Quantifying Lottery Choice Complexity*

Benjamin Enke Cassidy Shubatt

August 13, 2024

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. The most important complexity feature is the state-by-state dissimilarity of the lotteries in the set ("tradeoff complexity"). Using our complexity indices, we study behavioral responses to complexity out-of-sample across one million decisions in 11,000 unique binary choice problems. Complexity predicts strong attenuation of decisions to problem fundamentals. This can generate systematic biases in revealed preference measures such as spurious risk aversion. These effects are very large, to the degree that complexity explains a larger fraction of estimated choice errors than proximity to indifference. Accounting for complexity-driven attenuation in structural estimations improves model fit substantially. Complexity aversion explains a smaller fraction of the data.

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

*Sebastian Redl and Anna Valyogos provided outstanding research assistance. We thank Leeat Yariv, three very constructive refrees, Cary Frydman, Alex Imas, Ryan Oprea, Peter Robertson, Jesse Shapiro, Jeffrey Yang, Florian Zimmermann and many seminar and conference audiences 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 in both the lab and the field emphasizes that decision problems involving risk are often complex, meaning that they require a non-trivial degree of cognitive information processing to aggregate problem fundamentals into an expected-utility-maximizing decision. Understanding complexity is believed to be important because it is intuitively pervasive in real life, and may thus be a primitive that drives choice anomalies across multiple domains. Yet this also begs the question: which risky choice problems actually are complex, and how does complexity affect decisions? The typical approach in the literature is to proceed on a heuristic case-by-case basis: the researcher intuits a specific complexity feature and investigates how it 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.

In this paper, we make progress by developing an empirical mapping from choice set features to indices of choice complexity, which can be computed for any standard dataset. We use these indices to study behavioral responses to complexity and document that complexity-driven behavioral attenuation – an insensitivity of decisions to fundamentals – is quantitatively even more important in explaining choice than popular behavioral models such as prospect theory.

Development of indices of lottery choice complexity. We understand the complexity of a choice problem as the amount of information processing required to identify one's preferred choice option. This informal definition is broader than notions of complexity that exclusively focus on the "cardinality" or "dimensionality" of a problem.

Complexity is latent and unobserved. We, hence, develop a revealed measure of complexity that is based on errors (mistakes). In choice data, mistakes are generally unobserved: the researcher does not know the decision-maker's utility function and, as a result, cannot distinguish errors from preferences. To circumvent this problem, we propose to quantify the objective 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.

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

Quantifying complexity based on an EV task has two advantages. First, relative to revealed-preferences measures such as random choice, it does not impose a specific behavioral model of the nature of people's mistakes. For instance, the EV task captures mistakes that result from both noise and stable heuristics. Second, relative to subjective difficulty measures, it does not require awareness of mistakes and is incentivized.

To train our index, we implement experimental problems in which subjects' payout only depends on the EV of the lottery they select. 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 2,220 quasi-randomly generated problems.

In selecting the features that enter our objective complexity index (*OCI*), we balance the tradeoff between completeness and interpretability that is inherent to any predictive index. We implement exploratory LASSO regressions to identify the most predictive features, and then construct a handcrafted index using a sparse set of easily interpretable features that closely approximates the performance of machine learning algorithms.

OCI comprises two classes of features: (i) the proximity of the EVs of the lotteries in the set; and (ii) features that capture the difficulty of aggregating (and trading off) the constituent components of the problem. In this latter class, 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 the cumulative distribution functions of two lotteries are dissimilar from each other above and beyond their difference in expected value. This measure can intuitively be understood as capturing the strength of tradeoffs across different payout states ("tradeoff complexity") – problems are easier when they involve less pronounced tradeoffs across states. At the extreme, lotteries with a first-order stochastic dominance relationship have excess dissimilarity of zero because they do not involve any tradeoffs. Excess dissimilarity thus fundamentally captures a *comparative* evaluation process.

In binary choice sets that comprise a safe option, excess dissimilarity essentially captures the variance of the risky option. This implies that – if one option is safe – variance produces complexity.

In addition to the *objective* complexity of choice problems, we also quantify *subjective* complexity. To do this, we elicit subjects' cognitive uncertainty (*CU*) in the EV task and use this data to develop a subjective problem complexity index (*SCI*). In principle, *OCI* and *SCI* capture distinct concepts, but in practice they are very similar.

Evidence for construct validity. To provide evidence for our identifying assumption that there is a tight link between the EV task and choice problems, we implement standard binary lottery choice experiments and elicit subjects' *CU* about their choices (Enke and Graeber, 2023). Both *OCI* and *SCI* are strongly predictive of variation in *CU* across choice problems, which we view as encouraging evidence for the validity of our indices.

These correlations hold in both across- and within-subjects designs. Moreover, the effect of individual choice set features on *CU* is very similar in lottery choice and the EV task, which again suggests a tight link between the cognitive difficulty of the two problem types.

Behavioral responses I: Choice set complexity. With our complexity indices in hand, we study behavioral responses to complexity out-of-sample in traditional binary lottery choice problems. We both collect our own dataset 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.

We find that the most important complexity response by far is *behavioral attenuation*: an insensitivity of choice to problem fundamentals. Unlike attenuation bias in econometrics, attenuation in choice is not mechanically driven by mismeasured variables. Rather, it reflects noisy or heuristic decision-making processes. This attenuation effect is large: the elasticity of choice to the difference in EVs of the two lotteries (or the difference in estimated prospect-theory values) decreases by about 75% going from low to high complexity.

Behavioral attenuation could reflect either that complexity produces choice noise or that it leads to an increased reliance on stable heuristics. To investigate this, we link problem complexity to rich data on the frequency of within-subject choice inconsistencies in repeated elicitations of the same problem. We find that excess dissimilarity is strongly correlated with this proxy for choice noise ($r \approx 0.57$). In contrast, the number of distinct payout states is uncorrelated with across-trial variability (as in Arrieta and Nielsen, 2023). This tentatively suggests that different types of complexity trigger different simplification strategies.

We benchmark the magnitude of complexity effects against the proximity to indifference of the lotteries in the set. We find that *OCI* explains an order of a magnitude more of the across-problem variation in proxies for choice errors (such as choice inconsistency) than the estimated proximity to indifference in a prospect theory model.

Behavioral responses II: Lottery complexity. How do people respond to the complexity of individual lotteries? Ex ante, there are two plausible possibilities: stronger attenuation and systematically disliking complex options (complexity aversion). We find evidence for the importance of both. First, like previous work, we find pronounced complexity aversion to compound lotteries (e.g., Halevy, 2007; Gillen et al., 2019). The number of payout states, on the other hand, is largely uncorrelated with choice, reminiscent of the mixed evidence in the literature summarized by Wakker (2022).

Second, we find pronounced evidence for complexity-driven behavioral attenuation with respect to lottery variance. We document how such attenuation can produce systematic biases in revealed preference measures, such as spurious small-stakes risk aversion or risk love. Intuitively, because higher variance makes choice noisier, it pushes choice rates towards 50%, which can generate either spurious risk aversion (if the lottery is very attractive) or spurious risk love (if the lottery is unattractive). As a result, as in other recent work, complexity-driven attenuation (or noise) does not "cancel out" but produces systematic bias (e.g., Andersson et al., 2016; Gillen et al., 2019; McGranaghan et al., 2022). These effects of complexity are so strong that they can entirely override any true risk aversion that likely exists.

Structural estimations. To further assess the quantitative importance of complexity effects, we conduct structural estimations that allow for complexity aversion and / or complexity-dependent attenuation. To study complexity-driven attenuation, we allow the responsiveness 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 19%. In our dataset, this increase is even slightly larger than the combined increase resulting from all of prospect theory (reference dependence, value function curvature, loss aversion and probability weighting). Intuitively, the reason why models that ignore complexity exhibit considerably lower performance is that they dramatically underpredict the probability that people choose the (estimated) higher value option when complexity is low, but strongly overpredict it when complexity is high.

On the other hand, allowing for complexity aversion increases model fit by 3%.

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 problems. These indices are transparent and defined on objective choice set features, making them amenable to be computed in a standardized fashion across datasets. We make available code that automates this process. Part of this contribution is also to identify excess dissimilarity as the most important complexity feature. In Section 7, we discuss potential applications of the complexity indices.

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. We find some evidence for complexity aversion, and strong evidence for complexity-driven behavioral attenuation. Our results are related to other recent contributions that have documented links between an insensitivity of decisions and complexity, noise and / or cognitive uncertainty (Enke and Graeber, 2023; Enke et al., 2023, 2024; Oprea, 2022; Frydman and Jin, 2021, 2023; Vieider, 2021, 2022). Here, we offer an index that predicts the magnitude of attenuation across problems based on objective problem features.

