Simplicity Equivalents *

Ryan Oprea†

September 16, 2022

Abstract

We provide evidence that the signature empirical patterns of prospect theory are not special phenomena of risk. They also arise (and often with equal strength) when subjects evaluate deterministic 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 loss. Across subjects, the appearance of anomalies in these deterministic tasks strongly predicts their appearance in lotteries. These findings suggest that much of the behavior described by prospect theory may be driven by the complexity of evaluating lotteries rather than by risk or risk preferences.

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

JEL codes: C91, D91 G0,

*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 Reese-Jones, Dmitry Taubinsky, Ferdinand Vieider and Emanuel Vespa for valuable conversations. Dario Ochoa and Ravi Vora provided excellent research assistance. I am also grateful to audiences at the Cognitive Noise Workshop (Harvard), BRIQ, Harvard University, the Stanford-Ohio State Experimental Workshop, D-TEA Paris, SAET, the HKBU-NTU-Osaka-Kyoto-Sinica Theory Seminar and NYU. 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 provide evidence that the main empirical patterns motivating behavioral theories of risk like prospect theory are not in fact special phenomena of risk. They also arise (and often with equal strength) when we ask experimental subjects to value deterministic monetary payments that we describe in the same disaggregated way that we necessarily describe lotteries. Thus, we find "probability weighting" in valuation problems without probabilities and "loss aversion" in problems without risk of loss, and, across subjects, the appearance of these anomalies in deterministic valuations predicts their appearance in the valuation of true lotteries. Our findings therefore suggest that these anomalies may be, in large part, a response to the complexity of lotteries (i.e., the information processing required to value lotteries) rather than a response to their risk (e.g., risk preferences or failures to understand stochasticity).

Our study is centered on a suite of anomalies that researchers typically find in the *certainty equivalents* of lotteries – the deterministic monetary amounts subjects assess as equivalently valuable to lotteries. The "fourfold pattern" of risk, the "most distinctive implication of prospect theory" (Tversky & Kahneman 1992), is the tendency for certainty equivalents to be (i) lower than expected value (revealing risk aversion) for lotteries with high probabilities of gain, but (ii) higher than expected value (revealing risk seeking) for lotteries with low probabilities of gain and (iii) for exactly the reverse to occur in each of these cases for prospects of losses. These distinctive reversals in risk postures (between low vs. high probabilities and gains vs. losses) are generally interpreted as signatures of the two main distinctive components of prospect theory preferences: reference dependence (the tendency to evaluate changes in wealth rather than final wealth) and probability weighting (the tendency to value low probability prospects as if they are more likely and high probability prospects as if they are less likely than they actually are).

These anomalies, and the putative behavioral mechanisms driving them, are generally interpreted as special responses to the riskiness of lotteries (e.g., as consequences of non-EU risk preferences). But lotteries are not only risky – they are also necessarily *complex*, in the sense that they are *disaggregated* and therefore require potentially costly and difficult *information processing* to properly value them. To the degree the attentional and computational resources required to carefully evaluate a lottery are costly (or unavailable), decision makers may substitute to less careful (but less costly) valuation strategies instead.[1] As we discuss in more depth in Section 2.2, a number of recent theoretical models have described how the use of less precise valuation strategies can produce the distinctive fingerprint of the fourfold pattern, and for reasons that have nothing to do with risk.

Our key observation is that such complexity-based explanations apply not only to lotteries but also to *deterministic* objects that are disaggregated in a lottery-like way. By comparing the

[1]The idea that the complexity of a problem is a measure of the procedural *cost* of properly solving it is a direct adaptation of the definition of complexity used in computer science. See Footnote 2, below.

way people value lotteries to the way they value similarly complex deterministic prospects, we can measure how much of the pattern can be rationalized as a pure response to complexity (i.e., a pure response to the costs and difficulties of information processing).

To do this, we elicit not only subjects' certainty equivalents for a standard set of diagnostic lotteries, but also their valuations for what we call "deterministic mirrors" of the same lotteries – adaptations of lotteries in which we simply replace probabilities with deterministic payoff weights. These elicitations give us, not certainty equivalents (there is no uncertainty in these problems), but rather *simplicity equivalents* – simply-described monetary amounts that subjects believe to be of equivalent value to the relatively more complexly-described payoffs of mirrors. The *deterministic* values of mirrors are equal to the *expected* values of their source lotteries, so any appearance of standard anomalies in the simplicity equivalents of these objects are necessarily complexity-derived mistakes (because there is no risk, preferences over risk cannot apply). To put it differently, deterministic mirrors are lotteries in which preferences have been *induced* by the design to be linear so that deviations from risk neutrality in their simplicity equivalents cannot be rationalized by, e.g., risk preferences. By comparing the incidence and severity of anomalies in simplicity equivalents (where risk preferences cannot influence valuation) to those in certainty equivalents (where they can), we can benchmark how much of the empirical fingerprint of prospect theory can be reasonably attributed to complexity rather than risk.

We find that the full fourfold pattern appears in the simplicity equivalents of riskless mirrors, just as it does in the certainty equivalents of risky lotteries. That is, subjects tend to undervalue positive payoff components with large weights and overvalue positive payoff components with small weights, but do exactly the reverse when payoff components are negative, thus displaying the fourfold pattern. What's more, the pattern is roughly as severe in mirrors as it is in lotteries: the median subject's certainty and simplicity equivalents deviate from expected value to exactly the same degree for most of our elicitations and, in the aggregate, the fourfold pattern is 97% as severe in deterministic mirrors as in true lotteries.

We also apply these same methods to lotteries designed to measure *loss aversion*, the third and final distinctive component (alongside reference dependence and probability weighting) of prospect theory. Loss aversion is a regularity in which decision makers overweight negative payments relative to positive payments when evaluating probabilistic mixtures between the two. We find that the average subject displays apparent loss aversion when evaluating mirrors just as she does when evaluating lotteries, even though loss is not possible for the relevant choices in the mirrors we study. Subjects thus overweight negative payoff components when evaluating deterministic mirrors even when there is no risk of actually losing money. Overall "loss aversion" is 66% as severe in mirrors as in lotteries.

Crucially, we find significant evidence that these anomalies likely arise in lotteries and mirrors for similar reasons, driven by related behavioral mechanisms. Because we use a within-subjects design (all subjects are asked to value both lotteries and their deterministic mirrors in a random

order), our design allows us to evaluate how the severity of the pattern covaries in the two types of tasks across subjects. We find that the appearance of the pattern in simplicity equivalents strongly predicts its appearance in certainty equivalents – the correlation between the two across subjects ranges between 0.59 to 0.65 depending on the metric used and is highly statistically significant. What's more, in both deterministic and stochastic tasks, deviations are highly systematic and non-symmetric, running nearly uniformly in the direction of the fourfold pattern and loss aversion. This strong relationship suggests that these anomalies probably arise in lotteries for the same reasons that they arise in deterministic mirrors.

Additional treatments and controls affirm the robustness of these results and aid in their interpretation. Removing scope for treatment-to-treatment contagion, quintupling incentives, using a more sophisticated subject pool, intensifying training and decreasing the computational difficulty of evaluating lotteries/mirrors all have at most minor impacts on our results in both lotteries and mirrors. The nature of the design furthermore minimizes scope for subjects to "mis-import" probabilistic heuristics usually reserved for lotteries to deterministic mirrors (we deliberately describe mirrors in a frequentist way that allows us to avoid mention of probabilities or likelihoods altogether). Instead, demographic data, auxiliary behavioral data and post-experiment questions suggest that these anomalies arise because subjects consciously (perhaps even deliberately) elect to use imprecise, error-prone valuation procedures instead of the precise methods of evaluation we often implicitly assume when interpreting lottery choice.

In our concluding discussion, we interpret these results. There, we argue that because these anomalies in deterministic mirrors are clear mistakes, and mirrors are no more complex than their source lotteries, the principle of parsiomony suggests that the anomalies in lotteries are, to a similar degree, complexity-driven mistakes. This interpretation is significantly reinforced by the fact that the severity of anomalies strongly covaries across subjects, suggesting that they derive from the same source (which, recall, must be complexity since risk is not present in mirrors). Decomposing the data, we argue that in our main dataset roughly 91% of the overall pattern can be attributed to the complexity of aggregation (97% of the fourfold pattern and 66% of loss aversion). We emphasize, however, that this is a conservative estimate: there are strong reasons to believe based on recent evidence (Martinez-Marquina et al. 2019) that stochasticity makes information processing tasks like valuation *more* difficult. As such, this decomposition likely provides a lower bound estimate of the role complexity plays in driving prospect theoretic behavior in standard risky lotteries.

These findings, if borne out by future research, have several potentially important implications. First, they suggest that, to a great extent, the deviations from expected utility theory documented in behavioral economics may be a consequence of the costs and difficulties of information processing rather than of behavioral risk preferences. This has direct welfare implications, because it suggests that choices made in response to risk may not accurately reveal true risk preferences. This in turn has policy implications, suggesting as it does that these anomolous responses to risk may be correctible through training, nudging and judicious institutional framing and that such corrections

may be welfare-enhancing. Second, these results suggest that lottery anomalies and the theories constructed to describe them may have a far greater scope of application than is conventionally assumed. Complexity is in some sense more fundamental than risk, applying as it does to a much broader range of decision contexts. Our results suggest that theories like prospect theory may be directly descriptive of the way humans respond to one important and ubiquitous type of complexity in economics (the complexity of valuing objects with disaggregated components such as consumption bundles, production plans or strategies). As such, our results may point to a much wider range of descriptive and predictive uses for such theories in economics.

Our work relates to and builds on several literatures. First is a long literature documenting the signature anomalies of prospect theory empirically (e.g., Kahneman & Tversky 1979, Tversky & Kahneman 1992, Barberis 2013, Wakker 2010). Second is a nascent literature in economics that (i) follows computer science by defining complexity as the cost of implementing a decision rule and (ii) measures these algorithmic costs and their distortionary impacts on behavior directly (e.g., Oprea 2020, Banovetz & Oprea 2022, Camara 2021).[2] Third is a growing literature showing that as lotteries become more complex (i.e., as the number of elements in their supports grow), departures from expected value become more severe (e.g., Bernheim & Sprenger 2020, Puri 2020, Fudenberg & Puri 2022).[3] Fourth is a literature showing that measured risk aversion (e.g., Benjamin et al. 2013) and probability weighting (e.g., Choi et al. 2021) are strongly related to measures of cognitive ability, suggesting that information processing costs and therefore complexity likely influence lottery valuation (Stango & Zinman 2022). Fifth is a literature showing that measurements of risk posture (e.g., Friedman et al. 2017, 2022, Beauchamp et al. 2020), the fourfold pattern (Harbaugh et al. 2010), and prospect theory parameters (Bauermeister et al. 2018) are unstable, changing sometimes dramatically (even within-subject) when the method of elicitation is changed; Holzmeister & Stefan (2021) provides evidence that this instability is driven by the way complexity varies across choice environments. Finally, a recent literature (reviewed in detail in Section 2.2) shows theoretically how complexity (broadly writ) can produce classical anomalies like the fourfold pattern via a variety of boundedly rational mechanisms.[4] Methodologically, our paper is closely

[2] In computer science, complexity is defined as the cost (usually denominated in time or memory) of implementing an algorithm to properly solve a problem. We can define complexity analogously in humans: a problem is complex if it requires a human to use a mentally *costly* algorithm (i.e., procedure or rule) to properly solve it. One way to study the role of this complexity empirically is to attempt to measure it directly by measuring distaste for implementing algorithms, an approach followed by Oprea (2020). A second approach is to try to remove the costs of implementing algorithms and study whether this causes anomalies to disappear thus implicating complexity in the anomaly – an approach taken by Banovetz & Oprea (2022). Our paper suggests a third approach: remove putative drivers of anomalous behavior *other than complexity* from a decision problem, and study whether the anomalous behavior remains. To the degree it does, we have evidence that the anomaly was an outgrowth of complexity. Camara (2021) proposes an axiomatic framework for studying these computational costs theoretically and shows that they lead directly to heuristic behaviors like choice bracketing.

[3] Relatedly, Nielsen & Rehbeck (2022) show that subjects fail to comply with some of the central axioms of expected utility theory, not due to preferences but rather due to the complexity of applying axioms in lottery choice.

[4] Perhaps most closely related to our paper in this literature is Enke & Graeber (2021) who show empirically that subjects' stated uncertainty concerning the optimality of their valuations predicts the severity of the fourfold pattern.

related to Martinez-Marquina et al. (2019) who also compare stochastic and deterministic versions of otherwise isomorphic optimization problems and show that people are better at reasoning about the latter (the "power of certainty").

The remainder of the paper is organized as follows. In Section 2 we describe a collection of lottery anomalies that we collectively call the "classical pattern," discuss the relative role risk and complexity play in driving this pattern, and outline our methodology for separating the two classes of explanation. In Section 3, we discuss our experimental design and in Section 4, we present our results. We close with a discussion in Section 5.

2 Conceptual Background

In this section we discuss the key conceptual issues motivating our study. In subsection 2.1 we describe "the fourfold pattern of risk" and "loss aversion," the key empirical regularities underlying prospect theory, which we call collectively the "classical pattern." Next, in subsection 2.2 we discuss the complexity of lotteries and a recent literature that suggests that anomalies in lottery valuation may be a consequence of the complexity of lotteries rather than of their riskiness. Finally, in section 2.3 we describe how we propose to remove risk from lotteries while retaining their complexity. There, we argue that we can use this method to decompose the relative roles complexity versus risk play in driving the classical pattern.

2.1 The "Classical Pattern"

Consider a simple lottery $L = (p; X, Y)$ that pays out \$X with probability p and \$Y with probability $1-p$. Let \$C be the *certainty equivalent* of L: the certain (i.e., riskless) dollar amount the decision maker values equivalently to L. Let $E = pX + (1-p)Y$ be the expected value of lottery L. A decision maker is risk averse if $C - E$ is negative and risk seeking if it is positive. That is, a decision maker is risk averse if her certainty equivalent for the lottery is less than its expected value, and risk seeking if the reverse is true.

