

QUANTIFYING LOTTERY CHOICE COMPLEXITY*

Benjamin Enke Cassidy Shubatt

July 31, 2023

Abstract

We develop interpretable, quantitative indices of the objective and subjective complexity of lottery choice problems that can be computed for any standard dataset. These indices capture the predicted error rate in identifying the lottery with the highest expected value, where the predictions are computed as convex combinations of choice set features. The by far most important complexity feature is a measure of the excess dissimilarity of the lotteries' cumulative distribution functions. Using the indices, we study behavioral responses to complexity out-of-sample across one million decisions in 11,000 unique experimental choice problems. Complexity makes choices substantially noisier, which can generate systematic biases in revealed preference measures such as spurious risk aversion or spurious complexity aversion. These effects are very large, to the degree that complexity explains a larger fraction of estimated choice errors than proximity to indifference does. Accounting for complexity in structural estimations improves model fit substantially.

Keywords: Complexity, choice under risk, cognitive uncertainty, experiments

*Sebastian Redl and Anna Valyogos provided outstanding research assistance. We thank Cary Frydman, Ryan Oprea, Peter Robertson, Jesse Shapiro, Jeffrey Yang and Florian Zimmermann for helpful comments and discussions. Enke gratefully acknowledges funding from the Mind, Brain and Behavior Initiative at Harvard. Enke: Harvard University, Department of Economics, and NBER, enke@fas.harvard.edu; Shubatt: Harvard University, Department of Economics, cshubatt@g.harvard.edu.

1 Introduction

Much recent research emphasizes that decision problems involving risk are often complex, meaning that they require a non-trivial degree of cognitive processing to aggregate problem inputs into a decision. Understanding complexity is believed to be important for two broad reasons. First, people may systematically undervalue lotteries that are complex (e.g., Huck and Weizsäcker, 1999; Iyengar and Kamenica, 2010; Bernheim and Sprenger, 2020; Puri, 2022), or even entirely select out of complex decisions (e.g., Enke et al., 2023b). Second, complexity may cause cognitive noise and resulting systematic distortions such as probability weighting or violations of canonical axioms (e.g., Enke and Graeber, 2023; Oprea, 2022; Frydman and Jin, 2021, 2023; Khaw et al., 2021; Vieider, 2022; Nielsen and Rehbeck, 2022; McGranaghan et al., 2022; Hey, 1995).

While this literature argues for the importance of complexity, it also begs the question: which lottery choice problems are more or less complex? The typical approach in the literature – including in our own prior work – is to proceed on a heuristic case-by-case basis: the researcher intuits a specific complexity feature (e.g. number of distinct payouts) and investigates how that feature affects behavior. Yet, ideally, one would like to quantify the overall complexity of a choice set, and to study behavioral responses to such a composite notion of complexity. The ability to quantify complexity might help in (i) designing simple choice problems, or offering advice on complex ones; (ii) predicting complexity-driven choice errors and resulting biases in revealed preference measures; (iii) quantifying the importance of complexity relative to non-standard preferences; and (iv) comparing and evaluating empirical results across papers, when different researchers construct lottery choice problems with varying levels of complexity.

In this paper, we make progress by developing an empirical mapping from choice set features to indices of lottery choice complexity, which can be computed for any standard dataset. We use these indices to study behavioral responses to complexity and document that complexity-dependent noise is quantitatively even more important than the distortions captured by popular behavioral models such as prospect theory.

Development of indices of lottery choice complexity. We understand the complexity of a choice problem as the cognitive difficulty of identifying one's preferred choice option. Complexity is latent and unobserved. We, hence, develop measures of *revealed complexity* that are based on errors (mistakes). In choice data, even revealed complexity is generally unobserved: the researcher does not know the decision-maker's utility function and, as a result, cannot distinguish errors from non-standard preferences. To circumvent this problem, we propose to quantify the (revealed) complexity of a lottery choice set as *the predicted error rate in identifying the lottery with the highest expected*

value, where the prediction is computed as a function of choice set features.

To develop such a complexity index, we need to know which choice set features increase error rates. We implement experimental lottery problems in which subjects receive a fixed bonus if and only if they correctly indicate the lottery with the highest expected value. Because *ex ante* we do not know which choice set features drive complexity, we design a large-scale experiment in which subjects make decisions involving 2,200 quasi-randomly generated problems.

The complexity index that is derived from these data does not rest on the assumption that people are necessarily risk neutral. We also don't assume that maximizing expected value is as complex as maximizing expected utility. Rather, the thought is that those choice set features that make it difficult to gauge expected values also make it hard to gauge expected utilities (or other subjective values) because the cognitive process of aggregating probabilities and payouts (“aggregation complexity”) is similar across the two tasks. Based on this idea, we formally state an identifying assumption for the true revealed complexity of standard lottery choice problems to be a monotone function of our complexity index.

We develop two indices, both of which are based on the expected values task: (i) objective problem complexity (*OPC*), which captures predicted error rates; and (ii) subjective problem complexity (*SPC*), which captures predicted cognitive uncertainty (the predicted subjective probability of making an error). In selecting the choice set features that enter these indices, we need to balance the tradeoff between completeness and interpretability that is inherent to any predictive index. We use an iterative procedure according to which (i) we implement exploratory LASSO regressions to identify the most predictive features, and then (ii) construct handcrafted indices based on few and easily interpretable features that closely approximate the machine learning indices.

Both *OPC* and *SPC* comprise two classes of features: (i) the proximity of the (aggregated) expected values of the lotteries in the set; and (ii) features that affect the complexity of aggregation in the first place. Proximity to indifference only makes up a small share of our indices because some features that capture aggregation complexity turn out to matter substantially more. Both *OPC* and *SPC* increase in the absence of (first-order stochastic) dominance relationships, the magnitude of payouts, the lotteries' support, the presence of losses and compound probabilities.

While previous work on lottery complexity has often focused on these familiar features, it turns out that the most important feature by far is what we call the *excess dissimilarity* of the lotteries in a set, by which we mean the degree to which lotteries are dissimilar from each other above and beyond their difference in expected value. Intuitively, we compute dissimilarity by overlaying the cumulative distribution functions of two lotteries and calculating the area between the two. Excess dissimilarity is then given

by dissimilarity minus the (absolute) expected values difference. This measure could be interpreted as reflecting a cognitive process whereby people choose among lotteries by first putting them into a correlated common state-space (matching q -th quantile outcomes to define the common states) and then comparing the payoffs state-by-state. In our data, excess dissimilarity is strongly predictive of error rates and cognitive uncertainty, and drives a majority of the variation in both *OPC* and *SPC*.

We also construct indices of the complexity of *individual lotteries*. Again, the most important feature reflects the degree to which a lottery is disaggregated, captured by the lottery's variance.

To provide evidence that the application of the complexity indices to real lottery choice problems is meaningful, we implement standard choice tasks and measure subjects' cognitive uncertainty about each decision. Our complexity indices are strongly predictive of variation in cognitive uncertainty across problems, suggesting that – as our indices implicitly assume – there is indeed a tight link between the difficulty of maximizing expected value and the difficulty of maximizing expected utility.

Behavioral responses I: Choice sets. With our indices of problem complexity in hand, we study behavioral responses to complexity out-of-sample in traditional binary lottery choice problems. We both collect our own dataset on risky decisions and re-analyze the most comprehensive dataset on binary lottery choice ever collected (Peterson et al., 2021). In total, we evaluate one million decisions across 11,000 unique choice problems.

The indices of problem complexity are strongly predictive of proxies for choice noise. For example, the correlation between *OPC* and the frequency of within-subject choice inconsistencies in repeated elicitations of the same problem is $r = 0.57$. Problem complexity also strongly predicts the compression of choice rates to 50-50, which we interpret as reflecting choice noise. The magnitude of these effects is very large: both the frequency of picking the option with the lower expected value and the frequency of choice inconsistencies increase by more than 35 percentage points going from very low to very high complexity. We document that almost the entirety of these relationships reflect the effects of aggregation complexity rather than the proximity of expected values.

We benchmark magnitudes against a variable that drives the frequency of mistakes in random choice models: the expected utility difference between the lotteries in the set. We find that *OPC* explains 3-25 times more of the across-problem variation in choice inconsistencies and other proxies for choice errors than the estimated proximity to indifference in a prospect theory model does.

Behavioral responses II: Individual lotteries. How do people respond to the complexity of individual lotteries? Ex ante, there are at least two plausible possibilities: higher

noisiness (heteroscedasticity) and complexity aversion. Our evidence overwhelmingly points in the direction of complexity-dependent noise. We discuss how such complexity-dependent noise can spuriously generate or contribute to behavior that looks like small-stakes risk aversion or like risk seeking. The basic insight is that lottery features such as variance have both a direct effect on choice through people's preferences and an indirect effect because they affect the magnitude of decision noise. If sufficiently strong, this indirect effect can generate any arbitrary bias in favor of or against a lottery feature.

To illustrate, consider how lottery variance affects choice when (1) decision-makers are risk neutral and (2) complexity increases with lottery variance. First, consider the choice between a lottery and a *lower-value* safe payment. As the lottery variance increases, the choice becomes more complex (thus noisier), pushing the rate of choosing the lottery *down* towards 50%. Thus, we observe seeming risk-aversion. If, instead, the decision-maker must choose between a lottery and a *higher-value* safe payment, then higher lottery variance (and resulting noise) pushes the rate of choosing the lottery *up* towards 50%, creating seeming risk-seeking behavior. Analogous arguments can be made to show that people can spuriously appear complexity averse or complexity seeking, again purely as a result of complexity-driven heteroscedasticity in combination with an asymmetric selection of decision problems.

The takeaway from this analysis is that complexity-dependent noise does not “cancel out”. Instead, as in other recent work, *even mean-zero noise in expected utility can reliably produce bias* when researchers don't construct the set of choice problems in a symmetric way. We do not insist that genuine small-stakes risk aversion or complexity aversion do not exist. Rather, we emphasize that the differential noisiness caused by variation in complexity is so strong that it can entirely override any true aversion that may exist.

Structural estimations. How does complexity-dependent noise impact structural estimations? To estimate this, we allow the noise parameter in a logit model to be a function of complexity, which amounts to introducing one additional parameter. This generates an increase in model fit of 28%. In our dataset, this increase is even larger than the combined increase resulting from all of prospect theory (value function curvature, loss aversion and probability weighting).

We discuss for which types of decision problems standard approaches to estimating expected utility, prospect theory or salience theory deliver systematically wrong predictions because complexity is ignored. Intuitively, these models treat all choice problems as if they had the same level of complexity, such that they dramatically underpredict the probability that people choose the (estimated) higher value option when complexity is low, but strongly overpredicts it when complexity is high. Structurally incorporating our complexity index into the estimation almost entirely eliminates this problem.

Contribution and relation to prior work. We view this paper as making two main contributions. First, we develop the first comprehensive indices of the objective and subjective complexity of lottery choice sets and of individual lotteries. These indices are transparent and defined on objective choice set features, making them amenable to be computed in a standardized way across datasets. We make available both code and an online tool that automate this process. In Section 7, we discuss some of the potential applications of the indices: (i) the possibility to put experiments by different researchers on a common scale to evaluate their complexity, hence increasing transparency and comparability; (ii) to predict the direction and magnitude of biases in revealed preference measures; and (iii) to standardly account for complexity in structural estimations.

Our second contribution is to study behavioral responses to complexity in a dataset that is orders of magnitude larger and more comprehensive than typical experimental datasets. Our results suggest that the predominant consequence of complexity in binary choice is higher noise, rather than directional aversion. Because noise can cause bias, complexity nonetheless generates systematic anomalies.

Our paper ties into the interrelated literatures on complexity and noise in lottery choice. Appendix Table 5 presents an overview of experiments that isolate specific complexity features. The bottom line is that multiple papers have documented how features such as a lottery's support or payout magnitudes can produce aversion and / or higher noisiness. Such noise can confound the identification of risk preferences and their linkages with demographics (e.g., Andersson et al., 2016; Gillen et al., 2019; Vieider, 2018; McGranaghan et al., 2022). Our contribution to this literature is to offer a composite measure of complexity and to assess how people respond to overall complexity.

A small set of theoretical contributions have proposed that lottery complexity and / or noisiness depend on problem features such as entropy (Verstyuk, 2016), payout magnitudes (Khaw et al., 2021, 2022), support (Puri, 2022), or importance-weighted support (Gabaix and Graeber, 2023). In contrast to these contributions, we quantify complexity in a data-driven way. Our complexity metrics differ substantially from theirs, both because we construct a composite measure of complexity and because our index includes excess dissimilarity as a main feature.

Our empirical index of lottery choice complexity is also related to work that uses self-reported cognitive uncertainty to measure the cognitive difficulty of lottery choice (Enke and Graeber, 2023). In comparison to this measure, our indices are based on objective performance in an analogous value comparison task. This said, we find that our complexity indices are strongly predictive of cognitive uncertainty.

Section 2 lays out a conceptual framework. Section 3 describes the data we rely on and Section 4 develops the complexity indices. Section 5 discusses behavioral responses to complexity and Section 6 presents structural estimations. Section 7 concludes.

2 Conceptual Framework

2.1 Terminology and Identifying Assumption

Consider choice sets comprising two lotteries each and denote by $d_c \in \{A, B\}$ the decision maker's (DM's) choice. We denote by $EU(x)$ the DM's *true* expected utility from a lottery, where we allow "utility" to include any form of non-standard preferences such as loss aversion. We view distortions of objective probabilities as reflecting errors rather than non-standard preferences.

For the purposes of this paper, we informally define the objective complexity of a choice problem as the latent cognitive difficulty of identifying one's preferred object. Because this latent difficulty is unobservable, we define the *revealed* objective complexity (ROC) of a choice problem as the probability that the decision maker chooses the option that does not maximize her true expected utility.

$$ROC_{A,B} = P(d_c \notin \underset{x_A, x_B}{\operatorname{argmax}} EU(x)) \quad (1)$$

Because the researcher usually does not know the DM's objective function, revealed objective complexity cannot be directly inferred from choice data. To overcome this problem, we consider a second, ancillary decision problem in which the DM is tasked with identifying which lottery has the highest expected value. We denote the DM's decision by d_s . In this task, the objective function is known, thus we can identify errors. We define a second revealed complexity metric based on errors in an expected values task.

$$ROC_{A,B}^{EV} = P(d_s \notin \underset{x_A, x_B}{\operatorname{argmax}} EV(x)) \quad (2)$$

A main idea behind this paper is that choice errors arise in large part due to the latent complexity of aggregating probabilities and payouts ("aggregation complexity"). Therefore, we expect a close link between the complexity of a real choice problem and the complexity of the corresponding expected values problem. Aggregation complexity similarly arises in a lottery choice and an expected values problem because the aggregated value of a lottery is not transparent to real decision makers but requires cognitive processing to combine multiple probabilities and payouts into a decision (e.g., Oprea, 2022; Enke and Graeber, 2023). Our main identifying assumption is that the frequency of errors in the expected values task is predictive of errors in the choice task,

$$ROC_{A,B} = f(ROC_{A,B}^{EV}), \quad (3)$$

with $f(\cdot)$ a monotone increasing function. This identification assumption does not re-

quire that people are necessarily risk neutral. It also doesn't require that maximizing expected value is as difficult as maximizing expected utility. Instead, our identifying assumption is that the choice set features that make it difficult to gauge expected value also make it hard to gauge expected utility, producing a link between error rates in the two different problems.¹ Below, we will provide indirect evidence for this, by showing that error rates in expected values problems are strongly predictive of plausible proxies for choice errors.

2.2 Empirical Implementation

Indices of feature-predicted error rates. In the above, the revealed complexity of a choice problem is approximated by the error rate in the analogous expected values task. Because it is impractical for researchers to always implement an expected values problem to quantify the complexity of their choice problem of interest, we leverage the idea that error rates in the expected values problem can be *predicted* based on choice set features. This is attractive because once a complexity index is defined based on objective features of the set, it can be easily computed for any standard dataset.

Denote by $f^{A,B}$ an $(N+1)$ -dimensional vector of features of the choice set, where the zeroth element is a constant. Denote by $\epsilon_{A,B}$ a mean-zero disturbance term.

Definition 1. *The objective problem complexity of choice set $\{C,D\}$ is given by*

$$OPC_{C,D} := \sum_{i=0}^N \hat{\beta}_i f_i^{C,D} + \sum_{j=1}^2 \hat{\gamma}_j |EV(C) - EV(D)|^j, \quad (4)$$

where the vectors $\hat{\beta}$ and $\hat{\gamma}$ are estimated from the error rates in a sample of expected values problems using OLS:

$$ROC_{A,B}^{EV} = \sum_{i=0}^N \beta_i f_i^{A,B} + \sum_{j=1}^2 \gamma_j |EV(A) - EV(B)|^j + \epsilon_{A,B}. \quad (5)$$

This index has a simple interpretation once it is applied to standard lottery choice problems: it captures the error rate in the analogous expected values problem that is predicted by the choice set features.² OPC is, hence, an index of the predicted *revealed* complexity of a choice problem, rather than of the latent complexity of identifying one's preferred object as such.

¹An additional motivation is that if one accepts the premise that a complexity metric should be independent of the utility function, experimentally inducing risk neutrality should not affect the metric.

² Measurement error in the problem-level estimates of error rates (e.g. resulting from finite samples) does not affect the statistical unbiasedness of OPC because measurement error in the dependent variable does not create attenuation bias.

As discussed above, *OPC* comprises two components: features that capture the proximity of the aggregated (expected) values, and features that capture the complexity of aggregation. *OPC* does not allow the effect of the “aggregation complexity features” f_i to depend on the expected values distance. This is clearly a simplification. The aggregation complexity features may plausibly have a smaller effect on error rates either when the DM is very close to being indifferent or when the DM is far enough from indifference that the choice is obvious. We expect our simple index to perform well in problems for which the DM is neither extremely close nor very far from indifference. Below in Section 2.3, we develop an alternative index of problem complexity that is independent of the proximity of the aggregated values (but requires stronger structural assumptions on the decision errors).

Applying the complexity index to choice data. Researchers may be interested purely in predicting to what extent choice behavior y (e.g., choice inconsistencies) reflects errors, regardless of whether these errors reflect proximity of the aggregated values or aggregation complexity. In this case, *OPC* can be linked to choice behavior using univariate regressions. On the other hand, if the researcher is interested in isolating how a more specific component of *OPC* affects choices, multivariate regressions that control for relevant features are called for. In particular, we anticipate that researchers may be interested in understanding specifically the role of aggregation complexity. In this case, the appropriate estimating equation includes controls for proximity and is given by

$$y_{C,D} = \alpha + \theta \text{ } OPC_{C,D} + \sum_{j=1}^2 \omega_j |EV(C) - EV(D)|^j + \epsilon_{C,D}. \quad (6)$$

2.3 Structural Analyses: Indices of Predicted Logit Variance

In all of the above, we define revealed complexity as error rates, without taking a stance on whether errors reflect noise or systematic bias. In Section 4, we will show that the vast majority of the across-problem variation in errors in expected values problems reflects variation in noisiness. This enables researchers to incorporate complexity considerations into structural analyses, simply by allowing the error variance in a random choice model to depend on complexity. For instance, suppose that choice probabilities in a traditional binary choice problem can be characterized using the logit model:

$$P(A) = F(EU(A) - EU(B); \eta) = \frac{1}{1 + e^{-\eta [EU(A) - EU(B)]}}, \quad (7)$$

where η is the conventional responsiveness (precision) parameter. In this model, unlike in our reduced-form indices developed above, aggregation complexity and resulting

noise (captured by η) can be separated from proximity to indifference. To this end, we define the revealed aggregation complexity of a choice problem as the inverse of the decision-maker's precision for that problem. Inverting the logit CDF,

$$s_{A,B} := \frac{1}{\eta_{A,B}} = \frac{EU(B) - EU(A)}{\ln\left(\frac{1}{P(A)} - 1\right)}. \quad (8)$$

Again, we cannot observe revealed aggregation complexity in choice data since we do not know the utility function. Therefore, as before, we define an analogous measure in the expected values task, where we can directly observe the relevant value difference.

$$s_{A,B}^{EV} := \frac{1}{\eta_{A,B}^{EV}} = \frac{EV(B) - EV(A)}{\ln\left(\frac{1}{P(A)} - 1\right)}. \quad (9)$$

Denote by $\hat{s}_{A,B}^{EV}$ an empirical estimate of this object, calculated by replacing $P(A)$ with the observed rate of selecting A in the expected values task.³

Similarly to before, we develop an index of predicted complexity by positing a link between the level of aggregation complexity in the choice task and in the expected values task. Our identifying assumption is that the problem-specific precision in the choice task is a linear increasing function of the precision in the expected values task,

$$\eta_{A,B} = \eta_0 + \lambda \eta_{A,B}^{EV} \quad \text{with} \quad \lambda > 0. \quad (10)$$

Definition 2. *The objective aggregation complexity ($OAC_{C,D}$) of choice set $\{C,D\}$ under a logit model is given by the prediction of the regression*

$$\hat{s}_{A,B}^{EV} = \sum_{i=0}^N \alpha_i f_i^{A,B} + \epsilon_{A,B}. \quad (11)$$

Structural analyses can then be implemented by estimating a complexity-augmented logit model, in which the precision is heteroscedastic and specified as in eq. (10).

2.4 Complexity of Individual Lotteries

Sometimes, researchers are interested in the complexity of individual lotteries. In these cases, the relevant complexity component is that of aggregation complexity (because proximity is undefined for a single lottery). We, hence, construct the predicted complexity of an individual lottery in analogy to the aggregation complexity of a choice set.

³Empirically, we winsorize the across-problem distribution of $\hat{s}_{A,B}^{EV}$ at the 85th percentile because $s_{A,B}^{EV}$ explodes or is not defined when selection rates for the wrong lottery get close to, or exceed, 50%.

Definition 3. Objective lottery complexity (OLC_C) of lottery $\{C\}$ under a logit model is given by the prediction of the regression

$$\hat{s}_{A,B}^{EV} = \sum_{i=0}^N \zeta_i f_i^A + \epsilon_{A,B}, \quad (12)$$

where (i) lottery B is a safe payment and (ii) only features of the non-degenerate lottery A enter the feature vector f .

2.5 Subjective Complexity Indices

All of the above concerns *objective* choice complexity, as revealed through error rates. Yet in many economically-relevant situations, it is not so much objective errors that matter for behavior but instead the DM's subjective beliefs about her error rates, such as when a DM contemplates whether to seek advice on a problem. To capture this, we define analogous indices of *subjective* (revealed) complexity that are identical to the formulations above except that true probabilities of making errors are replaced with subjective probabilities.