A small number of theoretical contributions have proposed that lottery complexity depends on problem features such as entropy (Verstyuk, 2016; Mononen, 2021; Hu, 2023) or support (Puri, 2022; Gabaix and Graeber, 2023).[1] Our complexity metrics differ substantially from theirs, both because we construct composite measures and because our indices include excess dissimilarity as a main feature. Shubatt and Yang (2024) formalize the effects of dissimilarity on choice.

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, Approach and Identifying Assumption

Consider choice sets comprising two lotteries, A and B, 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. We are interested in quantifying the complexity of choosing between A and B.

Economists are yet to converge on a common definition of the term "complexity." Following recent work (e.g., Oprea, 2020, 2022; Enke and Graeber, 2023; Enke et al., 2023, 2024), we take an information-processing-based perspective. We understand a choice problem as more complex if the cognitive information processing that is required to identify one's preferred option is more difficult.

This definition is broader than those definitions that focus exclusively on the "size" or "cardinality" of the problem, as in approaches that operationalize complexity as a function of lottery support (e.g., Sonsino et al., 2002; Iyengar and Kamenica, 2010; Puri, 2022; Arrieta and Nielsen, 2023). Choice problems can vary in their cognitive difficulty even holding fixed its cardinality. For instance, a choice between two lotteries is often easier when a first-order stochastic dominance relationship exists.

Approach. Rather than exogenously impose what counts as complex, we wish to measure which choice set features contribute to complexity. Because latent complexity (required information processing) is fundamentally unobservable, we define the objective

[1]Appendix Table 4 provides an overview of the experimental literature on complexity in lottery choice. This literature has shown that different lottery features can produce either aversion, or the adoption of stable rules, or higher noisiness. Our comprehensive approach that relies on multiple complexity features allows us to highlight that there is no such thing as a "single complexity response". This said, we do find that behavioral attenuation is quantitatively by far the most important complexity effect.

complexity (*OC*) of a choice problem as the probability that the DM makes a mistake.

$$OC_{A,B} = P(d_c \notin \underset{x \in \{A,B\}}{argmax}\ EU(x)) \tag{1}$$

Because the researcher usually does not know the DM's objective function, this object still cannot be directly observed. To overcome this problem, we consider a second, ancillary decision problem in which the DM is tasked with identifying which lottery has the highest EV. We denote the decision by d_s. In this deterministic task, the objective function is known, thus we can directly observe *deterministic* objective complexity (*DOC*).

$$DOC_{A,B} = P(d_s \notin \underset{x \in \{A,B\}}{argmax}\ EV(x)) \tag{2}$$

This complexity metric has two features. First, it imposes no assumptions on the nature of the mistakes. Research in economics and psychology has emphasized at least two structurally different types of mistakes: those resulting from noisy evaluations and those that result from stable heuristics. As summarized in Appendix Table 4, multiple lottery choice experiments have shown that higher complexity can produce either the adoption of stable rules or higher noise. Because *DOC* is only defined on errors, we capture both of these literatures. For example, a more pronounced tendency to follow a stable rule such as "select lottery with highest minimum payout (max-min)" would generate more errors in the EV task. The fact that the EV task does not impose a specific behavioral model of a complexity response is an attractive feature relative to alternative conceivable approaches. For example, defining complexity based on observed choice inconsistencies would impose the assumption that complexity only triggers noise.

A second feature of the EV task is that, unlike subjective measures of the difficulty of choice (e.g., Enke and Graeber, 2023), it is objective and can be financially incentivized.

Identifying assumption. A main idea behind this paper is that choice errors arise in large part due to the latent difficulty of aggregating probabilities and payouts into a decision. Complexity arising from aggregation similarly arises in lottery choice and EV problems because the aggregated value of a lottery is not transparent to real decision makers (e.g., Oprea, 2022). Our main identifying assumption is that the frequency of errors in the EV task is predictive of errors in the choice task,

$$OC_{A,B} = f(DOC_{A,B}), \tag{3}$$

with $f(\cdot)$ a monotone increasing function. This identification assumption does not require that people are necessarily risk neutral. It also doesn't require that maximizing EV is as difficult as maximizing expected utility. Instead, our assumption is that those choice set features that make it *more* difficult to gauge EV also make it *more* difficult to gauge expected utility, producing a link between error rates in the two different problems.

As a result, our identifying assumption is a statement about how the *relative* difficulty of different choice problems is correlated with the relative difficulty of different EV problems, rather than about how the absolute difficulty of a choice problem compares to the absolute difficulty of the analogous EV problem. In particular, our identifying assumption is consistent with situations in which (i) choosing from the set $C_1 = \{A, B\}$ is hard and (ii) determining the highest EV option in C_1 is easy, as long as, for the set $C_2 = \{D, E\}$, a higher difficulty of choosing in C_2 compard to C_1 goes hand-in-hand with a higher difficulty of identifying the highest EV option in C_2 compared to C_1.[2]

2.2 Empirical Implementation

Because it is impractical for researchers to always implement an analogous EV problem to quantify the complexity of their choice problem of interest, we leverage the idea that error rates in the EV problem can be *predicted* based on choice set features. This is attractive because once complexity 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 choice set features, where the zeroth element is a constant. Denote by $\epsilon_{A,B}$ a mean-zero disturbance term.

Definition 1. *The objective complexity index for choice set {C,D} is given by*

$$OCI_{C,D} := \sum_{i=0}^{N} \hat{\beta}_i \, f_i^{C,D} + \hat{\gamma} \; ln(1 + |EV(C) - EV(D)|) \,, \tag{4}$$

where the vector $\hat{\beta}$ *and* $\hat{\gamma}$ *are estimated in a sample of EV problems using OLS:*

$$DOC_{A,B} = \sum_{i=0}^{N} \beta_i f_i^{A,B} + \gamma \; ln(1 + |EV(A) - EV(B)|) + \epsilon_{A,B} \,. \tag{5}$$

This index has a simple interpretation once it is applied to standard lottery choice problems: it captures the error rate in the analogous EV problem that is predicted by the choice set features. We allow the absolute EV difference to enter non-linearly because in random choice models such as logit, mistake rates are a concave function of the absolute EV difference.

[2]To illustrate, consider choice set C_1, deciding between (A) \$7 for sure and (B) a 50-50 lottery that pays \$20 or \$0. This choice problem is difficult for many people (e.g., Agranov and Ortoleva, 2017) even though the analogous EV problem is arguably easy. Yet now consider choice set C_2, comprising (D) \$15 for sure and (E) a 90-10 lottery that pays \$20 or \$0. As long as C_1 is more difficult than C_2 in terms of both choice and identifying the higher EV option, it is immaterial from our perspective that C_1 appears easy in the EV task and difficult in the choice task. In Section 4, we will see that this is indeed the case.

OCI comprises both features that capture the proximity of the aggregated (expected) values, and features that capture the complexity of aggregating payouts and probabilities conditional on the EV difference. *OCI* does not allow the effect of the features f_i to depend on the EV distance. This is clearly a simplification. We expect our index to perform well in problems for which the DM is neither extremely close nor very far from indifference. Below in Section 6.1, 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.

All of the above concerns the quantification of *objective* choice complexity. Yet in many economically-relevant situations, it is *subjective* complexity that matters for behavior. We define a subjective complexity index ($SCI_{C,D}$) analogously to Definition 1, except that in equation (5) objective mistake rates are replaced by the subjective probability of making a mistake in the EV problem.

3 Experimental Datasets

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

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 personally prefer, we instruct participants to indicate the lottery that has the highest EV. Each subject completed 50 decision problems. 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. Subjects' potential bonus equals $10 if they select the correct lottery, and nothing otherwise. This incentive scheme has two main upsides. First, it makes transparent the objective nature of the task. Second, it holds the incentives constant across problems.

We believe the cognitive difficulty of lottery choice problems mainly comes from the difficulty of aggregation: whether deliberately or not, people ought to somehow combine the probabilities and payouts of different states to reach a decision. An attractive feature

Table 1: Overview of experiments and data sources

Experiment	Description	Problems	Subjects	Decisions
EV Tasks	Indicate lottery with highest EV & *CU* elicitation	2,100 procedurally generated & 120 targeted	1,148	57k
Choice Tasks from *PEA*	Lottery choice problems	10,398 procedurally generated	15,151	973k
Choice Tasks	Lottery choice problems & *CU* elicitation	500 procedurally generated	250	12.5k
Within Subject	EV and lottery choice problems & *CU* elicitation	240 procedurally generated	300	12k

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

of our design is that this aspect of the decision problem is similar between our *EV Task* and real lottery choice problems.[3]

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 verfied subjects' understanding of this using a comprehension check. Second, the question on subjects' decision screen reads: "Which lottery has the highest average payout if the computer runs it many, many times?", rather than "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 still misunderstood our instructions, we would expect them to exhibit risk aversion. We find no evidence for this in our data.

Cognitive uncertainty elicitation. On each decision screen, we additionally 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 express how likely they think it is that each 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. (2023).[4] Appendix Figure 7 shows a screenshot of a decision screen.

