, you will add (100×$4)/100 = $4 to your initial money.

Set A		Set B
50 Boxes	50 Boxes	100 Boxes
$16.00	$0.00	$4.00

- Example: In the example below, there are also **100** boxes in each Set. For Set A, **50** boxes contain -$8.00 and **50** of them contain $0.00. If you choose Set A, the computer will pay you (-$8×50 + $0×50)/100 = -$4 for your choice; it will therefore subtract $4 from your initial amount of money (that is, you will lose $4). For Set B all **100** boxes contain -$6.00 so if you choose this set, the computer will pay you (-$6×100)/100 = -$6 for your choice. That is, you will have $6 subtracted from your initial money.

Set A		Set B
50 Boxes	50 Boxes	100 Boxes
- $8.00	$0.00	- $6.00

B.2 Comprehension Questions

Regardless of treatment, subjects are given 4 comprehension questions like the following which they must answer correctly before moving on. Crucially, although the questions are identical regardless of treatment, the correct answers to these questions depend on whether subjects are about to enter the Mirror or Lottery treatment. After subjects have completed the first treatment (Mirror or Lottery) and have read instructions for the next treatment, they are given the *same* 4 comprehension questions, now with different correct answers. This makes the difference between the payment schemes especially salient to subjects and is designed to prevent subjects from confusing payoffs in the two treatments.

Comprehension Questions

Set A		Set B
50 Boxes	50 Boxes	100 Boxes
$16.00	$0.00	$4.00

Suppose that the choice in the example above determines your payment, and you chose Set A.

- *What is the chance that $16 is added to your earnings?*

○ 0 in 100 (0%)
○ 50 in 100 (50%)
○ 100 in 100 (100%)

Submit Quiz

- *What is the chance that $8 is added to your earnings?*

○ 0 in 100 (0%)
○ 50 in 100 (50%)
○ 100 in 100 (100%)

Submit Quiz

- *What is the chance that $4 is added to your earnings?*

○ 0 in 100 (0%)
○ 50 in 100 (50%)
○ 100 in 100 (100%)

Submit Quiz

B.2.1 Final Part of Instructions

Choosing A Set of Boxes

- In the actual experiment, we will have you choose between between **MULTIPLE VERSIONS** of Set A and Set B. Each version will be shown as a **DIFFERENT ROW** of of the Table.

- Example: In the first row (**Version 1**) in the example below, Set A has **100** boxes containing $10 while Set B has **40** boxes containing $10 and 60 boxes containing $0. However in the second row (**Version 2**) is a different version in which Set A has **100** boxes containing $9, while Set B has **40** boxes containing $10 and **60** boxes containing $0. The other rows have other versions of Set A / Set B.

	Set A	Set B	
Version	100 Boxes	40 Boxes	60 Boxes
1	$10.00	$10.00	$0.00
2	$9.00	$10.00	$0.00
3	$8.00	$10.00	$0.00
4	$7.00	$10.00	$0.00
5	$6.00	$10.00	$0.00
6	$5.00	$10.00	$0.00
7	$4.00	$10.00	$0.00
8	$3.00	$10.00	$0.00
9	$2.00	$10.00	$0.00
10	$1.00	$10.00	$0.00

- You will make a choice for **EACH VERSION** of Set A / Set B by clicking on the Table and highlighting either Set A or Set B in each row of the Table.

	Set A	Set B	
Version	100 Boxes	40 Boxes	60 Boxes
1	$10.00	$10.00	$0.00
2	$9.00	$10.00	$0.00
3	$8.00	$10.00	$0.00
4	$7.00	$10.00	$0.00
5	$6.00	$10.00	$0.00
6	$5.00	$10.00	$0.00
7	$4.00	$10.00	$0.00
8	$3.00	$10.00	$0.00
9	$2.00	$10.00	$0.00
10	$1.00	$10.00	$0.00

- Example: In the example above, you selected Set A in Version 1, 2, 3, 4, 5, 6 and 7, and selected Set B in Version 8, 9 and 10.
- At the end of the experiment, the computer will randomly pick **ONE ROW** of the Table (one Version, with each row/version equally likely) and pay you based on your choice in that row. This means you should carefully consider your choice in **EACH ROW (EACH VERSION)** as any row/version could determine your payment.
- When you make your choices in the Table, the computer will put some limits on your choices. Specifically, you can only switch from choosing Set A to Set B at one point on the Table (though you are also welcome to choose Set A or only Set B in every row). You may click on the Table as many times as you like until you are happy with your choices. Then press the green button to finalize your choices.

