American Economic Review 2024, 114(12): 3789–3811
https://doi.org/10.1257/aer.20221227

Decisions under Risk Are Decisions under Complexity†

By Ryan Oprea*

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 predict 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. (*JEL* C91, D44, D81, D91)

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]

* Economics Department, University of California, Santa Barbara (email: roprea@gmail.com). Pietro Ortoleva was the coeditor for this article. 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, MIT, 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 under protocol 11-22-0691.

† Go to https://doi.org/10.1257/aer.20221227 to visit the article page for additional materials and author disclosure statement(s).

[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, i.e., that choices maximize overall welfare, given constraints such as information-processing costs.

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 probabilities 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 and Tversky 1979; Tversky and 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 \$25 with probability 0.9 (as a lottery does), the mirror of G90 pays $0.9 \times \$25 = \$22.50$ 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

(i) 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.

(ii) 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.

(iii) 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 nonstandard 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 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 I we describe our experimental design, and in Section II we report our results. In Section III we offer our interpretation of our findings, and in Section IV we discuss how our work connects to the prior literature. In Section V we conclude.

I. Experimental Design

In our experiment, we elicit valuations for a set of 12 lotteries (discussed in Section IA) known to produce some of the key valuation anomalies in the behavioral literature on risky choice. 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 percent chance.

[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 misvaluations. See Oprea (2020) for direct evidence that procedures produce complexity costs of this sort; Banovetz and 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.

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 1 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 subjects were assigned the 12 mirrors versus 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 online 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 selects 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 IB), 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 midpoint 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 online Appendix A.3 we report a robustness treatment in which we elicit values using the Becker-Degroot-Marschak (BDM) method (Becker, DeGroot, and Marschak 1964) instead. Screenshots are provided in online 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.

A. *Lotteries*

Eight of the 12 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 toward 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 zero, allowing for cleaner measurement (Hershey and Schoemaker 1985; Sprenger 2015). We include tasks with $\$Y$ equal to -10 and -15, giving us tasks A10 and A15, respectively.[3]

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

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 online 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.

B. *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 percent 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 online Appendix A.5, and the results are discussed in Section IIC. In addition, we collected data from robustness treatments in which we (i) elicited values using the BDM mechanism instead of MPLs ($N = 100$), (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 IIC and online Appendix A.

C. *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

[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 rerun the experiment. Data from that earlier run of the main design are reported in online Appendix A.6; the results are nearly identical to those from our main design.

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 is 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 and Tversky 1979; Tversky and 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 a reference point (e.g., 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.

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]

II. Main Results

In Figure 1, we plot raw data from the experiment. On the x-axis we plot the probability of the lottery's nonzero 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

[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 online Appendix A.7.

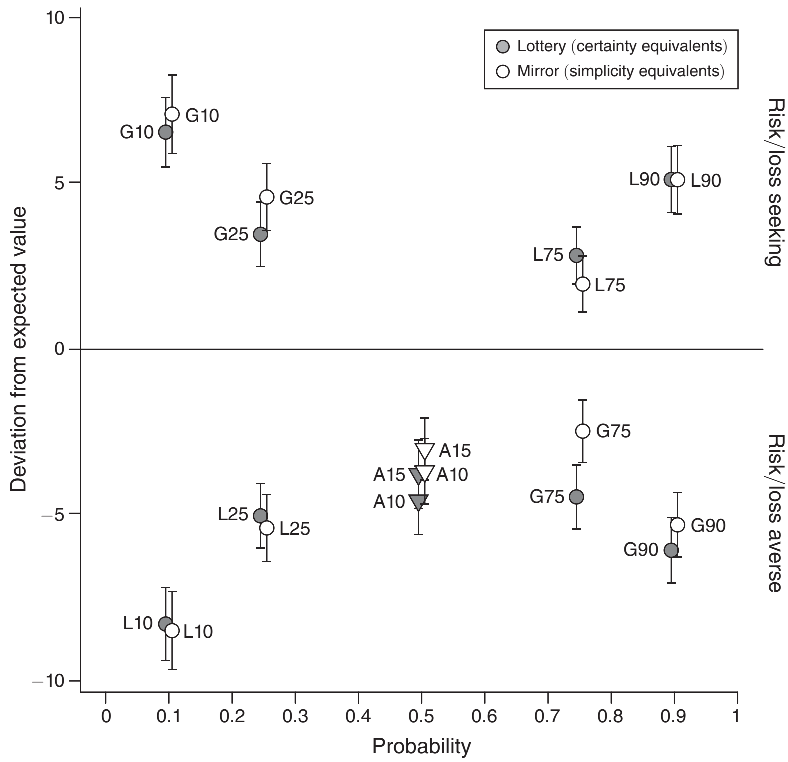

Figure 1. Mean Deviations from Expected Value in Lotteries (Solid Gray Dots) and Mirrors (Hollow Dots)

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 nonzero 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.