[3]At the end of the study, we asked subjects whether they had used a calculator or other help. 23% of subjects indicated they had. 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.

[4]Recent 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 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 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. The distributions of *CU* in these two sub-samples are very similar.

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 designed the experiment *EV Tasks* to comprise a total of 2,220 unique choice sets. A first set of 2,100 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. 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.

The remaining choice problems were 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 G, we report separate analyses that only make use of the smaller, targeted set.

We focus on two-item menus and discuss an extension to larger menus in Section 7. 95% of all tasks involve only two lotteries. In the two-item menus, 30% involve deciding between a two-state lottery and a safe payment. In the remaining 70% of problems, both lotteries are non-degenerate, and the number of states of both lotteries varies between two and seven. We collected data until each problem was completed by at least 20 subjects (median is 22 and average 26).

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 $r = 0.49$ ($p < 0.01$), suggesting subjects' beliefs are reasonably well-calibrated, on average. In 7.9% of two-item menus does a majority select the wrong option, yet this difference is statistically significant in only 1.4% of problems. Appendix Table 8 presents further summary statistics.

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 the 10,398 unique binary choice problems that we use.[5] 15,151 Amazon Mechanical Turk (AMT) workers completed an average of 13 problems

[5] We drop problems that involve ambiguity or that involve two safe payments. We combine identical problems with / without feedback.

five times each, for an average total of 65 decisions per subject. The dataset was designed to span a much larger space of choice problems than previous data-collection exercises, making the data well-suited for our purposes. Appendix Table 8 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) 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 next. Appendix C.3 replicates our main results excluding choice problems from PEA that involve losses.

Experiment Choice Tasks. 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 and deducted from a budget.

In addition to asking subjects to choose between the two lotteries, we also elicited their *CU*, asking how certain they are (in percentage terms) that they selected the option they actually prefer (Enke and Graeber, 2023). Appendix Figure 9 shows a screenshot. Each subject completed 50 choice problems.

Experiment Within Subject. To further verify that the difficulty of choice and of EV problems are correlated, we implemented an additional experiment in which each subject encountered 20 problems in two ways: first as a choice problem and then as EV task, or vice versa. Subjects first completed all 20 choice problems or all 20 EV problems, where the order of choice and EV task was randomized across subjects.[6] We generated 240 distinct choice sets, each of which was completed by 25 subjects, on average. To study robustness to payout procedures, in the EV task part of this experiment, subjects were paid out the EV of the lottery they selected, rather than receiving $10 if they made the correct decision.

3.3 Implementation

Our own experiments were conducted on Prolific. See Appendix H for screenshots of instructions and comprehension check quizzes. In *EV Tasks*, subjects earned a completion fee of $6. In addition, 1 in 2 subjects was randomly selected to be eligible for a bonus of $10 if they made the correct choice on a (uniformly) randomly selected decision.

[6] CU was measured using the following questions. EV task: "How certain are you that you actually selected the lottery with the highest average payout?" Choice task: "How certain are you that you actually prefer the lottery you chose above?"

In *Choice Tasks* and *Within Subject*, participants received a fixed payment of $3.50. In addition, 1 in 5 subjects were randomly selected to be eligible for a bonus wherein we randomly selected one decision to be payout-relevant.

We pre-registered the predictions and sample size for experiments *Choice Tasks* and *Within Subject* on aspredicted.org under #130662 and #173455. 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 to show that a specific feature would matter.

4 Development of the Complexity Indices

4.1 Choice Set Complexity

We train the complexity indices on a randomly selected subset of 75% of all EV problems (train set) and use the remaining 25% as a test set. A main question is which features should be included in the indices. To balance the typical tradeoff between interpretability and completeness, we strike a middle ground and proceed in three steps.

1. *Exploratory LASSO index:* 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 index:* We inspect the LASSO-generated index and approximate it based on only a handful of simple and easily interpretable features.

3. *Assess completeness:* We benchmark the handcrafted index against a machine learning ensemble.

Exploratory LASSO indices. Appendix B.2 provides a list of all 44 choice set features that we consider. When a feature is defined over a single lottery (such as a lottery's variance), we compute the average value in the set.[7] We also consider non-linear transformations of these averages (log and square).[8] We only consider primitive features of the lotteries, rather than also framing effects. We don't allow interactions between features (as we will see below, they do not matter much).[9]

[7]Computing averages appears justified because, in our data, the effect of lottery-specific features (such as variance) on mistake rates generally points in the same direction for both lotteries.

[8]Whenever we say we compute the log of x, we mean that we compute $ln(1+x)$.

[9]We also don't allow features of choice sets encountered in the past. This appears justified because, compared to the experimental efficient coding literature (e.g., Frydman and Jin, 2021, 2023), we (i) implement many fewer trials and (ii) vary many more features, both of which make it less likely that subjects differentially learn higher cognitive precision for some problems than others.

Some of the features 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 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 set the LASSO penalty parameter to the value that minimizes mean squared error in the train set. Appendix Table 6 reports the results of the LASSO regressions. 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 as 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 artifically 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.

At a formal level, Appendix Table 7 reports OLS estimates of eq. (5) using our hand-selected features.[10] Our indices *OCI* and *SCI* correspond to the fitted values of these regressions. To present the results in a more intuitive way, Figure 1 reports correlation coefficients between error rates (or *CU*) and those choice set features that enter our final indices.[11]

Excess dissimilarity: The strength of tradeoffs across states. The standout predictor of both objective and subjective 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. As illustrated in the top left panel of Figure 2, we compute dissimilarity by overlaying the cumulative distribution functions (CDFs) of the two lotteries and calculating the summed (absolute) area between the two. The so-called "Wasserstein 1-distance" between the CDFs of two lotteries is given by $\delta_{A,B} = \int_{\mathbb{R}} |F_A(x) - F_B(x)| dx$. We then define excess dissimilarity as:

$$d_{A,B} = \delta_{A,B} - |EV(A) - EV(B)|. \tag{6}$$

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, Op-

[10] Our procedure implicitly assumes that a single index captures complexity across all different types of menus. In ancillary analyses, we experimented with a finite mixture approach in which different types of menus are assigned different complexity indices. The results show that the resulting "within-type" indices are highly correlated ($\rho \geq 0.95$) with *OCI*.

[11] Appendix Figure 15 reports the analogous results for the *Within Subjects* experiment.