In standard expected utility theory a decision maker can be risk averse or risk seeking, but for conventional utility functions this should not change as p changes, holding the payoffs in L fixed. Instead, a decision maker should remain weakly risk averse or weakly risk seeking as p changes in the lottery. Similarly, under expected utility theory, decision makers' risk postures should not depend on what *direction* gambles shift wealth relative to the status quo, but should instead depend soley on the distribution of final wealths the lottery induces. Thus risk aversion in the valuation of lotteries should not depend on whether decision makers face potential losses versus gains (relative,

They also show that this cognitive uncertainty and the severity of the fourfold pattern rise in tandem as tasks become more complex. We show in Section 4.7 that cognitive uncertainty predicts the severity of anomalies in our data too, in both lotteries and mirrors.

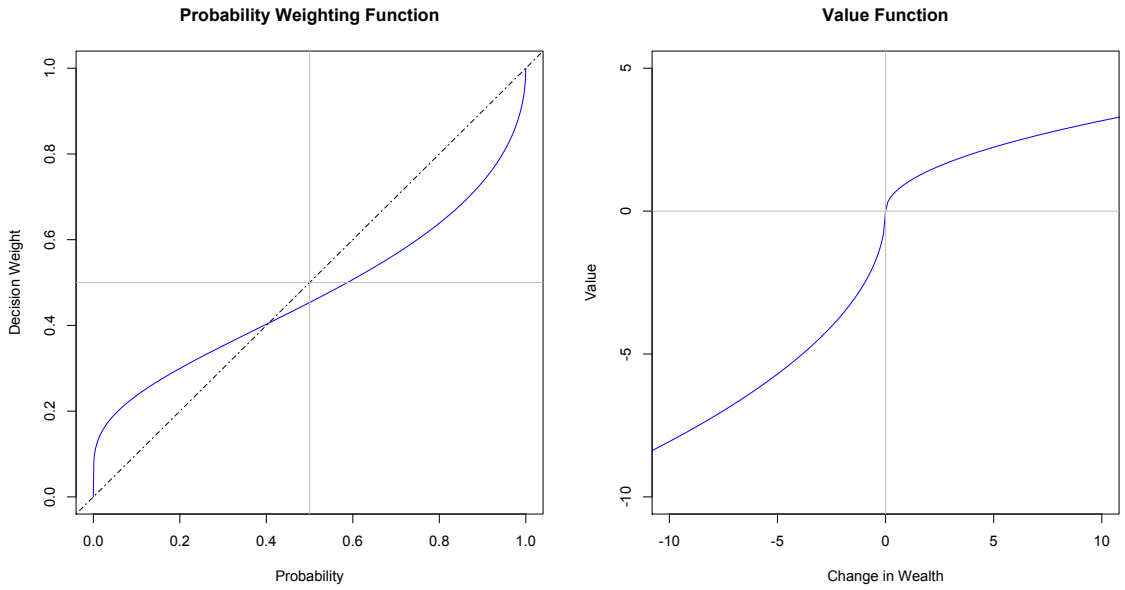


Figure 1: Prospect theory preference functions. *Notes: The left hand panel shows an example of the prospect theory probability weighting function which transforms probabilities into decision weight. The right panel shows an example of the value function which assigns value as a function of changes in wealth. The sharp kink at 0 and increased slope in negative values describes loss aversion.*

e.g., to the status quo) under expected utility theory.

However, a central finding of behavioral economics (encoded in influential behavioral theories like prospect theory) is that people tend to violate these broad predictions of expected utility theory in systematic ways. These anomalies of the certainty equivalent are summarized in the famous "fourfold pattern" of risk, consisting of four "effects" (assume, following typical measurement approaches, that $Y = 0$ in each case):

- **Certainty Effect in Gains**: When $X > 0$ and p is large (e.g, $p > 0.7$), decision makers are risk averse ($C - E < 0$).
- **Possibility Effect in Gains**: When $X > 0$ and p is small (e.g, $p < 0.3$), decision makers are risk seeking ($C - E > 0$).
- **Certainty Effect in Losses**: When $X < 0$ and p is large (e.g, $p > 0.7$), decision makers are risk seeking ($C - E > 0$).
- **Possibility Effect in Losses**: When $X < 0$ and p is small (e.g, $p < 0.3$), decision makers are risk averse ($C - E < 0$).

Thus, the fourfold pattern consists of a series of distinctive reversals in risk postures as probabilities move from low to high, and payoffs switch from gains to losses. When decision makers face the

possibility of gains, they seek risk at low probabilities and avoid it at high ones. When decision makers face losses, they do exactly the opposite.

Tversky & Kahneman (1992) call the fourfold pattern the "most distinctive implication of prospect theory," because it crisply displays two of the theory's three distinctive ingredients.[5] First prospect-theoretic decision makers display *probability weighting*, valuing lotteries as if small probabilities are larger and large probabilities are smaller than they actually are. This is usually summarized by a distinctive inverse s-shaped "probability weighting function," w, like the one pictured in the left hand panel of Figure 1 that transforms true probabilities into "decision weights." Second, prospect-theoretic decision makers are *reference dependent*, valuing *changes* in wealth relative to a reference point (conventionally zero in this setting) rather than (as in expected utility theory) final wealth. This is usually summarized by an s-shaped "value function," v, like the one pictured in the right hand panel of Figure 1. Prospect-theoretic agents are then assumed to make choices that maximize (in, e.g., binary settings like the ones described above) $w(p)v(X) + (1 - w(p))v(Y)$.[6] Applying probability weighting in this way to sign-preserving valuations of changes to wealth produces (and therefore rationalizes) the fourfold pattern.

While the fourfold pattern describes how prospect theoretic agents value prospects of gains relative to prospects of losses, *loss aversion*, describes how they value mixtures of the two. This third distinctive component of prospect theory (alongside probability weighting and reference dependence) describes a tendency for decision makers to put greater weight on losses than gains when evaluating lotteries containing both. A classic diagnostic case is the following:

- **Loss Aversion**: When $X > 0$, $Y < 0$ and $p = 0.5$, decision makers require $|X| > |Y|$ in order to be indifferent between L and a sure payoff of \$0. In particular there is some $\lambda > 1$ such that if $|X| = \lambda|Y|$, the decision maker is indifferent between L and \$0.

In prospect theory, loss aversion is described by a "kink" in the value function at the reference point of zero that steepens the impact of losses on valuations relative to symmetric gains (as in the right hand panel of Figure 1).

Together, these five anomalies in lottery valuation constitute the main empirical content of

[5] Prospect theory comes in two main forms: "original prospect theory" (Kahneman & Tversky (1979)) and "cumulative prospect theory" (Tversky & Kahneman (1992)). The distinctions between the two versions are unimportant for our analysis and so we ignore those distinctions throughout the paper.

[6] For example, Tversky & Kahneman (1992) propose the following simple functional forms of the probability weighting function

$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}} \tag{1}$$

and the value function

$$v(x) = \begin{cases} x^\alpha, x > 0 \\ -\lambda(-x)^\alpha, x < 0 \end{cases} \tag{2}$$

Many alternative functional forms have been proposed in the literature.

prospect theory, by far the most influential "behavioral" model of decision making under risk.[7] We will collectively refer to this group of empirical regularities as the "classical pattern" or simply "the pattern," and understanding its source is the goal of this paper.[8]

2.2 Risk and Complexity

The classical pattern is typically interpreted as a special response to to the risk inherent in lotteries. In particular, it is traditionally explained as an expression of non-expected utility *risk preferences*. For instance in prospect theory, probability weighting, reference dependence and loss aversion are traditionally understood as descriptions of how decision makers' preferences for lotteries respond to deviations from a reference point or to variation in probabilities.[9] Some accounts of the pattern also attribute some of its characteristics (e.g., some aspects of probability weighting) to mistakes in understanding stochasticity that produce systematic errors in lottery valuation (see, e.g., Wakker (2010)).

Importantly, however, lotteries are not only risky – they are also, unavoidably, *complex* in the sense that they are disaggregated into multiple payoff components that a decision maker must aggregate in order to discover its value. Indeed, "disaggregatedness" (the number of elements in a lottery's support) is the most common metric of lottery complexity in the recent literature on the topic (e.g., Bernheim & Sprenger (2020), Puri (2020), Fudenberg & Puri (2022)). As long literatures in psychology, neuroscience and computer science emphasize (e.g. Kool & Botvinick 2018), this kind of information processing is costly and such costs are precisely what we (following the operationalization used in computer science) mean by "complexity."[10] In this sense, even the seemingly simple two-outcome lotteries that are usually used to measure the classical pattern are complex relative to a simple dollar payment like a certainty equivalent (which, note, has only one payoff component and therefore requires minimal information processing to value).

[7] Some other well-known anomalies like the Allais paradox or the endowment effect are often also interpreted as part of the "main empirical content of prospect theory." However recent evidence calls this into question. McGranaghan et al. (2022) provides evidence that the common-ratio Allais effect, often interpreted as an outgrowth of probability weighting, is not consistent with measured probability weighting. Likewise, Chapman et al. (2021) provide evidence that the endowment effect is not statistically related to standard measurements of loss aversion.

[8] We will generally refer to this group of anomalies as the "classical pattern" rather than, e.g., "prospect-theoretic behavior" to emphasize that our interest is in understanding the roots of the empirical pattern that motivates theories like prospect theory rather than in testing any specific descriptive theory.

[9] Although the classical pattern is traditionally interpreted as a consequence of preferences, there is a great deal of ambivalence about this interpretation in behavioral economics. For instance, in a recent review O'Donoghue & Somerville (2018) conclude 'there is relatively limited discussion or consensus about the psychological principles that underlie probability weighting." Likewise, Camerer (2005), in a discussion of loss aversion, concludes "[a]n important open question about loss aversion is whether it is a judgement error or a genuine expression of preference."

[10] In computer science, a task or decision problem is described as "complex" to the degree that the least costly algorithm for correctly performing or solving it is complex (i.e., costly). Any lottery is complex in this sense to the degree that the procedure required to process information is more costly than the trivial procedure required to assess a simple, deterministic payment.

Complexity serves as an alternative explanation for the classical pattern, because it can produce systematic *distortions* in valuation that resemble the classical pattern for reasons unrelated to risk. Because a lottery is complex, its value is not immediate or transparent to a decision maker: she must first process information, combining the disaggregated payoff components of the lottery, in order to understand how she values it. And, because processing information with any precision is costly, complexity produces an immediate incentive to substitute to less costly (but less accurate) methods of information processing instead.[11] There are many ways this substitution to cheaper modes of information processing might produce biases in valuations, including biases resembling the classical pattern, and recent literatures in both psychology and economics have detailed a number of them in the last decade.

The aim of our investigation is *not* to differentiate between the many ways that economizing on information processing costs might produce the classical pattern via complexity; our experimental methods are largely agnostic to distinctions between the many possible proximal accounts. Nonetheless, we give a brief overview of some of the mechanisms that have been discussed in the recent literature in order to give the reader a flavor of how complexity, rather than risk, might produce these kinds of anomalies. First, DMs may economize by choosing to incompletely evaluate lotteries, ignoring or severely underweighting some relevant pieces of information and directly biasing valuations in the process. For instance, Bordalo et al. (2012) show that over-weighting of salient payoffs in lotteries can directly produce the fourfold pattern and apparent probability weighting. Likewise, decision makers can reduce information processing costs by evaluating changes to wealth directly instead of integrating lottery outcomes with final wealth, automatically producing apparent reference dependence (a key ingredient in the pattern). Second, a long literature in neuroscience and the psychophysics of perception suggests that precisely encoding or representing perceptual information is costly to the brain and that brains economize by coding information noisily (see Woodford (2020) and Glimcher (2022) for recent reviews). This will result in systematic bias if brains attempt to attenuate the costs of this "noisy coding" by efficiently biasing its evaluations when "decoding" noisy representations to inform choice, e.g., by shading evaluations towards prior beliefs in a Bayesian manner. A recent literature shows how this efficient biasing of noisily processed information can produce lottery anomalies including (i) reference dependence and small stakes risk aversion (e.g., Khaw et al. 2021, Frydman & Jin 2021), (ii) probability weighting and the fourfold pattern (e.g., Steiner & Stewart 2016, Vieider 2022) and even (iii) loss aversion (Khaw et al. 2021).[12]

[11] See Simon (1955) for an early discussion of this idea. See Oprea (2020) for evidence that humans find information processing costly and Banovetz & Oprea (2022) for evidence that they respond sensitively to these costs when choosing decision procedures. See Camara (2021) for a related theoretical discussion of this mechanism.

[12] The literature has gathered significant recent evidence for this class of mechanism. Khaw et al. (2021) shows that small-stakes risk aversion is intimately linked to noise in evaluations – a key implication of these models. Frydman & Jin (2021) show that subjects' beliefs about the range of lottery payoffs they are likely to encounter jointly impacts behavioral noise and risk aversion in a manner supportive of the hypothesis that "coding" is not only noisy but also efficiently adapted to the decision environment ("efficient coding"). Vieider (2022) provides evidence that such noisy coding models do a significantly better job of predicting and organizing the anomalies of the classical pattern than do standard noisy prospect theory models.

Third, and relatedly, imperfect processing of information (due to noisy coding or simply due to computational errors) might produce "cognitive uncertainty" about the optimality of valuations, leading decision makers to cautiously shade their valuations towards safe-seeming defaults; this, too, can produce probability weighting and the fourfold pattern.[13] Indeed, if a decision maker does nothing more sophisticated than force her noisy valuations to fall within the support of the lottery (either for rational reasons or because of bounds in the set of available valuations in the elicitation), this will result in apparent probability weighting in mean valuations (Blavatskyy (2007)).[14]

What is important for our purposes is that none of these complexity-based explanations rely in any special way on risk. Instead they rely on the fact that lotteries are disaggregated and their values are therefore not transparent to decision makers. In order to properly aggregate (value) them, the DM must implement a costly information processing procedure and may therefore opt to (or be forced to) use less costly but also less accurate valuation strategies instead. Our main observation is that because these complexity-based explanations do not rely on risk, they should also operate in valuations of riskless prospects that share the complexity (the information processing required) of a lottery.