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 5 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 5 percent level using Wilcoxon tests. In the left two panels of Figure 2, we replot these same data

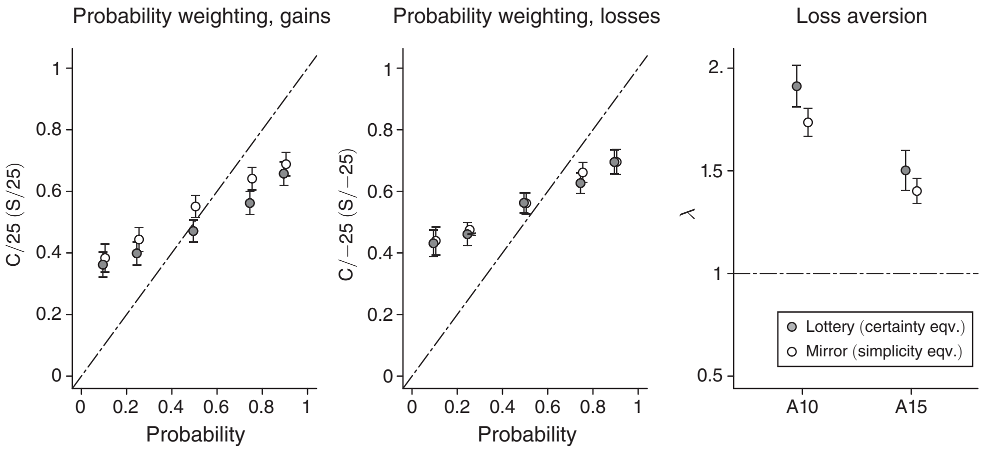

Figure 2. Naïve Visualization of the Probability Weighting Functions (Left Two Panels) and the Loss Aversion Parameter, λ

Notes: The first two panels plot a naïve estimate of the probability weighting function (following Tversky and Kahneman 1992) by plotting the ratio of the certainty/simplicity equivalent to the nonzero payment amount as a function of the probability of the nonzero payoff amount. The final panel plots a naïve estimate of λ, the standard linear parameter of loss aversion, under the assumption of a reference point of zero.

under the lens of probability weighting, following standard conventions from the literature (Tversky and Kahneman 1992). In these panels we plot (i) the probability of the nonzero payment on the x-axis and (ii) the ratio of the certainty/simplicity equivalent to the nonzero payment of the lottery on the y-axis. This provides a naïve 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 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 rightmost 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 toward 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 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 unmixed 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.*

A. *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 0 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 percent level in 9 out of our 10 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

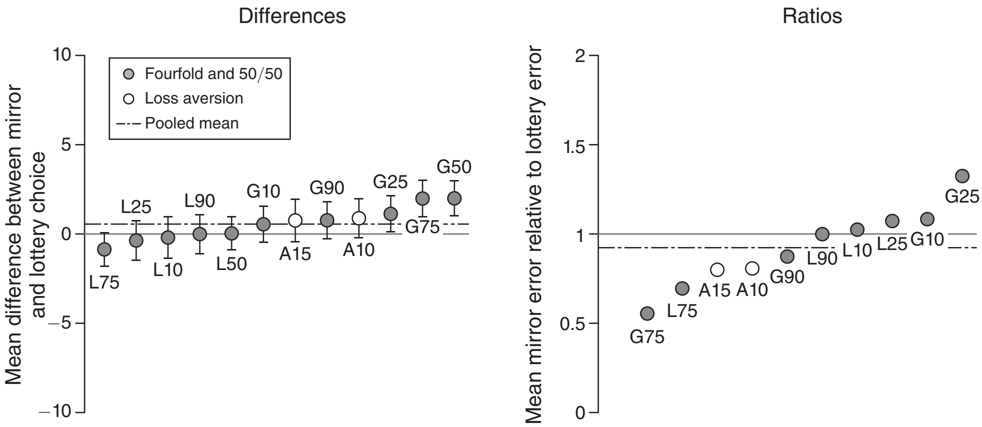

FIGURE 3. COMPARISONS BETWEEN LOTTERY AND MIRROR TASKS

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.