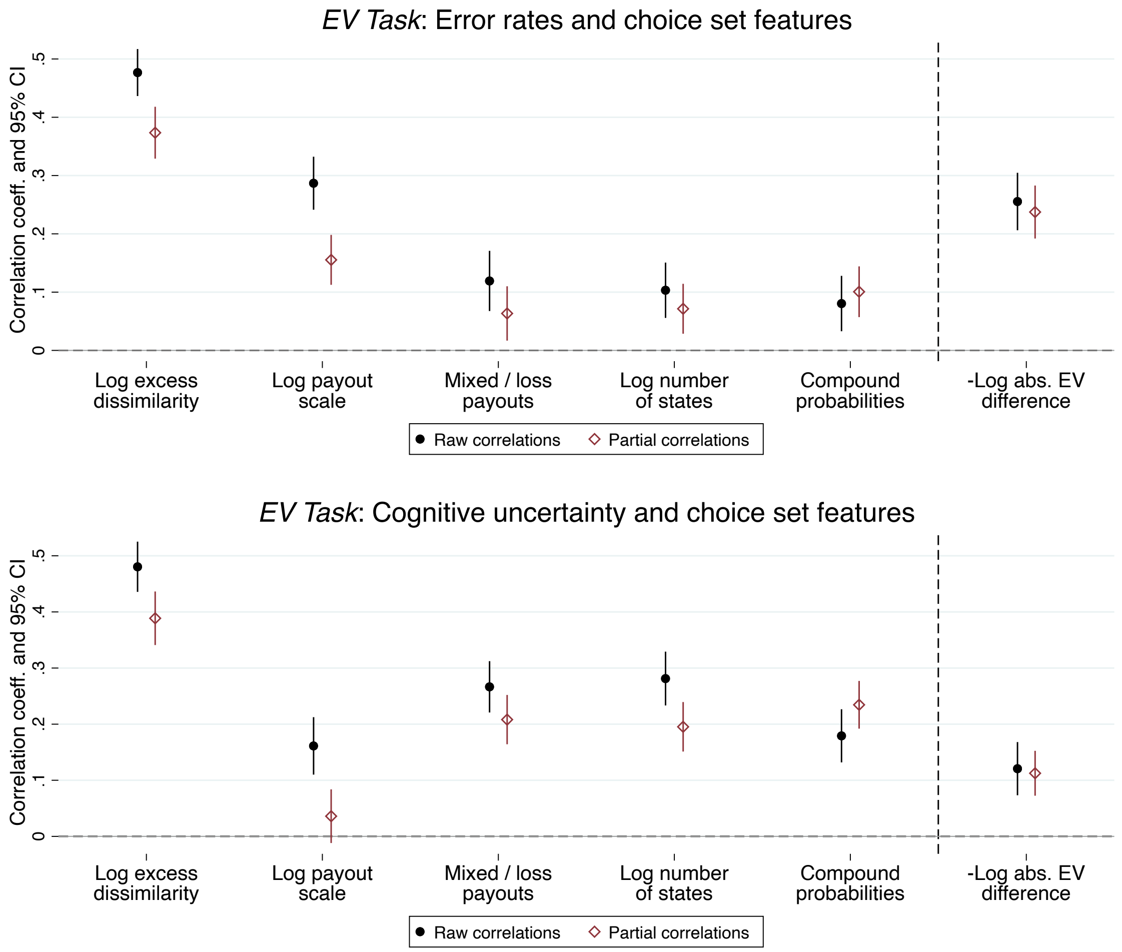


Figure 1: Raw and partial correlation coefficients between task-level error rates / average *CU* and choice set features in the train set in the *EV Tasks* experiment (1,587 unique problems). Whiskers show 95% confidence intervals. Partial correlations are calculated controlling for all of the other features in the figure. Log scale, mixed / loss payouts and log number of states are computed as averages across the lotteries in a set.

tion B': "Get \$70 with probability 21%" has high excess dissimilarity with Option A, even though B and B' have the same EV. Thus, choosing between A and B is predicted to be simpler than choosing between A and B'.

Excess dissimilarity is large when the lotteries have very different advantages and disadvantages. We find it helpful to think of this measure as capturing "tradeoff complexity" – complexity as it arises from the difficulty of aggregating tradeoffs across different payout states. Excess dissimilarity equals zero when there are no tradeoffs across states (i.e., when differences in EV 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 is fundamentally an object that captures the difficulty of *comparative* evaluations: how the two lotteries perform in their worst state, their best state, their "median" state, and so on.[12]

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.

Figure 1 shows the raw correlation between log excess dissimilarity 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 and thus the most important component of our complexity indices.

The top right and bottom panels of Figure 2 illustrate how variation in excess dissimilarity – as induced by variation in payouts and probabilities – predicts mistakes in the EV task. Each plot shows a binned scatter plot that controls for the absolute value difference between the two options in the set (because excess dissimilarity is dissimilarity net of absolute expected values difference).

In the top right panel, we consider the decision between a safe payment and a binary lottery, where (to illuminate the role of payouts) the probability of the upside is restricted to be in [0.4,0.6]. The plot shows that as the spread between the lottery upside and downside increases (which increases the dissimilarity between the lottery and the safe payment), mistake rates strongly increase. This insight will be important below because the spread of payouts (or lottery variance) is essential in the estimation of risk aversion.

In the bottom left panel, we illustrate the role of payout probabilities. Again, we consider the decision between a safe payment and a binary lottery, and vary the probability of the lottery upside. Dissimilarity between a lottery and a safe payment is small when the probabilities are very extreme. Consistent with the idea that low dissimilarity produces fewer mistakes, we see that mistake rates are strongly hump-shaped, with intermediate probabilities around 50% producing the highest error rates.

Finally, to illustrate that it is not generically true that intermediate probabilities contribute to dissimilarity and complexity, we consider the decision between two lotteries, A and B, where the probability of B's upside is in [0.4,0.6]. The bottom right panel shows mistake rates as a function of the probability of A's upside. In this case, dissimilarity is lowest when A's upside probability is intermediate. Indeed, mistake rates are now U-shaped. Viewed in combination, the bottom left and bottom right panels illustrate that complexity is fundamentally based on what a lottery is compared with.

The idea that dissimilarity contributes to choice complexity is prominent in perceptual psychology and has attracted some attention in economics (e.g., Rubinstein, 1988;

[12]We are implicitly putting the two lotteries into a common, perfectly correlated state space. Formally, we think of the state x as a draw from a Uniform distribution on $[0,1]$, and we say the lotteries A and B are "perfectly correlated" in that they return $F_A^{-1}(x)$ and $F_B^{-1}(x)$, respectively.

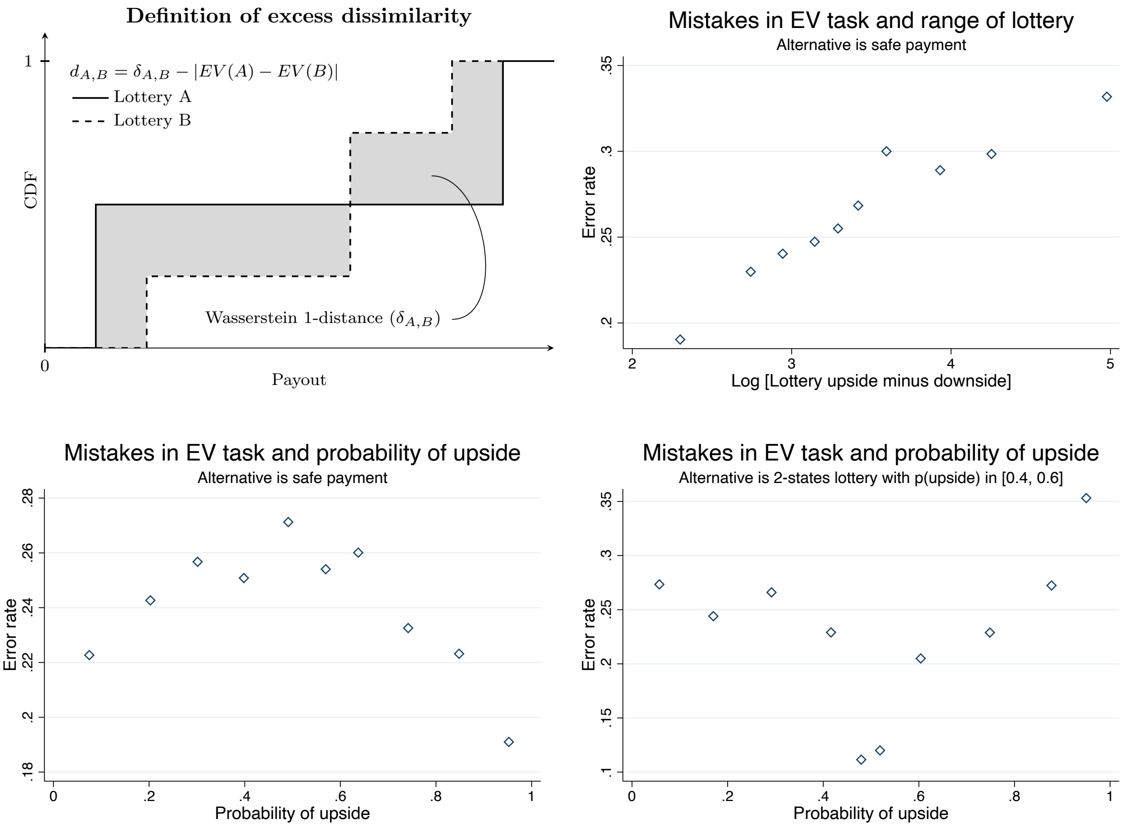


Figure 2: Top left panel: Definition of excess dissimilarity. Top right panel: Binned scatter plot of error rate in EV task as a function of difference between lottery upside and downside, when decision is between a binary lottery (with probability of upside in [0.4,0.6]) and a safe payment. Bottom left panel: Binned scatter plot of error rate in EV task as a function of probability of lottery upside, when decision is between binary lottery and safe payment. Bottom right panel: Binned scatter plot of error rate in EV task as a function of probability of lottery upside, when decision is between binary lottery and another binary lottery with probability of upside in [0.4,0.6]. All binned scatter plots control for the absolute expected values difference between the two options.

Fishburn, 1976, 1978; Natenzon, 2019; He and Natenzon, 2022; Shubatt and Yang, 2024). 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 EV difference, and we sometimes desire to separate effects stemming from aggregation complexity and proximity.

Payout scale. Much research on number perception suggests that people find it harder to process and transform larger numbers (Weber's law). 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 our data, both pure loss gambles and mixed gambles produce significantly higher error rates and *CU*, compared to pure gains menus. To keep our indices sparse, we generate one variable to capture these patterns, which is the fraction of lotteries in the set that includes at least one negative payout.

Support. Figure 1 shows that the average (log) number of states in the choice set is significantly correlated with error rates and *CU*. We include this variable in our complexity indices, but we note that – while a main focus of the literature – it 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 higher *CU*.[13]

Proximity of expected values. Unlike the aforementioned features, the difference in EV 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 work with the natural log in the construction of our handcrafted indices. 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 thoughout this paper: errors (and our complexity indices) largely reflect aggregation complexity rather than proximity.