2.3 Simplicity Equivalents

Our contribution is to propose a method for separating these two broad classes of explanations for anomalies like the classical pattern: risk-based explanations (that rely on behavioral risk preferences or confusions about stochasticity) versus complexity-based explanations (that stem from the generic costs and difficulties of evaluating disaggregated objects). Our method has the appealing advantage that it does not require us to commit to or even articulate any specific model of complexity (e.g., any one of the many proximal mechanisms discussed in the previous subsection) in order to accomplish this empirical separation.

Our proposal is to compare valuations of lotteries to valuations of objects that are complex (disaggregated) like a lottery but that contain no risk. Specifically, for any lottery L we can construct what we will call a "deterministic mirror," M_L, of L that replaces probability p with a deterministic weight. Thus, instead of paying X with probability p and Y with probability $1-p$ like a lottery, a mirror pays $pX+(1-p)Y$ with certainty. M_L is thus identically disaggregated and

[13] Consistent with such explanations, Enke & Graeber (2021) show that probability weighting and the fourfold pattern are significantly stronger for subjects who report in post-experiment questions that they are uncertain about the correctness of their valuations than they are for subjects who report relative confidence. They also show that this uncertainty and the severity of probability weighting jointly rise as the description of lotteries becomes more complex.

[14] Notice that this last kind of explanation highlights a relationship between "complexity" and recent discussions of the distorting effects of measurement error on measured preferences (e.g., Gillen et al. 2019). Blavatskyy (2007) can be interpreted as a description of how noisy implementation of preferences produces measurement error and thereby artificial evidence of probability weighting. To the degree this noise is a response to the costs of precisely interpreting lotteries, this kind of measurement error is a complexity effect.

therefore (in the sense described above) identically complex as L, but contains no risk. Because its certain value is the expected value of L, the cognitive act required to value it is precisely the same cognitive act that a risk neutral DM must perform in order to properly value L.

Because M_L is already a certain payment, the dollar value a decision maker assigns to it cannot be called a "certainty equivalent." Instead it is a "simplicity equivalent," – the *simply-described* dollar amount, S, the DM views as equivalently valuable to the complexly-described (but equally certain) payment M_L. While deviations of the certainty equivalent of L from expected value may be driven by, e.g., risk preferences, the same is not true of the simplicity equivalent of M_L. Since M_L pays off its expected value for sure, deviations of S from expected value are unambiguously complexity-driven mistakes in which the decision maker has transparently left money on the table.[15] Thus, to whatever degree we observe the fourfold pattern in simplicity equivalents, we have evidence that this complexity alone is sufficient to induce the pattern. To whatever degree DMs behave similarly in valuing M_L and L (i.e., to whatever degree errors in the simplicity equivalent predict errors in the certainty equivalent across DMs), we further have evidence suggesting that the same complexity that drives the pattern in M_L also drives the pattern in L.

In the same spirit, we can use deterministic mirrors to evaluate the role complexity plays in expressions of loss aversion, by examining which deterministic mixtures between positive and negative payments subjects treat as equivalent to a simply-described payment of \$0. Suppose M_L is the deterministic mirror of L, a lottery that mixes an $X > 0$ and $Y < 0$ with equal likelihood. As long as $X \geq |Y|$, M_L produces no loss, so if a decision maker demands a minimum X greater than $\lambda|Y|$ for some $\lambda > 1$ to counterbalance Y, it cannot be a rational expression of loss preferences. In mirrors, λ is instead a measure of the excess weight the decision maker mistakenly applies to negative numbers in the act of aggregation. To the degree $\lambda > 1$ in M_L, we have evidence that patterns resembling "loss aversion" can arise due to aggregation mistakes, even in the absence of true risk of loss.

A deterministic mirror is simply a lottery in which we have *induced* risk-neutral preferences by paying the expected value, following standard experimental economic methods (Smith 1976).[16] Risk neutrality is the natural benchmark for our purpose because under expected utility theory,

[15] Though, note, these "mistakes" may be optimal once the costs of information processing are accounted for.

[16] Preference induction is a standard technique in experimental economics, allowing researchers to study how individuals optimize and groups equilibrate without the confounding influence of unobserved preferences. For instance, instead of having subjects trade real goods in experimental markets, researchers create fictional goods and pay subjects reservation values and marginal costs for producing or acquiring them in trade directly (Smith 1962). Doing this removes subjects' unobserved "home grown" preferences (i.e., for real goods) from the experiment, allowing researchers to directly calculate and manipulate competitive equilibrium predictions by controlling the distribution of induced preferences in the market. Similar preference induction is used in experimental games for the same reason (e.g., we use dollar payments instead of jail time when studying Prisoner's Dilemmas). By removing unobserved preferences and replacing them with known preferences, induction allows the researcher to study how factors *other than preferences* shape behavior. Our methodology simply applies the same strategy to the choice tasks we usually use to measure preferences, inducing risk-neutral preferences in order to measure how factors other than risk preferences (i.e., complexity) shape valuation.

we should *expect* subjects to be approximately risk neutral given the stakes present in typical experiments (as emphasized, e.g., by Rabin (2000)). By studying how people value mirrors, we can uncover what errors in aggregation we should expect a risk neutral agent to make due to complexity alone. To the degree valuations of mirrors produce the same classical pattern that we observe in lotteries, we therefore have evidence that the pattern can arise as a consequence of risk-neutral but complexity-biased decision making. To the degree the pattern fails to appear in mirrors, we instead have evidence that either (i) risk/loss preferences or (ii) contributions to complexity that are special to risk are at the root the pattern.

Of course, this empirical strategy relies on an assumption that stochasticity introduces no additional complexity to the problem of valuation above and beyond the complexity already present in a similarly disaggregated mirror. That is, it assumes that it is no harder or more costly to process information and formulate and implement a strategy for valuing a risky object than a deterministic one. There are reasons to doubt this assumption. For instance in an important predecessor to our paper, Martinez-Marquina et al. (2019) show that subjects have more difficulty solving stochastic optimization problems than isomorphic deterministic problems, suggesting that stochasticity can make decision problems more complex. If risk adds to the difficulty of accurate valuation, it may drive decision makers to turn to less costly but more error-prone valuation strategies with greater frequency (or greater intensity) when valuing lotteries than when valuing mirrors. Crucially, however, to the degree this is true, our methodology will *underestimate* the role complexity plays in producing the pattern, meaning our method produces a *lower bound estimate* of the degree to which anomalies in lottery valuation are driven by complexity.

3 Experimental Design

We designed our experiment to

1. measure certainty equivalents for a set of standard lotteries usually used to document the "fourfold pattern" of risk,
2. compare these to simplicity equivalents for a set of deterministic mirrors of these same lotteries, *within subject* and
3. repeat the exercise for the measurement of loss aversion, again by comparing lotteries to their deterministic mirrors.

Our experiment is built around a series of "multiple price lists," the most popular tool for eliciting certainty equivalents and related measures of value in the literature. Figure 2 shows a screenshot from our software for one of the price lists we study (called G10). In each price list, subjects are shown a number of pairs of lotteries, A and B, one appearing on each row of the list. Lottery B is identical in all rows of the list, while lottery A changes in each row. For instance, in

Figure 2, A (in red) guarantees a sure payment amount ranging from \$25.00 to \$1 across rows; B (in blue), by contrast, offers the same lottery in each row: \$25.00 with 10% chance and \$0.00 with a 90% chance.

The subject's task is to choose a lottery for payment in each row of the list by clicking on the table, highlighting their choices in yellow. Subjects respecting first order stochastic dominance will choose A in early rows (or, with extreme preferences, may never choose A) and switch to B in exactly one row (possibly the first), never switching back to A again. We enforced this single switching property in the experimental design mostly to simplify our analysis and speed up the experiment.[17] The switching point between A and B reveals the A for which the subject is indifferent to fixed lottery B. In this example, this yields an estimate of the subject's dollar value (e.g., her certainty equivalent) for a lottery that pays \$25 with probability 0.1 and \$0 with probability 0.9.

After clicking on the table, the subject clicks a button to submit her choice before moving on to the next price list. The subject is told (truthfully) that one row from one price list from one treatment will be selected randomly at the end of the experiment to determine her payment.

These methods are standard in the literature and, as we discuss in Section 3.2, we use them to study very conventional lotteries. Our contribution is to include an additional version of each price list in which we pay subjects (in effect) the expected value of the lottery they select; in the language of Section 2, A and B are thereby transformed into "deterministic mirrors." This allows us to elicit "simplicity equivalents," as described in Section 2, which can be directly compared to certainty equivalents elicited for true stochastic lotteries.

3.1 Treatments

The main experiment consists of two treatments and every subject participated in *each*, in a random order (i.e., a *within-subjects* design). In the Lottery treatment, subjects are paid under a conventional stochastic incentive rule, as in standard lottery choice problems. If a row is selected for payment, the computer uses the likelihoods listed at the top of the table to randomly select a state from the lottery the subject chose (A or B) and pays the subject the monetary payout specified for that state. In the Mirror treatment, by contrast, the computer instead pays the subject the (deterministic) expected value of the lottery she selected for the paid row. This transforms lotteries into deterministic mirrors.

Crucially, we frame lotteries and mirrors in a nearly identical way by avoiding any reference to probabilities. Instead, we describe each lottery/mirror as consisting of a set of 100 "boxes," each of which contains some (possibly negative) payment. The description of any given lottery consists of a description of how many of these boxes contain each possible payment amount. In Figure 2, for instance, lottery A consists of 100 boxes, each of which contains the same amount (which changes

[17] Note that this means that our design prevents some mistakes from occurring and may therefore lead our design to *understate* the role bounded rationality plays in lottery valuation.

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 2: Screenshot from a mirror task (list G10). *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."*

systematically from row-to-row); B consists of 10 boxes each containing $25.00 and 90 boxes each containing $0.

This means that the *only* difference between lotteries and mirrors is in how these identically-described box-configurations determine the subject's payment. In lotteries, we explain to subjects that the computer will randomly open *one* of the 100 boxes randomly and uniformly and pay the subject the amount in the randomly selected box. In mirrors, by contrast, we explain that the computer will open *all* of the 100 boxes and pay them the sum (divided by 100).

We ran these two treatments within-subject (every subject experienced both treatments) in a random order (either Lottery followed by Mirror or the reverse). In each treatment, subjects were assigned all of the price lists discussed in Section 3.2, below, in an independent, random order. Prior to *each* treatment we clearly described the treatment's payoff rule and then gave subjects a set of quiz questions to give them practice with and feedback on the payment rule.[18] Thus, we made a serious effort prior to each treatment to highlight and provide salient insight into the difference between the two payoff rules in the two treatments.

Importantly, subjects who were initially assigned the Mirror (Lottery) treatment were not aware they would later be facing the Lottery (Mirror) treatment. This, combined with the frequentist way we framed the problem, means that we completely avoid "priming" stochasticity to the roughly half of subjects initially assigned to Mirror. For these subjects, the valuation task has nothing per se to do with probabilities or randomness but is instead a sort of reasoning task involving box-opening. As we will see, this produces a clean separation between Lotteries and Mirrors in the initially-assigned treatment, which will be important for interpreting the results.

3.2 Lists and Hypotheses

The core of the design is a series of ten price lists that includes two lists designed to measure each of the five components of the classical pattern (the fourfold pattern and loss aversion). We assign every subject each of these lists under stochastic (Lottery) and again under deterministic (Mirror) incentives (in a randomized order). We call these our "core lists."

Eight of these lists (G10, G25, G75, G90, L10, L25, L75, L90) we call "fourfold lists" and their purpose is to elicit certainty/simplicity equivalents that, when compared to expected value, allow us to test for the fourfold pattern. Lists beginning with G (for "gains") elicit the value (between $25 and $1 in $1 increments) of the lottery $(p; \$25, \$0)$ and we vary p across lists between 0.1 (G10),

[18] For instance before *both* the Lottery treatment and the Mirror treatment, we showed the subject the same example lottery/mirror $(0.5; \$16, \$0)$ and asked them the same question: "What is the chance that $8 is added to your earnings?" The correct answer is different in each treatment (0% in Mirror but 50% in Lottery), highlighting the difference in the two incentive rules. Likewise, we asked "What is the chance that $8 is added to your earnings" which has a correct answer of 100% in Mirror but 0% in Lottery. By asking the same question twice, we thus tried to make the difference in payoff rules highly salient to subjects prior to the change of incentives in the second-assigned treatment.

0.25 (G25), 0.75 (G75) and 0.9 (G90). Figure 2 pictures one of these lists (G10). As with all "G" lists, A's payoff falls monotonically in each row from \$25 to \$1 in \$1 increments, while B is a fixed lottery whose value the list is designed to elicit.[19] Lists beginning with L (for "losses") look for the monetary equivalent (between -\$1 and -\$25, again in \$1 increments) of the lottery $(p; -\$25, \$0)$ and across lists we vary p between 0.1 (L10), 0.25 (L25), 0.75 (L75) and 0.9 (L90). (Because some lists involve losses, all lists also come with an initial endowment added to the lottery payments from the list.[20])

Identifying the row at which subjects switch from A to B in Lotteries allows us to estimate the certainty equivalent C as the midpoint between the lowest sure amount at which A is selected and the highest sure amount at which A is rejected in favor of B. The fourfold pattern entails the following regularities for certainty equivalent estimates C in these lists (given the coarseness of the elicitation grid):[21]

- **Certainty Effect in Gains**: $C <$\$18.50 in G75, $C <$\$22.50 in G90
- **Possibility Effect in Gains**: $C >$\$2.50 in G10, $C >$\$6.50 in G25
- **Certainty Effect in Losses**: $C > -$\$18.50 in L75, $C > -$\$22.50 in L90
- **Possibility Effect in Losses**: $C < -$\$2.50 in L10, $C < -$\$6.50 in L25

Looking for the row at which subjects switch from A to B in the same lists in mirrors similarly allows us to elicit the simplicity equivalent, S. Our primary question is whether the fourfold pattern, a phenomenon of the certainty equivalent, arises also in the simplicity equivalent (i.e., whether the pattern continues to arise when we replace "C" with "S" in the description above).