Definition 4. Given choice set $\{C,D\}$, subjective problem complexity ($SPC_{C,D}$), subjective aggregation complexity ($SAC_{C,D}$) and subjective lottery complexity (SLC_C), are defined in analogy to Definitions 1, 2 and 3, except that in equations (5), (11) and (12) objective error rates are replaced by average cognitive uncertainty (the subjective probability of making an error) in the expected values problem.

An advantage of this notion of subjective complexity is that its main underlying identifying assumption is directly testable. In analogy to eq. (3), the identifying assumption is that the subjective probability of making an error in a choice task is a monotone increasing function of the subjective probability of making an error in the expected values task. Given that – unlike true choice errors – subjective error probabilities can be elicited in the form of cognitive uncertainty (e.g., Enke and Graeber, 2023), this identifying assumption is testable.

3 Experimental Datasets

Our main analysis is based on three experimental datasets, two of which we collected ourselves. Table 1 provides an overview.

Table 1: Overview of experiments and data sources

Experiment	Description	Problems	Subjects	Decisions
<i>EV Tasks</i>	Indicate lottery with highest EV & CU elicitation	2,100 procedurally generated & 120 targeted	1,148	57k
<i>Choice Tasks from PEA</i>	Lottery choice problems	10,493 procedurally generated	15,153	981k
<i>Choice Tasks</i>	Lottery choice problems & CU elicitation	500 procedurally generated	250	12.5k

Notes. *PEA* = Peterson et al. (2021). *CU* = Cognitive uncertainty.

3.1 Experiment *EV Tasks*

Decision task. We present experimental participants with two or more lotteries. Instead of asking them to choose the lottery that they would personally prefer, we instruct participants to indicate the lottery that has the highest expected value. This design has been used previously (e.g., Benjamin et al., 2013), albeit always on a very small set of distinct problems. The task is similar in spirit to the “deterministic mirrors” proposed by Oprea (2022) and Martínez-Marquina et al. (2019).

We avoid jargon and never speak of “expected value.” Rather, we instruct participants to select the lottery that has the highest average payout if each lottery is run many, many times (100,000 times). We explain that, in each run, we record the payout of the lottery and then compute the average payout across runs. To hold the incentives constant, and to make transparent the objective nature of the task, subjects’ potential bonus equals \$10 if they select the correct lottery, and nothing otherwise. Throughout this paper, we use the language “select a lottery”, with the implicit understanding that this selection does not refer to a traditional choice.

A potential concern with this design is that participants may misunderstand it and, instead, treat it as a standard choice task. We took the following measures to ensure that this was not the case. First, we deliberately designed the incentive scheme described above to make it clear that no lottery was ever actually being played out. We verified subjects’ understanding of this using a comprehension check question in which subjects were asked to indicate whether their bonus equals \$10 if they select the lottery with the highest average payout or whether their bonus is being determined by the outcome of the lottery they select. Second, to avoid confusion, the question on subjects’ decision screen reads: “Which lottery has the highest average payout if the computer runs it many, many times?”, rather than, for example “Which lottery do you select?”. Third, our instructions emphasized that the task has a mathematically correct solution. Fourth, if it was the case that some subjects had misunderstood our instructions and had passed

our comprehension checks due to pure luck, we would expect that – just like they do in our real choice experiments described below – subjects would exhibit some risk aversion. Instead, we find no evidence for this in our data, as seen in Appendix Table 10.⁴

Cognitive uncertainty elicitation. On each decision screen, we elicited both a subject’s discrete decision about which lottery has the highest average payout and their certainty they made the right choice. We asked “How certain are you that each lottery has the highest average payout?” and subjects distributed 100 “certainty points” across the lotteries in the choice set to indicate their probabilistic beliefs. Our instructions clarified that subjects should allocate the certainty points to each lottery according to how likely they think it is that the lottery has the highest average payout. This procedure is similar to the elicitation of cognitive uncertainty (*CU*) in Enke and Graeber (2023) and Enke et al. (2023a).⁵ Appendix Figure 9 shows a screenshot of a decision screen.

Discussion. The main aspect that we believe makes lottery choice problems cognitively difficult and costly is that of aggregation: combining the probabilities and payouts of different states and lotteries into a choice or valuation. An attractive feature of our design is that this aspect of the decision problem is similar between our *EV Task* and real lottery choice problems. Aggregation complexity includes – but is not limited to – computational complexity. For instance, it also includes the cognitive costs of figuring out what needs to be multiplied with what – for people without training in statistics, it is arguably non-trivial to understand how to compute an average payout in the first place.⁶

As noted in Section 2, we do not assume (or believe) that people are genuinely risk-neutral, as our design implicitly induces. Our assumption is that the complexity of estimating expected utilities is an increasing function of the complexity of estimating expected values, rather than that the levels of complexity are the same.

⁴First, when subjects decide between a lottery and a safe payment, if anything they select the lottery more often (52% of the time). Second, the variance of a lottery is uncorrelated with the probability that a subject indicated that lottery to have higher expected value, again at odds with risk aversion.

⁵Recent work has cast doubt on the effectiveness of canonical incentive schemes designed to elicit beliefs (Danz et al., 2022). To account for this, we incentivize *CU* in two different ways. In 15% of the sample, we deployed a standard binarized scoring rule with a prize of \$10 and a winning probability of $q = 1 - (1 - g)^2$, where g is the probability assigned to the correct option. In the remaining 85% of the sample, we instead paid subjects \$0.10 for each point they allocated to the correct lottery. This scoring rule is not proper but simple to understand. In either case, our instructions emphasize to subjects that “For example, if you think it is 60% likely that lottery A has the highest average payout, you should allocate 60 certainty points to lottery A.” The distribution of *CU* in these two sub-samples is essentially identical.

⁶In our online experiment, it is possible that subjects used external help. At the end of the study, we asked subjects whether they had done so, reassuring them that they would get paid either way. 23% of subjects indicated they had used some form of external help, 46% of which reported to have used it on less than half of all problems. We have verified that our complexity indices are very similar if we restrict attention to the sub-sample of subjects who report not having used external help.

Generation of problems. We desire our complexity indices to be applicable across different datasets. It is, hence, crucial for us to develop them on a dataset that includes as many commonly-encountered lottery features as possible. We, hence, designed the experiment *EV Tasks* to comprise a total of 2,220 unique choice sets. A first set of 2,080 unique choice problems was generated using a quasi-random procedure, meaning that the lotteries are random conditional on a set of parameters that we impose to (i) make the problems non-trivial and (ii) ensure variation across a large set of features. Appendix J.1 describes the procedure in detail. This random procedure is called for because we as researchers do not know *ex ante* which features matter most for complexity, and because we do not want our own intuitions to constrain the development of the indices.

A second set of 120 choice problems was devised by following the typical approach in lab experiments of designing a relatively small number of problems that are targeted at identifying some specific effect of interest. Our main analyses will leverage all 2,220 choice problems in the dataset. In Appendix K, we report separate analyses that only make use of the smaller, targeted set.

In the development of our complexity indices, we focus on two-item menus and discuss an extension to larger menus below. 95% of all tasks involve only two lotteries and the remaining 5% involve choice sets with between three and five lotteries. In the two-item menus, 58% involve deciding between a two-state lottery and a safe payment. In the other 42%, the number of states of both lotteries varied between three and seven. The problems exhibit large variation in the difference in expected values between the lotteries. We collected data until each problem was completed by at least 20 subjects (median is 22 and average 26). Appendix Table 7 presents further summary statistics.

Summary statistics. The average problem-level error rate in the *EV Tasks* experiment is 27%, with a median of 25% and $IQR = [14\%, 38\%]$. The average problem-level subjective error rate (average *CU*) is 18%, with a median of 17% and $IQR = [13\%, 22\%]$. The correlation between problem-level error rates and average *CU* is $r = 0.49$.

3.2 Lottery Choice Problems

Dataset of Peterson et al. (2021). Peterson et al. collected by far the largest and most comprehensive binary lottery choice dataset in the literature. The authors used a quasi-random procedure to generate 10,493 unique binary choice problems.⁷ 15,153 Amazon Mechanical Turk (AMT) workers completed an average of 16 choice problems five times each, for a total of 80 decisions per subject, on average. As discussed in Peterson et al. (2021), the dataset was deliberately designed to span a much larger space of choice

⁷We drop problems that involve ambiguity. We combine identical problems with / without feedback.

problems than previous data-collection exercises, making the data well-suited for our purposes. Another feature of the dataset that we make use of below is that almost 50% of the choice problems are such that the riskier option has a lower expected value. In contrast, in typical lab experimental datasets, the riskier option usually has a higher expected value. Appendix Table 7 presents summary statistics.

While the richness and size of this dataset provide many advantages, it has the downside that Peterson et al. (2021) did not pay out losses and instead truncated all payouts from below at zero. While this is a shortcoming, we view it as ultimately inconsequential: (i) Appendix F documents that the results shown below are robust to restricting attention to choice problems that only involve gains and (ii) the results are very similar in our own incentivized experiments described below.

Experiment Choice Tasks. To overcome this limitation, and as a robustness check, we implemented our own lottery choice experiments. We generated 500 choice problems using a similar automated quasi-random procedure as in the *EV Tasks* experiment, except that we only implemented binary choice sets. Losses were incentivized: subjects were informed that in every decision they encounter they receive a budget that equals the largest possible loss in that choice set, and that potential losses would be deducted from that budget. No budget was provided in choice sets that only involved gains.

In addition to asking subjects to choose between the two lotteries, we also elicited their *CU*, by asking how certain they are (in percent) that they selected the option that they actually prefer (Enke and Graeber, 2023). Appendix Figure 10 shows a screenshot.

3.3 Implementation

Our own experiments were conducted on Prolific. See Appendix L for screenshots of instructions and comprehension check quizzes. Each subject encountered 50 randomly-selected decision problems. In *EV Tasks* (median completion time 42 minutes), subjects earned a completion fee of \$6. In addition, 1 in 2 subjects was randomly selected to be eligible for a bonus, which was determined by randomly selecting one decision, with equal probability. In *Choice Tasks* (median completion time 22 minutes), subjects received a fixed payment of \$3.50. In addition, 1 in 5 subjects were randomly selected to be eligible for a bonus, which was determined by randomly selecting one decision.

We pre-registered the predictions and sample size for experiment *Choice Tasks* on aspredicted.org under #130662. We didn't pre-register the *EV Tasks* experiment because there was no specific hypothesis: we use these data to create complexity indices, rather than because we desired to show that a specific feature would matter.

4 Development of the Complexity Indices

We begin by developing the complexity indices based on the *EV Tasks* data. A main question is which features should be included in the indices, for example in the regression in eq. (5). There is a tradeoff between interpretability and completeness: an index that is based on a large number of features may be more complete (predictive) but less interpretable. We strike a middle ground and proceed in three steps.

1. *Exploratory LASSO indices*: We assemble a large vector of features. Because many of these features will be intra-correlated (giving rise to multicollinearity), we estimate LASSO regressions, such that only a relatively small number of features will have non-zero coefficients.
2. *Handcrafted indices*: We inspect the LASSO-generated indices and approximate them based on only a handful of simple and easily interpretable features.
3. *Assess completeness*: We benchmark the handcrafted indices against a non-parametric black-box machine learning benchmark (a convolutional neural net).

Exploratory LASSO indices. Appendix B provides a list of all 56 choice set features that we consider. When a feature is defined over a single lottery (such as a lottery's variance), we compute the feature as its average value in the choice set. We also consider non-linear transformations of these averages (log, square root and square). We only consider objective features of the lotteries, rather than also framing effects stemming from colors and other display effects. We don't allow interactions between features (as we will see below, they do not matter much). Finally, we also don't allow features of choice sets encountered in the past (as in Frydman and Jin, 2021, 2023).

Some of the features that we consider may affect choice for standard expected utility reasons, such as a lottery's variance. However, as emphasized in Section 5.3, the way in which features like variance affect choice through complexity is often distinct from what expected utility theory prescribes. Moreover, because our indices are developed based on the *EV Task*, utility curvature cannot drive any feature's inclusion in an index.

We run the LASSO regressions on a randomly selected subset of 75% of all problems (train set) and use the remaining 25% as a test set. We set the LASSO penalty parameter to the value that minimizes mean squared error in the train set. In Appendix Tables 8 and 9, we report the results of the LASSO regressions of error rates (or *CU*) on all features. Because many of the features are highly intra-correlated (e.g., range and variance of payouts), the particular features that get selected by the LASSO should not be viewed as uniquely important, but instead representative of broad classes of important features.

Handcrafted indices. This observation motivates us to develop handcrafted versions of the complexity indices that are based on fewer features, each of which represents a broad class of features that we now discuss. While we recognize that manually selecting features raises potential concerns over artificially generating “desired” results, we view these as ultimately inconsequential because the handcrafted indices turn out to be almost perfectly correlated with the exploratory LASSO indices, which means that all results on actual choice data that we discuss in Sections 5 and 6 hold in a quantitatively almost identical fashion when we use the LASSO-generated indices instead.

Table 2 shows the features that enter our objective complexity indices (the ones that are based on observed error rates rather than CU). Appendix Table 11 shows the results for the subjective complexity indices, which look very similar. We report the results of OLS estimates of eq. (5), (11) and (12). As with the LASSO index, we run the regressions on a randomly selected subset of 75% of all problems and use the remaining 25% as a test set to assess completeness. In all specifications, an observation is a unique decision problem. In columns (1) and (2), we develop the indices of the complexity of a choice set. In column (3), we develop the complexity of individual lotteries, meaning that the sample is restricted to problems in which one option is a safe payment. Accordingly, in columns (1) and (2) the features apply to choice sets, while in column (3) they apply to one specific lottery. We compute our complexity indices as the predictions of these regressions, censoring from below at zero.

To complement this table, Figure 1 reports pairwise correlation coefficients between error rates (or CU) and choice set features. We organize the figure such that those features that capture aggregation complexity rather than proximity appear to the left of the dashed vertical line.

Dissimilarity and variability. The standout predictor of both objective and subjective choice set complexity is the excess dissimilarity between the lotteries in a set, by which we mean the degree to which lotteries are dissimilar from each other above and beyond their difference in expected value. Intuitively, we compute dissimilarity by overlaying the cumulative distribution functions (CDFs) of the two lotteries and calculating the area between the two. Excess dissimilarity is then given by dissimilarity minus the (absolute) expected values difference.

The measure could describe a cognitive process whereby people choose among lotteries by first putting them into a correlated common state-space (matching q -th quantile outcomes to define the common states), and then comparing the payoffs state-by-state, rather than by assessing overall value separately for each lottery and then comparing those (e.g., Arieli et al., 2011). More formally, suppose that states s are ordered by payouts, and denote by p_s a probability and by x_s a payout. We bring lotteries into common

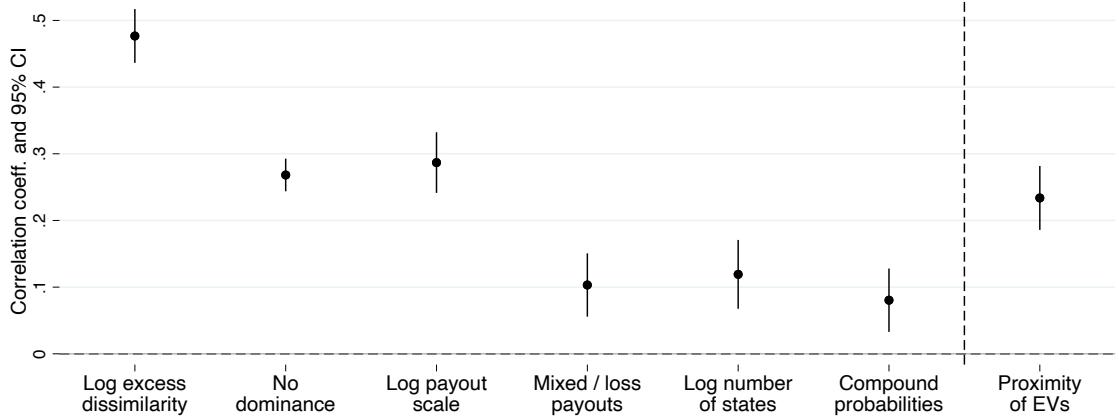
Table 2: Coefficients of features in objective complexity indices

Index:	Dependent variable:		
	Observed error rate	Implied obj. logit imprecision \hat{s}	
		OPC	OAC
	(1)	(2)	(3)
Log excess dissimilarity	0.052*** (0.00)	1.70*** (0.15)	
No dominance	0.068*** (0.01)	0.66* (0.35)	
Average log payout magnitude	0.036*** (0.00)	1.12*** (0.18)	
Average log number of states	0.045*** (0.02)	2.07*** (0.59)	
Frac. lotteries involving loss	0.029*** (0.01)	1.17*** (0.32)	
Frac. lotteries involving compound prob.	0.15*** (0.03)	4.93*** (1.26)	
Absolute EV difference	-0.032*** (0.01)		
Absolute EV difference sqr.	0.0014*** (0.00)		
Log variance			1.06*** (0.11)
Log payout magnitude			0.90*** (0.24)
Log number of states			2.10*** (0.48)
1 if involves loss			0.96*** (0.36)
1 if involves compound prob.			2.14*** (0.74)
Constant	0.027 (0.03)	-4.12*** (0.81)	-5.67*** (0.91)
Observations	1587	1587	935
R ²	0.31	0.20	0.23

Notes. OLS estimates, robust standard errors in parentheses. An observation is a decision problem from the train set in the *EV Tasks* experiment. In columns (2) and (3), the dependent variable is the implied logit precision $\hat{s}_{A,B}^{EV}$ as defined in eq. (9), winsorized at the 85th percentile. In column (3), the sample is restricted to problems in which one option is a safe payment, and the independent variables are features of the non-degenerate lottery. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

state space by assuming that they are perfectly correlated. The so-called “Wasserstein 1-distance” between the CDFs of two lotteries is given by $\delta_{A,B} = \sum_s p_s |x_s^A - x_s^B|$. Then, we define excess dissimilarity as:

EV Task: Error rates and choice set features



EV Task: Cognitive uncertainty and choice set features

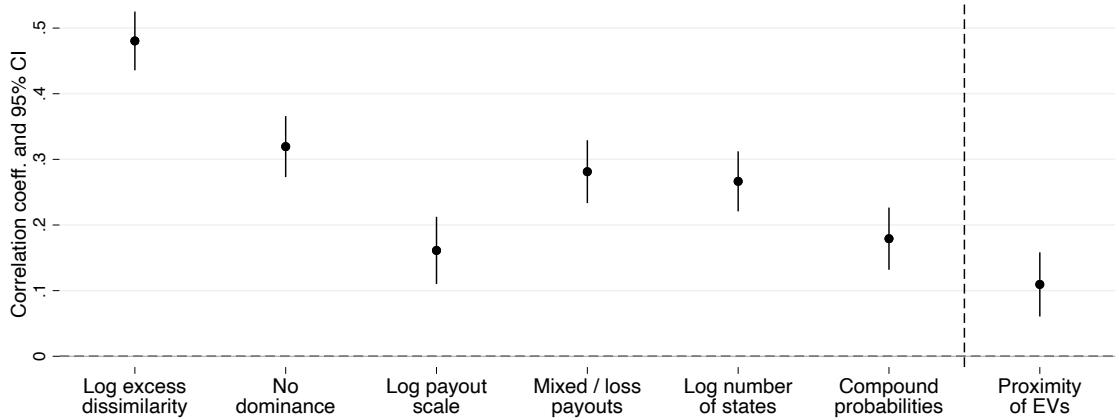


Figure 1: Pairwise correlation coefficients between task-level error rates / average cognitive uncertainty and choice set features in the train set in the *EV Tasks* experiment. Whiskers show 95% confidence intervals. Log scale, mixed / loss payouts, log number of states and compound probabilities are computed as averages across the lotteries in a set.

$$d_{A,B} = \delta_{A,B} - |EV(A) - EV(B)| = \sum_s p_s [|x_s^A - x_s^B| - (x_s^A - x_s^B)]. \quad (13)$$

Excess dissimilarity is large when the lotteries have very different advantages and disadvantages. Excess dissimilarity equals zero when there are no tradeoffs across the common, correlated states (i.e., when differences in expected value arise due to first-order stochastic dominance). We can loosely think of excess dissimilarity as a measure of how “close” lotteries A and B are to having a dominance relationship.

Excess dissimilarity depends on scale. While we have verified that $d_{A,B}$ is also strongly predictive of error rates when it is normalized by the average absolute payout, we view the dependence on scale as a feature because it appears intuitively true that two lotteries are more dissimilar when all payouts are multiplied by a constant.

To illustrate, the following lotteries have low excess dissimilarity. Option A: “Get

\$20 with probability 80%", and Option B: "Get \$21 with probability 70%". In contrast, Option B': "Get \$70 with probability 21%" has high excess dissimilarity with Option A (note that B and B' have the same expected value). Thus, choosing between A and B is predicted to be simpler than choosing between A and B'.

It may appear surprising that dissimilarity adds to complexity because researchers often think of "similar" as "difficult". The key distinction is that here "similarity" does not refer to the proximity to indifference (similarity of *aggregated* values) but, instead, to the similarity of the *disaggregated* objects, netting out the similarity in aggregate value.

The idea that similarity affects lottery choice complexity is also present in earlier work (e.g., Rubinstein, 1988). A small psychology literature has shown that $\delta_{A,B}$ in our notation above is predictive of choice noise (Buschena and Zilberman, 2000; Erev et al., 2002, 2010). We work with $d_{A,B}$ because $\delta_{A,B}$ is mechanically correlated with the expected values difference, and we sometimes desire to separate effects stemming from aggregation complexity and from proximity.

Figure 1 shows the raw correlation between log excess dissimilarity⁸ and error rates as well as average CU , which is approximately $r = 0.5$ in both cases. Log excess dissimilarity is the strongest predictor of both errors and CU in our data. In fact, because the correlations are so much stronger than with other features, both OPC and SPC heavily load on this variable, a point we quantify below.⁹

It is easy to see that excess dissimilarity will typically increase in the variance of the lotteries, for example when one option in the choice set is a safe payment. This is important to note because measures of variability are strongly correlated with both error rates and CU . For example, the correlation between average log variance and error rates is $r = 0.27$, yet measures of variability always lose statistical significance and collapse in magnitude once excess dissimilarity is accounted for in regressions. This is unsurprising given that the correlation between average log variance and log excess dissimilarity in our data is $r = 0.62$. For this reason, we do not include direct measures of variability in our choice set complexity indices.

However, we return to direct measures of variability in constructing our lottery-specific indices, OLC and SLC . Excess dissimilarity cannot enter these indices since it is defined on the choice set. Thus we include log variance as the most predictive feature in these indices. The correlation of log lottery variance with error rates and average CU is $r \approx 0.42$.