Composite complexity indices. *OCI* and *SCI* consist of linear combinations of the features listed above, with weights given by the OLS coefficients in Appendix Table 7.[14] In the *EV Tasks* test set data, *OCI* and *SCI* exhibit a raw correlation of $r = 0.87$, ($p < 0.01$). 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. To assess completeness, we follow Fudenberg et al. (2022) and benchmark the out-of-sample performance of our complexity indices against that of a machine learning ensemble, including a black-box convolutional neural net. Appendix B discusses the details. In short, we find that *OCI* explains 95% of the variation relative to a machine learning benchmark, and that excess dissimilarity alone is 65% complete.

[13]For compound probability choice problems, one option involves an unknown probability p, which subjects know is drawn uniformly from some specified range.

[14]We winsorize all complexity indices from below at zero and from above at 0.5 because they capture predicted mistake rates.

4.2 Lottery-Specific Complexity

Our main index quantifies the complexity of a choice problem rather than of a single lottery. We believe this is important because – by the logic of dissimilarity causing complexity – the complexity of an individual lottery will depend on what it is compared against. Still, one useful special case that attracts much attention in theoretical and empirical practice is to quantify the complexity of a single lottery when the alternative is a safe payment. In particular, a natural benchmark is to compute the complexity of a lottery as the difficulty of choosing between the lottery and a safe payment that is just below (or above) its own EV.

We construct indices of objective and subjective lottery complexity (*OLCI* and *SLCI*) in an analogous fashion to *OCI* and *SCI*, except that they are computed based on a two-item 'choice set' that comprises a lottery and its own EV. Thus the indices of lottery complexity comprise as features a lottery's support, scale, negative payouts, compound probabilities, and excess dissimilarity between a lottery and its own EV.

Given that the dissimilarity between a lottery and its own EV basically amounts to a lottery's variance, this means that – when the alternative is a safe payment – variance is a key predictor of a lottery's complexity. Indeed, in our *EV Task*, the correlation between log lottery variance and error rates (when the alternative is a safe payment) is $r = 0.42$ ($p < 0.01$).

4.3 Evidence on Identifying Assumption

Our main identifying assumption is that the difficulty of a choice problem is monotone in the difficulty of the analogous EV problem. We can make progress on assessing the validity of this assumption by assessing the distribution of *CU* in real choice tasks, in particular how it varies with the complexity features derived from the EV task. We do so both within and across subjects.

First, in our *Choice Tasks* data, we correlate average *CU* in a choice problem with *SCI* (which captures predicted *CU* in the EV problem). The correlation is $r = 0.63$ ($p < 0.01$). Similarly, the correlation between average *CU* in choice and *OCI* is $r = 0.47$ ($p < 0.01$). See Appendix Figure 10. This is fundamentally 'between' evidence, both in the sense that (i) no subject completed both choice and EV problems; and (ii) the choice problems in the *Choice Tasks* experiment are distinct from the ones that were used to develop the indices.

Second, in our *Within Subject* experiment, we can study the same question from a 'within' angle, both in the sense that (i) the same subject completes both choice and EV problems; and (ii) these problems are in fact identical, which allows us to directly link *CU* in choice and *CU* in the same EV task. We find that the correlation between *CU* across the two tasks – controlling for subject fixed effects – is $r = 0.30$ ($p < 0.01$). See Appendix Figure 14.[15] These results are encouraging because they strongly suggest that the same features determine how difficult it is to gauge expected utilities on the one hand and expected values on the other hand.

To assess whether it is indeed true that any given complexity feature (rather than the composite index) affects *CU* in choice and in the EV task in similar ways, Appendix Table 11 compares the correlations between the features in our indices with *CU* in choice and *CU* in the EV task. The results show that the correlations between *CU* and features are usually very similar (in particular for excess dissimilarity, support and losses), which further suggests that the sources of complexity are similar in the two types of problems.[16]

5 Behavioral Responses to Complexity

We now deploy the complexity indices to explain 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.

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 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 index of objective complexity, *OCI*. Given the similarity between *OCI* and *SCI*, the results using *SCI* are essentially identical. In a third step, we study separate lottery features.

5.1 Examples

Table 2 presents six example choice problems. Five of them are selected to have similar EV differences but varying *OCI*. All problems are such that Option A (the lottery with the weakly larger number of states) has a lower EV.

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

[15]The correlation between *CU* in choice and a binary indicator for making a mistake in the EV task – again controlling for subject fixed effects – is $r = 0.08$ ($p < 0.01$). That the correlation with objective mistakes is smaller in this within-subjects exercise is unsurprising and mechanical because when we look within a given subject, mistakes are a binary indicator while *CU* is a more continuous measure of difficulty.

[16]The most notable exception are compound probabilities, which are significantly positively correlated with *CU* in the EV task but not in the choice task.

Table 2: Example choice problems

#	Probabilities A	Payouts A	Probabilities B	Payouts B	EV(A)-EV(B)	*OCI*	Frac. chose A
1	0.25, 0.375, 0.375	-21, 0.5, 1.5	1	2	-6.5	0.06	4%
2	0.99, 0.01	18, 33	1	25	-6.9	0.10	9%
3	0.95, 0.05	-2, 28	0.75, 0.25	1, 22	-6.8	0.12	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	0.4, 0.3, 0.15, 0.15	-28, 31, 29, 25	1	7	-0.8	0.42	41%

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 and the alternative is a safe payment. For instance, in problem 2, one can intuitively see that lottery A "is worth approximately $18," which makes B's payout of $25 look transparently superior.

In contrast, in problems 4 and 5, excess dissimilarity is high because the options have different advantages and disadvantages. For example, in problem 5, heuristic pairwise comparisons of the lottery upsides and downsides is difficult.

Finally, problem 6 illustrates a very high complexity problem in which the drivers of complexity are high excess dissimilarity, large support, a presence of losses, and also very similar EVs.

5.2 Choice Complexity, Behavioral Attenuation and Noise

A first potential implication of complexity – highlighted by a recent literature – is a form of attenuation, according to which decisions become less elastic to variation in problem fundamentals. Unlike attenuation bias in econometrics, such 'behavioral attenuation' is not driven by mismeasured variables (after all, we as researcher perfectly observe the gambles participants are exposed to). Rather, it is driven by a noisy or heuristic decision process – by a noisy or heuristic mapping of problem fundamentals into a decision.

The top left panel of Figure 3 shows choice rates for lottery A as a function of the EV difference between A and B, separately for choice problems that are above or below median *OCI*. 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).

The figure shows a binned scatter plot because of the large number of underlying choice problems. These plots are constructed such that each dot represents an equal number of choice problems (about 109 problems per dot), and show the average fraction of subjects choosing lottery A across all of the choice problems in a bin. We see that choice rates in problems that are predicted to be more complex are substantially more compressed towards 50% and, hence, more attenuated.[17]

A potential concern is that these differences between high- and low-complexity problems are confounded by the fact that the appropriate x-axis is not the EV difference but the true expected utility difference. To gauge this, the top right panel shows analogous results by plotting empirical choice rates against the estimated value difference in a cumulative prospect theory (CPT) model.[18] The results are very similar.

The middle panels provide a complementary perspective, by showing choice rates for lottery A as a function of the continuous *OCI* index. In the left panel, the red dots correspond to cases where $EV(A) > EV(B)$ and the blue dots to cases where $EV(A) < EV(B)$. The middle right panel is constructed analogously but splits by CPT value difference. Thus, for an EV (or CPT) maximizer, the choice rates should be 0% and 100%. More interestingly, even standard random choice models would predict that choice rates are constant in *OCI*. 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 EV decrease by about 40 percentage points going from very low to very high complexity. These results hold regardless of whether or not we control for the EV difference (or the estimated CPT value difference). This suggests that the vast majority of the explanatory power of *OCI* for behavioral attenuation is due to the effects of aggregation complexity (chiefly excess dissimilarity) rather than proximity to indifference. We further quantify this point below.

Noise or stable heuristics? While the previous results suggest that complexity predicts attenuation (an insensitivity of decisions to fundamentals), this could occur through at least two different mechanisms. First, higher complexity could make people's evaluations more noisy, producing attenuation akin to the effects in random choice models. Second, attenuation could reflect the use of stable heuristics. For instance, usage of stable rules such as max-min can produce behavior that is only weakly correlated with maximizing EV or CPT and can, hence, produce compressed choice functions such as those in Figure 3.

To study mechanisms, we link *OCI* to a standard measure of random choice: across-

[17]One 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. We ran 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". This treatment has a larger effect on choice for more complex problems. To conserve space, these experiments are summarized in Appendix F of our earlier NBER working paper (#31677).

[18]As discussed in detail in Section 6 and Appendix D, we estimate CPT as a representative agent model with the specifications that – according to a recent meta-analysis (Brown et al., 2024, Table 3) – are most common in the literature: a reference point of zero and CRRA utility.