Two additional core lists, LA10 and LA15 measure loss aversion, the fifth main empirical regularity prospect theory is designed to explain. In these lists B is fixed at \$0 while A yields a fixed payment -\$Y (-\$10 in LA10 or -\$15 in LA15) with 50% chance and otherwise pays a positive amount \$X that declines from \$50 to \$2 in \$2 steps from the highest to lowest row in the list. By looking at the row at which subjects switch from A to B we can observe the positive payment a subject demands to compensate for a negative payment of -\$10 (LA10) or -\$15 (LA15) in a 50/50 gamble. Loss aversion entails the following regularities in Lotteries:

- **Loss Aversion**: $X >$ \$10 in LA10 and $X >$ \$15 in LA15.

[19] The G lists are used in Bernheim and Sprenger (2020); similar price lists are also used in Gonzalez and Wu (1999) and Bruhin et al. (2010). The L lists are negative reflections of these standard lists.

[20] This payment is \$5 for the G lists, \$15 for LA10, \$20 for LA15, \$30 for the L lists.

[21] Given the coarseness of the elicitation grid, in practice we will look for evidence of deviations in the specified direction that are at least \$1 (one price list row) away from these thresholds in establishing evidence of the pattern.

Once again, our interest is in comparing choices in mirrors to choices in lotteries. In mirrors (unlike lotteries) there are *no losses* at the expected value maximizing value of $X. Thus, evidence of loss aversion here cannot be a rational response to distaste for losses.

Finally, we included a pair of lists, G50 and L50 that elicit certainty/simplicity equivalents for a lottery mixing $0 and $25 with equal chance, included mostly to allow us to draw typical plots used to visualize probability weighting. For half of our subjects in our main treatment, we *repeated* each of these lists in both lotteries and mirrors, giving us a measure of how consistent subjects' decisions are. This gives us a measure of the noisiness of subjects' decisions, which we discuss in our analysis of the results.

3.3 Implementation and Variations

We ran the experiment on a total of 389 subjects using custom Javascript software programmed by the author and deployed via Qualtrics. Our main design (described above), consisted of 186 subjects recruited on Prolific in May 2022, who were paid a $6 base payment and, with 20% chance, were additionally paid the outcome from a randomly selected list and row. Subjects in this design spent an average of 31.2 minutes in the experiment and earned an average of $9.85.

We also ran an additional diagnostic treatment called "Easier" on Prolific using 90 subjects under the same protocol. This treatment assigned subjects lists G25, G50, G75, L25, L50, L75 and LA10 but described the underlying probabilities using 4 boxes rather than 100 (i.e., a 0.25 probability of earning $25 was described as a 1-out-of-4 chance rather than a 25-out-of-100 chance). To these, we added list sL25 and sG75 which were identical to L25 and G75 but paid out $20 instead of $25. We motivate and discuss this treatment in Section 4.6 and Online Appendix C.

Finally, we ran a robustness version of our main design using 113 undergraduate student subjects at UC Santa Barbara. This experiment (discussed in Online Appendix C) included longer training (in particular more comprehension questions), payment to all subjects (rather than 20% as in our Prolific dataset) and was run in traditional sessions (with subjects recruited via ORSEE, Greiner (2015)) on Zoom in which subjects were allowed to ask the experimenter clarifying questions throughout these experiments as in a typical laboratory session.

At the end of the experiment, we included a short battery of three cognitive reflections tasks, a short demographic survey (focused on the subject's technical education) and (in the Prolific samples) a number of questions about subjects' strategies and beliefs during the experiment. We discuss these in more depth in Section 4.7 and Online Appendix B.2.

4 Results

In this Section we report our empirical results. In Section 4.1 we show that the classical pattern arises with nearly equal strength in both risky lotteries and riskless mirrors, suggesting that the pattern is not primarily a phenomenon of risk. In Section 4.2 we show that the severity of the pattern in the two cases are strongly correlated with one another across subjects, suggesting they are driven by a common behavioral mechanism. In Section 4.3 we propose a simple taxonomy of subjects, suggesting that the pattern is far more sensitive to complexity than it is to risk in the subject population. In Section 4.4 we show that the results are not due to order effects or contagion across treatments. In Section 4.5 we show that similar results occur using a different subject pools with more intensive instructions and stronger incentives. In Section 4.6 we present results from a diagnostic treatment that suggests that the pattern is not driven primarily by computational difficulties or computational errors. Finally, in Section 4.7 we examine predictors of the pattern.

Throughout the analysis we will analyze deviations from the benchmark of the expected earnings maximizing choice, which we will call *EvMax*. For the fourfold lists (the "G" and "L" lists) EvMax is simply the lottery's expected value; for the loss aversion lists (the "LA" lists) EvMax is the lottery that equally weights gains and losses.[22] Throughout, we will refer to the absolute size of deviations from EvMax as our measure of *error*. We will instead refer to deviations that we have *normalized to be positive if they go in the direction of the classical pattern* as our measure of pattern-consistent *bias*. Thus, in our usage, high average *error* is evidence of severe departures from EvMax, while high average *bias* is evidence that these departures are those of the classical pattern. We use these terms (bias and error) in a statistical sense not in a normative sense, and their use should not be read as judgements on the optimality of choices made in lotteries (where deviations from expected payoff maximization can, ex ante, be rationalized as optimal responses to risk preferences).

4.1 Main Findings: The Pattern in Lotteries and Mirrors

Figure 3 presents our main results by plotting the difference between subjects' mean elicited certainty/simplicity equivalent and the lottery's expected value (on the y-axis) for each of our "fourfold lists" as a function of the probability of the lottery's non-zero payment (on the x-axis).[23] Deviations above zero are evidence of risk seeking and below zero evidence of risk averse behavior. List names are plotted next to dots and error bars represent two standard errors. Light gray labels describe the regions (e.g., north-east, north-west etc.) in which we expect each of the four parts of the fourfold pattern to occur (i.e., the certainty effect and possibility effect, in gains and losses). Data from lotteries are plotted as solid blue dots while data from mirrors are plotted as hollow red

[22] In calculating EvMax and deviations from it we use, not the lottery's/mirror's expected value, but the midpoint between price list rows that is closest to expected value. This prevents us from recording small deviations from expected value that are unavoidable given the coarseness of the design.

[23] Certainty/simplicity equivalents are calculated as the midpoint between the lowest certain/simple amount the subject preferred to the lottery/mirror and the largest certain/simple amount she rejected in favor of the lottery/mirror.

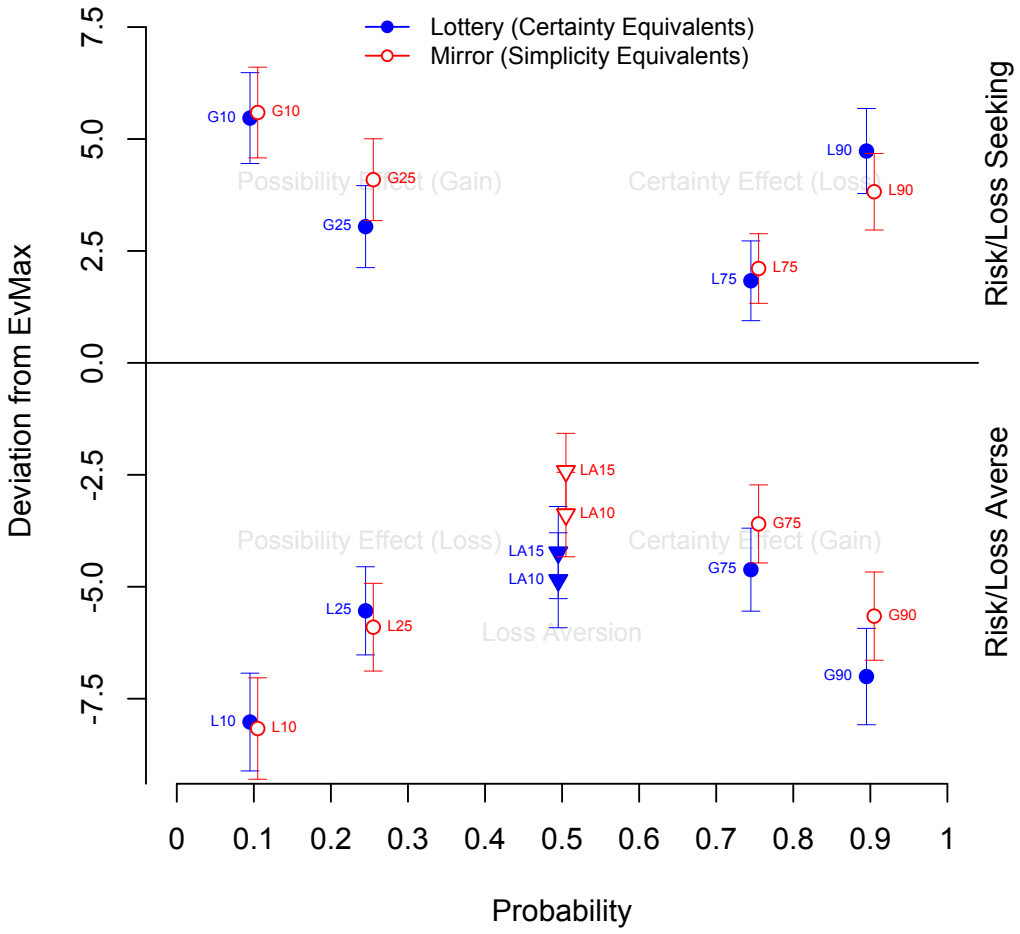


Figure 3: Mean deviations from EvMax in lotteries (solid blue dots) and mirrors (hollow red dots) for all core lists. *Notes: For fourfold lists, the y-axis measures the difference between subjects' certainty/simplicity equivalent and expected value (as stated in the axis label): positive values display risk-seeking and negative values risk-averse valuation. The x-axis is the probability of the non-zero payoff. For loss aversion lists, the y-axis measures instead the difference between zero and the expected value of the lottery subjects evaluate equivalent to zero: positive values are evidence of loss-seeking and negative values loss-averse valuation. Two-standard-error bars are included for every list. Light gray background labels list and give the expected region of deviation for each component of the classical pattern.*

dots.

The figure shows, as in previous work, that valuations of lotteries follow the fourfold pattern: subjects are risk averse towards gains and risk seeking towards losses at high probabilities (the certainty effects) and risk seeking towards gains and risk averse towards losses at low probabilites (the possibility effects). In the left two panels of Figure 4 we plot the same data using an alternative visualization (used, e.g., in Tversky & Kahneman (1992)) by plotting the probability of the non-zero payoff on the x-axis and the ratio of the certainty equaivalent to the lottery's non-zero payoff

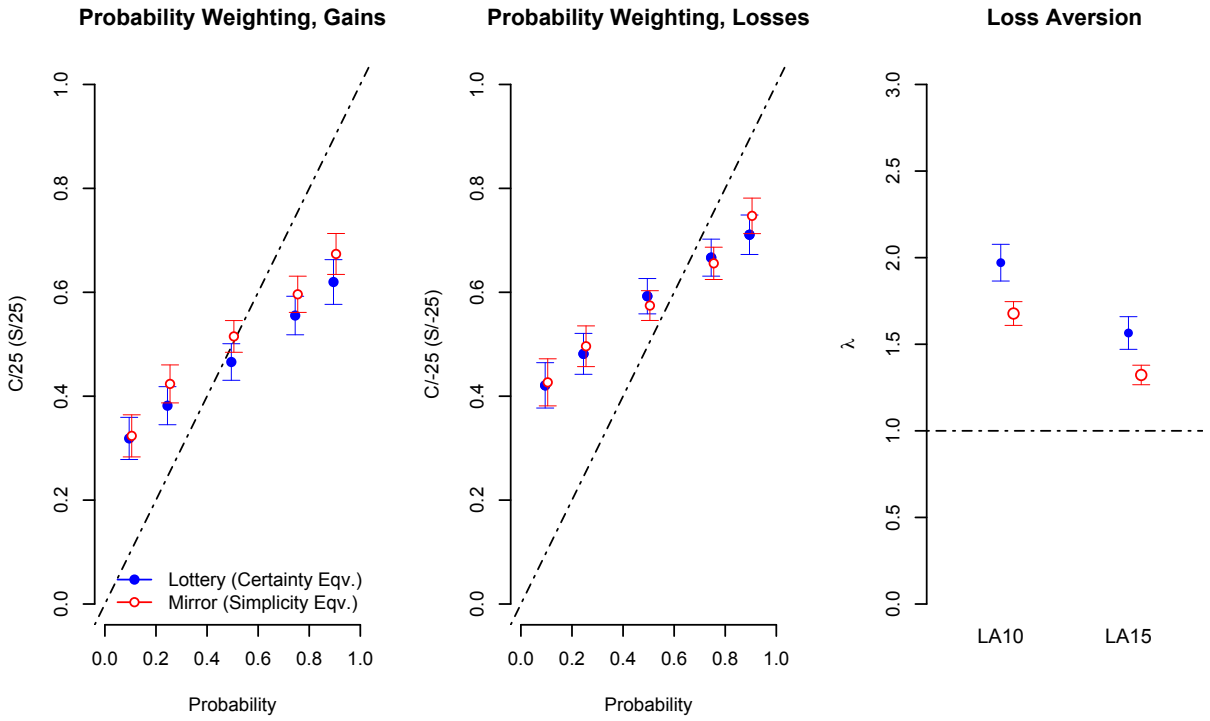


Figure 4: Naive visualization of the probability weighting functions (left two panels) and the loss aversion parameter, λ (rightmost panel) for lotteries and mirrors. *Notes: The first two panels plot a naive estimate of the probability weighting function (following Tversky & Kahneman (1992)) by plotting 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. Under the lens of prospect theory, the estimate is naive in that it doesn't correct for possible value function curvature. The final panel plots the absolute value of the ratio of the negative payment Y and positive payment X in the lottery treated as equivalently valuable to a sure payment of 0. This is a naive estimate of the λ parameter of loss aversion – naive because it does not correct for potential asymmetries in the probability weighting function (which we see little evidence of) or in the value function in gains versus losses (which the prior literature tends to show little evidence of).*

(i.e., C/E(v)) on the y-axis. The results produce a naive non-parameteric estimate of the prospect theory probability weighting function,[24] and the results we observe are typical: subjects value low probability lotteries as if they are more likely and high probability lotteries as if they are less likely than they really are.