Dominance relationships. The second choice set feature that enters our OPC and SPC indices is the absence of weak stochastic dominance: lottery A dominates B if $F_A(x) \leq$

⁸Throughout the paper, whenever we say we compute the log of x , we mean that we compute $\ln(1+x)$.

⁹Dissimilarity is also the strongest predictor of errors in the targeted EV problems, see Appendix K.

$F_B(x) \forall x$, where $F(\cdot)$ denotes a CDF. This is an intuitively plausible (and empirically important) feature of aggregation complexity because the presence of dominance reduces the need to aggregate probabilities and payouts.

Payout scale. Much research on number perception suggests that people find it harder to process and transform larger numbers (Weber’s law). Again, this intuitively adds to the difficulty of integrating different payouts and probabilities. Our preferred measure of payout scale for an individual lottery is the log average absolute payout; for the choice-set measure, we average this individual measure across the set.

Mixed and loss gambles. It appears cognitively harder for people to process negative payouts in aggregation problems. In our data, both pure loss gambles and mixed gambles produce significantly higher error rates and CU , compared to pure gains lotteries. To keep our indices sparse, we generate one variable to capture these patterns, which is the fraction of lotteries in a choice that includes at least one negative payout.

Number of distinct payout states. Figure 1 shows that the average (log) number of states in the choice set is significantly correlated with error rates and CU . Again, this is plausible from a perspective of aggregation complexity because a larger number of components needs to be aggregated (e.g., Puri, 2022; Gabaix and Graeber, 2023). We include this variable in our complexity indices, but we note that it is clear from Figure 1 that lottery support – while a main focus of the literature – is a considerably less important determinant of overall complexity than some of the other features.

Compound probabilities. The presence of compound lotteries intuitively leads to higher aggregation complexity because they require an additional computational step (reduction). We find that compound probabilities are associated with both higher error rates and average CU , and include this variable in OPC and SPC .

Proximity of expected values. Unlike the aforementioned features, the difference in expected values between the two options *does* affect the magnitude of errors in standard random choice models. Figure 1 shows that proximity to indifference is indeed meaningfully correlated with errors and CU . Because this relationship is concave, we include both linear and squared terms in the construction of our handcrafted indices, see Table 2. Importantly, however, the link between proximity and errors (or CU) is relatively small compared to some of the features that capture aggregation complexity. This is a first indication of what we repeatedly emphasize throughout this paper: errors (and our complexity indices) largely reflect aggregation complexity rather than proximity.

Completeness of indices. Appendix Figure 12 shows histograms of the complexity indices, though the distribution will of course depend on the dataset under consideration. In the *EV Tasks* test set data, *OPC* and *SPC* exhibit a raw correlation of $r = 0.87$. An important question is how complete the complexity indices are. In principle, they could be incomplete for two reasons. First, the list of features that the LASSO was based on may be incomplete. Second, interactions of features may play an important role. As is well-understood, the appropriate metric for assessing completeness is not simply the variance explained of the regression in Table 2 due to the presence of irreducible error. For instance, problem-level error rates are only noisily estimated given our sample size. To assess completeness, we follow Fudenberg et al. (2022) and benchmark the out-of-sample performance of our complexity indices against that of a convolutional neural net (CNN). This algorithm is trained to predict error rates in a non-parametric fashion based on both (i) the raw lottery features (payouts and probabilities) and (ii) all hand-coded features that we include in the LASSO regressions. If we were missing important complexity features, or if interactions played an important role, we would expect the CNN to have substantially better model fit than our interpretable indices. Formally, completeness is defined as the ratio of the variance explained of the index and the variance explained by the CNN (in the test set).

Figure 2 shows the completeness of the main choice set indices, *OPC* and *SPC*. To comprehensively study completeness, we show the results for four different types of indices. First, an index that only captures proximity to indifference: the absolute difference of the expected values and its square. Second, an index that only consists of the (by far) most important aggregation complexity feature, log excess dissimilarity. Third, our handcrafted *OPC* and *SPC* indices. Fourth, the analogous exploratory LASSO indices.

There are three main takeaways. First, proximity to indifference is highly incomplete. Second, log excess dissimilarity is more than three times as complete as proximity to indifference. Third, our handcrafted indices are almost as complete as the exploratory LASSO indices, and both of these indices are 85-95% complete. We conclude from this discussion that our indices capture a large fraction of the predictable component of revealed complexity. This strongly suggests that (i) we are not overlooking important features and (ii) interactions between features do not play an important role.

What do the indices capture? In principle, the indices could reflect two different types of errors: bias and noise. The easiest way to see this is to consider the *OLC* index that captures the complexity of the lottery in a choice set when the alternative is a safe payment. If this index would mostly reflect *bias*, subjects would have systematically selected higher *OLC* lotteries less often (or more often), regardless of whether the lottery is objectively the correct choice. On the other hand, if the index mostly reflected *noise*, subjects

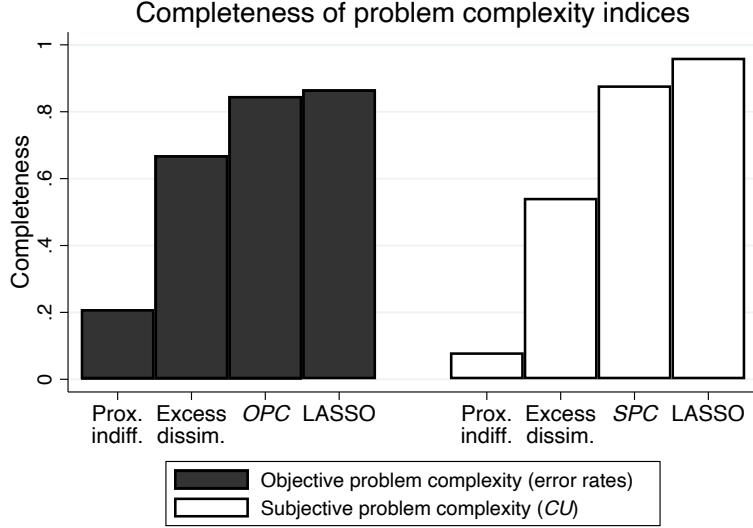


Figure 2: Completeness of choice set complexity indices relative to performance of a convolutional neural net. Completeness is defined as ratio of variance explained vis-a-vis the CNN in the test set. Leftmost model includes linear and squared terms of absolute expected values difference. Second model includes log excess dissimilarity. Third model includes *OPC* and fourth model the exploratory LASSO index that predicts error rates. Fifth through eight model defined analogously.

would have selected higher *OLC* lotteries less often when the lottery is the correct decision but more often when it is the incorrect decision. In our train data, the correlation between *OLC* and the frequency of selecting the lottery is $r = 0.00$ ($p = 0.99$), meaning that we cannot reject the null hypothesis that the index reflects zero bias.¹⁰ In contrast, in the sub-sample of problems in which the lottery is the correct decision, the correlation between selecting the lottery and *OLC* is $r = -0.43$, while in the sub-sample in which the lottery is the incorrect decision it is $r = 0.44$. These correlations document that our complexity indices almost entirely reflect predicted noisiness rather than predicted bias.

Evidence on identifying assumption. As noted in Section 2, we can directly test the identifying assumption that underlies the construction of *SPC*: that *CU* in the *EV Task* is predictive of *CU* in real choice tasks. We can test this assumption because in our own *Choice Tasks* experiment, we elicited subjects' *CU* for every decision they made. The correlation of average *CU* in a choice problem with *SPC* (computed out-of-sample) is $r = 0.62$.¹¹ Similarly, the correlation with *OPC* is $r = 0.49$. These results are encouraging because they strongly suggests that the same features determine how difficult it is to gauge expected utilities on the one hand and expected values on the other hand.

¹⁰The main feature that produces bias in the *EV Task* is compound probabilities, see Appendix E.

¹¹While large, this correlations is most likely even biased downward due to finite-sample measurement error in the calculation of problem-level average *CU*.

5 Behavioral Responses to Complexity

We now deploy the complexity indices to explain standard choice behavior. Unless noted otherwise, we pool the data from our own *Choice Tasks* experiment with those collected by Peterson et al. (2021). In our analyses, the level of observation is not an individual decision but, instead, choice rates in a unique choice problem. Because the underlying number of decisions is very large (almost one million), the statistical significance of the results will almost always be trivial. For each choice problem in the pooled data, we compute the complexity indices as implied by Table 2. Appendix Figure 13 shows histograms of all indices in the pooled data.

We study the link between complexity and choice in three steps. First, we provide a few illustrative examples of low- and high-complexity problems and associated choice patterns. While cherry-picked, these examples are arguably helpful in building intuition for the results in the very large set of problems that we use for our main analysis.

Second, we systematically study the role of the indices of problem complexity, *OPC* and *SPC*. Because these indices apply to a choice set as a whole, the natural prediction is that higher complexity is linked to more noisy decisions.

In a third step, we study the indices of the complexity of individual lotteries, *OLC* and *SLC*. As summarized in Appendix A, the prior literature has argued that lottery complexity can have two different consequences: higher noisiness and / or systematic complexity aversion. We will identify whether noise or aversion is the dominant pattern in the data, and how complexity-dependent noise can spuriously produce behavior that looks like complexity aversion (or small-stakes risk aversion) in some sets of problems.

5.1 Examples

Table 3 presents seven example choice problems. Six of them are selected to have similar expected value differences but varying *OPC*. This will allow us to intuitively attribute variation in complexity to features that drive aggregation complexity rather than proximity. All problems are such that Option A (the lottery with the weakly larger number of states) has a lower expected value.

Problem 1 is very simple even though lottery A has three distinct payout states. This is because it features dominance (thus excess dissimilarity is 0), but also because the payout scale is relatively low.

To see the role of excess dissimilarity more clearly, consider problems 2 and 3. In both problems, there is no dominance but the problems are intuitively simple. The reason is that excess dissimilarity is very low because the payout probabilities in lottery A are relatively extreme. For instance, in problem 2, one can intuitively see that lottery

Table 3: Example choice problems

#	Probabilities A	Payouts A	Probabilities B	Payouts B	EV(A)-EV(B)	OPC	Frac. chose A
1	0.25, 0.375, 0.375	-21, 0.5, 1.5	1	2	-6.5	0	4%
2	0.99, 0.01	18, 33	1	25	-6.9	0.11	9%
3	0.95, 0.05	-2, 28	0.75, 0.25	1, 22	-6.8	0.13	16%
4	0.75, 0.25	8, 52	1	26	-7	0.24	42%
5	0.8, 0.2	-30, 34	0.1, 0.9	7, -12	-7.1	0.27	43%
6	0.9, 0.05, 0.025, 0.025	-17, 34, 36, 40	0.5, 0.5	19, -29	-6.7	0.30	43%
7	0.4, 0.3, 0.15, 0.15	-28, 31, 29, 25	1	7	-0.8	0.41	41%

Notes. Each row is a choice problem.

A “is worth approximately \$18,” which makes B’s payout of \$25 look transparently superior. In this example, *OPC* is small because, intuitively, there is relatively little need to precisely aggregate payout probabilities and payouts across different states because the probabilities of A and B are sufficiently similar that a heuristic comparison suffices. Indeed, the choice rates in these problems are overwhelmingly in favor of lottery B.

In contrast, in problems 4 and 5, excess dissimilarity is high. The reasons are (i) that the probabilities are less extreme and (ii) that the options have different advantages and disadvantages. For example, in problem 5, heuristic pairwise comparisons of the lottery upsides and downsides is difficult, both because the probabilities aren’t identical and because the payouts are relatively spread out. Presumably as a result, the choice rates are close to 50-50.

Problem 6 provides another illustration. Here, excess dissimilarity is high. Moreover, both lotteries have at least two separate payout states, all payouts are relatively large, and a loss is involved, creating additional complexity. Finally, problem 7 illustrates a very high complexity problem. in which the driver of complexity is not just high aggregation complexity (e.g., high excess dissimilarity) but also very similar expected values.

5.2 Choice Set Complexity and Choice Noise

In linking complexity to choice noise, we show the results for both *OPC* and *SPC* because we are agnostic over whether the patterns reflect genuine choice errors that increase in complexity (in which case *OPC* is the appropriate tool) or deliberate randomization that increases in subjectively perceived complexity (in which case *SPC* is more appropriate).

Choice inconsistencies. First, we link problem complexity to across-trial variability in repetitions of the same problem (within-subject choice inconsistencies). Recall that in Peterson et al. (2021), each subject that completed any given binary choice problem did so five times (consecutively). For each choice problem, we compute the fraction of subjects who are inconsistent at least once, i.e., who do not make the same decision in all five iterations. In these analyses we restrict attention to problems in which the absolute expected values difference is at least \$0.20 to reduce concerns that “choice inconsistencies” simply reflect indifference (the results are identical in the full sample).

Figure 3 shows binned scatter plots of problem-level choice inconsistencies against the complexity indices. We work with binned scatter plots because of the large number of underlying choice problems. In these binned scatter plots, each dot represents an equal number of choice problems (roughly 105 choice problems and 1,995 associated decisions per dot). The left panels show raw correlations and the right panels partial correlations, controlling for linear and squared terms of the absolute expected values difference. Thus, as discussed in Section 2.2, the right panels isolate that component of *OPC* and *SPC* that reflects aggregation complexity rather than proximity.

Moving from $OPC = 0$ to $OPC = 0.5$ is associated with an increase in the frequency of choice inconsistencies of about 35 percentage points. The raw and partial correlations are almost identical ($r = 0.55$ and $r = 0.57$ with *OPC* and $r = 0.56$ and $r = 0.55$ with *SPC*). This suggests that a large majority of the explanatory power of the complexity indices for choice inconsistencies reflects the impact of aggregation complexity features rather than proximity. We further quantify this point below.

Compression of choice rates to 50-50. A second potential implication of the idea that choice set complexity creates noise is that the link between differences in expected values and choice rates should be attenuated for more complex problems, meaning that higher complexity reduces the probability that the DM will choose the higher value option. Intuitively, if a problem is extremely complex, people may (consciously or subconsciously) choose uniformly at random.

To study this prediction, the top left panel of Figure 4 shows choice rates for lottery A as a function of the difference in expected values between A and B, separately for choice problems that are above or below median *OPC*. Again, we show a binned scatter plot. We label the lotteries such that lottery A is always the one with a weakly larger number of distinct payout states (lottery B is often a safe payment). We see that choice rates in problems that are predicted to be more complex are substantially more compressed towards 50%, as we’d expect with fully random choice.¹² The top right panel shows

¹²One interpretation of the compression towards choice rates of 50-50 is that subjects have a prior belief over the expected utility associated with Options A and B, and that this prior is uninformative. In

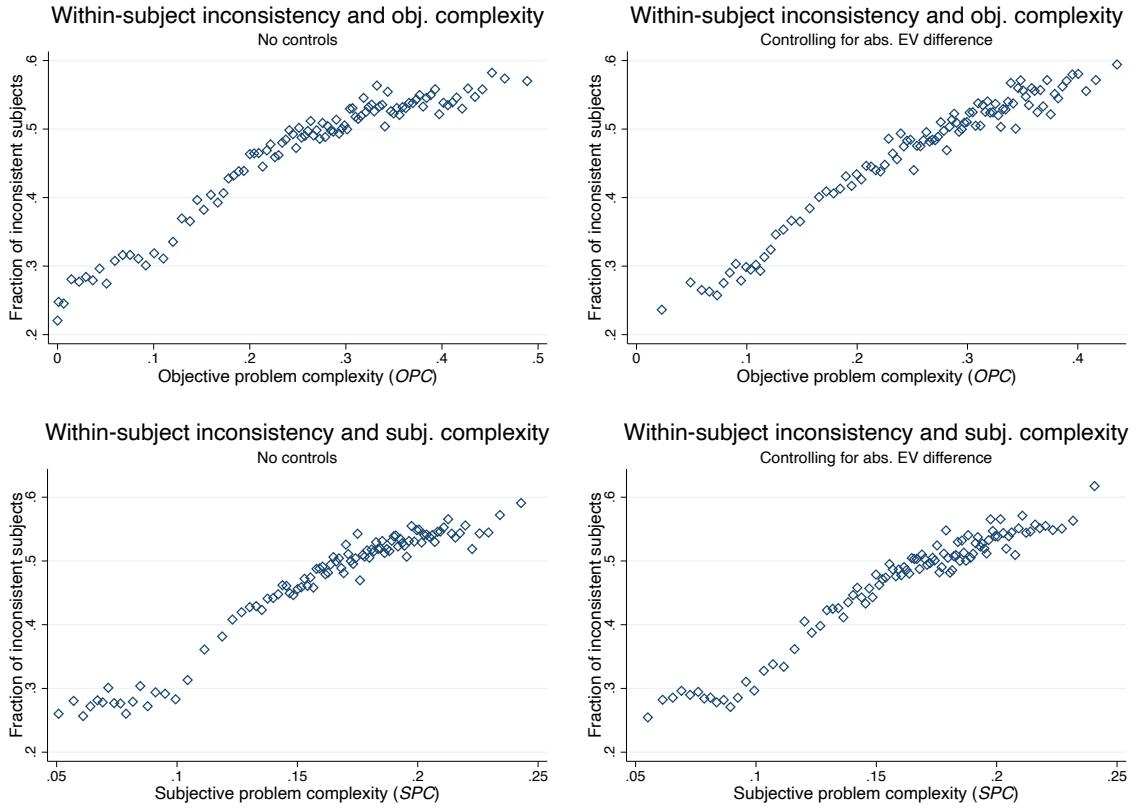


Figure 3: Binned scatter plots. The y-axis represents the problem-level fraction of subjects who are inconsistent at least once (in five repetitions). The x-axes reflect OPC (top panels) and SPC (bottom panels). The left panels show raw correlations and the right panel partial correlations, controlling for linear and squared absolute expected values difference. Based on 10,448 choice problems in which the absolute expected values difference is at least \$0.20.

analogous results for SPC .

The middle left panel provides a complementary perspective that does not rely on a median split of OPC but, instead, shows choice rates for lottery A as a function of the continuous OPC index. The red dots correspond to cases where $EV(A) > EV(B)$ and the blue dots to cases where $EV(A) < EV(B)$. Thus, for a noiseless expected value maximizer, the choice rates should be 0% and 100%. Yet even a model that does feature homoscedastic noise would predict that choice rates are constant in OPC . Instead, we see that choice rates monotonically approach 50% as complexity increases. The magnitude of this effect is very large: choice rates for the option with the higher expected value decrease by 40 percentage points going from very low to very high complexity.

An immediate question is whether these compression patterns “only” reflect the effects of proximity, given that the proximity of expected values is a component of OPC .

Appendix H, we report on additional pre-registered experiments that study the role of prior beliefs and how they interact with problem complexity. In these experiments, we experimentally manipulate prior beliefs over which option is “better”. As we pre-registered, this treatment has a larger effect on choice for more complex problems, an effect that is significant at 5% using SPC and at 10% using OPC .

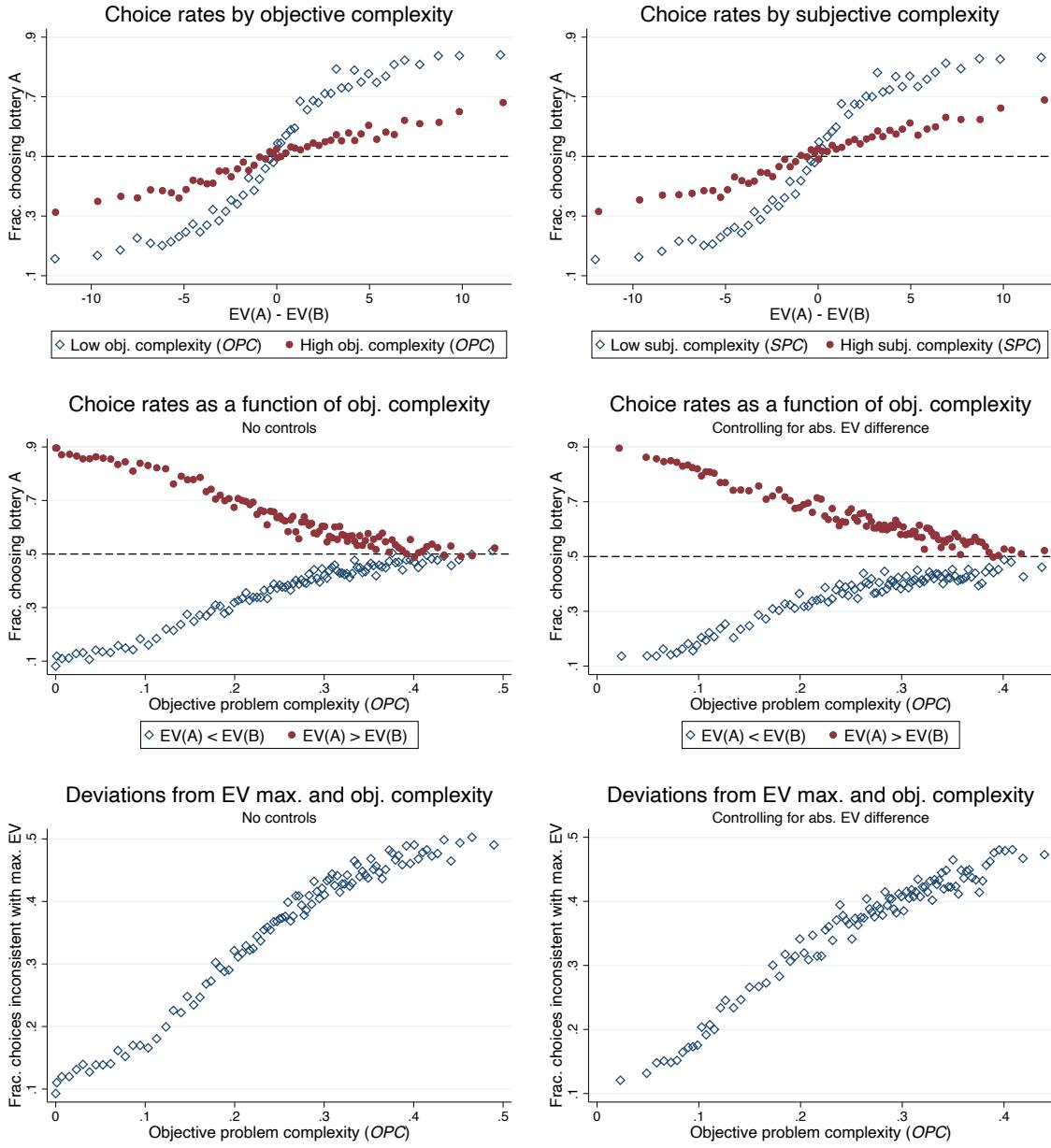


Figure 4: Complexity and deviations from expected value maximization in lottery choice. The top panels implement a median split by *OPC* / *SPC*, separately within each percentile of the EV difference between A and B. The middle right and bottom right panels show partial correlation plots that control for linear and squared terms of the absolute EV difference. Top panels constructed from 10,993 choice problems, mid and bottom panels omit problems with absolute EV difference of less than \$0.20, hence constructed from 10,448 choice problems.