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 percent across tasks, suggesting that, overall, the fourfold pattern is about 97 percent as strong in mirrors as in lotteries. For our two loss aversion tasks, the mean ratio is 80 percent, 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 are, 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 dataset 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 percent as strong) in mirrors than it is in lotteries.[7]

[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

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.*

B. *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 panel 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 online 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 online 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 underestimated due to attenuation bias (Gillen, Snowberg, and Yariv 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

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.

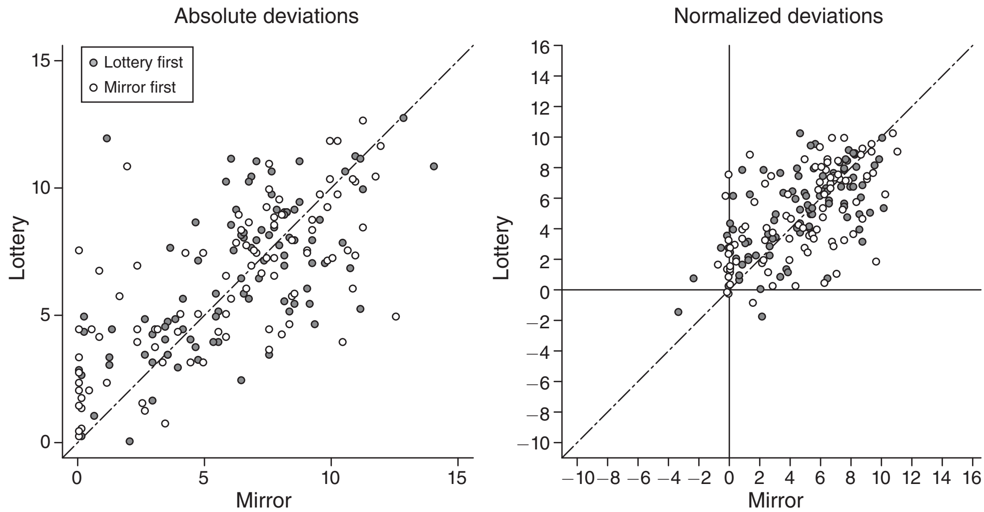

FIGURE 4. DEVIATIONS FROM EXPECTED VALUE–MAXIMIZING CHOICES IN MIRRORS (*x-axis*) VERSUS LOTTERIES (*y-axis*), BY SUBJECT

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.

to report whether they used completely/mostly different strategies in lotteries and mirrors or identical/mostly similar strategies; 75 percent 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 percent of subjects can be classified this way, exhibiting the pattern in lotteries but not in mirrors. Most subjects therefore deviate from the most basic prediction of risk preference–based theories. By contrast, the vast majority of subjects (82 percent) can be classified as complexity sensitive, showing clear evidence of the classical pattern even in the absence of risk. Indeed, the modal subject (59 percent 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 percent 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 IV, the latter interpretation is consistent with Martinez-Marquina, Niederle, and Vespa (2019), who provide evidence that already-complex problems become more complex in the presence of risk.

C. *Robustness and Additional Evidence*

In the online 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 online 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 versus Mirror treatments, we find virtually identical results.

Second, in online Appendix A.3, we show that these results are not a special outgrowth of our use of multiple price lists but are instead a broader phenomenon of valuation. There, we report a variation on our main design using the BDM mechanism (Becker, DeGroot, and Marschak 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 IV), we find that the "shape" of these anomalies changes 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 versus MPL) has a far larger effect on behavior than does varying the objective function itself (lottery versus 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 online Appendix A.4 we provide evidence that these results are not due to simple math errors or an aversion to mathematical difficulty. 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 percent 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 online Appendix A.2, we show that these results are slightly weakened but are 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 percent as strong and loss aversion 54 percent 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 online 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 and 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 the pattern 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 online 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.