Figure 3 also shows our main finding: the pattern appears in a virtually identical fashion in subjects' valuations of mirrors. Indeed, Figure 3 shows that the entire fourfold pattern arises in simplicity equivalents, including each of the distinctive reversals in apparent risk posture it predicts. Subjects over- and under-value mirrors in such as way to produce apparent risk postures (though there is no risk) that reverse precisely when their source lotteries do. Likewise, Figure 4 shows strong evidence of "probability weighting" in mirrors (though there are no probabilities). What's

[24]Under the lens of prospect theory, these are naive estimates of probability weighting in the sense that they implicitly assume a piecewise-linear value function (whose curvature isn't separately measured in this dataset). This is consistent with typical findings of near-linearity in the value function in much of the literature.

more, the size of these effects and the deviations from expected value they produce in mirrors are roughly of the same magnitude as those in lotteries.

Result 1 *The fourfold pattern of risk arises in settings without risk. Apparent probability weighting arises in settings without probabilities.*

For loss aversion (LA) lists, we plot in Figure 3 the difference between 0 and the expected value of the lottery which the subject evaluates as equivalent to a sure payoff of zero on the y-axis: values below zero are evidence of loss aversion. Once again, in lotteries we find evidence consistent with standard loss aversion: subjects require lotteries with positive payoffs strictly larger (in absolute value) than the negative payoffs they are mixed with to be indifferent to a payoff of \$0. In the third panel of Figure 4 we plot the same data in the standard way by calculating naive estimates of λ: the excess weight attached to losses relative to gains in evaluating mixtures between the two.[25] Here we find estimates of λ in lotteries that vary across our two lists but average to $\lambda = 1.77$ – within the typical range reported in meta-analysis of estimates from the literature (Brown et al. (2022)).

Again, our main finding here is that we also observe strong evidence of "loss aversion" in mirrors even though there are no losses possible in the relevant region of the choice space in mirrors. Estimates of λ, pictured in Figure 4, are clearly somewhat smaller in mirrors than in lotteries (though this difference is much smaller than the differences in λ we observe between elicitations, i.e., between L10 and L15). Nonetheless our estimate of $\lambda = 1.68$ for the LA10 mirror is close to the median estimate in the literature as reported in a recent metanalysis (Brown et al. 2022), and our estimate of 1.32 in the LA15 mirror is roughly equal to the modal aggregate estimate (and is within the overall interquartile range).

Result 2 *Apparent loss aversion occurs in settings without risk of loss.*

Wilcoxon signed rank tests confirm that valuations are significantly different from EvMax (at the 1% level) for all of our core lists for both lotteries and mirrors. Paired Wilcoxon tests reveal statistically significant differences between lottery and mirror choices in only half of the lists (and in one of these cases, G25, EvMax deviations are statistically *greater* in mirrors than in lotteries). However, none of the treatment differences are terribly economically significant: in every one of our core lists, the median within-subject difference between lottery and mirror valuations is *zero*.

We can summarize the relative severity of the pattern in the two types of problems by comparing the degree to which the deviations from EvMax characteristic of the pattern that occur in lotteries also occur in mirrors. Normalizing deviations so that they are positive if they run in the direction

[25] Here we follow the literature by assuming that λ is a linear weight on negative payments. Thus we calculate it as the ratio $-X/Y$ for the lottery/mirror the subject values as equivalent to \$0. In prospect theory, this is a naive estimate in that it assumes symmetric probability weighting in gains and losses at $p = 0.5$ (which is roughly true in most datasets including ours) and symmetric curvature of the value function in gains and losses (which is unmeasured in our data but is found true in many previous investigations, e.g., Tversky & Kahneman (1992)).

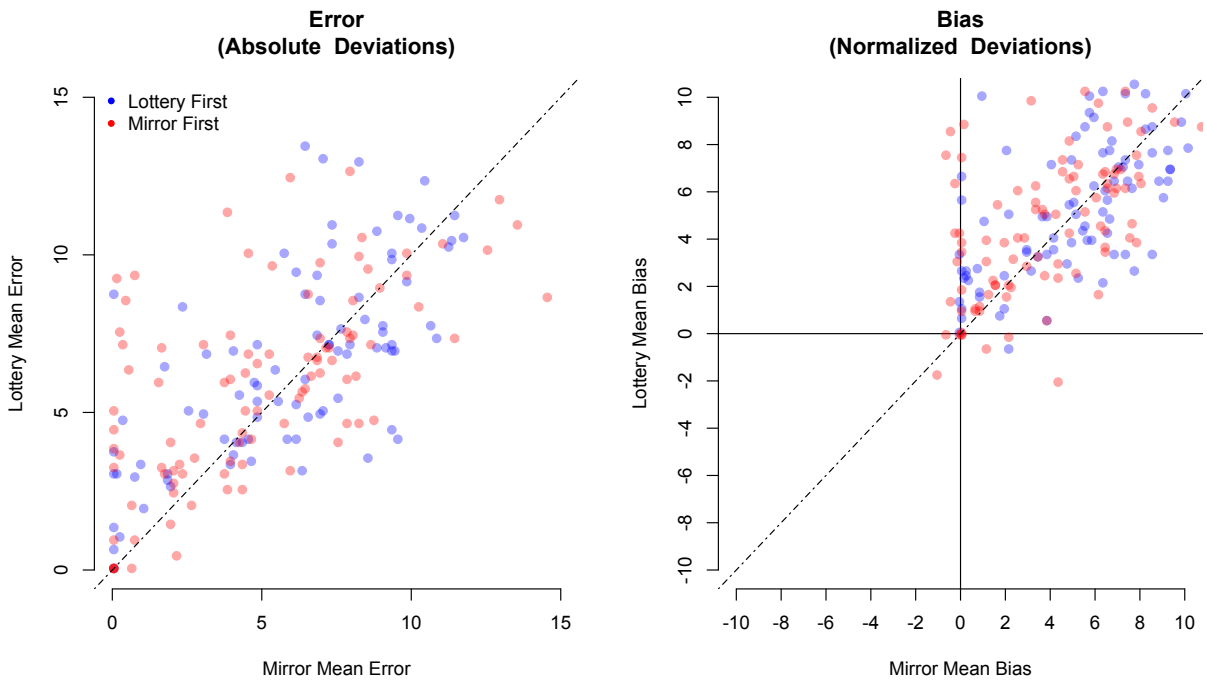


Figure 5: Deviations from expected value maximizing choices (in core lists) in mirrors (x-axis) versus lotteries (y-axis), by subject. *Notes: Each dot represents a subject. "Lottery First" designates subjects who were initially assigned lotteries (Mirror First is the reverse). The left panel plots "error" (mean absolute deviations from EvMax); the right panel plots mean "bias" (mean deviations normalized to be positive if they are in the direction of the classical pattern).*

of the fourfold pattern and loss aversion and summing, we find that for fourfold lists valuation bias is 97% as severe in Mirrors as in Lotteries (s.e. 5 percentage points) and for loss aversion lists they are 64% as severe (s.e. 9.5 percentage points).[26] Overall, pattern-consistent bias is 91% as severe in mirrors as in lotteries (s.e. 4 percentage points), suggesting that the vast majority of the pattern that appears in risky valuation occurs also in merely complex (but riskless) valuation.

Result 3 *Overall, 91% of the bias in the direction of the classical pattern that occurs in lotteries occurs also in mirrors.*

4.2 Relationship Between Risky and Riskless Valuations

We find very similar evidence of the classical pattern in lotteries and mirrors and a first, natural question is whether this similarity is coincidental. After all, it is possible that distinct behavioral mechanisms underlie the appearance of the pattern in risky and riskless settings and that the apparent relationship between the two is therefore an illusion. It was with this possibility in mind that we used a within-subjects design in our experiment. Using this design, we can determine

[26]Standard errors are derived from subject-wise calculations of the ratio of normalized deviations in mirrors and lotteries, Winsorized at 5% and 95%.

whether the magnitude of the pattern in mirrors *predicts* it's magnitude in lotteries. To the degree the two phenomena are correlated across subjects in this way, there is reason to believe the two appearances of the pattern derive from a similar or closely related behavioral mechanism.

The left panel of Figure 5 plots a separate dot for each subject, with the x-axis representing that subject's mean error (mean absolute deviation from EvMax) in our core mirrors and the y-axis the mean error in our core 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 0.65 ($p < 0.001$): errors in mirrors strongly predict errors in lotteries. The right panel of the same Figure instead examines mean pattern-consistent *bias* (deviations normalized to be positive if they run in the direction of the classical pattern), 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 bias is concentrated in the northeast quadrant suggesting that subjects make highly asymmetric errors on net in the distinctive direction of the pattern in both lotteries and mirrors. Second, although the severity of this bias is highly heterogeneous, there is again a very strong correlation (0.6, $p < 0.001$) between mirror and lottery bias, suggesting the two biases 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.[27,28,29]

Result 4 *The severity of the pattern in mirrors strongly predicts the severity of the pattern in lotteries.*

In Online Appendix B.1 , we present separate versions of Figure 5 for lists measuring the fourfold pattern and for lists measuring loss aversion. We find strong and highly significant correlations between lotteries and mirrors for both the fourfold pattern ($\rho = 0.62$) and loss aversion ($\rho = 0.41$).[30]

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. 77% of subjects reported using identical or mostly similar strategies in the two treatments, strongly matching our behavioral findings.

[27]For error (the left panel) the correlation is 0.62 when mirrors come first and 0.68 when lotteries come first; for bias (the right panel) it is 0.57 when mirrors come first and 0.61 when lotteries come first.

[28]This analysis uses subject-wise means to focus on subject-level tendencies. We find a similarly strong correlation between lottery and mirror deviations when using disaggregated individual choices instead ($\rho = 0.57$).

[29]These subject-wise estimates are likely noisily estimated and therefore suffer from measurement error (Gillen et al. 2019). Distortions in magnitudes resulting from this are of minimal concern for us because our main purpose is to compare lotteries and mirrors which are arguably similarly afflicted by measurement error. However, we should expect measurement error to artificially weaken correlations between mirrors and lotteries. We therefore should view the correlations reported in this section as lower bound estimates of the relationship between the appearance of the pattern in the two cases.

[30]The weaker correlation we find for loss aversion may signify a weaker latent relationship or may, instead, reflect greater attenuation due to increased measurement error (bias for the fourfold pattern is measured using eight choices, while bias for loss aversion is measured only using two).

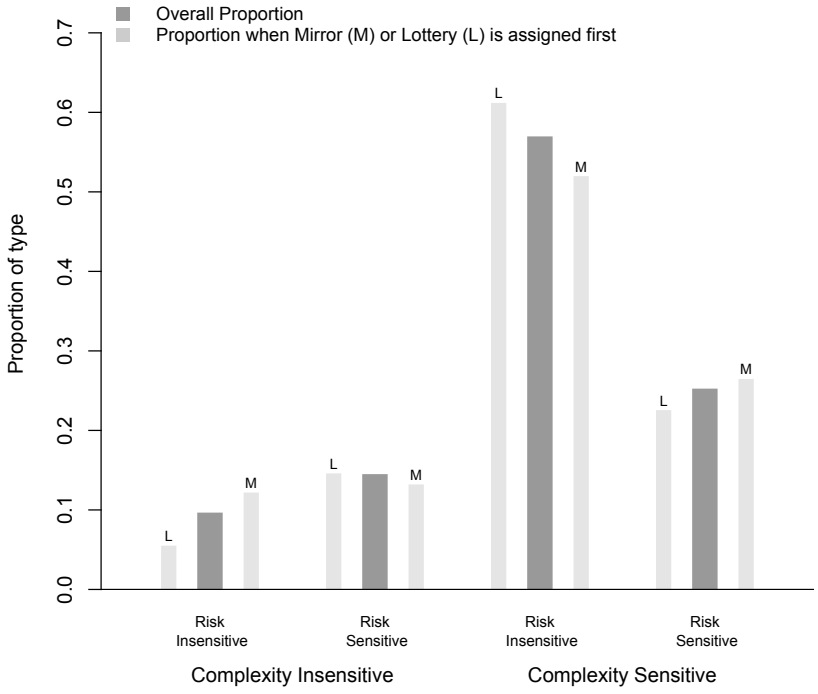


Figure 6: Simple taxonomy of subjects, with proportion of subjects of each type plotted on the y-axis. *Notes: Dark gray bars show raw proportions, light bars show proportions for subjects initially assigned to mirrors (M) vs. lotteries (L).*

4.3 Complexity Sensitivity and Risk Sensitivity

To better understand subject heterogeneity and how it relates to our main motivating question (the relative role of complexity and risk in driving valuation anomalies), we propose a simple taxonomy of subjects that, again, makes use of our within-subjects design. We classify each subject in our dataset according to each of the following two criteria:

- **Complexity Sensitive/Insensitive**: We say a subject is "complexity sensitive" if she shows evidence of the classical pattern (deviates in the direction of the pattern by at least one row of the price list) in her *average* mirror valuation. Otherwise we classify the subject as "complexity insensitive."
- **Risk Sensitive/Insensitive**: We say a subject is "risk sensitive" if she shows *additional evidence* of the classical pattern (deviates in the direction of the pattern by at least one row *more* in lotteries than in mirrors) in her average lottery decision than in her average mirror decision. Otherwise we classify the subject as "risk insensitive."