To address this, the middle right panel shows a partial correlation plot that residualizes both choice rates and *OPC* from the absolute expected values difference and its square. The results are almost identical. Thus, again, the results suggest that the vast majority of the explanatory power of *OPC* is due to the effects of aggregation complexity rather than proximity.

Finally, the bottom panels put these pieces together by showing the fraction of choices

that do not correspond to choosing the lottery with the higher expected value, as a function of *OPC*. In both raw ($r = 0.64$) and partial ($r = 0.56$) correlation plots we see strikingly strong relationships.

We do not explicitly show analogous results for *SPC* because they are visually and quantitatively very similar to those based on *OPC*.¹³

Benchmarking. A natural question is how quantitatively important our complexity indices are for understanding choice behavior. Here, a natural point of comparison is the proximity to indifference (in the choice task), which is the main driver of choice errors in standard random choice models. We, hence, compare the predictive power of *OPC* with that of the subjective value difference between the two choice options, estimated from a full prospect theory model as described in Appendix G. Because *OPC* includes the proximity of the expected values of the two options, we also benchmark *OPC* against the expected values difference.

For our benchmarking analysis, we desire proxies for the problem-level frequency of choice errors. Building on the analysis above, we work with three such proxies: (i) the fraction of subjects who are inconsistent at least once in repetitions of the same problem; (ii) the fraction of decisions that does not correspond to choosing the lottery that – according to an estimated prospect theory model – delivers higher value (“non-expected utility”); and (iii) average self-reported cognitive uncertainty in a choice problem.

Table 4 shows the results. The main metric of interest is the variance explained in each regression. There are two main takeaways. First, for all dependent variables, *OPC* explains a much larger fraction of the variation than the proximity of the expected values of the two options. This is of interest because it again shows that the vast majority of the predictive power of our complexity index for potential choice errors reflects *aggregation complexity* rather than complexity resulting from *proximity of the expected values*.

Second, again for all dependent variables, *OPC* explains a substantially larger fraction of the variation than the estimated prospect theory value difference between the two options. For example, in column (7), the estimated value difference explains a sizable share (21%) of the frequency of decisions that do not maximize prospect theory subjective value. Yet the variance explained by *OPC* is even larger (33%, column (5)). Perhaps most strikingly, *OPC* alone explains 28% of the variation in self-reported *CU* (a subject’s subjective probability of making a mistake), while the proximity of estimated subjective values explains almost none of the variation (column (11)).

We conclude from this analysis that choice set complexity appears to be a quantitatively important driver of choice behavior.

¹³Appendix Figure 14 replicates Figure 4, excluding dominance problems. The results are very similar.

Table 4: Benchmarking

	Dependent variable:											
	Frac. subjects inconsistent				Deviation rate from PT prediction				Avg. cognitive uncertainty			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
OPC	0.71*** (0.01)		0.73*** (0.01)	0.80*** (0.01)				0.66*** (0.01)	49.7*** (3.79)			53.9*** (3.99)
Abs. EV diff.		-0.0078*** (0.00)			-0.018*** (0.00)					0.039 (0.18)		
Abs. PT value diff.			-0.015*** (0.00)	0.0036*** (0.00)			-0.043*** (0.00)	-0.026*** (0.00)			0.065 (0.31)	0.97*** (0.27)
Constant	0.28*** (0.00)	0.50*** (0.00)	0.50*** (0.00)	0.27*** (0.00)	0.14*** (0.00)	0.42*** (0.00)	0.45*** (0.00)	0.24*** (0.00)	3.09*** (0.87)	15.6*** (0.99)	15.7*** (0.86)	-0.31 (1.29)
Observations	10493	10493	10493	10493	10993	10993	10993	10993	500	500	500	500
R ²	0.30	0.02	0.03	0.30	0.34	0.12	0.21	0.41	0.24	0.00	0.00	0.26

Notes. OLS estimates, robust standard errors in parentheses. An observation is a choice problem. The dependent variable in columns (1)–(4) is the fraction of subjects who is inconsistent at least once in the five repetitions of the choice problem. In columns (5)–(8) it is the fraction of decisions that does not equal the prediction of a full prospect theory model, see Appendix G. In columns (9)–(12) the dependent variable is self-reported cognitive uncertainty in the choice experiments (in percent). The absolute PT value difference is the absolute difference in “non-expected utilities”, as estimated from a full prospect theory model, see Appendix G. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

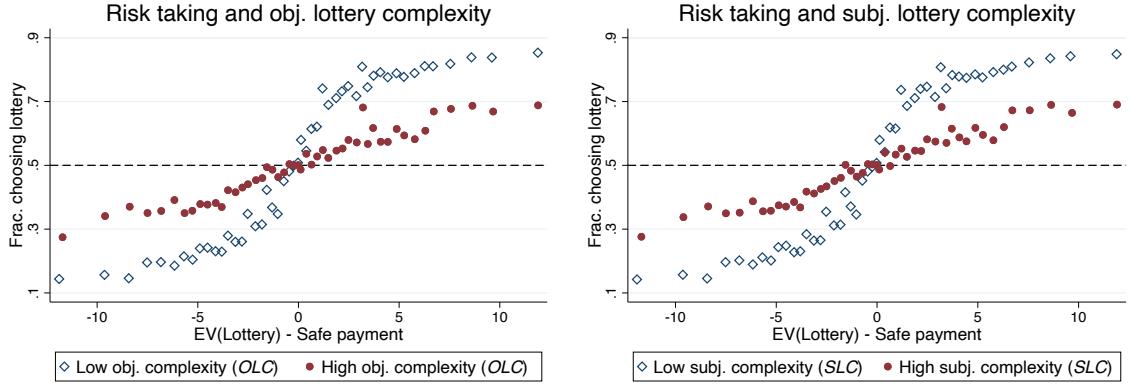


Figure 5: Binned scatter plots of choice rates for lottery vs. safe payment (based on 6,466 choice problems). Figures show median splits by *OLC* / *SLC*, separately within each percentile of the EV difference.

5.3 Lottery Complexity, Noise and Aversions

Up to this point, we focused on the complexity of a choice set. We now turn to studying the complexity of individual lotteries. As discussed above, various features of lotteries could have two distinct impacts on choice that we refer to as “direct” and “indirect” effects. The direct (preferences-based) effect is that people may be genuinely averse to a lottery feature, making them less likely to choose the lottery. For example, people may be systematically averse to lotteries with higher variance because they are risk averse. The indirect effect, on the other hand, is that the same feature may increase noisiness.

To cleanly study the effects of the complexity of individual lotteries, we focus on problems in which one option is a safe payment and the other one a non-degenerate lottery. Given the usually obvious statistical significance of the results, we only present figures in the main text and relegate econometric analyses to Appendix Table 12.

The left panel of Figure 5 plots the frequency of choosing the lottery as a function of the expected value difference, split by median *OLC*. Complexity aversion predicts a *uniform downward shift* of choice fractions as complexity increases. Complexity-dependent noise, on the other hand, predicts that the choice function becomes *flatter*, such that choice rates involving more complex lotteries are more compressed to 50-50.

We see no evidence of complexity aversion in the binary choice data. While people choose the complex lotteries less often when the lottery has a higher expected value than the safe payment, the opposite is true when the lottery is relatively unattractive. Thus, people can spuriously appear “complexity-averse” (to the right of zero) or “complexity-seeking” (to the left of zero), purely as a result of complexity-dependent heteroscedasticity and the specific ways in which the researcher designs the decision problems.

The right panel of Figure 5 shows that we find very similar results for the *subjective* complexity of lotteries. Again, people spuriously look like they “love” subjectively complex lotteries when these are relatively unattractive.

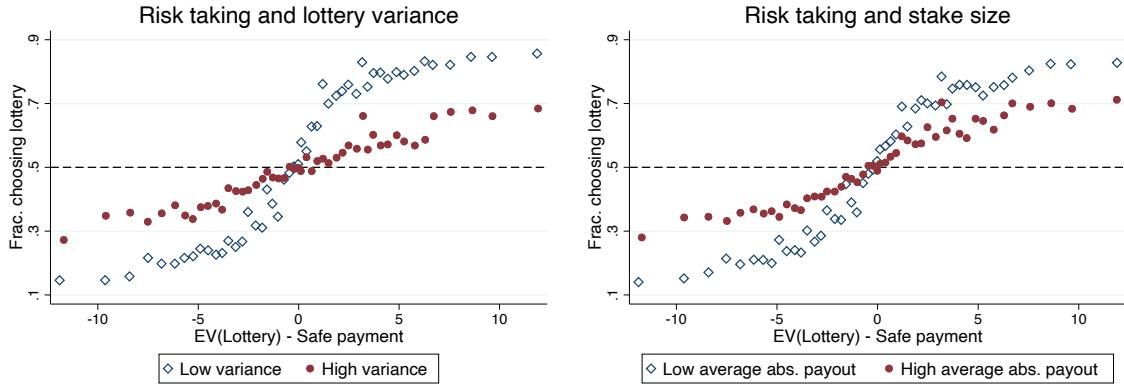


Figure 6: Binned scatter plots of choice rates for lottery vs. safe payment (based on 6,466 choice problems). Figures show median splits, separately within each percentile of the EV difference.

Small-stakes risk aversion. The insight that complexity-driven heteroscedasticity can generate spurious aversions is also relevant for our understanding of other preference anomalies. As we saw in Section 4, many lottery features that are commonly believed to induce a preferences-based response also induce higher complexity.¹⁴ In particular, we saw that lottery variance and the magnitude of payouts both contribute to higher error rates in the *EV Task*. Given that these features are intimately linked to conventional definitions of small-stakes risk aversion and increasing absolute risk aversion (Holt and Laury, 2002), this raises the question if complexity can confound or spuriously generate behavior that looks like it reflects certain types of risk preferences.

The left panel of Figure 6 again shows a binned scatter plot of choice rates for the lottery as a function of the difference between its expected value and the safe payment, this time split by median lottery variance. If people are risk averse and noise was independent of complexity, we should expect choice rates for the high-variance lotteries to be lower than for the low-variance lotteries everywhere (a vertical shift). In contrast, cognitive noise that increases in variance again predicts a compression (or “flipping”) pattern, according to which observed risk taking can even increase in the lottery’s variance when the lottery is very unattractive (the left part of each panel).

In the data, we indeed see a pronounced compression pattern. This implies that people *look* risk averse when the lottery is attractive (to the right of zero), but risk loving when the lottery is unattractive (to the left of zero).

To highlight the confound this poses for estimating risk preferences, we restrict attention to the 2,676 problems for which all payouts are weakly positive (and one option

¹⁴Recent work has focused on the number of distinct payout states. As shown in Appendix Figure 15, in our binary choice data, there is no evidence that people choose lotteries with a larger number of payout states less often. If anything, people are *more* likely to choose a lottery over a safe payment if it has more states, in particular when the lottery is unattractive. This pattern looks almost identical in the Peterson et al. (2021) and our own choice data.

is a safe payment), where we can estimate a standard CRRA expected utility model, $EU(x) = E[x^\alpha]$. When we estimate this model on the sub-sample in which the lottery has a higher expected value than the safe payment, we estimate $\hat{\alpha} = 0.79$ – a typical estimate suggesting small-stakes risk aversion. In contrast, when we estimate on the sub-sample in which the lottery has a lower expected value than the safe payment, we estimate $\hat{\alpha} = 1.06$ – suggesting apparent risk loving preferences. This exercise shows that complexity-dependent noise can predictably bias the estimation of preference parameters depending on how the researcher constructs the problems.

It is worth pointing out that the vast majority of binary lottery choice experiments implement problems in which the expected value of the lottery is at least as large as the safe payment. Because it is precisely in this space of problems that complexity-dependent heteroscedasticity spuriously produces risk aversion, we suspect that conventional estimates of small-stakes risk aversion in the literature may be upward biased.

Increasing absolute risk aversion. The right panel of Figure 6 shows an analogous analysis for the lottery’s stake size. This is of interest because a widely-cited result in experimental economics is that observed risk aversion increases in the magnitude of payouts. The typical way in which this result is derived is by showing that multiplying all payouts with a large constant produces higher absolute risk aversion (e.g., Holt and Laury, 2002). However, multiplying payouts with a constant also increases variance. Thus, the previous discussion on the complexity effects of variance already suggest that higher stakes may produce similar effects. The right panel of Figure 6 shows that we indeed see the familiar pattern: lotteries that have higher payouts produce choice rates that are more compressed to 50-50. Thus, decisions can *look like* risk aversion that increases in stakes when the lottery is attractive but like risk aversion that decreases in stakes when the lottery is unattractive, purely as a result of heteroscedastic noise.

We emphasize that we do not wish to claim based on our results that genuine small-stakes risk aversion or complexity aversion do not exist. Rather, the point is that the indirect effect generated by complexity-dependent heteroscedasticity is so strong that it can either confound or entirely override any true aversion that may exist.

6 Structural Estimations Accounting for Complexity

How important is complexity-dependent heteroscedasticity in structural estimations? In order to explain classical choice irregularities, almost all canonical behavioral economics models augment the DM’s objective function. Popular modifications include allowances for probability weighting, loss aversion, regret, salience, relative thinking and so on. We,

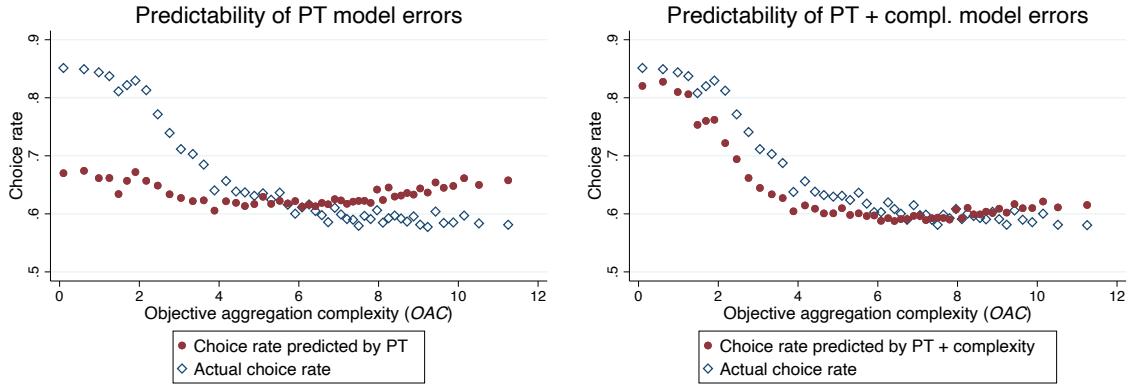


Figure 7: Model prediction errors in pooled choice data as a function of objective aggregation complexity. Left panel: Prospect theory model. Right panel: Prospect theory model with complexity-dependent noise. Both panels plot the actual and predicted choice rates for the choice option that the respective model predicts has higher expected utility.

instead, proceed by allowing for complexity-dependent noise. Recall the choice model in Section 2, where the index of objective aggregation complexity, OAC , scales the precision in the logit model:

$$P(A) = F(EU(A) - EU(B); \eta) = \frac{1}{1 + e^{-(\eta_0 + \lambda 1/OAC_{A,B})[EU(A) - EU(B)]}}. \quad (14)$$

We estimate eq. (14) using maximum likelihood for different combinations of (i) the specification of the DM's objective function (expected value, expected utility, prospect theory, salience theory) and (ii) the presence of complexity-dependent heteroscedasticity (i.e., whether λ is estimated or forced to be zero). Appendix G presents details for the estimating equation for each model as well as the resulting parameter estimates.

To start out, consider prospect theory (separate utility curvature for gains and losses, loss aversion and probability weighting). The left panel of Figure 7 plots the actual choice rate for the lottery that – according to an estimated prospect theory model – has higher value, as a function of OAC . In addition, we plot the model-predicted choice rates in a prospect theory model. As is clear from this figure, prospect theory has highly systematic prediction errors in our data: it strongly underpredicts how often people choose the estimated higher value option when complexity is low but overpredicts it when complexity is high. Intuitively, this happens because the model treats all choice problems as if they had the same level of complexity. Overall, the correlation between prediction errors and OAC is $r = 0.51$.

The right panel of Figure 7 shows the prediction errors of a prospect theory model augmented by a complexity-dependent noise term. We see that predicted and actual choice rates track each other much more closely.

To systematically assess model fit, Figure 8 plots the variance explained across each

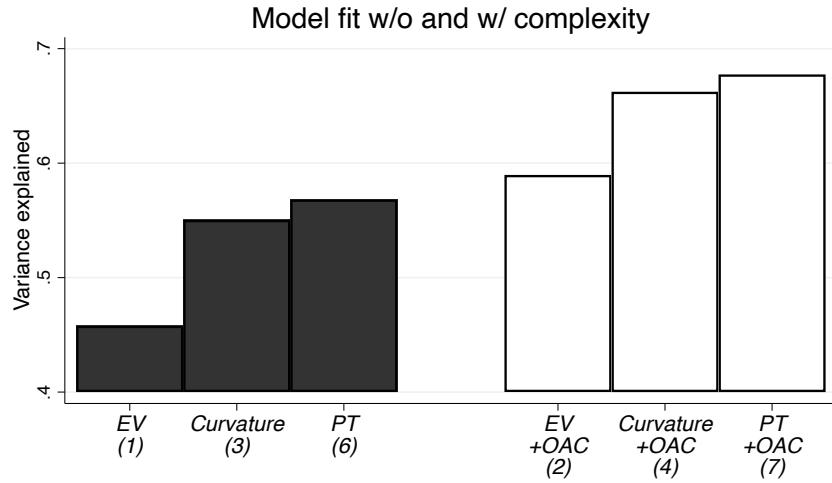


Figure 8: Variance explained of different models. Number of estimated model parameters in parentheses. The first model assumes EV maximization and a constant error variance. The second model adds utility curvature parameters for gains and losses, and the third model loss aversion and two probability weighting parameters. The fourth through sixth models are analogous except that they also estimate the parameter that maps OAC into logit responsiveness, see eq. (14). See Appendix G for all estimating equations and estimated parameter values.

of six models. The first model assumes expected value maximization and only estimates a constant error variance. The second model adds separate utility curvature parameters for gains and losses (and hence includes reference-dependence relative to a reference point of zero). The third model adds loss aversion and two probability weighting parameters. The fourth through sixth models are identical to the first three except that they also estimate the parameter that maps the complexity index into error variance.

The expected values model has an R-squared of 46%, which increases to 55% when two utility curvature parameters are introduced. The full prospect theory model does not perform much better (57%). This is consistent with the recurring finding in the literature that prospect theory substantially improves performance over expected utility in valuation experiments (which involve multiple price lists or BDM elicitations), but adds little in binary choice tasks (e.g., Harbaugh et al., 2010; Bouchouicha et al., 2023; Peterson et al., 2021).

Introducing one parameter that maps problem complexity into error variance brings an expected values model to $R^2 = 59\%$, larger than the full prospect theory. In other words, in our dataset, complexity-dependent noise alone is quantitatively more important than utility curvature, loss aversion and probability weighting combined. The variance explained further increases to 67% when utility curvature is accounted for and to 68% under the full prospect theory specification. The results are very similar when we estimate a salience model (Bordalo et al., 2012) instead of prospect theory.

We conclude from this analysis that allowing for complexity-dependent heteroscedas-

ticity is quantitatively important. This resonates with a literature in psychology that finds that allowing the noise term in a stochastic choice model to depend on lottery dissimilarity ($\delta_{A,B}$ rather than $\ln(d_{A,B})$, see Section 4) yields the best model fit relative to other models proposed in the psychology literature (Erev et al., 2010).

7 Discussion

This paper has made two contributions. First, we developed indices of the objective and subjective complexity of individual lotteries and of lottery choice sets. A significant practical advantage of these indices is that they consist of simple linear combinations of a handful of choice set features and can, hence, be computed for any standard lottery choice dataset. Our interpretable complexity indices are highly complete, meaning that they perform almost as well as a black-box neural net. A single feature – the excess dissimilarity between the lotteries in a set – captures the bulk of variation in complexity.

Our second contribution is to comprehensively study behavioral responses to complexity, which also allows us to illustrate the large predictive power of the complexity indices. We find that the most important consequence of complexity in binary choice is heteroscedasticity. This increased noisiness does not “cancel out” but produces consequential deviations from normative benchmarks, including choice inconsistencies, accepting unattractive gambles, and spurious “aversions”. Allowing for complexity-dependent noise is quantitatively important. First, complexity explains much more of the variation in proxies for choice noise than proximity to indifference does. Second, a single parameter that maps complexity into error variance adds more explanatory power to structural estimations than prospect theory or salience theory parameters.

We interpret these results as suggesting that behavioral economists need to devote more attention to heteroscedastic noise rather than the “directional” distortions that we often emphasize. We now discuss what we believe to be fruitful next steps.

A common complexity scale across papers. A common criticism of lab experiments is that researchers have many degrees of freedom in constructing the choice problems they use to document an effect of interest. We believe that if our complexity indices were standardly computed in lottery choice experiments going forward, they would provide a standardized metric along which papers can be compared and assessed.

Assessing the direction and magnitude of bias and noise in revealed preference measures. A key takeaway of the recent literature on bounded rationality is that choice does not transparently reveal preferences but, instead, to a large degree reflects complexity responses. Yet, using conventional methods it is difficult for researchers and read-

ers to know in which direction and to what degree revealed preference measures will be biased. In this paper, we have developed a blueprint for such analyses, by showing how the complexity indices allow us to predict the direction and magnitude of bias.

Real-world assets. The complexity indices we develop in this paper could be used to quantify the complexity of real-world financial assets such as stocks, bonds and mutual funds. All that would be required to do so is (i) information about the assets' return profiles (or information about what people know about these return profiles) and (ii) information about people's choice sets.

Larger menus and valuation tasks. A natural question is how our indices can be applied to (i) larger choice sets and (ii) continuous valuation tasks, such as elicitations of certainty equivalents. Regarding the latter, no work is required: because we also developed indices of the complexity of individual lotteries, these can directly be used to predict the complexity of valuation problems. Regarding larger discrete menus, however, our indices require some work to be generalized. In our *EV Tasks* experiment, we also included 102 menus with between three and five options. Appendix I discusses the results. Menu size is strongly linked to both error rates and cognitive uncertainty. This suggests that incorporating menu size into our indices would be productive. The main challenge we see is that extending to larger menus would necessitate averaging measures such as excess dissimilarity across multiple different lottery pairs in the set.