III. 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 IV for a caution concerning this interpretation). Likewise, our data suggest 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, Niederle, and Vespa 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 toward prior beliefs in a Bayesian manner, this can produce the fourfold pattern (Steiner and Stewart 2016; Khaw, Li, and Woodford 2022; Vieider 2023; Frydman and Jin 2023) and even loss aversion (Woodford 2012) under some assumptions. Enke and 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, Chater, and Brown 2006). Another literature shows that the pattern can arise instead from *inattentive valuation strategies*. Bordalo, Gennaioli, and Shleifer (2012) show that if decision-makers put excess weight on salient components of lotteries when valuing them, this can generate the fourfold pattern; given widespread evidence in psychology on the greater salience of negative relative to

[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, Slovic, and Zink 1969; Montgomery and 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.

positive information (Baumeister et al. 2001; Rozin and Royzman 2001), it is easy to see how a similar mechanism might produce apparent loss aversion as well (e.g., Bhatia and 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 IIC (and online 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 and Graeber 2023; Frydman and Jin 2023; Vieider 2023; Khaw, Li, and Woodford 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, Bohren, and Imas 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.

IV. Connections to the Literature

Methodologically, the closest paper to ours is Martinez-Marquina, Niederle, and Vespa (2019), who, like us, compare behavior in risky and deterministic versions of the same 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, Niederle, and Vespa (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 decision 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 20 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, Niederle, and Vespa (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, Niederle, and Vespa (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, Brown, and Shapiro 2013; Choi et al. 2021), cognitive load (Benjamin, Brown, and Shapiro 2013; Deck and Jahedi 2015; Gerhardt et al. 2016), choice inconsistency (Khaw, Li, and Woodford 2021; Enke and Graeber 2023), and inattention (Pachur et al. 2018), all of which suggest that many such departures may be driven by judgment 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 and Weizsacker 1999; Bernheim and Sprenger 2020; Puri 2023); Enke and Shubatt (2023) identify a number of lottery characteristics that predict subjects' failures to correctly infer the expected value of lotteries and show 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, Hermann, and Musshof 2018; Harbaugh, Krause, and Vesterlund 2010; Holzmeister and 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 and Haruvy 2017; Charness, Chemaya, and Trujano-Ochoa 2023), probability weighting (Van de Kuilen 2009), and the Allais paradox (Van de Kuilen and 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 and Rehbeck 2022; Benjamin, Fontana, and Kimball 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 and Tversky 1979; Tversky and 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-relevant tastes for risk and loss or whether it instead describes judgment 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 insensitive to features of decision problems that matter for optimal choice. The classical pattern can be interpreted, in large part, as an outgrowth of just this sort of insensitivity, an observation that goes back at least to Tversky and Kahneman (1992). Enke and 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, Bohren, and Imas (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 and Jäger (2015) show that increasing the complexity of taxes results in insensitivities to marginal tax rates. Enke, Graeber, and Oprea (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.

V. 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 accommodated 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, Johannes, and Simon Jäger. 2015. "Complex Tax Incentives." *American Economic Journal: Economic Policy* 7 (3): 1–28.

Ba, Cuimin, J. Aislinn Bohren, and Alex Imas. 2023. "Over- and Underreaction to Information." Unpublished.

Banovetz, James, and Ryan Oprea. 2022. "Complexity and Procedural Choice." Unpublished.

Barberis, Nicholas C. 2013. "Thirty Years of Prospect Theory in Economics: A Review and Assessment." *Journal of Economic Perspectives* 27 (1): 173–96.

Bauermeister, Golo-Friedrich, Daniel Hermann, and Oliver Musshof. 2018. "Consistency of Determined Risk Attitudes and Probability Weightings Across Different Elicitation Methods." *Theory and Decision* 84 (4): 627–44.

Baumeister, R. F., E. Bratslavsky, C. Finkenauer, and K. D. Vohs. 2001. "Bad is Stronger than Good." *Review of General Psychology* 5 (4): 323–70.

Beauchamp, Jonathan P., Daniel J. Benjamin, David I. Laibson, and Christopher F. Chabris. 2020. "Measuring and Controlling for the Compromise Effect when Estimating Risk Preference Parameters." *Experimental Economics* 23 (4): 1069–99.