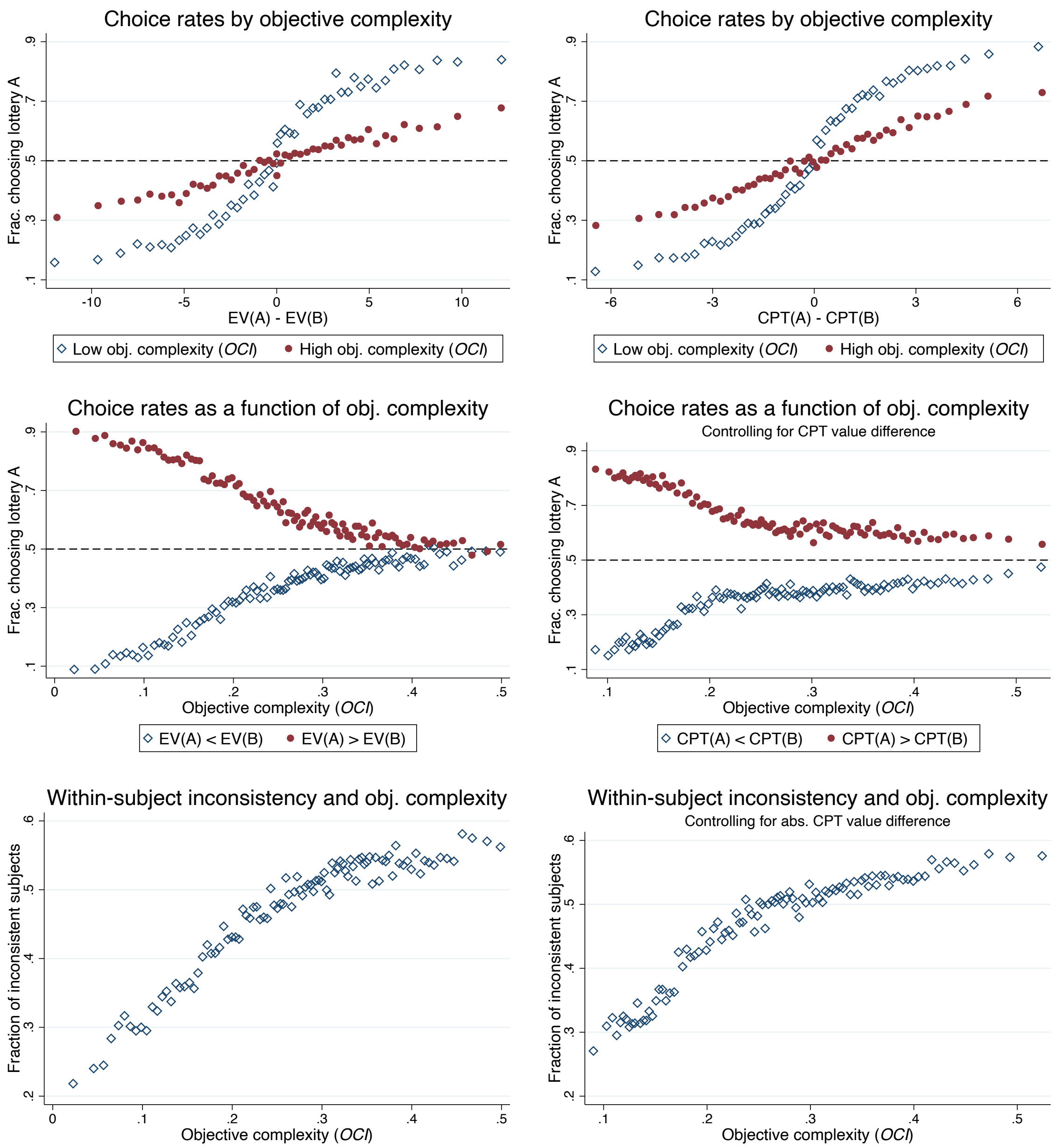


Figure 3: Objective complexity and behavioral attenuation in lottery choice. In the top and middle panels, the y-axis is the fraction of subjects choosing lottery A. The top panels implement a median split by *OCI*, separately within each percentile of the EV difference (or the estimated CPT value difference) between A and B. The middle panels split the choice problems according to whether A or B has a higher EV (or a higher estimated CPT value difference). In the bottom panels, the y-axis shows the fraction of subjects who are inconsistent at least once. The middle panel controls for the estimated CPT value difference and its square, and the bottom right panel for the absolute CPT value difference and its square. All panels show binned scatter plots. Top panels constructed from 10,898 choice problems, middle and bottom panels omit problems with absolute EV difference of less than $0.20, hence constructed from 10,366 choice problems.

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 choice problem did so five times (consecutively). For each choice problem, we compute the fraction of subjects who are inconsistent at least once. In these analyses, we restrict attention

Table 3: Benchmarking *OCI* and proximity to indifference

	Dependent variable:								
	Frac. subjects inconsistent			Deviation rate from CPT prediction			Avg. cognitive uncertainty		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
OCI	0.70***			0.81***			50.0***		
	(0.01)			(0.01)			(3.97)		
Abs. EV diff.		-0.0078***			-0.017***			0.039	
		(0.00)			(0.00)			(0.18)	
Abs. CPT value diff.			-0.017***			-0.045***			0.17
			(0.00)			(0.00)			(0.32)
Constant	0.27***	0.50***	0.50***	0.12***	0.41***	0.44***	2.94***	15.6***	15.4***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.93)	(0.99)	(0.84)
Observations	10398	10398	10398	10898	10898	10898	500	500	500
R^2	0.26	0.02	0.03	0.32	0.11	0.21	0.24	0.00	0.00

Notes. OLS estimates, robust standard errors in parentheses. An observation is a choice problem. The dependent variable in columns (1)–(3) is the fraction of subjects who are inconsistent at least once in the five repetitions of the choice problem. In columns (4)–(6) it is the fraction of decisions that does not equal the prediction of a full CPT model, see Appendix D. In columns (7)–(9) the dependent variable is average self-reported *CU* in the choice experiments (in percent). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

to problems in which the absolute EV difference is at least $0.20 to reduce concerns that "inconsistencies" simply reflect indifference (the results are identical in the full sample).

The bottom panels of Figure 3 show binned scatter plots of problem-level choice inconsistencies against *OCI*, either the raw correlation (left panel) or controlling for the absolute estimated CPT value difference (right panel). Moving from $OCI = 0$ to $OCI = 0.5$ is associated with an increase in the frequency of choice inconsistencies of 35 percentage points. The raw and partial correlations are always around $r \approx 0.49 - 0.53$ ($p < 0.01$).

Benchmarking. A natural question is how quantitatively important complexity is for understanding choice behavior. Here, a natural point of comparison is the proximity to indifference (in the choice task), which is a main driver of choice errors in standard random choice models. We, hence, compare the predictive power of *OCI* with that of the subjective value difference in a CPT model, estimated as described in Appendix D. Because *OCI* includes the proximity of the EVs of the two options, we also benchmark *OCI* against the EV difference.

For our benchmarking analysis, we desire proxies for the 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 CPT model – delivers higher value; and (iii) average self-reported *CU* in a choice problem.

Table 3 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, *OCI* explains a considerably larger fraction of the variation than the EV difference. This again shows that the vast majority of the predictive power of *OCI* reflects the difficulty of aggregating the constituent components of a lottery choice problem, rather than proximity of the expected values.

Second, again for all dependent variables, *OCI* explains a substantially larger fraction of the variation than the estimated CPT value difference. For example, in column (6), the estimated value difference in a CPT model almost by construction explains a sizable share (21%) of the frequency of decisions that do not maximize CPT value. Yet the variance explained by *OCI* is substantially larger (36%, column (4)).

5.3 Separate Lottery Features

Up to this point, we focused on the overall complexity of a choice set. We now disaggregate the previous results in two ways. First, we study the complexity of individual lotteries rather than of a choice set as a whole. This allows us to study not only behavioral attenuation but also potential complexity aversion. Second, we analyze the separate features that jointly make up the complexity index. After all, given that excess dissimilarity drives the vast majority of the variation in *OCI*, it is not at all obvious that the results emphasized earlier – behavioral attenuation and choice noise – apply to every feature that is a part of the index.

To transparently study the complexity of individual lotteries, we restrict attention to decisions between a lottery (option A) and a safe payment (option B). This can be thought of as fixing the complexity of B at zero and varying the complexity of A.

We focus on three lottery features given their disproportionate importance and prominence in the literature: (i) lottery variance because, as discussed in Section 4.2, it is the main driver of excess dissimilarity when the alternative is a safe payment; (ii) lottery support; and (iii) compound probabilities.