Several Tables

- Over the course of the experiment, we will show you several Tables. Each Table has a different **initial amount of money** and **different Versions** displayed in rows. You must make a choice for each Version in every Table. At the end of the experiment the computer will **RANDOMLY** select **ONE** Table and then **RANDOMLY** select **ONE** Version (row) from that Table and determine your payment based on your choice in that Version.

  Since you do not know which choice will be selected, you should make each choice as if it alone determines your payment.

B.3 BDM Treatment

B.3.1 Beginning of Instructions

The first part of the instructions are given at the beginning of the session, regardless of whether subjects are assigned mirrors or lotteries first.

Boxes With Money

- In each of several tasks, we will give you an **INITIAL** sum of money.

- You will then evaluate a **set of 100 BOXES** which the computer may open to either increase or decrease this initial sum.

- Each box contains either a **POSITIVE** or **NEGATIVE** amount of money (or nothing). When the computer opens one or more boxes from a set, the amount of money in the opened boxes will be added to (or subtracted from) your **INITIAL** money to determine your **BONUS** (if you are randomly selected to be paid a bonus).

The Decision Table

- Each set of boxes will be described in a **TABLE** like the one below. For each set, one or more counts of boxes (for instance 75, 25 or 100 boxes) are listed at the top, and the positive or negative amount of money in that number of boxes (for instance $20, $0, $7) is shown in the row of the Table.

75 Boxes	25 Boxes
$20.00	$0.00

- In the example above, the set consists of **75** boxes each containing $20 and **25** boxes each containing $0.
- In the example below, the set consists of **25** boxes with -$12 (negative $12) in each box and **75** boxes with $0 in each box.

25 Boxes	75 Boxes
- $12.00	$0.00

- Depending on the task, your job will be to decide how much you'd be willing to pay to either **cause** the computer to open boxes from the set to modify your **BONUS** or **prevent** the computer from opening the boxes to modify your bonus.

B.3.2 Treatment Instructions

Next, one of the following two pages of instructions is given, depending on whether subjects are assigned mirrors or lotteries first. After subjects have completed making choices the first treatment (Mirror or Lottery), they are given the *other* page from the Treatment Instructions, below.

A Random Box

- In the upcoming tasks, if the computer opens boxes, it will pay you by **RANDOMLY** selecting one of the 100 boxes (each box in the set is **EQUALLY** likely to be selected by the computer). If the amount in the box is positive, it will be **ADDED** to your initial money. If the amount is negative, it will be **SUBTRACTED** from your initial money.

- Example: In the example below, there are **100** boxes in the set. For this set, **50** boxes contain $16.00 and **50** of them contain $0.00. If the computer opens the boxes, there is therefore **50% chance** $16 will be added to your initial amount of money and a **50% chance** $0 will be added.

50 Boxes	50 Boxes
$16.00	$0.00

- Example: In the example below, there are also **100** boxes in the set. For this set, **50** boxes contain -$8.00 and **50** of them contain $0.00. If the computer opens the boxes, there is therefore a **50% chance** you will have $8 subtracted from your initial amount of money (you lose $8) and a **50% chance** you have $0 subtracted.

50 Boxes	50 Boxes
- $8.00	$0.00

The Average Box

- In the upcoming tasks, if the computer opens boxes, it will pay you by calculating the **AVERAGE** amount of money across all 100 boxes. That is, it will add up the amount of money from each of the 100 boxes and divide that sum by 100. If the amount is positive, that amount will be **ADDED** to your initial money. If the amount is negative, it will be **SUBTRACTED** from your initial money.

- Example: In the example below, there are **100** boxes in the set. For this set, **50** boxes contain $16.00 and **50** of them contain $0.00. If the computer opens these boxes, it will therefore add (50×$16 + 50×$0)/100 = $8 to your initial amount of money.

50 Boxes	50 Boxes
$16.00	$0.00

- Example: In the example below, there are also **100** boxes in the set. In this set, **50** boxes contain -$8.00 and **50** of them contain $0.00. If the computer opens these boxes, the computer will pay you (-$8×50 + $0×50)/100 = -$4 for your choice; it will therefore subtract $4 from your initial amount of money (that is, you will lose $4).

50 Boxes	50 Boxes
- $8.00	$0.00

B.4 Comprehension Questions

We next gave subjects the same comprehension questions used in the MPL treatment.

B.4.1 BDM Mechanism

Paying <u>for</u> a Set of Boxes

- In the experiment, we will ask you **the maximum amount you would be willing to pay** either to **cause** or **prevent** the computer from opening the set of boxes on your screen to modify your Initial earnings.