This produces a simple 2x2 taxonomy of subject types that focuses on the relative role of our

two key lottery characteristics – complexity and risk – in driving the pattern. While mirrors are *only* complex (in the sense discussed in Section 2), lotteries are both risky and complex. Evidence that a subject shows some systematic evidence of the pattern in mirrors is evidence that complexity is *sufficient* to drive her to express the pattern. On the other hand, a systematic intensification of the pattern in lotteries relative to mirrors is evidence that, for that subject, risk has an additional role (above and beyond complexity) in rationalizing the pattern.

Figure 6 plots the proportions of each type in this taxonomy using dark gray bars. The results show that nearly 80% of subjects are complexity sensitive in our data, while less than half as many (37%) are risk sensitive. What's more, over half of subjects (55%) are *only* complexity sensitive, showing no more evidence of the pattern in Lotteries than in Mirrors. By contrast only 1/4 as many subjects (13%) are purely risk sensitive (i.e., complexity-insensitive/risk-sensitive), complying with the standard interpretation of the pattern as a pure response to risk. Most risk sensitive subjects (65%) are also complexity sensitive but most complexity sensitive subjects (70%) are not additionally risk sensitive. A minority of subjects (8%) are sensitive to neither complexity nor risk (i.e., show little evidence of the pattern).[31]

In light gray we overlay the same taxonomy calculated only for subjects who were assigned mirrors (marked with 'M') or lotteries (marked with 'L') in their first treatment. The results show that these relative sensitivities are only minorly affected by the order in which subjects experience the treatments. This symmetry strongly suggests that these conclusions about the relative role of risk and complexity in driving the pattern is not an artifact of our within-subjects design.

Result 5 *Subjects are more than twice as likely to be complexity-sensitive as risk-sensitive in displaying the pattern. Most subjects that are complexity sensitive are not risk sensitive, while most subjects that are risk sensitive are also complexity sensitive. Subjects are four times more likely to be purely complexity-sensitive than to be purely risk-sensitive.*

4.4 Robustness: Cross-Treatment Contagion

One natural concern about these results is that they may be a consequence of contagion between mirrors and lotteries resulting from our 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 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 transfor-

[31] As we might expect given the analyses in the preceding sections, these results are much stronger for the fourfold pattern than for loss aversion. If we repeat the exercise using only loss aversion elicitations, we find that subjects are equally likely to be classified as "risk sensitive" and "complexity sensitive," rather than overwhelmingly more complexity sensitive.

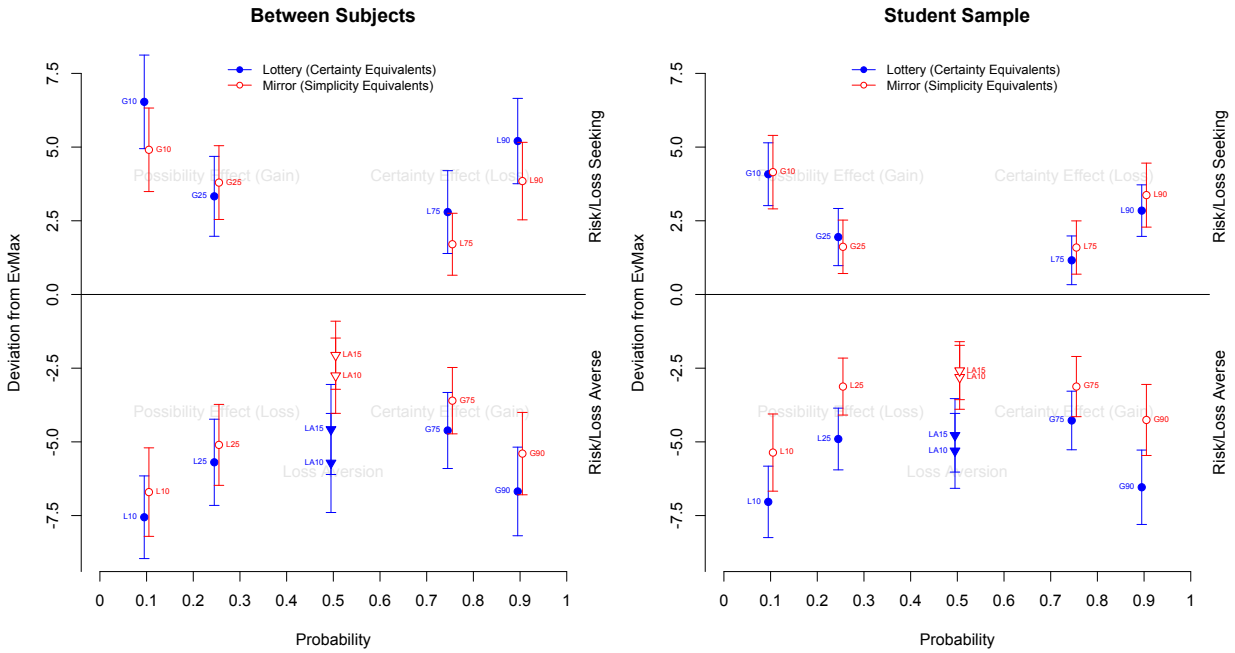


Figure 7: The pattern in between-subjects comparison (left panel) and student lab sample (right panel). *Notes: Details of the Figure are as described in caption to Figure 3.*

mation 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. The left panel of Figure 7 reconstructs Figure 3 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 the 1% level via Wilcoxon tests) from expected value in the direction of the pattern in both lotteries and mirrors for all lists, and we continue to find using Wilcoxon tests that for most lists these deviations are not significantly greater in lotteries than in mirrors at conventional levels.[32]

A related concern is that the correlations between lotteries and mirrors visualized in Figure 5 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 (i) as just discussed, nearly identical initial behavior occurs across the two treatments before subjects know the other treatment exists and (ii) as discussed above, the correlations between the two treatments in Figure 5 are also nearly identical regardless of the order of treatments. 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

[32]The only exceptions are the two LA lists where deviations are greater for lotteries than for mirrors.

place.

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

Result 6 *The 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 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 7 shows) these subjects continue to display the pattern strongly suggests that such "mis-importation" of probabilistic 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 in economics as well.

4.5 Robustness: Stakes and Subject Pool

A second potential concern is that these results might be 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 forced by 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. 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). Additional details on this experiment and minor differences between it and our main treatment are provided in Online Appendix C.

The right hand panel of Figure 7 plots the results of 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 EvMax 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 there is a somewhat larger "gap" between the strength of the pattern in mirrors and lotteries: errors in the direction of the pattern are overall 75% as large in Mirrors as Lotteries (compared to 91% in the main dataset). Recalculating the taxonomy from Section 4.3, we find that the difference is driven by a modest decrease in subjects' sensitivity to complexity and a modest increase in sensitivity to risk in this sample. Slightly fewer (66% of subjects, down from 80%) are typed as complexity-sensitive, and slightly more (48%, up from 36%) are risk-sensitive, producing an increase in the gap between the severity of the pattern in lotteries and mirrors. Nonetheless, our main finding – that the pattern arises with strength in mirrors and that this predicts the pattern in lotteries – continues to hold in this sample.

Result 7 *A robustness sample of university students with increased training and quintupled incentives produces results similar to those in the main dataset.*

4.6 Robustness: Computational Difficulty

A third explanation for our results is that they are an artifact of the computational difficulty of evaluating the specific set of lotteries/mirrors we implemented in the main design. For instance, we describe lotteries/mirrors using 100 states (i.e., 100 boxes) and perhaps it is difficult to reason about this many outcomes. Likewise, the non-zero payment in our fourfold lists was \$25 which does not produce whole-number expected values in any of our lotteries – perhaps this makes it unnecessarily difficult to assess true value in mirrors and the expected value in lotteries.

To evaluate the role this kind of computational difficulty plays in driving our results, we ran a robustness treatment we call "Easier" (with 90 subjects, details are provided in Online Appendix C) in which we reduce or remove these mathematical difficulties. 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 Easier treatment we shrink the state space from 100 boxes to 4 boxes *without changing the underlying probabilities.* Doing

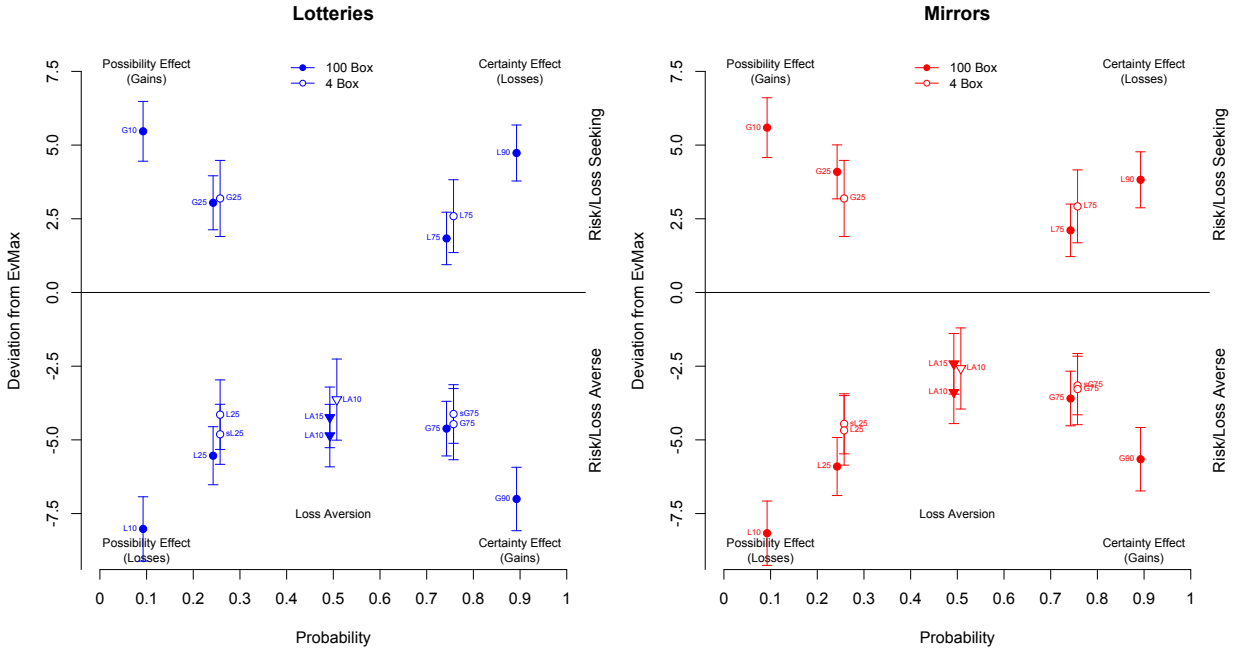


Figure 8: Results from the Easier treatment, overlaid on on results from the main sample. *Notes: Separate panels are provided for Lotteries and Mirrors. Details of the Figure are as described in caption to Figure 3.*

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 Easier treatment we repeat the G25, G50, G75, L25, L50, L75 and LA10 lists but describe them using 4 boxes instead of 100.[33]

Figure 8 plots the results. It includes one panel for lotteries and another for mirrors and in these panels repeats the data pictured in Figure 3 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 LA10 lists from the Easier 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 EvMax 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 lists for only one of the ten comparisons (L25 mirrors). We conclude that the number of "states" has at most a secondary effect on the appearance and severity of the

[33]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 (2020) 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.

pattern.

A second potential source of computational difficulties in the main dataset is the use of a non-zero payoff of \$25 in the fourfold lists, 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 Easier treatment a repetition of lists 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 lists sL25 and sG75 and plot valuations from these lists in Figure 8. We find no overall reduction in the severity of the pattern. Indeed the largest difference is a slight worsening of the pattern in sL25 lotteries relative to L25 lotteries. Again, this suggests that mere mathematical 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 computationally 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 simple valuation tasks.

Result 8 *Making valuation tasks computationally easier has only minor effects on the severity of the pattern in mirrors or lotteries.*

4.7 Correlates of the Pattern

We hypothesize that the reason that computational difficulty has so little effect on the severity of the pattern is that subjects are not attempting precise computation in the first place. Subjects, instead, make errors because formally processing the information in a lottery (or mirror) is generically *costly* (or difficult), and many subjects respond to these costs by substituting to casual, intuitive, error-prone approaches to evaluation instead. For instance, instead of calculating expected value precisely (a relatively mentally laborious task), many subjects instead *intuitively* approximate the relative value of the lottery/mirror and the menu of certainty/simplicity equivalents they are asked to consider. Recent work provides direct evidence that complexity produces these kinds of substitutions to simpler procedures: Oprea (2020) shows that humans, indeed, suffer significant costs from implementing formal decision procedures, and Banovetz & Oprea (2022) provides evidence that decision makers respond sensitively to these costs by abandoning relatively complex optimal procedures in favor of simpler-than-optimal alternative procedures.

In Online Appendix B.2, we report an exploratory analysis of auxiliary behavioral data, post-

experiment questions and demographic data that provides some provisional support for this broad interpretation of our findings. There we show that subjects make highly noisy valuations (as measured by inconsistent valuations across repetitions of the same task) and that variation in this noisiness across subjects strongly predicts the severity with which they display the classical pattern. Likewise, measures of inattentiveness (i.e., poor performance in cognitive reflection tests); hasty decision making (i.e., short response times) and high private costs of conducting precise calculation (i.e., relatively weak technical backgrounds) are all positively correlated with the severity of the pattern. Finally, evidence from unincentivized post-experiment questions suggest that subjects are aware of (and perhaps even deliberately choose) these imprecise methods of valuation, and this awareness too predicts prospect-theoretic behavior. Self-reported measures of (i) cognitive uncertainty (uncertainty about the optimality of choices made in Mirrors, Enke & Graeber (2021)), (ii) use of intuitive (rather than precise) valuation strategies (measured using a 100-point Likert scale) and (iii) imperfect attention to probabilities in the descriptions of lotteries/mirrors (again using a 100-point Likert scale) are significantly correlated with the magnitude of the pattern.