Lottery variance: Behavioral attenuation and small-stakes risk aversion. The left panel of Figure 4 shows a binned scatter plot of choice rates for the lottery as a function of the difference between the estimated CPT values of the lottery and the safe payment. To study the role of complexity, we now split by median lottery variance. In this plot, aversion to variance is indicated by a *uniform downward shift* of choice fractions. Thus, if people are risk averse and the degree of attenuation was independent of variance, we should expect choice rates for the high-variance lotteries to be lower than for the low-variance lotteries everywhere. Complexity-driven attenuation, on the other hand, is again indicated by a compression (or "flipping") pattern, according to which observed

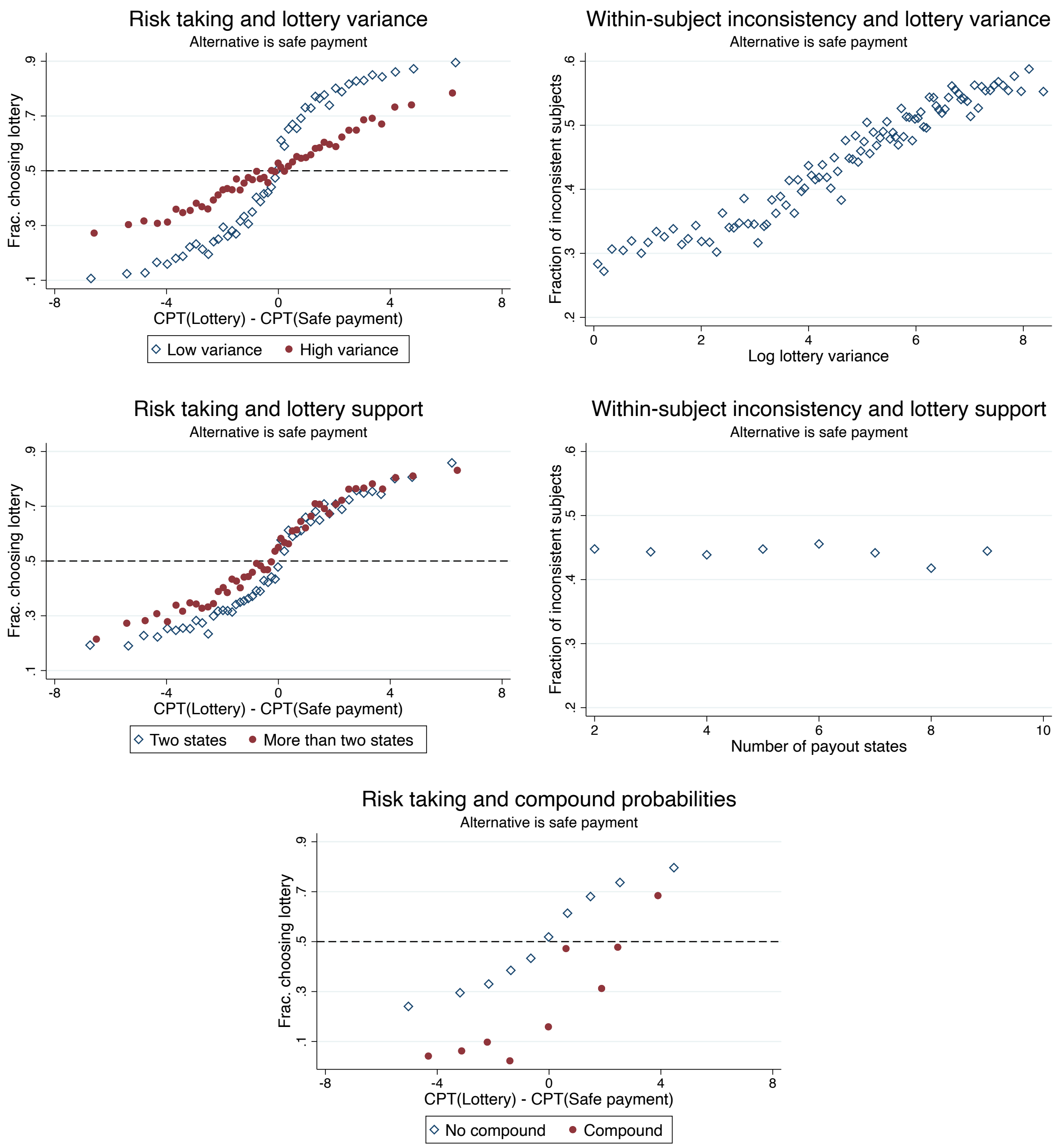


Figure 4: Binned scatter plots of choice rates (left panels) and choice inconsistency (right panels) for problems that involve one non-degenerate lottery and one safe payment. In the left panels, the y-axis is the fraction of subjects choosing lottery A. These panels implement a median split by lottery variance (separately within each percentile of the estimated CPT value difference) or by the number of payout states. In the right panels, the y-axis shows the fraction of subjects who are inconsistent at least once. Left panels constructed from 6,476 choice problems, right panels omit problems with absolute EV difference of less than \$0.20, hence constructed from 6,137 choice problems.

risk taking can even increase in the lottery's variance when the lottery is very unattractive (the left part of the figure).

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 2,678 problems for which all payouts are weakly positive (and one option 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 EV than the safe payment, we estimate $\hat{\alpha} = 0.77$ ($s.e. = 0.01$) – a typical estimate suggesting small-stakes risk aversion. In contrast, when we estimate on the sub-sample in which the lottery has a lower EV than the safe payment, we estimate $\hat{\alpha} = 1.05$ ($s.e. = 0.01$) – if anything, suggesting apparent risk loving preferences.[19] This exercise shows that complexity-dependent attenuation can predictably bias the estimation of preference parameters.

What is more, we can again provide evidence on the underlying mechanism. The top right panel of Figure 4 shows that lottery variance strongly predicts choice randomness (within-subject choice inconistencies). Thus the top left and top right panels paint a consistent picture: as variance increases, choice becomes more random (and hence more compressed to 50-50), which produces behavior that spuriously contributes to estimated risk aversion or risk love.[20]

We do not claim that genuine small-stakes risk aversion does not exist. Rather, the point is that the indirect effect generated by complexity-dependent attenuation is so strong that it can either amplify or entirely override any true aversion that likely exists.

Lottery support. The middle left panel of Figure 4 again shows choice rates for the lottery (when the alternative is a safe payment) as a function of the estimated CPT value difference, now split by whether the lottery has two or more payout states. Again, aversion to complexity would be indicated by a horizontal downward shift, while attenuation is indicated by a compression effect.

If anything, we see mild evidence of a compression effect and / or complexity seeking behavior. These patterns do not hinge on the specific sample split but look very similar when we instead split the sample at 3, 4 or 5 states. One reason for this null result is plausibly the mechanical effect of a heuristic of treating all payout states roughly equally (i.e., probability weighting). Intuitively, as is well-known in the literature (e.g. Wakker, 2022), if people treat all payout states the same to some degree, splitting up payout

[19]We get similar results when we estimate a CARA model instead, $EU(x) = E[-e^{-\gamma x}]$. In the sub-sample of problems in which the lottery has a higher EV than the safe payment, we estimate $\hat{\gamma} = 0.0093$ ($s.e. = 0.0004$), and in the sub-sample in which the lottery has a lower EV than the safe payment, we estimate $\hat{\gamma} = -0.0008$ ($s.e. = 0.0003$).

[20]An alternative interpretation of the compression pattern in the top left panel of Figure 4 is that subjects evaluate the lottery using CPT preferences and use the safe payment as a reference point, in which case prospect theory would predict the results. However, this does not explain the strong effect of lottery variance on random choice, which mechanically generates the compressed choice patterns. Moreover, this alternative explanation is conceptually unattractive because – as shown in a recent meta-analysis (Brown et al., 2024, Table 3) – prospect theory is almost always specified by assuming a reference point of zero or of the expected value across all options (rather than of the alternative).

states will either produce seeming complexity aversion or complexity lovingness, purely depending on whether the split states are attractive or unattractive (relative to the other payout states of the respective lottery). Indeed, as we unpack in Appendix Table 9, we find that lottery support has a positive effect on choice rates for the lottery when the low-probability events have relatively high payouts, yet a negative effect when the low-probability events have relatively low payouts. The overall effect of weak 'complexity seeking' in Figure 4 may thus simply reflect that in our dataset low-probability events tend to have higher payouts, on average.

The middle right panel shows the lack of a correlation between the number of payout states and choice inconsistencies ($r = -0.01$, $p = 0.32$). This is consistent with the results in Arrieta and Nielsen (2023) who also document that a larger number of payout states does not increase choice noise.

Compound probabilities. The bottom panel again shows the familiar figure for choice rates, now split by whether the lottery involves compound probabilities. Here, we see a strong and statistically significant complexity aversion effect – for any given estimated CPT value difference, choice rates for the lottery are considerably smaller when it involves compound probabilities. This result of compound aversion is consistent with various contributions in the literature (e.g., Halevy, 2007; Gillen et al., 2019). We cannot consider the link between compound probabilities and choice inconsistencies because the PEA experiment did not include compound probabilities.