References

- Agranov, Marina, Paul J Healy, and Kirby Nielsen**, “Stable Randomization,” *Available at SSRN*, 2020.
- Andersson, Ola, Håkan J Holm, Jean-Robert Tyran, and Erik Wengström**, “Risk aversion relates to cognitive ability: Preferences or noise?,” *Journal of the European Economic Association*, 2016, 14 (5), 1129–1154.
- Arieli, Amos, Yaniv Ben-Ami, and Ariel Rubinstein**, “Tracking decision makers under uncertainty,” *American Economic Journal: Microeconomics*, 2011, 3 (4), 68–76.
- Benjamin, Daniel J, Sebastian A Brown, and Jesse M Shapiro**, “Who is ‘behavioral’? Cognitive ability and anomalous preferences,” *Journal of the European Economic Association*, 2013, 11 (6), 1231–1255.
- Bernheim, B Douglas and Charles Sprenger**, “On the empirical validity of cumulative prospect theory: Experimental evidence of rank-independent probability weighting,” *Econometrica*, 2020, 88 (4), 1363–1409.
- Birnbaum, Michael H**, “Three new tests of independence that differentiate models of risky decision making,” *Management Science*, 2005, 51 (9), 1346–1358.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer**, “Salience theory of choice under risk,” *Quarterly Journal of Economics*, 2012, 127 (3), 1243–1285.
- Bouchouicha, Ranoua, Jilong Wu, and Ferdinand M Vieider**, “Choice lists and ‘standard patterns’ of risk-taking,” Technical Report, Ghent University, Faculty of Economics and Business Administration 2023.
- Bruhin, Adrian, Helga Fehr-Duda, and Thomas Epper**, “Risk and rationality: Uncovering heterogeneity in probability distortion,” *Econometrica*, 2010, 78 (4), 1375–1412.
- Buschena, David and David Zilberman**, “Generalized expected utility, heteroscedastic error, and path dependence in risky choice,” *Journal of Risk and Uncertainty*, 2000, 20, 67–88.
- Carvalho, Leandro and Dan Silverman**, “Complexity and sophistication,” Technical Report, National Bureau of Economic Research 2019.
- Danz, David, Lise Vesterlund, and Alistair J Wilson**, “Belief elicitation and behavioral incentive compatibility,” *American Economic Review*, 2022.

Enke, Benjamin and Thomas Graeber, “Cognitive Uncertainty,” *Quarterly Journal of Economics*, 2023.

— , — , and **Ryan Oprea**, “Complexity and Time,” *Working Paper*, 2023.

— , — , and — , “Confidence, self-selection and Bias in the Aggregate,” *American Economic Review*, 2023.

Erev, Ido, Alvin E Roth, Robert Slonim, and Greg Barron, “Combining a theoretical prediction with experimental evidence,” *Available at SSRN 1111712*, 2002.

— and **Greg Barron**, “On adaptation, maximization, and reinforcement learning among cognitive strategies.,” *Psychological review*, 2005, 112 (4), 912.

— , **Eyal Ert, Alvin E Roth, Ernan Haruvy, Stefan M Herzog, Robin Hau, Ralph Hertwig, Terrence Stewart, Robert West, and Christian Lebiere**, “A choice prediction competition: Choices from experience and from description,” *Journal of Behavioral Decision Making*, 2010, 23 (1), 15–47.

— , — , **Ori Plonsky, Doron Cohen, and Oded Cohen**, “From anomalies to forecasts: Toward a descriptive model of decisions under risk, under ambiguity, and from experience.,” *Psychological review*, 2017, 124 (4), 369.

Frydman, Cary and Lawrence J Jin, “On the Source and Instability of Probability Weighting,” *Working Paper*, 2023.

— and **Lawrence Jin**, “Efficient Coding and Risky Choice,” *Quarterly Journal of Economics*, 2021.

Fudenberg, Drew and Indira Puri, “Evaluating and extending theories of choice under risk,” *Working paper; MIT Economics*, 2021.

— , **Jon Kleinberg, Annie Liang, and Sendhil Mullainathan**, “Measuring the completeness of economic models,” *Journal of Political Economy*, 2022, 130 (4), 956–990.

Gabaix, Xavier, “Behavioral inattention,” in Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *Handbook of Behavioral Economics-Foundations and Applications* 2, Elsevier, 2019, p. 261.

— and **Thomas Graeber**, “The Complexity of Economic Decisions: Theory and Experimental Evidence,” *Work in progress*, 2023.

Gillen, Ben, Erik Snowberg, and Leeat Yariv, “Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study,” *Journal of Political Economy*, 2019.

Halevy, Yoram, “Ellsberg revisited: An experimental study,” *Econometrica*, 2007, 75 (2), 503–536.

Harbaugh, William T, Kate Krause, and Lise Vesterlund, “The fourfold pattern of risk attitudes in choice and pricing tasks,” *The Economic Journal*, 2010, 120 (545), 595–611.

Hey, John D, “Experimental investigations of errors in decision making under risk,” *European Economic Review*, 1995, 39 (3-4), 633–640.

Holt, Charles A and Susan K Laury, “Risk aversion and incentive effects,” *American economic review*, 2002, 92 (5), 1644–1655.

Huck, Steffen and Georg Weizsäcker, “Risk, complexity, and deviations from expected-value maximization: Results of a lottery choice experiment,” *Journal of Economic Psychology*, 1999, 20 (6), 699–715.

Iyengar, Sheena S and Emir Kamenica, “Choice proliferation, simplicity seeking, and asset allocation,” *Journal of Public Economics*, 2010, 94 (7-8), 530–539.

Khaw, Mel Win, Ziang Li, and Michael Woodford, “Cognitive imprecision and small-stakes risk aversion,” *The review of economic studies*, 2021, 88 (4), 1979–2013.

— , — , and — , “Cognitive Imprecision and Stake-Dependent Risk Attitudes,” Technical Report, National Bureau of Economic Research 2022.

Martínez-Marquina, Alejandro, Muriel Niederle, and Emanuel Vespa, “Failures in contingent reasoning: The role of uncertainty,” *American Economic Review*, 2019, 109 (10), 3437–74.

McGranaghan, Christina, Kirby Nielsen, Ted O’Donoghue, Jason Somerville, and Charles D Sprenger, “Distinguishing common ratio preferences from common ratio effects using paired valuation tasks,” *Unpublished Manuscript*, 2022.

Nielsen, Kirby and John Rehbeck, “When choices are mistakes,” *American Economic Review*, 2022, 112 (7), 2237–68.

Oprea, Ryan, “Simplicity Equivalents,” Technical Report, Working Paper 2022.

Peterson, Joshua C, David D Bourgin, Mayank Agrawal, Daniel Reichman, and Thomas L Griffiths, “Using large-scale experiments and machine learning to discover theories of human decision-making,” *Science*, 2021, 372 (6547), 1209–1214.

Puri, Indira, “Simplicity and Risk,” Available at SSRN 3253494, 2022.

- Rubinstein, Ariel**, “Similarity and decision-making under risk (Is there a utility theory resolution to the Allais paradox?),” *Journal of economic theory*, 1988, 46 (1), 145–153.
- Sonsino, Doron, Uri Benzion, and Galit Mador**, “The complexity effects on choice with uncertainty—Experimental evidence,” *The Economic Journal*, 2002, 112 (482), 936–965.
- Verstyuk, Sergiy**, “Log, Stock and Two Simple Lotteries,” *Available at SSRN 2763959*, 2016.
- Vieider, Ferdinand M**, “Violence and risk preference: experimental evidence from Afghanistan: comment,” *American Economic Review*, 2018, 108 (8), 2366–82.
- , “Decisions under Uncertainty as Bayesian Inference on Choice Options,” 2022.
- Wakker, Peter P**, “A Criticism of Bernheim & Sprenger’s (2020) Tests of Rank Dependence,” *Available at SSRN 4188758*, 2022.
- Webb, Ryan, Paul W Glimcher, and Kenway Louie**, “The normalization of consumer valuations: Context-dependent preferences from neurobiological constraints,” *Management Science*, 2021, 67 (1), 93–125.
- Woodford, Michael**, “Modeling imprecision in perception, valuation, and choice,” *Annual Review of Economics*, 2020, 12, 579–601.
- Zilker, Veronika, Ralph Hertwig, and Thorsten Pachur**, “Age differences in risk attitude are shaped by option complexity.,” *Journal of Experimental Psychology: General*, 2020, 149 (9), 1644.

ONLINE APPENDIX

A Previous Literature

Table 5 provides an overview of prior studies that have investigated either (i) how noisiness in lottery choice depends on choice set features or (ii) systematic complexity-averse or -seeking behavior.

Table 5: Experimental literature on how lottery complexity depends on choice set features

Choice set feature	Result	Papers
Number of states	Aversion	Huck and Weizsäcker (1999), Sonsino et al. (2002), Iyengar and Kamenica (2010), Carvalho and Silverman (2019), Bernheim and Sprenger (2020), Puri (2022), Fudenberg and Puri (2021)
Number of states	Seeking	Birnbaum (2005), Erev et al. (2017), see Wakker (2022) for additional references
Number of states	Higher noise	Hey (1995), Huck and Weizsäcker (1999), Sonsino et al. (2002), Zilker et al. (2020)
Absolute dist. b/w CDFs	Higher noise	Buschena and Zilberman (2000), Erev et al. (2002), Erev et al. (2010)
Compound prob.	Aversion	Halevy (2007), Gillen et al. (2019)
Compound prob.	Higher noise	Enke and Graeber (2023)
Opaque payouts / prob.	Higher noise	Enke and Graeber (2023), Zilker et al. (2020)
Payout range	Higher noise	Bruhin et al. (2010)
Payout magnitude	Higher noise	Webb et al. (2021)
Dominance	Lower noise	Agranov et al. (2020)
Payout variance (decisions from experience)	Higher noise	Erev and Barron (2005)

B Potential Complexity Features

Consider a choice between two lotteries indexed by j and denoted by letters A, B etc. Each lottery is characterized by payout probabilities $(p_1^j, \dots, p_{k_j}^j)$ and payoffs $(x_1^j, \dots, x_{k_j}^j)$, where k_j denotes the number of distinct payout states of lottery j . For features defined at the level of the choice set, we first “couple” the lotteries to put them in a common state-space with states $1, \dots, k$, such that in the “worst” state, both A and B pay out their worst outcomes, in the “best” state, both A and B pay out their best outcomes, and so on. In the construction of our complexity indices, we include the features listed in Table 6. Whenever a feature is defined for a single lottery rather than a choice set, we include the average feature in the set. For continuous features f , we include the linear term (f), square (f^2), and the natural log ($\ln(f + 1)$, where the added 1 ensures bounded values for f ranging between 0 and 1). In total, the set of features contains 56 elements.

Table 6: Potential complexity features

Feature	Defined on	Formal definition
Number of states	Option	k_j
Payout range	Option	$\max\{x_1^j, \dots, x_{k_j}^j\} - \min\{x_1^j, \dots, x_{k_j}^j\}$
Variance	Option	$\sum_{s=1}^{k_j} P_s^j (x_s^j)^2 - (\sum_{s=1}^{k_j} P_s^j x_s^j)^2$
Payout variance	Option	$1/k_j \sum_{s=1}^{k_j} (x_s^j - \bar{x}^j)^2$
Probability variance	Option	$1/k_j \sum_{s=1}^{k_j} (p_s^j - \bar{p}^j)^2$
Magnitude	Option	$1/k_j \sum_{s=1}^{k_j} x_s^j $
Pure Gains	Option	$\mathbb{1}\{x_s^j \geq 0 \forall j\}$
Mixed	Option	$\mathbb{1}\{\exists x_s^j > 0 \wedge \exists x_s^j < 0\}$
Pure Loss	Option	$\mathbb{1}\{x_s^j < 0 \forall j\}$
Distance to certainty	Option	$1/k_j \sum_{s=1}^{k_j} \min\{p_s^j; 1 - p_s^j\}$
Payout-weighted dist. to certainty	Option	$1/k_j \sum_{s=1}^{k_j} x_s^j \min\{p_s^j; 1 - p_s^j\}$
Entropy	Option	$\sum_{s=1}^{k_j} P_s^j (-\ln(p_s^j))$
Normalized payout dispersion	Option	$1/k_j \sum_{s=1}^{k_j} \frac{ x_s^j - \bar{x}^j }{ \bar{x}^j }$
Normalized Variance	Option	$(1/\text{Magn.}^2) \cdot \text{Var}$, with Magn., Var. as defined above
Irregular probabilities	Option	$\mathbb{1}(p_s^j \in \{0.01, 0.05, 0.1, \dots, 0.9, 0.95, 0.99\} \text{ for } s = 1, \dots, k_j)$
CDF self-distance	Option	$\sum_{s=1}^{k_j} x_s^j - EV(j) p_s^j$
Compound	Option	
Weak dominance	Choice set	$\mathbb{1}\{F_A(x) \leq F_B(x) \forall x\}$
Excess dissimilarity	Choice set	$\sum_{s=1}^k x_s^A - x_s^B p_s - EV(A) - EV(B) $, where states s are ordered by payouts
Absolute payoff difference	Choice set	$\sum_{s=1}^k x_s^A - x_s^B $
Probability difference	Choice set	$\sum_{x \in X} f_A(x) - f_B(x) $, where $X = \{x_1^A, \dots, x_{k_A}^A\} \cup \{x_1^B, \dots, x_{k_B}^B\}$

C Additional Tables

Table 7: Summary statistics for problems across experiments

Experiment		# options	Safe payment	# states	Var	Scale	Mixed	Dominance
<i>EV Tasks</i>	Mean, median	2, 1, 2	56%	3, 2, 2	507, 130	31, 25	44%	5%
	Range	2, 5		2, 7	0, 28934	1, 213		
<i>Choice Tasks</i>	Mean	2	80%	3, 3	725	31	47%	7%
	IQR	2, 2		2, 5	13, 630	14, 41		
<i>Choice Tasks</i> from PEA	Mean	2	58%	3, 7	454	29	53%	17%
	IQR	2, 2		2, 5	19, 530	15, 40		
<i>Prior Manipulation</i>	Mean	2	81%	3, 2	728	31	47%	7%
	IQR	2, 2		2, 4	7, 718	12, 43		

Notes. PEA = Peterson et al. (2021). Scale = absolute average payout. We display information for the lottery with the largest number of distinct payout states.

Table 8: LASSO coefficients for *OPC* and *SPC*

Feature	Coefficients		Coefficients		
	OPC	SPC	Feature	OPC	SPC
Intercept	1.5E-01	1.1E-03	Log Payout Dispersion		-9.0E-03
Abs EV Difference	-3.3E-02	-3.4E-03	Log Norm. Variance		-1.3E-01
Abs EV Difference Sq	1.4E-03	9.5E-05	Log CDF Self-Dist		6.8E-02
Scale		-7.7E-04	Sq Scale	-9.4E-07	2.1E-06
Range		-2.1E-04	Sq Range		1.2E-07
Variance		2.0E-05	Sq Variance		-5.3E-10
Payout Variance		-2.4E-06	Sq Payout Variance		2.2E-10
Dist to Certainty (DC)	-2.2E-01		Sq Num States		1.1E-03
Pay-wtd DC	-2.0E-04	-9.1E-04	Sq DC		-2.4E-01
Entropy		1.2E-04	Sq Pay-wtd DC	-4.6E-06	-6.7E-06
Probability Variance		-1.8E-01	Sq Entropy		-3.1E-02
Payout Dispersion	4.3E-03	4.8E-04	Sq Prob. Variance		2.9E+00
Norm. Variance		7.9E-02	Sq Payout Dispersion		2.5E-03
CDF Self-Dist	-7.6E-04	-2.9E-03	Sq Norm. Variance	4.5E-03	8.9E-03
Log Scale	3.3E-02	1.9E-02	Sq CDF Self Dist		1.4E-05
Log Range		1.9E-02	Gains	-1.0E-02	-2.2E-02
Log Variance		-3.0E-02	Irregular Probabilities	3.5E-02	-4.1E-04
Log Payout Variance	1.4E-02	-5.3E-04	Excess Dissimilarity		-1.2E-03
Log Num States	1.3E-02	-2.9E-02	Log Excess Dissimilarity	5.2E-02	3.4E-02
Log DC	-3.4E-03	1.7E-01	Dominance	-5.4E-02	-2.5E-02
Log Pay-wtd DC		1.0E-02	Compound	7.6E-02	5.6E-02
Log Entropy		1.7E-01	Safe Option		4.5E-02
Log Prob. Variance		-1.9E-01			

Notes. Coefficients of LASSO regression of problem-level errors rates or cognitive uncertainty on choice set features in the *EV Tasks* experiment.

Table 9: LASSO coefficients for *OLC* and *SLC*

Feature	Coefficients	
	OPC	SPC
Intercept	7.0E-02	8.2E-02
Abs. EV Difference	-4.1E-02	-7.7E-03
Sq Abs EV Difference	1.7E-03	2.8E-04
Expected Value	-4.7E-05	-2.9E-04
Mixed	2.1E-02	
Pure Gains		-3.7E-03
Pure Losses	-2.0E-02	1.8E-02
Irregular Probabilities		-1.3E-04
Scale		-2.2E-04
Variance	-4.7E-06	-5.9E-06
Norm Variance	-1.9E-03	
Payout Variance	-7.0E-07	-4.1E-06
Pay-wtd DC		4.3E-04
Log Scale	2.0E-02	8.7E-03
Sq Scale	-1.3E-06	
Log Range		7.1E-03
Log Variance	9.2E-03	
Log Payout Variance	2.9E-02	1.8E-03
Sq Payout Variance		3.6E-10
Log Num States	3.4E-02	
Sq Num States		1.5E-04
Sq DC	-4.6E-02	
Log Pay-wtd DC	1.2E-02	
Sq Pay-wtd DC	-1.7E-07	
Log Entropy		7.5E-03
Log Prob Variance		-8.6E-02
Sq Payout Dispersion	5.5E-03	-5.2E-04
Log Norm Variance	-4.8E-04	
Sq Norm Variance		7.4E-03
Log CDF Self-Dist		2.3E-02

Notes. Coefficients of LASSO regression of problem-level errors rates or cognitive uncertainty on lottery features in the *EV Tasks* experiment. The sample is restricted to problems in which one option is a safe payment, and only features of the lottery are used.

Table 10: No risk aversion in *EV Tasks*

	<i>Dependent variable:</i> 1 if selected lottery		
	(1)	(2)	(3)
Log variance of A	0.0025 (0.00)	0.0033 (0.00)	
Constant	0.52*** (0.01)	0.51*** (0.03)	0.51*** (0.02)
Controls for EV diff.	No	No	Yes
Observations	33586	33586	33586
R^2	-0.00	0.00	0.22

Notes. OLS estimates, standard errors (two-way clustered at subject and problem level) in parentheses. Each observation is a subject-decision. The sample is restricted to problems in which one option is a lottery and the other option a safe payment. Controls for EV difference include linear, second- and third-order terms. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Coefficients of features in subjective complexity indices

Index:	<i>Dependent variable:</i>		
	<i>Cognitive uncertainty</i>	<i>Implied subj. logit imprecision \hat{s}</i>	
		<i>OPC</i>	<i>OAC</i>
	(1)	(2)	(3)
Log excess dissimilarity	0.018*** (0.00)	1.79*** (0.16)	
No dominance	0.043*** (0.01)	0.83** (0.41)	
Average log payout magnitude	0.0036* (0.00)	1.43*** (0.20)	
Average log number of states	0.052*** (0.01)	3.10*** (0.63)	
Frac. lotteries involving loss	0.028*** (0.00)	2.24*** (0.35)	
Frac. lotteries involving compound prob.	0.12*** (0.01)	6.84*** (1.35)	
Absolute EV difference	-0.0037** (0.00)		
Absolute EV difference sqr.	0.00010 (0.00)		
Log variance			1.17*** (0.13)
Log payout magnitude			1.05*** (0.29)
Log number of states			2.95*** (0.53)
1 if involves loss			1.71*** (0.39)
1 if involves compound prob.			3.14*** (0.84)
Constant	0.021** (0.01)	-5.35*** (0.88)	-6.75*** (1.03)
Observations	1587	1587	935
R ²	0.38	0.24	0.26

Notes. OLS estimates, robust standard errors in parentheses. An observation is a decision problem from the train set in the *EV Tasks* experiment. In columns (2) and (3), the dependent variable is the implied logit precision $\hat{s}_{A,B}^{EV}$ as defined in eq. (9). In column (3), the sample is restricted to problems in which one option is a safe payment, and the independent variables are features of the non-degenerate lottery.
 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: Noise and aversions in risky choice

Sample:	<i>Dependent variable:</i> Fraction of subjects who chose lottery			
	EV(Lottery) > Safe payment		EV(Lottery) < Safe payment	
	(1)	(2)	(3)	(4)
Log variance	-0.052*** (0.00)		0.029*** (0.00)	
Log scale		-0.089*** (0.01)		0.091*** (0.00)
Controls for EV diff.	Yes	Yes	Yes	Yes
Observations	2378	2378	2485	2485
R ²	0.49	0.19	0.35	0.37

Notes. OLS estimates, robust standard errors in parentheses. Each observation is a choice problem between a non-degenerate lottery and a safe payment. The sample in columns (4) and (8) is restricted to choice problems in which the safe payment is weakly positive. Controls for EV difference include linear, second- and third-order terms. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D Additional Figures

D.1 Screenshots of Experimental Decision Screens

Decision 1/50

Which lottery has the highest average payout if the computer runs it many, many times?
Please select one.

Lottery A Prob. 5%: Get \$26 Prob. 10%: Get \$24 Prob. 15%: Get \$23 Prob. 10%: Get \$17 Prob. 5%: Get \$11 Prob. 15%: Get \$5 Prob. 40%: Lose \$15	Lottery B Prob. 100%: Get \$11
--	-----------------------------------

How certain are you that each lottery has the highest average payout?
Please allocate 100 certainty points.

<input type="text"/> points	<input type="text"/> points
-----------------------------	-----------------------------

100 points left to allocate.

Remember, the "Next" button will only appear if you selected a lottery and entered a number into each textbox!