Together, these findings suggest that subjects display the pattern in both lotteries and mirrors in large part because they economize on information processing costs by using imprecise, inattentive and noisy valuation strategies. Importantly we find strikingly similar evidence of this substitution in lotteries and mirrors: the correlation of correlation coefficients between pattern-consistent bias and the 13 measures we examine in Online Appendix B.2 in lotteries and mirrors is 0.93, reinforcing our conclusion that the pattern occurs in the two settings due to the same mechanism.

Result 9 *Analysis of auxiliary data suggests that the pattern arises due to subjects' (possibly deliberate) use of informal, imprecise procedures for valuing lotteries and mirrors.*

Of course, this leaves open the question of what exact procedures subjects use to produce the systematic distortions of the classical pattern. Here we reach the limits of what our data can decisively tell us, but as Section 2.2 details, the literature suggests a number of possibilities. Among these, particularly promising-seeming given our auxiliary findings is a class of recently proposed (and so far empirically successful) "noisy coding" models in which agents (i) imprecisely "encode" (i.e., represent) numerical quantities when casually assessing them and (ii) compensate for this noise efficiently by shading valuations towards prior beliefs when "decoding" them again to inform choice, producing (iii) systematic evaluative biases that can resemble phenomena like probability weighting, the fourfold pattern and, even, loss aversion (e.g., Khaw et al. (2021), Steiner & Stewart (2016), Vieider (2022), Enke & Graeber (2021), Frydman & Jin (2021), Glimcher (2022)). Because this class of models explains the pattern as a direct outgrowth of evaluative noise, it seems consistent with our finding that (in both lotteries and mirrors) the pattern is strongly correlated with valuation-noise and and other indices of noisy behavior. However, we caution that (as Section 2.2 suggests) there are other noise-based explanations available and we conclude that further research is needed to fully understand the mechanism by which complexity produces these twin distortions in lotteries and mirrors.

5 Discussion

We can view our results through three equivalent lenses. First, if we remove risk from a lottery but retain its other characteristics, the distinctive empirical fingerprint of prospect theory remains. Thus, risk is not a *necessary* ingredient for producing prospect-theoretic behavior. Second, when we disaggregate a deterministic monetary amount so that the information processing required to evaluate it is like that of a lottery, we find strong evidence of the fourfold pattern and loss aversion. Thus, complexity (the information processing required for valuation) is a *sufficient* ingredient for producing prospect-theoretic behavior. Finally, when we conduct a standard risk-preference elicitation experiment but use standard experimental tools to *induce* risk-neutral preferences, the signature empirical anomalies usually rationalized by prospect theory continue to arise. Thus, we should expect to find significant prospect-theoretic behavior even in perfectly risk neutral decision makers.

Our results therefore suggest that complexity (the information processing required in valuation), rather than risk, is the primary driver of the most important lottery valuation anomalies we've uncovered in behavioral economics. Decomposing the data, we find that the vast majority (91%) of the anomalous valuation occurring in lotteries (which are both risky and complex) occurs also in deterministic mirrors (which are only complex).[34] What's more, the intensity of this pattern covaries strongly between lotteries and mirrors, suggesting that they derive (in large part) from a common source. That common source cannot be risk, which is absent in mirrors, and therefore must be the complexity of information processing, the characteristic the two valuation tasks have in common.[35]

The methodological idea behind our experiment has the advantage that it can crisply benchmark the relative role complexity plays in driving lottery anomalies without committing to a specific

[34] It is possible that this small excess severity of the pattern in lotteries relative to mirrors is evidence of some secondary role for, e.g., risk preferences in driving the pattern. However, as we discuss in Section 2.3, there is good reason to think that this decomposition provides only a lower bound estimate of the role complexity plays in these anomalies. Recent evidence (Martinez-Marquina et al. (2019)) suggests that stochasticity makes information processing more difficult and the task of valuation therefore more complex. To the degree this is true, the residual excess severity of the pattern we find in lotteries relative to mirrors may in fact be a consequence of the fact that risk itself introduces additional complexity to the task of valuation.

[35] A natural question is whether we should expect our findings regarding anomalies in lottery valuation to extend to anomalies in other types of lottery choice (e.g., in binary choices between lotteries). Our data does not tell us directly, but two considerations suggest that we should expect complexity to be similarly implicated in, e.g., binary choice anomalies. First, the method of valuation we use (price lists) are nothing other than collections of binary lottery choices, presented together in an orderly way. Allowing subjects to simultaneously evaluate lottery pairs in this way seems likely (if anything) to *reduce* the scope for boundedly rationality decision-making because it makes it easier to make consistent choices across lottery pairs. Second, and consistent with this, there is evidence that binary lottery choices produce noisier behavior than elicited valuations (McGranaghan et al. 2022) – a tendency that would seem to suggest that complexity has no smaller effect on behavior in these settings than in elicited valuation. Nonetheless, extending our methods to directly study how complexity impacts other types of lottery choice seems like a natural next step in this line of research.

model of complexity. But the model-agnosticity of this method also means it is not well-suited to identifying the precise mechanism by which complexity produces systematic lottery anomalies – a task we leave largely to future work. Nonetheless, our data provides some clues, suggesting that subjects consciously (perhaps even deliberately) use intuitive, error-prone evaluation procedures instead of the precise, deliberate evaluation we often assume in economics – and that evidence of the use of such procedures is predictive of the severity of prospect-theoretic behavior. Strong correlations between (i) apparent prospect-theoretic behavior and (ii) indices of behavioral noise are highly suggestive of a recent (and, so far, empirically fruitful) class of "noisy coding" models but we emphasize that further research is needed to pin down the exact mechanism by which complexity induces these common patterns in risky and riskless settings.

Regardless of the precise mechanism at work, there are two important direct implications of our findings.

First, because our results suggest that these anomalies are mostly not driven by risk, they also suggest they are unlikely to be expressions of true risk preferences – a finding with obvious welfare implications. To the degree they are outgrowths of complexity, behaviors like probability weighting and reference dependence do not primarily reveal the structure of subjects' preferences for risk but instead their aversion to (or incapacity for) careful information processing. Indeed, our results directly suggest that we should *expect* these kinds of behaviors to arise even for vanilla risk neutral decision makers due purely to the distorting influence of complexity.[36] This suggests that policies designed to accommodate these types of behaviors likely enshrine mistakes rather than respond in a welfare-enhancing way to preferences. It also suggests that cognitively-inspired interventions that manage to remove these types of behaviors from risky choice would likely be welfare-enhancing.[37]

Second, the anomalies we've uncovered in lottery valuation in recent decades and the descriptive theories we've built to explain them likely have a far broader scope of application than we've so

[36]Strictly speaking, our methods induce not only risk-neutral preferences but also *loss-neutral* preferences in mirrors. Thus, one interpretation of our finding of loss aversion in mirrors is that loss aversion (like, e.g., probability weighting) is not a preference, but instead a complexity-derived valuation mistake. This interpretation is certainly consistent with recent evidence suggesting that directly-measured loss aversion is uncorrelated with key anomalies often explained by loss averse preferences like the endowment effect (Chapman et al. 2021). However, interestingly, there are some aggregation mistakes that might allow true distaste for loss to contribute to the errors we observe in mirrors. Suppose the mistake subjects make in our loss aversion mirrors is to assess the value as $0.5v(X) + 0.5v(y)$ (where v is a loss-averse value function) rather than $v(0.5X + 0.5Y)$, as is optimal. This is an error in aggregation but it is one that will cause subjects to mistakenly express actual loss averse preferences (actual excess distaste for losses relative to gains) even in settings in which there is no risk of loss. We cannot rule out this alternative possibility, meaning that while it is possible to read our results as evidence against the existence of latent loss-averse preferences, we should be cautious in drawing such conclusions based on our evidence alone.

[37]Of course, our results do not (and are not meant to) rule out the existence of risk preferences or a role for them in explaining choice, particularly at the much larger stakes sizes at which we should expect expected utility preferences to begin to apply (Rabin (2000). Rather, our results show how strongly complexity influences risky choice and that it is this complexity, rather than non-standard preferences, that likely produces many of the most important lottery anomalies we've discovered in our efforts to measure human responses to risk.

far recognized. Valuation of disaggregated objects is, in some sense, the primordial cognitive act described by economic theory, applying in settings ranging from simple consumer choice (where disaggregated bundles have to be compressed into indices of value) to strategic interaction (where disaggregated contingencies need to be similarly compressed) to lottery choice (where it is states that have to be aggregated). Our finding that it is this disaggregation (not risk) that drives anomalous lottery valuations, suggests that systematic distortions like those we've documented in lotteries over the last half century may be generic in economic behavior, applying across many deterministic choice domains as well. It may, in other words, be that the descriptive theories we've built of human evaluation of risk in the last few decades are, to a large extent, in fact first steps in the task of building descriptive models of human evaluation of complex things. As such, their value to economics may be more significant than we've so far appereciated.

References

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.

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.

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.

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.

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.

Brown, A., Imai, T., Vieider, F. & Camerer, C. F. (2022), 'Meta-analysis of empirical estimates of loss-aversion', *Journal of Economic Literature* p. forthcoming.

Camara, M. (2021), 'Computationally tractable choice', *Unpublished Manuscript* .

Camerer, C. (2005), 'Three cheers – psychological, theoretical, empirical – for loss aversion', *Journal of Marketing Research* **42**, 129–133.

Chapman, J., Dean, M., Ortoleva, P., Snowberg, E. & Camerer, C. (2021), 'On the relation between willingness to accept and willingness to pay', *Unpublished Manuscript* .

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

Enke, B. & Graeber, T. (2021), 'Cognitive uncertainty', **manuscript**.

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

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. (2021), 'Efficient coding and risky choice', *Quarterly Journal of Economics* **forthcoming**.

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

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

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.

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

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.

Kool, W. & Botvinick, M. (2018), 'Mental Labour', *Nature Human Behavior* **2**, 899–908.

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

McGranaghan, C., Nielsen, K., O'Donoghue, T., Somerville, J. & Sprenger, C. D. (2022), 'Distinguishing common ratio preferences from common ratio effects using paired valuation tasks', *Unpublished Manuscript* .

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

O'Donoghue, T. & Somerville, J. (2018), 'Modeling risk aversion in economics', *Journal of Economic Perspectives* **32**(2), 91–114.

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

Puri, I. (2020), 'Preference for simplicity'. Mimeo, Available at SSRN: https://ssrn.com/abstract=3253494.

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

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

Smith, V. (1962), 'An experimental study of competitive market behavior', *Journal of Political Economy* **70**(3), 111–137.

Smith, V. (1976), 'Experimental economics: Induced value theory', *American Economic Review, P and P* **66**(2), 274–279.

Stango, V. & Zinman, J. (2022), 'We are all behavioral, more or less: a taxonomy of consumer decision-making.', *Review of Economic Studies* .

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

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

Vieider, F. (2022), 'Decisions under uncertainty as bayesian inference', *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. (2020), 'Modeling imprecision perception, valuation and choice', *Annual Review of Economics* **12**.

Online Appendices

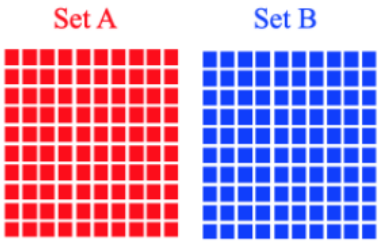
A Instructions to Subjects

A.1 Beginning of Instructions

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

A.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 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

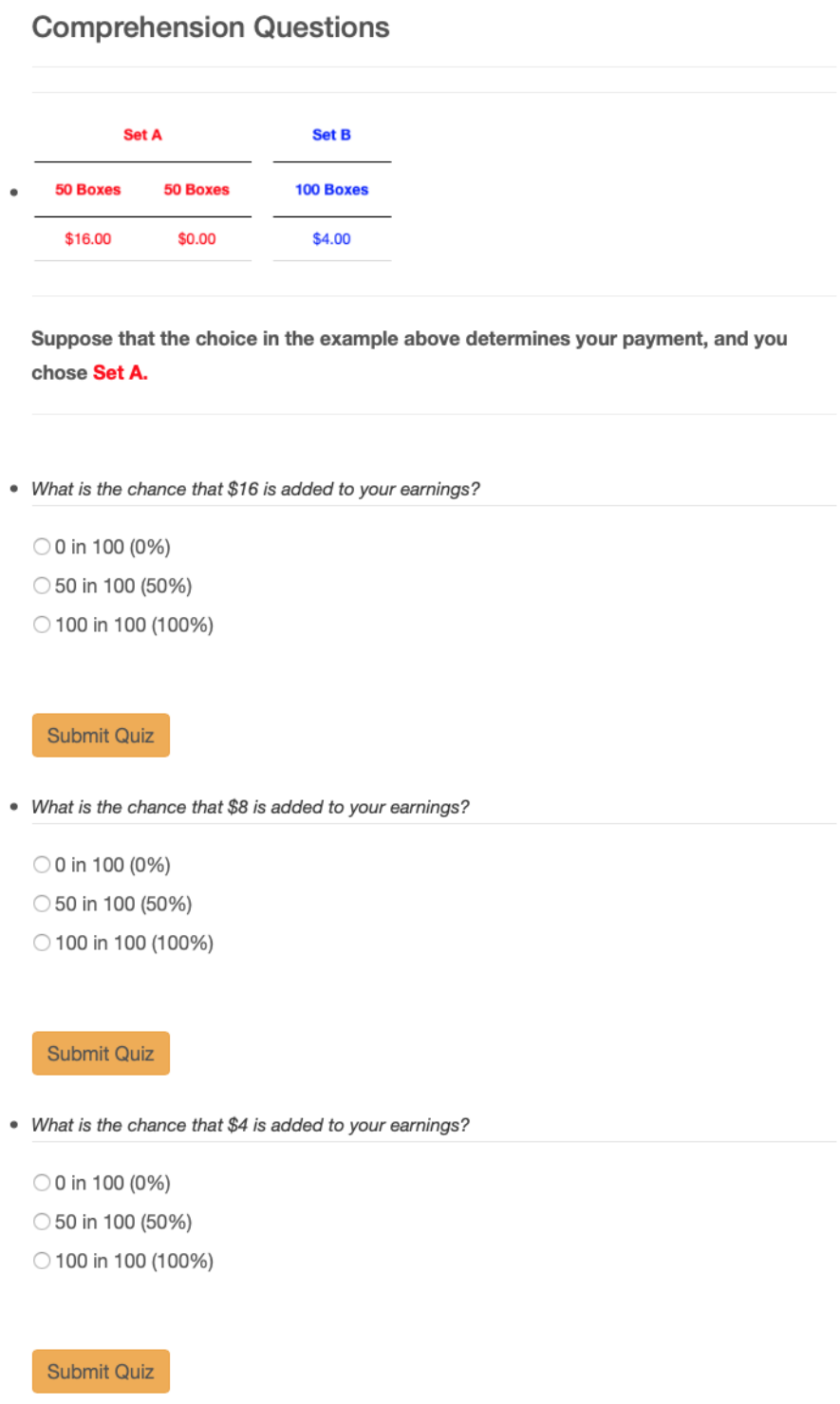
- Example: In the example below, there are also **100** boxes in each Set. For Set A, **50** boxes contain -$8.00 and **50** of them contain $0.00. If you choose Set A, the computer will pay you (-$8×50 + $0×50)/100 = -$4 for your choice; it will therefore subtract $4 from your initial amount of money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

A.3 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.