- In some tasks (colored in green) we will show you a set that contains positive amounts of money, and ask you to tell us how many dollars you would (**at the very maximum**) be willing to pay to **cause** the computer to open boxes from the set to increase your earnings. Or, in other words, we will ask you how much do you think it is worth to you to have these boxes influence your earnings?

- <u>Example</u>: On your screen, we will show you a text box like the one below. Just enter the amount of money you think the set is worth to you (the maximum amount you'd be willing to pay for the set to be opened - the screen will give you the range you can enter):

I would be willing to pay **a maximum of**:

$

(enter a number between $0 and $25)

- To **reward you** for giving an **honest answer**, we are going to use a special set of rules to determine your payments in these tasks. We will randomly pick a **price** (equally likely between 0 and the maximum value you are allowed to enter) for the set of boxes (you won't know the price when you make your choice). If the amount you entered is **greater than or equal to** that random price, the computer will open the set of boxes on the screen to modify your Initial earnings as described **and** you will pay the amount of the random price (**not** the amount you entered) from your total earnings. If your maximum amount is less than the random price, the computer will not open the boxes on your screen and you will simply earn your initial amount (and you will not pay the random price).

- <u>Important</u>: If the computer uses boxes from the set to modify your earnings, you will not have to pay the maximum amount you enter, but instead will pay the random price. The maximum amount you enter just lets you tell us the range of random prices you are willing to pay for the set of boxes.

- If this sounds confusing, it is **actually very simple**. We've designed the payments so it is in your best interest to **tell us honestly** the most you would be willing to pay to have the set of boxes opened to influence your bonus. So just think about how much at a maximum you'd be willing to give up to have the computer modify your bonus based on the set of boxes on your screen, and enter this amount truthfully.

Paying to <u>Avoid</u> a Set of Boxes

- In other tasks (colored in red) we will show you a set that contains negative amounts of money, and ask you to tell us how many dollars you would (**at the very maximum**) be willing to pay to **prevent** the computer from opening boxes from the set to decrease your earnings. Or, in other words, we will ask you how much do you think it is worth to you to prevent these boxes from influencing your earnings?

- <u>Example</u>: On your screen, we will, again, show you a text box like the one below. Just enter the amount of money you think it is worth to <u>prevent</u> the computer from using that set to modify your bonus (the maximum amount you'd be willing to pay to prevent it - the screen will give you the range you can enter):

I would be willing to pay **a maximum of**:

$

(enter a number between $0 and $25)

- To **reward you** for giving an **honest answer**, we are, again, going to use a special set of rules to determine your payments in these tasks. We will randomly pick a **price** (equally likely between 0 and the maximum value you are allowed to enter) required to **prevent** the set of boxes from influencing your earnings (you won't know the price when you make your choice). If the amount you entered is **greater than or equal to** that random price, the computer will **not open** the set of boxes on the screen to modify your Initial earnings as described **and** you will pay the amount of the random price (**not** the amount you entered) from your total earnings. If your maximum amount is less than the random price, the computer **will** open the set of boxes on your screen to modify your initial earnings (but you will not pay the random price).

- <u>Important</u>: If you prevent the computer from opening boxes from the set, you will not have to pay the maximum amount you enter, but instead will pay the random price. The maximum amount you enter just lets you tell us the range of random prices you are willing to pay to avoid the set of boxes.

- If this sounds confusing, it is **actually very simple**. We've designed the payments so it is in your best interest to **tell us honestly** the most you would be willing to pay to prevent the set of boxes from being opened to influence your bonus. So just think about how much at a maximum you'd be willing to give up to prevent the computer from modifying your bonus based on the set of boxes on your screen, and enter this amount truthfully.

B.4.2 Final Part of Instructions

Several Sets of Boxes

- Over the course of the experiment, we will show you several sets of boxes. Each may have a different **initial amount of money** and **different** amounts of money distributed across the boxes.

- Important:. Make sure you pay attention to the type of question we are asking in each task. In some tasks colored in green we are asking you to tell us how much you'd be willing to pay to **cause** the boxes to influence your earnings. In other tasks negative we are asking you to tell us how much you'd be willing to pay to **prevent** the boxes from influencing your earnings.

- **One out of five (1/5 of)** participants will be randomly selected by the computer to be paid a **BONUS** based on their choices. If you are one of these participants, at the end of the experiment the computer will **RANDOMLY** select **ONE** Task and then **RANDOMLY** select a **PRICE** to determine your payment based on how much you said you're willing to pay.

  Since you do not know which choice will be selected, you should make each choice as if it alone determines your payment.