Figure 9: Example decision screen in *EV Tasks*

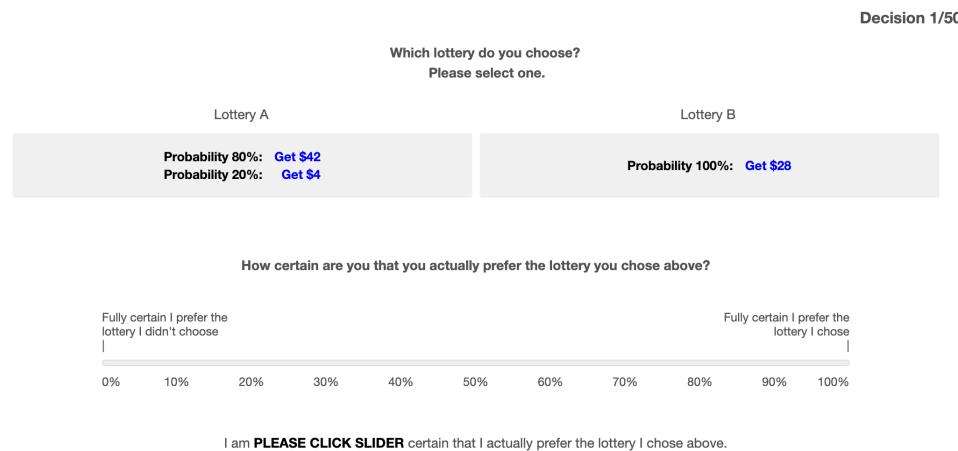


Figure 10: Example decision screen in *Choice Tasks*

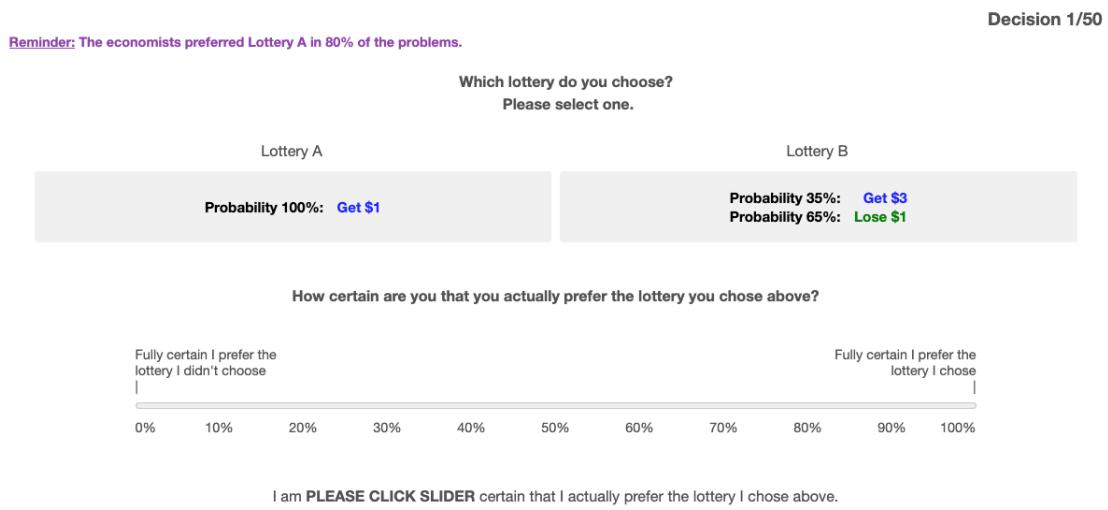


Figure 11: Example decision screen in *Prior Manipulation*

D.2 Figures for *EV Tasks* experiment

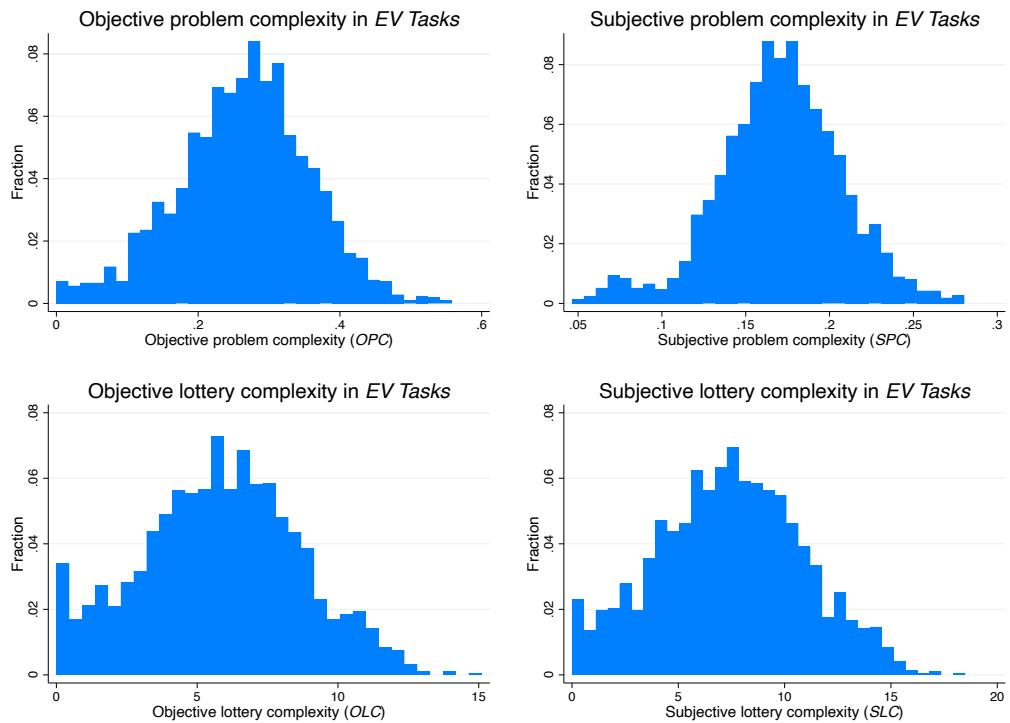


Figure 12: Histograms of complexity indices in *EV Tasks* data

D.3 Analyses for Risky Choice Data

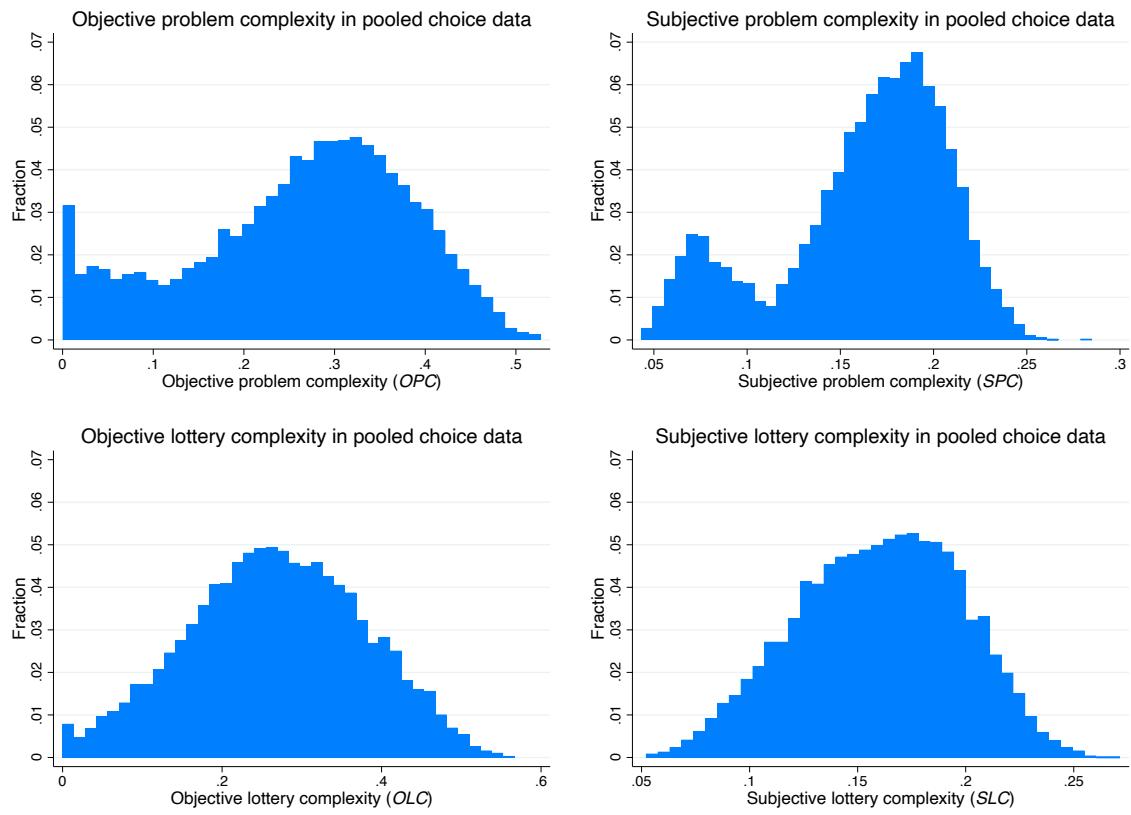


Figure 13: Histograms of complexity indices in pooled choice data

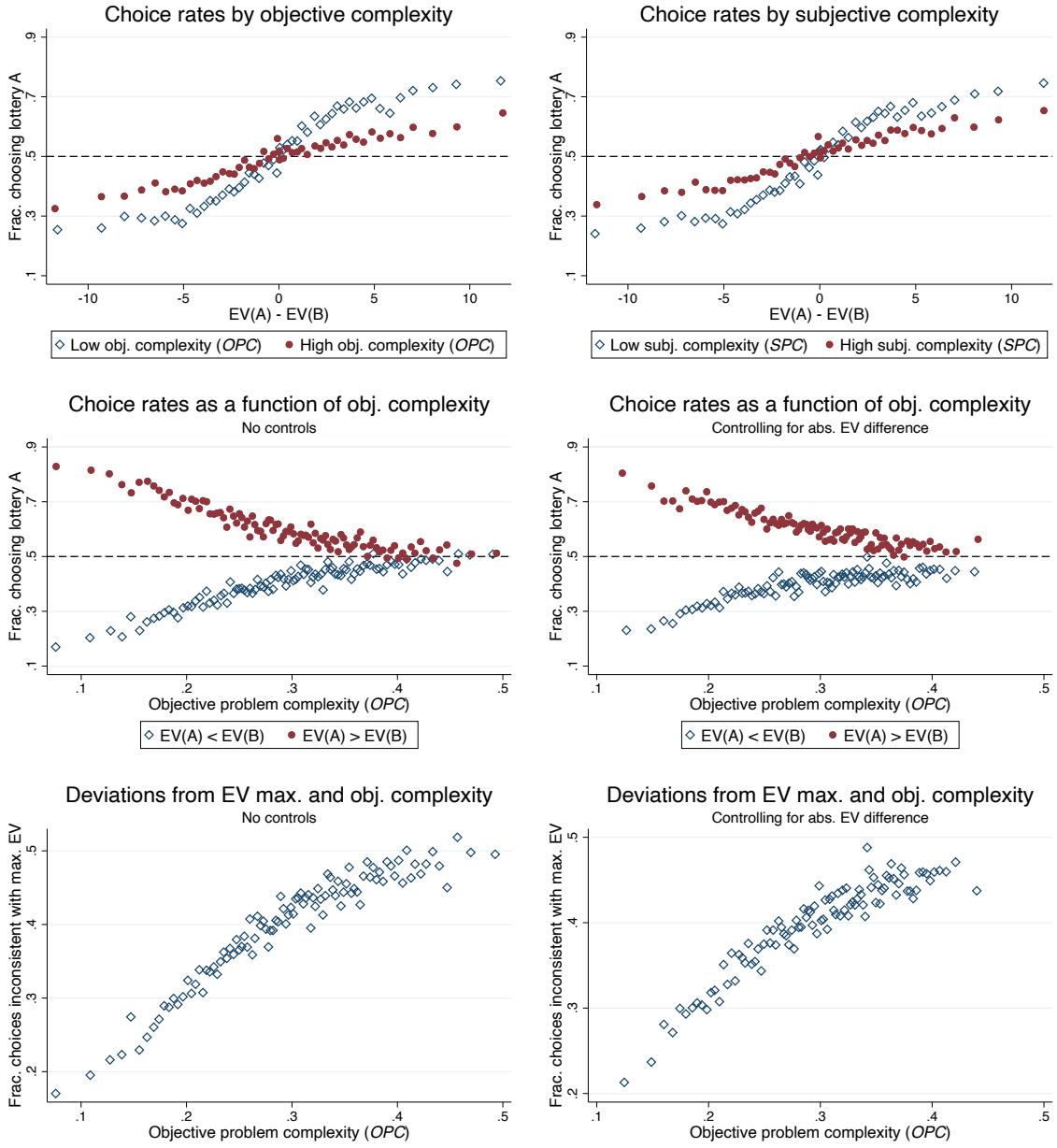


Figure 14: Complexity and deviations from expected value maximization in lottery choice, excluding choice problems that feature dominance. The top panels implement a median split by OPC / SPC , separately within each percentile of the EV difference between A and B. The middle right and bottom right panels show partial correlation plots that control for linear and squared terms of the absolute EV difference. Top panels constructed from 9,187 choice problems, mid and bottom panels omit problems with absolute EV difference of less than \$0.20, hence constructed from 8,698 choice problems.

E Effect of Compound Lotteries

This Appendix continues the analysis from Section 5.3 on the link between choice set features, noise and aversions. Figure 16 studies compound lotteries. As we can see in the bottom left panel, compound probabilities are the only feature that induce strong and essentially uniform ‘aversion’ – subjects are substantially less likely to choose compound

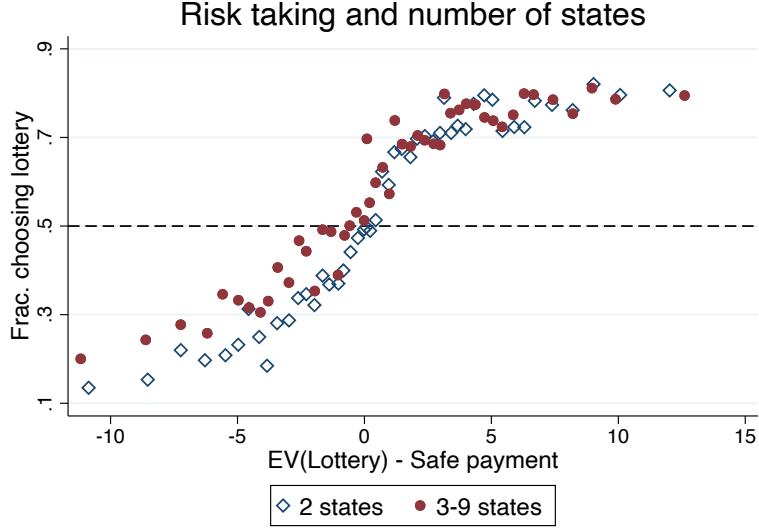


Figure 15: Binned scatter plots of choice rates for lottery vs. safe payment as a function of number of states. Based on the 6,466 choice problems in which one option is a safe payment.

lottery than reduced ones, regardless of whether the lottery is relatively attractive or unattractive. A widespread interpretation in the literature is that this pattern reflects “complexity aversion” (e.g., Gillen et al., 2019). While this is descriptively accurate, it is interesting to note that the same “aversion” is also visible in the (preferences-free) *EV Task*. This strongly suggests that the “aversion” in the choice data rather reflects “underestimation” – it appears that subjects do not choose compound lotteries because they systematically underestimate how much they pay out, rather than because they are averse to the complexity introduced by compound probabilities.

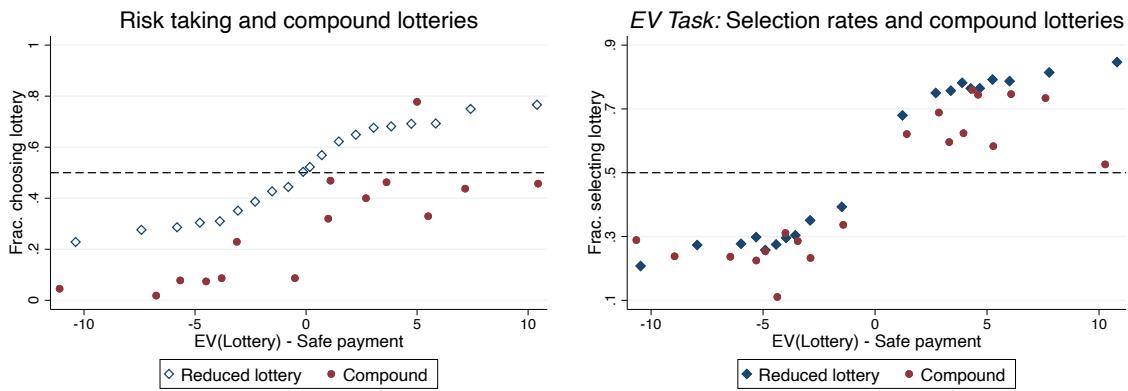


Figure 16: Left panels: Binned scatter plots of choice rates for lottery vs. safe payment (6,466 choice problems, 575,404 decisions). Right panels: Binned scatter plots of selection rates of lottery vs. safe payment in *EV Task* experiment (1,230 choice problems, 33,586 decisions).

F Replication of Results in Peterson et al. Data Excl. Losses

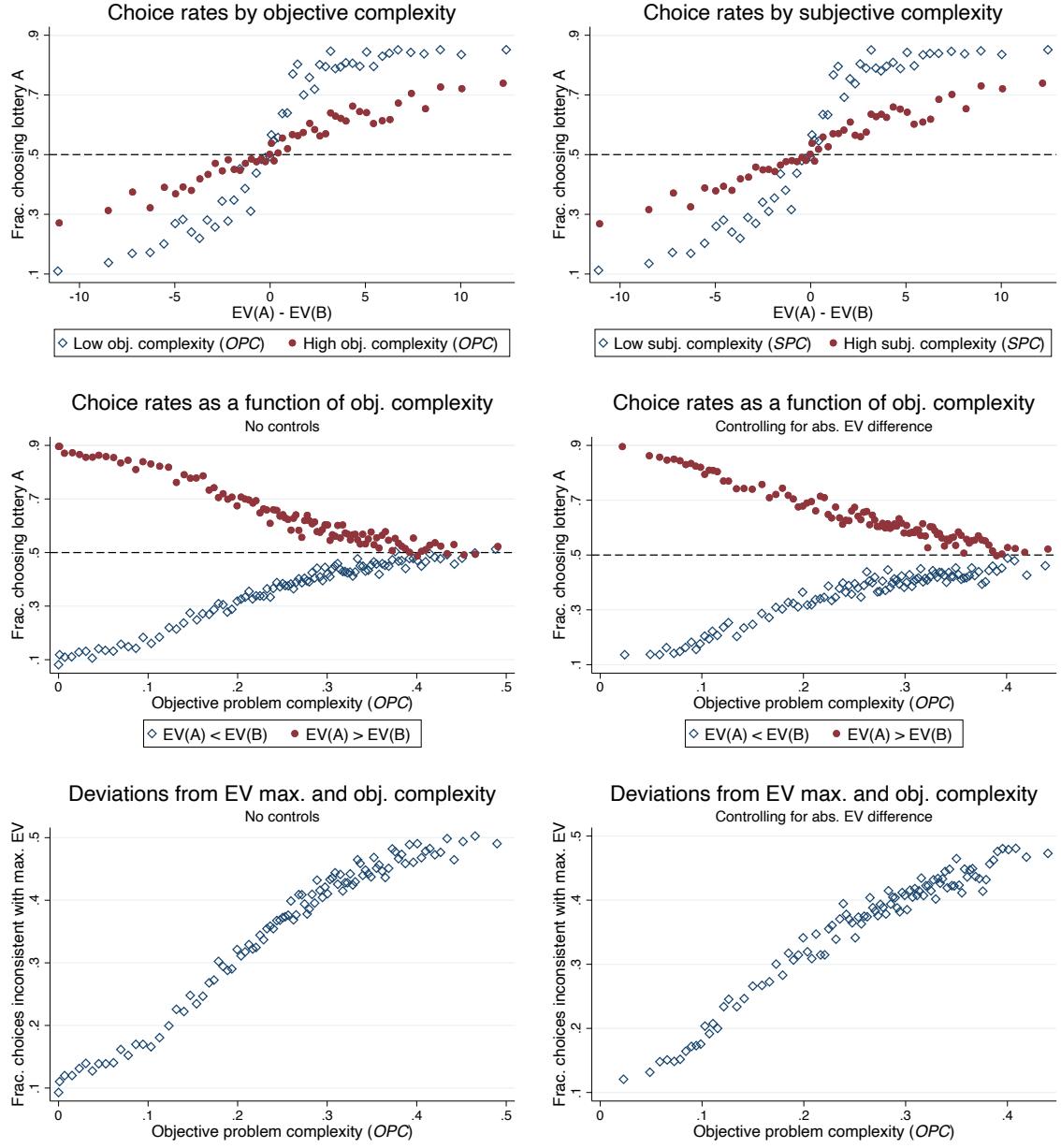


Figure 17: Complexity and deviations from expected value maximization in lottery choice in the Peterson et al. (2021) dataset, excluding choice problems that feature losses. The top panels implement a median split by *OPC* / *SPC*. The top panels implement a median split by *OPC* / *SPC*, separately within each percentile of the EV difference between A and B. The middle right and bottom right panels show partial correlation plots that control for linear and squared terms of the absolute EV difference. Top panels constructed from 3,598 choice problems, mid and bottom panels omit problems with absolute EV difference of less than \$0.20, hence constructed from 3,378 choice problems.

G Details on Structural Estimations

Prospect theory. We allow up to five additional parameters:

$$EU_{PT}(x) = \sum \frac{\chi p^\gamma}{\chi p^\gamma + (1-p)^\gamma} u(x), \quad (15)$$

where

$$u(x) = \begin{cases} x^\alpha & \text{if } x \geq 0 \\ -\lambda(-x)^\beta & \text{if } x < 0 \end{cases} \quad (16)$$

Salience theory. We first put both lotteries into a common state space by “correlating” the payoffs. Then we specify salience values for each option as follows:

$$\sum_s \delta_s p_s u(x_s), \quad (17)$$

where ω_s is the salience weight of state s given by

$$\omega_s = \frac{\delta^{k_s}}{\sum p_s \delta^{k_s}}, \quad (18)$$

and k_s is the salience rank of state s , which is given by the ordering of the salience function

$$\sigma(x_s^A, x_s^B) = \frac{|x_s^A - x_s^B|}{|x_s^A| + |x_s^B| + \theta} \quad (19)$$

with lower ranks k_s indicating a higher value of the salience function. The primitives we estimate for the salience model are the usual utility parameters (α , β , and λ), the logit noise parameters η , and the two salience parameters θ and δ .

Estimation results. Table 13 summarizes the parameter estimates across the different models. We do not report standard errors because they are almost always close to zero given the large sample.