Becker, Gordon M., Morris H. DeGroot, and Jacob Marschak. 1964. "Measuring Utility by a Single-Response Sequential Method." *Behavioral Science* 9 (3): 226–32.

Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro. 2013. "Who is 'Behavioral'? Cognitive Ability and Anomalous Preferences." *Journal of the European Economic Association* 11 (6): 1231–55.

Benjamin, Daniel J., Mark Alan Fontana, and Miles S. Kimball. 2023. "Reconsidering Risk Aversion." Unpublished.

Bernheim, B. Douglas, and Charles Sprenger. 2020. "On the Empirical Validity of Cumulative Prospect Theory: Experimental Evidence of Rank-Independent Probability Weighting." *Econometrica* 88 (4): 1363–1409.

Bhatia, Sudeep, and Russell Golman. 2019. "Attention and Reference Dependence." *Decision* 6 (2): 145–70.

Blavatskyy, Pavlo. 2007. "Stochastic Expected Utility Theory." *Journal of Risk and Uncertainty* 34: 259–86.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2012. "Salience Theory of Choice under Risk." *Quarterly Journal of Economics* 127 (3): 1243–85.

Camara, Modibo. 2023. "Computationally Tractable Choice." Unpublished.

Chapman, Jonathan, Erik Snowberg, Stephanie W. Wang, and Colin Camerer. 2022. "Looming Large or Seeming Small? Attitudes Towards Losses in a Representative Sample." Unpublished.

Charness, Gary, Nir Chemaya, and Dario Trujano-Ochoa. 2023. "Learning Your Own Risk Preferences." Unpublished.

Choi, Syngjoo, Jeongbin Kim, Eungik Lee, and Jungmin Lee. 2022. "Probability Weighting and Cognitive Ability." *Management Science* 68 (7): 5201–15.

Deck, Cary, and Salar Jahedi. 2015. "The Effect of Cognitive Load on Economic Decision Making: A Survey and New Experiments." *European Economic Review* 78: 97–119.

Enke, Benjamin, and Thomas Graeber. 2023. "Cognitive Uncertainty." *Quarterly Journal of Economics* 138 (4): 2021–67.

Enke, Benjamin, Thomas Graeber, and Ryan Oprea. 2023. "Complexity and Time." Unpublished.

Enke, Benjamin, and Cassidy Shubatt. 2023. "Quantifying Lottery Choice Complexity." Unpublished.

Ert, Eyal, and Ernan Haruvy. 2017. "Revisiting Risk Aversion: Can Risk Preferences Change with Experience." *Economic Letters* 151: 91–95.

Frederick, Shane. 2005. "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives* 19 (4): 25–42.

Friedman, Daniel. 1989. "The S-Shaped Value Function as a Constrained Optimum." *American Economic Review* 79 (5): 1243–48.

Friedman, Daniel, Sameh Habib, Duncan James, and Brett Williams. 2022. "Varieties of Risk Preference Elicitation." *Games and Economic Behavior* 133: 58–76.

Friedman, Daniel, R. Mark Isaac, Duncan James, and Shyam Sunder. 2017. *Risky Curves: On The Empirical Failures of Expected Utility*. London: Routledge.

Frydman, Cary, and Lawrence J. Jin. 2023. "On the Source and Instability of Probability Weighting." Unpublished.

Fudenberg, Drew, and Indira Puri. 2022. "Evaluating and Extending Theories of Choice Under Risk." Unpublished.

Gerhardt, Holger, Guido P. Biele, Hauke R. Heekeren, and Harald Uhlig. 2016. "Cognitive Load Increases Risk Aversion." Unpublished.

Gillen, Ben, Erik Snowberg, and Leeat Yariv. 2019. "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study." *Journal of Political Economy* 127 (4): 1826–63.

Glimcher, Paul. 2022. "Efficiently Irrational: Illuminating the Riddle of Human Choice." *Trends in Cognitive Science* 26 (8): 669–87.

Harbaugh, William T., Kate Krause, and Lise Vesterlund. 2010. "The Fourfold Pattern of Risk Attitudes in Choice and Pricing Tasks." *Economic Journal* 120 (545): 595–611.

Hershey, John, and Paul Schoemaker. 1985. "Probability versus Certainty Equivalence Methods in Utility Measurement: Are They Equivalent?" *Management Science* 31 (10): 1213–31.