A.4 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 **50** boxes containing $10 and **50** 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 Additional Analysis

B.1 More on Correlations Between Lotteries and Mirrors

Figure 9 repeats the analysis reported in the right-hand panel of Figure 5 separately for the fourfold pattern and loss aversion. In particular, the left hand panel plots mean bias measured in "fourfold

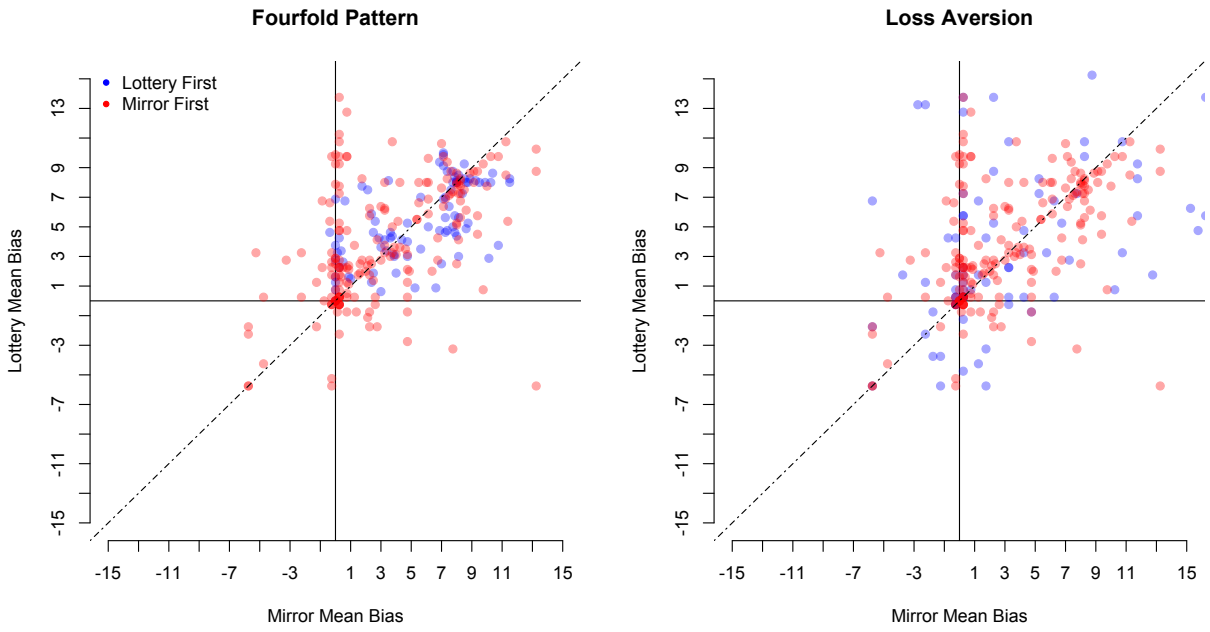


Figure 9: Deviations from expected value maximizing choices (in core lists) in mirrors (x-axis) versus lotteries (y-axis), by subject. *Notes: Each dot represents a subject. "Lottery First" designates subjects who were initially assigned lotteries (Mirror First is the reverse). The left panel plots mean "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" (LA10 and LA15).*

lists" (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" (LA10 and LA15). For fourfold lists (left hand panel) we measure a lottery-mirror correlation of $\rho = 0.62$ ($p < 0.001$) and for loss aversion (right hand panel) we measure $\rho = 0.40$ ($p < 0.001$).

B.2 Correlates of the Pattern

In order to get 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 the pattern 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 the classical pattern in our main treatment.

First, we gathered several behavioral measures. Most importantly, we *repeated* the G50 and L50 lists in both lotteries and mirrors (for half of our subjects), 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 (though difficult to interpret) measure of effort. Similarly, we studied how many **revisions** subjects made (the

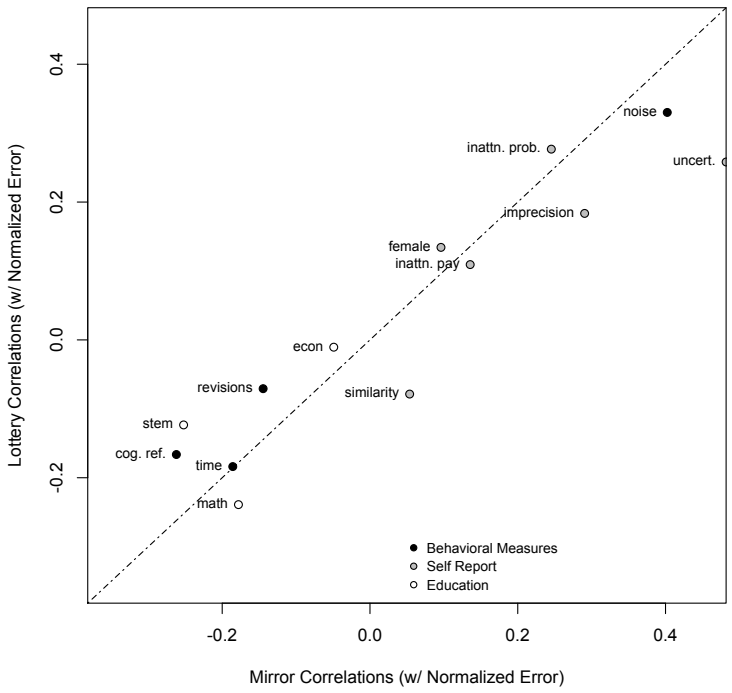


Figure 10: Correlates predictors (labeled dots) with pattern-consistent bias in mirrors (x-axis) and lotteries (y-axis). *Notes: Each dot is a different predictor and the x- and y-axes show the correlation of each predictor with pattern-consistent bias in mirrors and lotteries, respectively.*

number of times they changed their decision before submission) in the average task, a possible behavioral measure of the care subjects took in decision-making. Finally, after the main experiment 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 informed subjects that one choice in each Mirror task was optimal in that it maximized earnings and we asked subjects how confident they were (in percentage terms) that they made this payoff-maximizing choice. Measures of **cognitive uncertainty** like this have proved predictive of the fourfold pattern and other anomalies in recent work (e.g., Enke & Graeber (2021)).[38] 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

[38]We also included a question about subjects' cognitive uncertainty about lottery choices, but unavoidably this was a much more vaguely worded question. Instead of asking whether subjects chose a payoff-maximizing decision (a relatively objective question), we had to ask them the chances they made the "best choice." Unsurprisingly perhaps, we found this was a much noisier measure and much more weakly predictive of behavior in both lotteries and mirrors and so we do not make use of that question here.

evaluating lotteries/mirrors. This gives us measures of **inattention to payoffs** and **inattention to probabilities** for each subjects 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" in their valuations, again for both lotteries and mirrors. This gives us a self-reported measure of **imprecision** of decisions.[39] Finally, we include a four point Likert scale (discussed in the paper) describing how **similar** subjects' strategies were in lotteries and mirrors.

Third, we gathered several demographic measures, focused on measures that proxy for cognitive sophistication and perhaps subjects' *ability* to, e.g., set up and compute an expected value calculation. 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 10, we estimate the Pearson correlation between each measure and the mean pattern-consistent bias (errors coded to be positive if they run in a prospect-theoretic direction and negative otherwise) in mirrors and lotteries, plotting the correlation coefficient ρ for (i) mirrors on the x-axis and (ii) lotteries on the y-axis. 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 =0.93 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 not special to risk.

Result 10 *There is a strong similarity in the predictors of the pattern in lotteries and mirrors.*

Second, the strongest and most consistent predictor of the pattern in lotteries and mirrors is simply the noisiness of subjects' own decisions in a separate valuation task. The mean subject makes valuations that are \$2.60 different on average when repeating the same valuation choice in lotteries L50 and G50, suggesting that the average subject makes noisy, imprecise valuations. This is remarkable given that 50/50 lotteries seem particularly computationally simple and this seems to point to the deliberate use of intuitive rather than explicit valuation strategies. The degree of this noise (the mean size of the absolute deviation between choices in identical tasks) strongly predicts the severity of the pattern. Indeed, as Figure 10 shows, subjects who on average make more inconsistent choices in identical problems are significantly more likely to display prospect-theoretic

[39] A coding error caused the sliders in this task to be mislabeled "Little Attention" for choices of 0 and "A Lot of Attention" for choices of 100. However the instructions above the slider are very clear that scores towards zero represent guesses and to the right precise decisions, meaning subjects were unlikely to be confused. Nonetheless this mislabeling is worth bearing in mind in interpreting the results.

behavior in both mirrors and lotteries. (Importantly, this is an out-of-sample test – our measure of prospect-theoretic error does not include data from the L50 and G50 lists.)

Third, there is evidence from other behavioral measures that this noisy behavior may be a consequence of the use of intuitive or inattentive (rather than precise, deliberate) valuation strategies. Performance on the cognitive reflection test – which is meant to measure deliberative over intuitive decision making – is significantly negatively correlated to the intensity of the pattern: careful, reflective subjects are less prospect-theoretic. Moreover, subjects who make slower choices (higher response time) display the pattern less intensively, perhaps suggesting that the pattern derives from the use of lower effort valuation strategies.

Fourth, Figure 10 shows that self-reported imperfection and imprecision in choice is strongly predictive of the pattern in both lotteries and mirrors. The average subject believes there is a 43% chance she made a valuation mistake in the average mirror, but there is great variation in this cognitive uncertainty measure and this variation is strongly related to the pattern in both lotteries and mirrors. This suggests that subjects are not only using error-prone strategies, but they also seem to be aware of this fact and this knowledge too is correlated with prospect-theoretic behavior. Similarly, there is evidence that subjects knowingly use imperfectly precise strategies in valuation and that they rely on intuitive guesswork to some degree in their decision-making. On a 100-point scale between "guessed" and "made a precise (exact) decision" the average subject described her decisions as only 76% as precise as they might have been in lotteries (80% in mirrors). Variation in this self-report of precision of valuations is significantly predictive of prospect-theoretic behavior, with more imprecise subjects more severely prospect-theoretic in both lotteries and mirrors. Together these results suggest that many subjects are aware of imperfections in their choice and may even be using imperfectly precise strategies deliberately.

Fifth, subjects' qualitative reports of the amount of attention they paid to probabilities in lotteries is significantly negatively correlated with the severity of the pattern. However, the relationship is much weaker and not statistically significant for self-reported "attention to payments." This suggests that inattention is a potentially important driver of prospect-theoretic behavior and that careless attention to variation in probabilities may be the more consequential inattention.

Finally, we find that mathematically sophisticated people – people who majored in technical STEM fields or who were exposed to college-level mathematical training – are less likely to show evidence of the pattern in both lotteries and mirrors. Such subjects may be better practiced at and therefore suffer lower costs from implementing precise mathematical evaluations of lotteries and mirrors, influencing the decision to use these strategies rather than intuitive valuation strategies. By contrast, economics training has no significant predictive power. Likewise, gender, the similarity of choices subjects make across lotteries and mirrors and the number of revisions subjects made in the choice process have little predictive power.

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

Result 11 *Noise in decision making is the strongest predictor of the pattern in both lotteries and mirrors. Subjects' self reports suggest that beliefs about one's decision quality, imprecision in subjects' decision-making strategy and inattention to probabilities the pattern in both lotteries and mirrors. Mathematically sophisticated subjects are less likely to show evidence of the pattern in both lotteries and mirrors.*

Together, these results seem consistent with the idea that subjects knowingly use imprecise, intuitive strategies to value disaggregated objects like lotteries and mirrors and that this is decision is an important driver of the classical pattern.

C Robustness Treatments

C.1 UCSB Student Sessions

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 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.

As highlighted in the body of the paper, 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 Online Appendix A 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 gain 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 Online Appendix B.2). 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.

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

C.2 Easier Treatment

The Easier treatment was conducted in May of 2022 using 90 subjects on Prolific. The treatment repeated Lotteries G25, G50, G75, L25, L50, L75 and LA10. 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 states (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 Easier 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 11 gives an example of how this framing looked on decision screens by providing a screenshot from the G75 list.

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 **one** of the boxes from the Set you selected, **randomly** chosen by the computer.

	Set A	Set B	
Version	4 Boxes	3 Boxes	1 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 11: Screenshot from a Mirror task (list G75) in the Easier treatment. *Notes: In Lottery tasks, the screen is identical except for the text in green which would instead read "...plus the value of one of the boxes from the Set you selected, randomly chosen by the computer."*