Table 13: Parameter estimates for model estimations

Parameter	EV	Curvature	PT	EV salience	Curv. sal.	EV + C	Curvature + C	PT + C	EV salience + C	Curv. salience + C	Model	
α	0.83	0.80		0.81		0.79		0.78		0.80		
β	0.79	0.74		0.77		0.74		0.68		0.75		
λ		1.12						1.34				
χ		0.92						0.89				
γ		0.98						0.96				
θ		0.92		1.15					-0.01		0.92	
δ		0.98		1.95					0.88		0.98	
η_0	0.12	0.24	0.25	0.12	0.26	0.05	0.14	0.14	0.04	0.13		
η_1					0.31	0.58	0.55	0.55	0.32	0.57		

Notes. Parameter estimates for model estimations described in Section 6. “+ C” = complexity-dependent heteroscedasticity.

H Complexity and Reliance on a Prior

Experimental Design. The analyses reported in the main text showed that choice rates in more complex problems are more strongly compressed to 50-50. Through the lens of a Bayesian cognitive noise model (Woodford, 2020; Gabaix, 2019), an interpretation of these patterns is that subjects' prior belief does not systematically favor Lottery A or B. This is a very natural assumption because the label and location of the lotteries on subjects' decision screen were randomized.

To study the role of priors, we exogenously manipulate subjects' prior beliefs. We implement two treatments, *Low A Prior* and *High A Prior*. These treatments build on the *Choice Tasks* experiment. In both treatments, the instructions inform subjects that some economists completed the same choice problems as the ones they will encounter, and that the economists preferred Lottery A in 80% (*High A Prior* treatment) or 20% (*Low A Prior* treatment) of all choice problems. This information is also placed on each decision screen as a reminder, see Appendix Figure 11. This information is truthful and based on the economists (us) always preferring the lottery with the higher expected value.

The problems are then arranged such that subjects in both treatments actually work on exactly the same choice problems for the first 20 out of 50 rounds ("target problems"). The problems in the last 30 rounds are filler tasks that differ across treatments to make the information provided truthful. In the first 20 rounds, Lottery A actually has a higher expected value in 50% of all problems. In the last 30 rounds, Lottery A always has a higher expected value in the *High A Prior* treatment and a lower expected value in the *Low A Prior* treatment. Because we only study the target problems in the first 20 rounds, the only aspect that differs across treatments is the prior belief.

Our main object of interest is not so much the raw treatment effect of this manipulation but, rather, whether the treatment effect increases in complexity. We pre-registered the sample size and the prediction that it does. 501 subjects participated in this experiment, for a total of 10,020 decisions in the target problems. We randomly generated 200 choice problems that exhibit large variation in complexity (as measured by *OPC* and *SPC*), and randomly assigned choice problems to subjects. The experiment was pre-registered on aspredicted.org under #130662.

Results. Table 14 summarizes the results. Column (1) shows that the treatment shifts choice rates by 5.7 percentage points. Our main interest is in how this effect depends on complexity. Column (2) interacts both the prior manipulation treatment and the expected value difference with *OPC*. We find that the treatment effect increases in complexity, an effect that is marginally statistically significant. Moreover, as in the main text,

Table 14: Effect of complexity on treatment effect of manipulating prior

	Dependent variable: 1 if chose lottery A				
	(1)	(2)	(3)	(4)	(5)
High A Prior vs. High B Prior	0.057*** (0.01)	0.025 (0.02)	-0.0057 (0.03)	0.022 (0.02)	0.0010 (0.02)
EV(A) - EV(B)	0.036*** (0.00)	0.084*** (0.01)	0.11*** (0.01)		
High A Prior vs. High B Prior \times Objective problem complexity		0.12* (0.07)		0.11* (0.06)	
EV(A) - EV(B) \times Objective problem complexity		-0.20*** (0.03)			
High A Prior vs. High B Prior \times Subjective problem complexity			0.37** (0.16)	0.30** (0.14)	
EV(A) - EV(B) \times Subjective problem complexity			-0.47*** (0.06)		
Objective problem complexity			-0.12 (0.14)		
Subjective problem complexity				-0.34 (0.34)	
Round FE	Yes	Yes	Yes	Yes	Yes
Problem FE	No	No	No	Yes	Yes
Observations	10020	10020	10020	10020	10020
R ²	0.16	0.21	0.21	0.40	0.40

Notes. OLS estimates, standard errors (two-way clustered at subject and problem level) in parentheses. An observation is a binary decision. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

we see that higher complexity makes choice rates less responsive to differences in expected values. Together, these two interaction effects paint a picture according to which complexity produces stronger compression to a prior.

Column (3) shows that these results are, if anything, stronger using *SPC*, regarding both their quantitative magnitude and their statistical significance. While this is an ex post interpretation, we believe that the stronger results for *SPC* make intuitive sense. The reason is that in our experiment the prior is very explicitly induced, such that subjects may be more likely to use this external information (“advice”) if they find the problem subjectively difficult (rather than when they objectively make many decision errors). After all, one’s tendency to accept advice should be governed by one’s subjective probability of making a mistake (as captured by *SPC*) rather than the corresponding objective probability (as captured by *OPC*).

Finally, columns (4) and (5) present alternative regression specifications in which we control for fixed effects for each choice problem. In these analyses, we, hence, only compare choice rates on identical problems. Again, reliance on a prior is more pronounced with higher complexity.

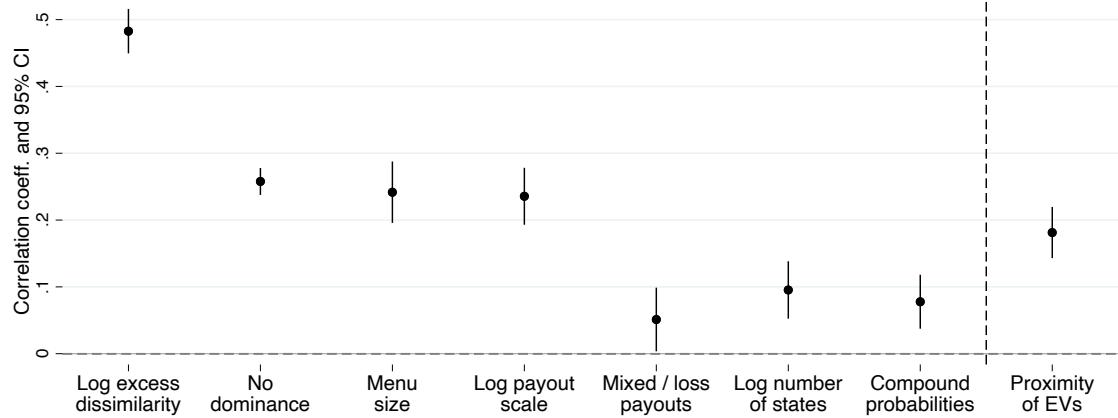
I Extension: Larger Menus

In our *EV Tasks* experiment, we also included 102 menus with between three and five options. This allows us to study how error rates and cognitive uncertainty (i) depend on menu size and (ii) how they depend on the other features we discussed in the main text when the menus are not binary.

Figure 18 shows the link between choice set features and errors / cognitive uncertainty. With menus larger than two, calculating excess dissimilarity and the presence of dominance relationships is less trivial. We proceed by calculating both of these variables for each pair of options in the set, and then average across all pairs. Similarly, for the lottery-specific features such as payout scale, we average across all lotteries in the set. We have verified that we get very similar results if we, instead, work with the maximum in the set.

Figure 18 shows that menu size has a large effect on both errors and *CU*, though the effect is still smaller than that of excess dissimilarity.

EV Task: Error rates and choice set features



EV Task: Cognitive uncertainty and choice set features

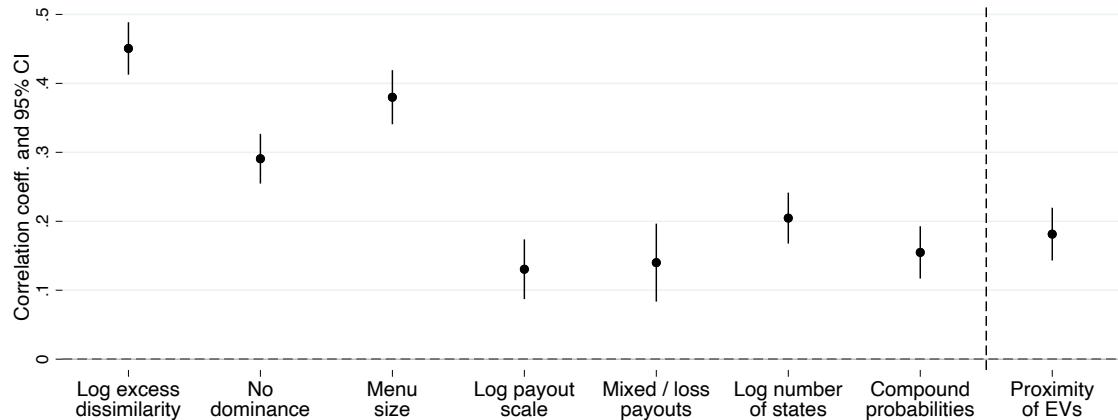


Figure 18: Correlations between choice set features and errors / cognitive uncertainty in the full dataset in the *EV Task*, including menus with more than two options. Pairwise partial correlation coefficients between task-level error rates / average cognitive uncertainty and choice set features, controlling for the absolute difference in expected values, squared and cubic terms. Excess dissimilarity, dominance and proximity of expected values are computed for each pair in the set and then averaged across pairs. Log scale, mixed / loss payouts, log number of states and compound probabilities are computed separately for each lottery and then averaged across the lotteries in a choice set.

J Quasi-Random Problem-Generation Procedures

J.1 Quasi-Random Problems in EV Tasks

We quasi-randomly generate 1,998 pairs of lotteries for our *EV Task*. While we could in principle generate a wide variety of lotteries deterministically, we introduce randomness in order to better span the space of *potential* lottery pairs. In particular, we do not want to pre-ordain which features of lotteries may matter for complexity; by allowing variation in the problems beyond what we as experimenters explicitly think to vary, we produce a rich set of lotteries. Where possible, we have designed our algorithm to maximize orthogonality between lottery features that we anticipate will be important contributors to complexity. When this is not possible, we instead try to ensure that candidate complexity features are positively correlated with one another.

J.1.1 Complex Lotteries

We begin by generating 3,200 “base lotteries,” each with two states; these will be our “complex” options. We will leave some of these as two-state lotteries, and others will be inputs to an iterative “event-splitting” procedure in order to construct n-state lotteries. The first payoff x_1 is drawn from a positive exponential distribution with $\lambda = 0.04$ ($\mathbb{E}[x_1] = 25$). With probability 0.3, x_1 is set to be negative, and with probability 0.7, it remains positive. The second payoff x_2 is independently drawn identically to x_1 with probability 0.5; and with probability 0.5, x_2 is drawn from a normal distribution with mean equal to x_1 , standard deviation equal to 20. The probability p_1 associated with x_1 is in general drawn uniformly from a set of “nice” probabilities

$$P \equiv \{0.01, 0.05, 0.1, 0.15, \dots, 0.85, 0.9, 0.95, 0.99\},$$

and p_2 is necessarily equal to $1 - p_1$. In 5% of all problems, we allow for “strange” probabilities (always truncating at two digits). If a problem allows “strange” probabilities, then p_1 is instead calculated by a draw from a truncated normal distribution with mean 0.5 and standard deviation 0.25 (truncation is achieved by redrawing when the draw is less than 0.01 or larger than 0.99).

After generating the “base lotteries,” we randomly assign each problem to a number of states $k \in \{2, 3, 4, 5, 6, 7\}$. We assign values of k such that $\mathbb{P}(k = 2) = 0.5$, and $\mathbb{P}(k = y) = 0.1$ for $y \in \{3, 4, 5, 6, 7\}$. For problems with $k = 2$, the lottery is exactly the base lottery. For problems with $k > 2$, we iteratively “split” states of the base lottery until we reach the desired number of states. The splitting process proceeds as follows.

We first randomly select a state to split, from the states that have probability of at

least 0.1. The sampling process is a 50-50 mixture of uniform sampling, and probability-proportional sampling; that is, the sampling weights for the states are proportional to the probabilities associated with those states. This mixture makes it so that larger probability states are generally more likely to be split, but it's still reasonably likely that we can have a 6 or 7 state lottery with a high-probability state. Once a state x_i is selected for splitting, we will replace it with two different states x_{i1} and x_{i2} , and split the probability p_i across the two new states such that expected value of the lottery remains roughly constant. In practice, we achieve this by randomly drawing x_{i1} from a normal distribution with mean x_i , standard deviation 0.25 times the range of payoffs of the lottery. We set p_{i1} by dividing p_i by 2, and then adding 0.025 if this division yields a probability outside of P . We then set $p_{i2} = p_i - p_{i1}$, and choose x_{i2} such that $p_i x_i \approx p_{i1} x_{i1} + p_{i2} x_{i2}$. If the problem is allowed to have “strange” probabilities, then the procedure is virtually the same, except we allow p_{i1} to be a perturbation of $p_i/2$.

J.1.2 Simple Lotteries

We generate simple lotteries according to a similar process. Before generating these lotteries, we first randomly set the expected value for each one, based on the expected value of a paired complex lottery. Specifically, we draw the difference in expected value (ΔEV) from a mixture of uniform distributions with the following weights:

- with probability 0.2, $\Delta EV \sim \text{Unif}(-12, -6)$,
- with probability 0.3, $\Delta EV \sim \text{Unif}(-6, -3)$,
- with probability 0.3, $\Delta EV \sim \text{Unif}(3, 6)$, and
- with probability 0.2, $\Delta EV \sim \text{Unif}(6, 12)$.

Expected value for the simple lottery ($EV(S)$) is then set by adding ΔEV to the expected value of the paired complex lottery.

We again construct 2-state “base lotteries,” this time adding a restriction on the expected value. We draw x_1 from a normal distribution with mean given by the desired expected value, standard deviation of 20. We draw p_1 exactly as we did for the complex lotteries, and calculate p_2 accordingly. Finally, we set x_2 such that the expected value of the lottery is as desired.

We then randomly draw the number of states for the simple lottery. With probability 0.6, the simple lottery has one state, i.e. is a certain payment equal to the determined expected value. With probability 0.4, the number of states is drawn uniformly between

2 and the number of states in the paired complex lottery. That is, the states are drawn such that the simple lottery always has weakly fewer states than the complex lottery.

For the simple lotteries assigned to have more than one state, the base lotteries are split to achieve the desired number of states using exactly the same procedure as for the complex lotteries.

J.1.3 Dominance Problems Without Certain Option

Because we are also interested in studying dominance problems (where one lottery first order stochastically dominates the other), which don't arise very frequently in the algorithm described previously, we also intentionally construct pairs of lotteries with a dominance relationship. In particular, we construct pairs of lotteries with "non-obvious" dominance relationships, where it is not the case that all the payouts of one lottery are weakly larger than all the payouts of the other. We begin by randomly selecting 12% of our generated complex problems: for the selected problems, we will replace the simple lottery with an alternative 2-state lottery that has a dominance relationship with the paired complex lottery.

First, we randomly determine whether the new simple lottery will be dominated or dominating (each with probability 0.5). If the simple lottery will be dominated, then we begin by drawing an x_1 that is strictly less than the smallest payout in the complex lottery. Specifically, we uniformly draw a whole number between 1 and 5 and subtract this number from the minimum complex payout to determine x_1 . We then assign a probability $p_1 \in P$ to x_1 , ensuring that the probability is strictly larger than the probability assigned to the smallest payout in the complex lottery. Then, we choose x_2 to be an *intermediate value* for the complex lottery; that is, we ensure that x_2 is larger than the least payout and smaller than the greatest payout. We also enforce dominance, so x_2 is bounded above by the p_1 th quantile of the complex lottery. Because of how we chose p_1 , we can be sure that the p_1 th quantile is strictly larger than the minimum payout – that is, we will always be able to choose a strictly intermediate x_2 that satisfies these bounds. In particular, we set x_2 by uniformly drawing a whole number between the minimum complex payout and the p_1 th quantile of the complex lottery. If instead the simple lottery will be dominating, we follow a mirror image of this procedure, first drawing x_1 greater than the maximum complex payout, choosing an associated probability p_1 , and then drawing an intermediate x_2 that satisfies the dominance relationship.

In the generation of the dominance problems, we do not strictly specify the EV difference, and by construction, we will get EV differences that are higher on average in these problems than the rest of the problems. We don't see this as an issue, since it is natural that strictly dominating problems should also tend to have a larger EV advantage.

J.1.4 Multi-Option Menus

Until now, we have only discussed the construction of two-option menus; but we also generate prediction problems with three, four, and five lottery options. To do this, we randomly select 120 of the binary choice problems with some restrictions. For each of these problems, we create three, four, and five option counterparts by sequentially adding lotteries to the menus. The sequentially added lotteries generally follow the construction procedure of the simple lotteries, with a few adjustments.

We restrict the initial sample of problems to those where the complex lottery has at most three states, the simple lottery has at most two states, there are no “strange” probabilities, and there is no dominance relationship. These restrictions are motivated by a desire to keep the larger-menu problems from getting too complex; if we still allowed lotteries to have as many as 7 states and asked people to identify the best of 5 options, we anticipate that choice behavior would be too random to draw meaningful conclusions.

Then, we set expected values for the third, fourth, and fifth options by drawing a value uniformly between 5 and 13, and subtracting this from the maximum expected value of the first two options. That is, we are enforcing that one of the initial lotteries from the binary problem remain the correct answer, even with larger menus; and we are requiring that the added choices be substantially worse in terms of expected value. Note that of course, the options are presented to participants in random order; so the “ordering” here is not relevant to the experiment, but is nonetheless helpful for understanding the structure of our generated problems. Number of states for options three, four, and five are all set to two if the original complex problem has only states; otherwise, it is set to two with probability 0.7 and three with probability 0.3. Once the expected value and number of states are set, each of the extra options is generated according to exactly the same procedure as the original simple problems.

To finish up the set of larger menu problems, we drop all problems that exhibit a dominance relationship between any of the options. Of the remaining five-option problems, we randomly select 34 problems and use these along with the associated three- and four-option problems for a total of 102 large menu prediction problems.

J.1.5 Compound Lotteries

Of the binary problems where the complex lottery has two states, we randomly select 5% in which to replace the complex lottery with a compound lottery that reduces to the same lottery as the original. We restrict the potential sample for “compounding” to lotteries where the probabilities are no more extreme than 80-20. For the compound problems, we tell participants they will receive x_1 with probability q , which is drawn uniformly from some known range. In generating these problems, we assign this range

by uniformly sampling from $\{4, 10, 20, 40\}$. So, for example, if a problem original had $p_1 = 30$, and we draw a range of 10, then a participant in the experiment would see a lottery that promises x_1 with probability q , where q is between 20 and 40, and x_2 with probability $1 - q$.

J.1.6 Ex-post Selection of Lotteries

After completing this process of generating lottery problems, we also systematically drop some problems to ensure that our final set of problems is relatively well-balanced. The drops are motivated by not wanting to have too many problems that are very easy, as we anticipate that we would not be able to learn a lot from decision patterns in these problems. First, we drop all problems with an expected value difference larger than 16; this is only relevant for the specially constructed dominance problems. Then, we drop 90% of problems where the simple lottery is a certain payment and there is a dominance relationship. Essentially, the algorithm as designed produces more dominance problems than we would like, since we anticipate these will be relatively easy to participants. After trimming some of these dominance problems, 6.5% of the remaining problems have a dominance relationship. Of these, about 60% are two-state simple lotteries, and 40% are single-state simple lotteries. We also drop a handful of problems which contain a lottery wherein any of the payments in different states are equal; so the number of states for a lottery always captures the true number of distinct states. After this selection process, we are left with 1,998 problems.

J.2 Quasi-Random Problems in *Choice Tasks*

The problems were generated using a very similar procedure as in *EV Tasks*, except that (i) we only generate two-item menus and (ii) we do not generate problems that only consist of losses.

K Targeted Problems in EV Tasks

K.1 Design of Targeted Problems

This Appendix K presents all 120 “targeted” *EV Tasks* problems that we manually devised. They broadly fall into two categories.

1. In a first category of 96 problems, one option is a safe payment and the other one a non-degenerate lottery. Here, we designed three “sets,” each of which is defined by a base lottery. Across choice problems within each set, we manipulate specific features of the base lottery: scale (average absolute payout), variance, the presence of mixed gain-loss payouts, number of states, extremity of probabilities (distance to certainty), and the presence of compound probabilities. In designing these manipulations, we were careful to hold other aspects of the lotteries constant to the greatest degree possible. Of course, it is logically impossible to hold *all* features but one constant but we tried to the degree possible.
2. In a second category of 24 problems, both options in a choice set consisted of non-degenerate lotteries. Here, we manipulated the similarity of the CDFs of the two options while holding features such as expected value and variance constant.

Problems designed for lottery-specific features. We designed 96 choice problems in which one option was a non-degenerate lottery and the other one a safe payment. Table 15 provides an overview. The lotteries were organized in three sets, within each of which we test for effects of specific complexity features. Each set consists of the following types of lotteries:

1. Base: Lottery against which other lotteries are compared
2. Scale: Increase average absolute payout, holding payout range, mixed-ness and probabilities constant
3. Range / variance: Increase payout range / variance, holding EV, mixed-ness and probabilities constant
4. Mixed / scale: Shift all payouts up / down by a constant, such that lottery becomes mixed while payout range and probabilities are held constant
5. Mixed / range: Increase payout range / variance to make lottery mixed, holding EV and probabilities constant
6. Number of states: Increase number of states through event splits, holding EV constant and changing variance / range as little as possible

Table 15: Targeted EV problems in *EV Tasks* (1/2)

Lottery type	Set 1	Set 2	Set 3
Base	$p = (0.3, 0.7)$ $x = (32, 12)$	$p = (0.6, 0.4)$ $x = (45, 25)$	$p = (0.5, 0.5)$ $x = (-30, -56)$
Scale	$p = (0.3, 0.7)$ $x = (192, 172)$	$p = (0.6, 0.4)$ $x = (195, 175)$	$p = (0.5, 0.5)$ $x = (-200, -226)$
Range / variance	$p = (0.3, 0.7)$ $x = (53, 3)$	$p = (0.6, 0.4)$ $x = (61, 1)$	$p = (0.5, 0.5)$ $x = (-3, -83)$
Mixed / scale	$p = (0.3, 0.7)$ $x = (8, -12)$	$p = (0.6, 0.4)$ $x = (15, -5)$	$p = (0.5, 0.5)$ $x = (21, -5)$
Mixed / range	$p = (0.3, 0.7)$ $x = (151, -39)$	$p = (0.6, 0.4)$ $x = (91, -44)$	$p = (0.5, 0.5)$ $x = (83, -169)$
# states	$p = (0.15, 0.15, 0.2, 0.3, 0.2)$ $x = (38, 28, 16, 12, 8)$	$p = (0.3, 0.3, 0.1, 0.2, 0.1)$ $x = (50, 40, 35, 25, 15)$	$p = (0.25, 0.25, 0.2, 0.1, 0.2)$ $x = (-20, -40, -48, -56, -64)$
Distance certainty	$p = (0.15, 0.85)$ $x = (35, 15)$	$p = (0.85, 0.15)$ $x = (40, 20)$	$p = (0.15, 0.85)$ $x = (-21, -47)$
Compound	$p = (0.2 - 0.4, 0.6 - 0.8)$ $x = (32, 12)$	$p = (0.4 - 0.8, 0.2 - 0.6)$ $x = (32, 12)$	$p = (0.45 - 0.55)$ $x = (-30, -56)$

Notes. For compound problems, the probability ranges indicate which probabilities were equally likely to get drawn.