Hertwig, Ralph, Greg Barron, Elke Weber, and Ido Erev. 2004. "Decisions from Experience and the Effect of Rare Events in Risky Choice." *Psychological Science* 15 (8): 534–39.

Holzmeister, Felix, and Matthias Stefan. 2021. "The Risk Elicitation Puzzle Revisited: Across Methods (In)Consistency?" *Experimental Economics* 24: 593–616.

Huck, Steffen, and Georg Weizsäcker. 1999. "Risk, Complexity and Deviations from Expected-Value Maximization: Results of a Lottery Choice Experiment." *Journal of Economic Psychology* 149 (9): 1644–83.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47 (2): 263–91.

Khaw, Mel Win, Ziang Li, and Michael Woodford. 2021. "Cognitive Imprecision and Small-Stakes Risk Aversion." *Review of Economic Studies* 88 (4): 1979–2013.

Khaw, Mel Win, Ziang Li, and Michael Woodford. 2022. "Cognitive Imprecision and Stake-Dependent Risk Attitudes." Unpublished.

Lichtenstein, Sarah, Paul Slovic, and Donald Zink. 1969. "Effect of Instruction in Expected Value on Optimality of Gambling Decisions." *Journal of Experimental Psychology* 79 (2): 236–40.

Martinez-Marquina, Alejandro, Muriel Niederle, and Emanuel Vespa. 2019. "Failures in Contingent Reasoning: The Role of Uncertainty." *American Economic Review* 109 (10): 3437–74.

Montgomery, Henry, and Thomas Adelbratt. 1982. "Gambling Decisions and Information about Expected Value." *Organizational Behavior and Human Performance* 29 (1): 39–57.

Nielsen, Kirby, and John Rehbeck. 2022. "When Choices are Mistakes." *American Economic Review* 112 (7): 2237–68.

Oprea, Ryan D. 2020. "What Makes a Rule Complex." *American Economic Review* 110 (12): 3913–51.

Oprea, Ryan. 2024. *Data and Code for: "Decisions under Risk Are Decisions under Complexity."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E208268V1.

Pachur, Thorsten, Michael Schulte-Mecklenbeck, Ryan O. Murphy, and Ralph Hertwig. 2018. "Prospect Theory Reflects Selective Allocation of Attention." *Journal of Experimental Psychology: General* 147 (2): 147–69.

Puri, Indira. 2023. "Simplicity and Risk." Unpublished.

Rozin, Paul, and Edward B. Royzman. 2001. "Negativity Bias, Negativity Dominance, and Contagion." *Personality and Social Psychology Review* 5 (4): 296–320.

Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics* 69 (1): 99–118.

Sprenger, Charles. 2015. "An Endowment Effect for Risk: Experimental Tests of Stochastic Reference Points." *Journal of Political Economy* 123 (6): 1456–99.

Steiner, Jakub, and Colin Stewart. 2016. "Perceiving Prospects Properly." *American Economic Review* 106 (7): 1601–31.

Stewart, Neil, Nick Chater, and Gordon D. Brown. 2006. "Decision by Sampling." *Cognitive Psychology* 53 (1): 1–26.

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty* 5 (4): 297–323.

Van de Kuilen, Gijs. 2009. "Subjective Probability Weighting and the Discovered Preference Hypothesis." *Theory and Decision* 67 (1): 1–22.

Van de Kuilen, Gjis, and Peter P. Wakker. 2006. "Learning in the Allais Paradox." *Theory and Decision* 33 (3): 155–64.

Vieider, Ferdinand. 2023. "Decisions under Uncertainty as Bayesian Inference on Choice Options." Unpublished.

Wakker, Peter P. 2010. *Prospect Theory: For Risk and Ambiguity*. Cambridge: Cambridge University Press.

Wakker, Peter, and Daniel Deneffe. 1996. "Eliciting von Neumann-Morgenstern Utilities when Probabilities are Distorted or Unknown." *Management Science* 42 (8): 1131–50.

Woodford, Michael. 2012. "Prospect Theory as Efficient Perceptual Distortion." *American Economic Review Papers and Proceedings* 102 (3): 41–46.

Woodford, Michael. 2020. "Modeling Imprecision Perception, Valuation and Choice." *Annual Review of Economics* 12: 579–601.