7. Distance to certainty: make probabilities closer to certainty, holding EV constant
8. Compound: turn base lottery into compound lottery by letting payout probability be drawn from a uniform distribution over a known support

In our experiments, each of the lotteries in Table 15 was combined with safe payments of: (i) $EV - 4$, (ii) $EV - 2$, (iii) $EV + 2$ and (iv) $EV + 4$, where EV refers to the expected value of the lottery. This gives rise to a total of 96 unique choice problems.

Problems designed for lottery dissimilarity. We designed 24 problems that are partitioned into three sets. Table 16 summarizes the problems. A problem consists of a combination of lottery A and a lottery B. Within each set, we designed a lottery B that is “similar” to A in the sense that the distance between the two CDFs is relatively small, and a lottery B that is dissimilar. In constructing similar and dissimilar lotteries, we paid special attention to ensuring that (i) similar and dissimilar lotteries have almost exactly the same expected values; (ii) no dominance relationship is present; and (iii) if anything, the variance of the similar lottery is higher than that of the corresponding dissimilar lottery. The reason is that in our quasi-randomly generated lottery set, the dissimilarity between two lotteries is highly correlated with the variance of the two lotteries. Thus, in devising targeted problems, we sought to investigate whether lottery similarity also matters independently of variance (or even if variance predicts a result in the other direction).

Table 16: Targeted EV problems in *EV Tasks* (2/2)

Lottery	Set 1	Set 2	Set 3
A	$p = (0.5, 0.5)$ $x = (37, 57)$	$p = (0.5, 0.5)$ $x = (75, 35)$	$p = (0.7, 0.3)$ $x = (36, 66)$
Similar B ($EV = EV(A) + 6$)	$p = (0.5, 0.5)$ $x = (36, 70)$	$p = (0.55, 0.45)$ $x = (83, 34)$	$p = (0.65, 0.35)$ $x = (35, 81)$
Similar B ($EV = EV(A) + 3$)	$p = (0.5, 0.5)$ $x = (36, 64)$	$p = (0.55, 0.45)$ $x = (78, 34)$	$p = (0.65, 0.35)$ $x = (35, 72)$
Similar B ($EV = EV(A) - 3$)	$p = (0.5, 0.5)$ $x = (30, 58)$	$p = (0.55, 0.45)$ $x = (65, 36)$	$p = (0.65, 0.35)$ $x = (37, 51)$
Similar B ($EV = EV(A) - 6$)	$p = (0.5, 0.5)$ $x = (24, 58)$	$p = (0.55, 0.45)$ $x = (60, 36)$	$p = (0.65, 0.35)$ $x = (37, 43)$
Dissimilar B ($EV = EV(A) + 6$)	$p = (0.1, 0.9)$ $x = (17, 57)$	$p = (0.8, 0.2)$ $x = (72, 17)$	$p = (0.9, 0.1)$ $x = (45, 105)$
Dissimilar B ($EV = EV(A) + 3$)	$p = (0.1, 0.9)$ $x = (14, 54)$	$p = (0.8, 0.2)$ $x = (69, 14)$	$p = (0.9, 0.1)$ $x = (42, 102)$
Dissimilar B ($EV = EV(A) - 3$)	$p = (0.1, 0.9)$ $x = (8, 48)$	$p = (0.8, 0.2)$ $x = (63, 8)$	$p = (0.9, 0.1)$ $x = (36, 96)$
Dissimilar B ($EV = EV(A) - 6$)	$p = (0.1, 0.9)$ $x = (5, 45)$	$p = (0.8, 0.2)$ $x = (60, 5)$	$p = (0.9, 0.1)$ $x = (33, 93)$

Notes. Each column shows eight problems, all of which involve A and one of the B's.

K.2 Results for Targeted EV Problems

Table 17 summarizes the results. We present OLS regressions in which each observation is a subject-decision. The dependent variable in columns (1)–(2) equals 1 if the subject selected the incorrect lottery. In columns (3)–(4), it is CU. All regressions include problem set fixed effects.

In columns (1) and (3), we see that the following lottery features significantly impact error rates and CU: higher variability (variance or range); mixed payout profiles (in particular if they are associated with a higher range); and compound lotteries. A higher number of states also leads to more errors and CU, though this relationship is statistically significant only for CU.

Columns (2) and (4) document that the dissimilarity of the lotteries' CDFs likewise exerts a strong effect on errors and CU. Notably, the variance explained in column (2) is relatively large, larger than that explained by all features in column (1). This is consistent with our result – reported in the main text – that in our full dataset, including the randomly-generated problems, excess dissimilarity is the strongest predictor of error rates.

These results are broadly consistent with those from the full set of (randomly-generated) problems reported in Figure 1. The main difference is that in the targeted problems we

find little indication that the magnitude of payouts itself affects errors.

Table 17: Results for targeted *EV tasks*

	Dependent variable:			
	1 if error		Cognitive uncertainty	
	(1)	(2)	(3)	(4)
Higher variance / payout range	0.095*** (0.02)		0.028*** (0.01)	
Higher average abs. payout	0.033 (0.02)		-0.0076 (0.01)	
Mixed lottery (constant range)	0.033 (0.04)		0.021* (0.01)	
Mixed lottery (larger range)	0.18*** (0.02)		0.038*** (0.01)	
Higher number of states	0.069 (0.04)		0.034*** (0.01)	
Lower distance to certainty	-0.0088 (0.03)		-0.024* (0.01)	
Compound lottery	0.11*** (0.03)		0.074*** (0.01)	
Higher distance between CDFs		0.18*** (0.04)		0.020*** (0.01)
Controls for EV diff.	Yes	Yes	Yes	Yes
Problem set FE	Yes	Yes	Yes	Yes
Observations	8087	2003	8087	2003
R ²	0.02	0.06	0.03	0.02

Notes. OLS regressions, standard errors (clustered at subject level) in parentheses. An observation is a decision. Each independent variable is a binary dummy for a problem type, as listed in Appendix Tables 15 and 16. The omitted category comprises the base problems. In columns (1) and (3), the decision always involves a lottery and a safe payment; in columns (2) and (4) it always involves two lotteries. Controls for EV difference include linear, second- and third-order terms. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

L Experimental Instructions and Comprehension Checks

L.1 Experiment EV Tasks

Instructions 1/3

Please read these instructions carefully. There will be comprehension checks. If you fail these checks, you will be excluded from the study and you will not receive the completion payment.

In this study, you will make multiple decisions.

Your payment will consist of two components:

- Completion fee:

If you pass all our **comprehension checks and complete the study**, you will receive a **completion fee of \$6**.

- Additional bonus:

On each of 50 decision screens, you will make two decisions. One of the decision screens will be selected at random by the computer, with equal probability. The computer will then flip a coin to determine which of your two decisions on that screen will determine your **bonus**. The **maximal bonus** you can earn in this study is **\$10**. Your bonus will actually be paid out to you with probability 1/2.

Instructions 2/3

Guessing task: which lottery has the highest average payout?

On each decision screen, you will be presented with multiple lotteries that pay out different amounts of money with different probabilities.

Your task is NOT to indicate which lottery you would personally prefer to receive.

Instead, you will be **asked to guess which lottery has the highest average payout**, in the following sense:

- **Each lottery is run many, many times** (100,000 times). In each run, we **record the payout** of the lottery. In the end, **for each lottery**, we **compute the average payout** across all runs.
 - For some lotteries, you will not know the precise probabilities. Instead, we will tell you how the probabilities are determined.
- On each decision screen, you will make two decisions that reflect which lottery you think has the highest average payout.

Step 1: Guess the lottery with the highest average payout

- We will ask you to indicate which lottery you think produces the highest average payout. You need to **select exactly one lottery**.
- This question has a mathematically correct solution.
- If this decision is randomly chosen for payment, you will receive \$10 if your answer is correct, and nothing otherwise.

Step 2: Indicate your certainty about your guess

- We will ask you to **allocate 100 "certainty points"** between the **different lotteries** to indicate how likely you think it is that each lottery has the highest average payout.
 - If you are uncertain, you should allocate some points to each lottery you think could have the highest average payout.
 - For example, if you think it is 80% likely that lottery A has the highest average payout, you should allocate 80 certainty points to Lottery A.
 - If you are certain which lottery has the highest average payout, you should allocate all 100 points to that lottery.
- If this decision is randomly chosen for payment, you will receive 10 cents for each certainty point that you allocate to the lottery that actually delivers the highest average payout. Certainty points allocated to incorrect lotteries earn nothing.

Example screen:

Which lottery has the highest average payout if the computer runs it many, many times?
Please select one.

Lottery A Lottery B

Prob. 50%: Get \$30
Prob. 50%: Get \$0

Prob. 100%: Get \$9

How certain are you that each lottery has the highest average payout?
Please allocate 100 certainty points.

[] points [] points

100 points left to allocate.

In this example, if Lottery A was run many, many times, it would have a higher average payout (\$15) than lottery B if it was run many, many times (\$9).

Here is how your bonus would be determined in this example:

- If Step 1 was selected for payment, you'd get a \$10 bonus if you indicated Lottery A, and nothing otherwise.
- If Step 2 was selected for payment, you'd get 10 cents for each certainty point that you allocated to Lottery A. For example, if you had allocated 80 points to A and 20 points to B, your bonus would be \$8.

Instructions 3/3

In total, you will complete 50 decision tasks. You may take as much time for each task as you'd like, though remember that the study was advertised for 35 minutes and you will only be paid on that basis. If you find that you don't have much time, you may look at the lotteries and make an informed guess about which one has the highest average payout. But again, it is up to you how you work on the tasks.

Once you click the next button, the comprehension check questions will start!

Comprehension check

To verify your understanding of the instructions, please answer the comprehension questions below. If you get one or more of them wrong twice in a row, you will not be allowed to participate in the study and earn a completion payment. In each question, exactly one response option is correct.

You can review the instructions [here](#).

1. Which of the following statements is correct?

One of my tasks is to indicate which lottery I would personally prefer to receive.

One of my tasks is to guess which lottery has the highest average payout if the computer runs it many, many times.

2. On each decision screen, you are asked to indicate which lottery has the highest average payout. How is your potential bonus for this part determined?

My bonus will be \$10 if I indicated the lottery with the highest average payout and \$0 otherwise.

My bonus depends on the outcome of the lottery that I selected. Thus, if the lottery has negative values, it's possible for me to make a loss.

3. Which of the following statements is correct?

On each decision screen I will be asked to allocate 100 certainty points to indicate how certain I am about which lottery I would personally prefer to receive.

On each decision screen I will be asked to allocate 100 certainty points to indicate how likely I think it is that a lottery has the highest average payout.

4. Please select the statement that is true about the certainty points.

If I think it is 60% likely that Lottery A has the highest average payout, I should allocate 60 certainty points to A.

I should always allocate all certainty points to one lottery, even if I'm not certain which one has the highest average payout.

I should always allocate some certainty points to each lottery, even if I'm certain which one has the highest average payout.

If I think it is 60% likely that Lottery A has the highest average payout, I should allocate 100 certainty points to A.

L.2 Experiment Choice Tasks

Instructions 1/2

Please read these instructions carefully. There will be comprehension checks. If you fail these checks, you will be excluded from the study and you will not receive the completion payment.

In this study, you will make multiple decisions.

Your payment will consist of two components:

- Completion fee:

If you pass all our **comprehension checks and complete the study**, you will receive a **completion fee of \$3.50**.

- Additional bonus:

On each of 50 decision screens, you will make a decision. One of the decision screens will be selected at random by the computer, with equal probability, and will determine your **bonus**. The **maximal bonus** you can earn in this study is **\$320**. Your bonus will actually be paid out to you with probability 1/5.

Instructions 2/2

Choice task: which lottery would you like to receive?

On each decision screen, you will be presented with two lotteries that pay out different amounts of money with different probabilities. The computer actually plays out these lotteries according to the probabilities we specify, and pays you accordingly.

On each decision screen, you will be asked to indicate which lottery you prefer to receive.

Step 1: Choose the lottery you prefer

- We will ask you to indicate which lottery you prefer to receive.
- If this decision is randomly chosen for payment, the computer will play out the selected lottery. The outcome of this lottery will determine your potential bonus.

Step 2: Indicate your certainty about your choice

- You might feel uncertain about which lottery you actually prefer. Therefore, we will ask you to indicate how certain you are (in percent) that you actually prefer the lottery that you chose.
 - For example, if you think it is 70% likely that you actually prefer the lottery that you chose, you should set the slider to 70%.
 - If you are certain that you prefer the lottery you chose, you should set the slider to 100%.

For each choice problem in which you can potentially incur a loss, you will receive a budget. This budget will always be given by the largest amount of money you can lose in that choice problem. The budget cannot be transferred across decisions. If a decision of yours is selected to determine your bonus and it happens to be a loss, then this loss will be subtracted from the budget in that particular task, thus giving your final bonus. If you don't incur a loss, the budget will simply be added to your payout from the lottery.

For some lotteries, you will not know the precise probabilities. Instead, we will tell you how the probabilities are determined.

Example screen:

The screenshot shows a user interface for a choice task. At the top, a question reads "Which lottery do you choose? Please select one." Below this, there are two lottery options: "Lottery A" and "Lottery B". Lottery A has a probability of 50% to get \$30 and 50% to get \$0. Lottery B has a probability of 100% to get \$9. At the bottom, a question asks "How certain are you that you actually prefer the lottery you chose above?" with a horizontal slider scale from 0% to 100%. The slider is positioned at 50%. A note below the slider says "I am PLEASE CLICK SLIDER certain that I actually prefer the lottery I chose above."

Here is how your bonus would be determined in this example:

- If you selected Lottery A, the computer would actually play out Lottery A, so your bonus would be \$30 with 50% probability and nothing with 50% probability.
- If you selected Lottery B, your bonus would be \$9.

After your bonus is determined, the computer will randomly determine whether or not it will be paid out to you. You will actually receive your bonus 1/5 of the time.

Once you click the next button, the comprehension check questions will start!

Comprehension check

To verify your understanding of the instructions, please answer the comprehension questions below. If you get one or more of them wrong twice in a row, you will not be allowed to participate in the study and earn a completion payment. In each question, exactly one response option is correct.

You can review the instructions [here](#).

1. How is your bonus determined?

I will make multiple decisions, and every one of them will get paid. Thus, I can strategize across decisions.

I will make multiple decisions. The computer will randomly select one of them, and my potential bonus will depend on my decision in this one question. Thus, there is no point for me in strategizing across decisions.

2. Suppose that you picked Lottery A in one of the tasks.

Lottery A

Probability 60%: Get \$20
Probability 40%: Get \$3

Which of the following statements is correct?

I know that my payoff from this lottery will be either a profit of \$20 OR a profit of \$3, but not both.

I know that my payoff from this lottery will be a profit of \$20 AND a profit of \$3.

3. Which of the following statements is correct?

On each decision screen I will be asked to indicate how certain I am that my bonus will be determined by each lottery.

On each decision screen I will be asked to indicate how certain I am that I actually prefer the lottery that I chose.

4. Please select the statement that is true about the certainty slider.

If I think it is 60% likely that I actually prefer the lottery that I chose, I should set the certainty slider to 60%.

If I think it is 60% likely that I actually prefer the lottery that I chose, I should nonetheless set the certainty slider to 100% because I chose a specific lottery.

L.3 Experiment Prior Manipulation

Instructions 1/3

Please read these instructions carefully. There will be comprehension checks. If you fail these checks, you will be excluded from the study and you will not receive the completion payment.

In this study, you will make multiple decisions.

Your payment will consist of two components:

- Completion fee:

If you pass all our **comprehension checks and complete the study**, you will receive a **completion fee of \$3.50**.

- Additional bonus:

On each of 50 decision screens, you will make a decision. One of the decision screens will be selected at random by the computer, with equal probability, and will determine your **bonus**. The **maximal bonus** you can earn in this study is **\$354**. Your bonus will actually be paid out to you with probability 1/5.

Instructions 2/3

Choice task: which lottery would you like to receive?

On each decision screen, you will be presented with two lotteries that pay out different amounts of money with different probabilities. The computer actually plays out these lotteries according to the probabilities we specify, and pays you accordingly.

On each decision screen, you will be asked to indicate which lottery you prefer to receive.

Step 1: Choose the lottery you prefer

- We will ask you to indicate which lottery you prefer to receive.
- If this decision is randomly chosen for payment, the computer will play out the selected lottery. The outcome of this lottery will determine your potential bonus.

Step 2: Indicate your certainty about your choice

- You might feel uncertain about which lottery you actually prefer. Therefore, we will ask you to indicate how certain you are (in percent) that you actually prefer the lottery that you chose.
 - For example, if you think it is 70% likely that you actually prefer the lottery that you chose, you should set the slider to 70%.
 - If you are certain that you prefer the lottery you chose, you should set the slider to 100%.

For each choice problem in which you can potentially incur a loss, you will receive a budget. This budget will always be given by the largest amount of money you can lose in that choice problem. The budget cannot be transferred across decisions. If a decision of yours is selected to determine your bonus and it happens to be a loss, then this loss will be subtracted from the budget in that particular task, thus giving your final bonus. If you don't incur a loss, the budget will simply be added to your payout from the lottery.

For some lotteries, you will not know the precise probabilities. Instead, we will tell you how the probabilities are determined.

Example screen:

The screenshot shows a decision screen with the following content:

Which lottery do you choose?
Please select one.

Lottery A Lottery B

Probability 50%: Get \$30
Probability 50%: Get \$0

Probability 100%: Get \$9

How certain are you that you actually prefer the lottery you chose above?

Fully certain I prefer the lottery I didn't choose Fully certain I prefer the lottery I chose

A horizontal slider with tick marks at 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100%.

I am PLEASE CLICK SLIDER certain that I actually prefer the lottery I chose above.

Here is how your bonus would be determined in this example:

- If you selected Lottery A, the computer would actually play out Lottery A, so your bonus would be \$30 with 50% probability and nothing with 50% probability.
- If you selected Lottery B, your bonus would be \$9.

After your bonus is determined, the computer will randomly determine whether or not it will be paid out to you. You will actually receive your bonus 1/5 of the time.

Instructions 3/3

Additional information about the problems

Before the study, we showed a few economists the problems you will encounter and polled them about which option they would choose.

The polled economists preferred Lottery A in 80% of the problems (and Lottery B in 20% of the problems). In other words, in 80% of the problems they picked the lottery on the left-hand side and in 20% of the problems the lottery on the right-hand side.

It is entirely up to you how you make your choices and whether or how you use the information provided above.

Once you click the next button, the comprehension check questions will start!

Comprehension check

To verify your understanding of the instructions, please answer the comprehension questions below. If you get one or more of them wrong twice in a row, you will not be allowed to participate in the study and earn a completion payment. In each question, exactly one response option is correct.

You can review the instructions [here](#).

1. How is your bonus determined?

I will make multiple decisions, and every one of them will get paid. Thus, I can strategize across decisions.

I will make multiple decisions. The computer will randomly select one of them, and my potential bonus will depend on my decision in this one question. Thus, there is no point for me in strategizing across decisions.

2. Suppose that you picked Lottery A in one of the tasks.

Lottery A

Prob. 60%: Get \$20

Prob. 40%: Get \$3

Which of the following statements is correct?

I know that my payoff from this lottery will be either a profit of \$20 OR a profit of \$3, but not both.

I know that my payoff from this lottery will be a profit of \$20 AND a profit of \$3.

3. Which of the following statements is correct?

On each decision screen I will be asked to indicate how certain I am that I actually prefer the lottery that I chose.

On each decision screen I will be asked to indicate how certain I am that my bonus will be determined by each lottery.

4. Please select the statement that is true about the certainty slider.

If I think it is 60% likely that I actually prefer the lottery that I chose, I should set the certainty slider to 60%.

If I think it is 60% likely that I actually prefer the lottery that I chose, I should nonetheless set the certainty slider to 100% because I chose a specific lottery.

5. Which statement is true?

A group of economists were polled on 80% of Lottery A (on the left side of the screen) and they generally liked them.

In the problems I will encounter, a group of polled economists selected Lottery A (on the left side of the screen) 80% of the time.

M Pre-Registration on aspredicted.org

'Quantifying lottery choice complexity'
(AsPredicted #130662)

Created: 04/30/2023 02:00 PM (PT)

Author(s)

Benjamin Enke (Harvard University) - enke@fas.harvard.edu
Cassidy Shubatt (Harvard University) - cshubatt@fas.harvard.edu

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

- 1) Does problem complexity predict compression of choice rates between two lotteries to 50%?
- 2) Does problem complexity predict the magnitude of self-reported cognitive uncertainty?
- 3) Does problem complexity predict the magnitude of the treatment effect when subjects' prior beliefs about which of two lotteries is better are manipulated?

3) Describe the key dependent variable(s) specifying how they will be measured.

For 1 and 3: Choice rate between two lotteries

For 2: Self-reported uncertainty about decision

4) How many and which conditions will participants be assigned to?

3 Treatments:

- Baseline treatment to test hypotheses 1 and 2. Subjects simply choose between two lotteries
- "Low prior" and "high prior" treatments to test hypothesis 3. In these treatments, subjects are informed ahead of time that in 20% (80%) of the choice problems they will encounter, economists chose Option A

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

- 1) Regress choice rate for lottery A on difference in expected value, interacted with problem complexity index
- 2) Regress cognitive uncertainty on problem complexity index
- 3) Regress choice rate for lottery A on treatment dummy, interacted with problem complexity index

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

n/a

7) How many observations will be collected or what will determine sample size?

No need to justify decision, but be precise about exactly how the number will be determined.

250 subjects in each treatment

8) Anything else you would like to pre-register?

(e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

Nothing else to pre-register.