6 Structural Estimations

6.1 Incorporating Complexity into Structural Analyses

The reduced-form analysis in Section 5 provided evidence for both complexity-driven attenuation and complexity aversion. A natural question is thus how complexity can be incorporated into structural estimation to quantify magnitudes.

Incorporating complexity aversion into structural analyses is relatively straightforward, by adding a cost function to the decision maker's objective (e.g., Puri, 2022; Fudenberg and Puri, 2021). To incorporate complexity-driven attenuation, we allow the responsiveness parameter (or error variance) in a random choice model to depend on complexity. Suppose that choice probabilities are given by the logit model:

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

where η is the conventional responsiveness (precision) parameter. In this model, unlike in the reduced-form *OCI* index developed earlier, aggregation complexity and resulting attenuation (captured by η) can be separated from proximity to indifference. Appendix D develops a 'structural' complexity index that allows researchers to let the precision η in their estimated logit model to depend on complexity. This index of objective aggregation complexity (*OAC*) captures the predicted logit noisiness in the EV task, where the prediction is computed as a function of the same choice set features as *OCI*. Structural analyses can then be implemented by estimating a complexity-augmented logit model, in which the precision is heteroscedastic and specified as

$$\eta_{C,D} = \eta_0 + \eta_1 / OAC_{C,D} + \epsilon_{C,D} \,, \tag{8}$$

where we supply $OAC_{C,D}$ and the researcher estimates η_0 and η_1.

This exercise relies on the same logic and identifying assumptions as the composite complexity indices discussed above, except that we need to invoke the additional assumption that complexity impacts choice through the responsiveness parameter in a logit model.

6.2 Evidence

We estimate eq. (7) and (8) using maximum likelihood for different combinations of (i) the specification of the DM's objective function and (ii) the presence of complexity-dependent heteroscedasticity / attenuation (i.e., whether η_1 is estimated or forced to be zero). Appendix D presents details for the estimating equation for each model as well as the resulting parameter estimates.

To start out, consider CPT. For each choice problem, the left panel of Figure 5 plots the actual choice rate for the lottery that has higher value (according to an estimated CPT model), as a function of *OAC*. In addition, we plot the model-predicted choice rates in a CPT model. As is clear from this figure, CPT 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.

The right panel of Figure 5 shows the prediction errors of a CPT model augmented by a complexity-dependent precision term. We see that predicted and actual choice rates track each other much more closely.

To systematically assess model fit, Figure 6 plots the variance explained across each of six models. To assess model completeness, the dashed horizontal line corresponds to the variance explained of a machine learning ensemble ($R^2 = 74\%$).[21]

The first model assumes EV maximization and only estimates a constant precision parameter. The second model estimates CPT and includes separate utility curvature

[21] This ensemble prediction is computed in an analogous fashion to the completeness analyses in Appendix B, i.e., as a combination of a convolutional neural net and several alternative models.

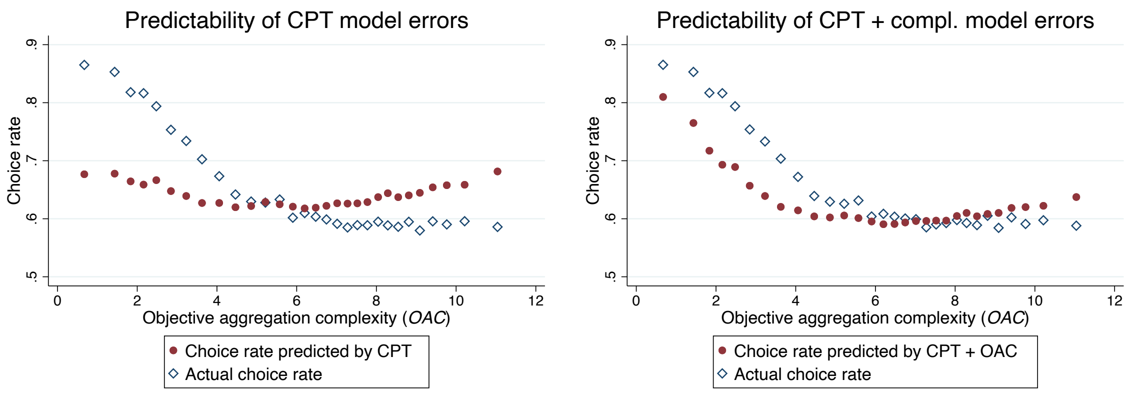


Figure 5: Model prediction errors in pooled choice data as a function of objective aggregation complexity (10,898 choice problems). Left panel: CPT model. Right panel: CPT model with complexity-dependent precision parameter. Both panels plot the actual and predicted choice rates for the choice option that the respective model predicts has higher CPT value.

parameters for gains and losses (and hence includes reference-dependence relative to a reference point of zero), loss aversion and probability weighting. The third and fourth models are identical to the first two except that they also estimate the parameter that maps the complexity index into logit precision. The fifth and sixth models are identical to the first two except that they also estimate two complexity aversion parameters, one for compound lotteries and one for support.

The EV model has an R-squared of 44%, which increases to 57% in the full CPT model. Introducing one parameter that maps problem complexity into logit precision brings an EV model to $R^2 = 58\%$, slightly larger than the full CPT model. In other words, in our dataset, complexity-dependent attenuation alone is quantitatively more important than reference dependence, utility curvature, loss aversion and probability weighting combined. The variance explained further increases to 68% under the full complexity-dependent CPT specification.[22]

Complexity aversion has small effects on model fit. Consistent with the discussion in Section 5.3, we usually estimate significant aversion against compound lotteries but not against lotteries with many states (see Appendix Table 10 for parameter estimates). The small increase in model fit (about 3% relative to CPT) is partly driven by the construction of the dataset (only 0.4% of all choice problems involve compound lotteries).

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

[22]These estimations assume CRRA utility and that the reference point is zero. Appendix Figure 13 shows that we get similar results when we assume CARA utility or a reference point that is given by the expected value of all lotteries in an experiment. Regardless of how we specify CPT, we always find that the variance explained increases substantially when the logit precision term is allowed to depend on complexity.

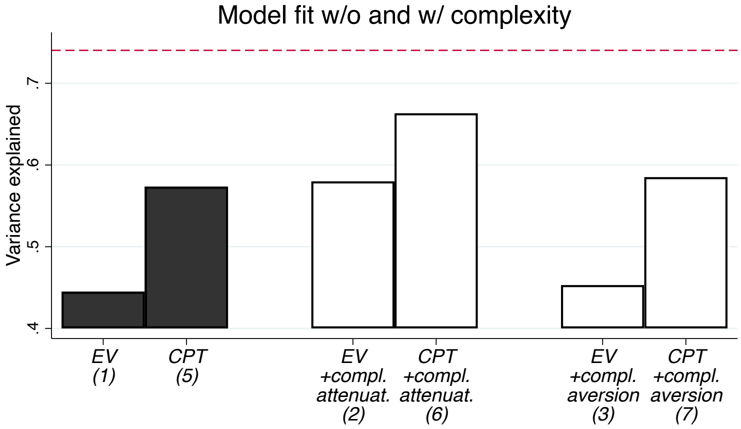


Figure 6: Variance explained of different choice models. Number of estimated model parameters in parentheses. R^2 is computed by first estimating each model, then using estimated model parameters to predict choice rates, and then regressing actual on predicted choice rates in the test set. The first model assumes EV maximization and a constant logit responsiveness term. The second model adds utility curvature parameters for gains and losses, loss aversion and probability weighting. The third and fourth models are analogous except that they also estimate the parameter that maps *OAC* into logit responsiveness, see eq. (8). The fifth and sixth models are identical to the first two except that they also allow for aversion to compound lotteries and lottery support. See Appendix D for all estimating equations and estimated parameter values. For all models, we show the performance in a test set of 2,726 choice problems after the models were estimated in a train sample of 8,172 choice problems. The dashed line corresponds to the performance of a machine learning ensemble.

($\delta_{A,B}$ rather than $ln(d_{A,B})$) 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 objective and subjective lottery choice complexity. 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 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 that captures the tradeoff complexity 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 attenuation: an insensitivity of choice to problem fundamentals. This complexity-driven attenuation explains considerably more of the variation in proxies for choice errors than proximity to indifference. Moreover, in structural estimations, a single parameter that maps complexity into logit responsiveness (or logit error variance) adds more explanatory power than all prospect theory parameters combined. 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.

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, bolds 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. A natural question is how our indices can be applied to larger choice sets. In our *EV Tasks* experiment, we also included menus with between three and five options. Appendix F 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 generalizing measures such as excess dissimilarity to larger choice sets (see Natenzon (2019) and Shubatt and Yang (2024) for steps in this direction).

The importance of dissimilarity/ tradeoff complexity. Economists frequently equate the term "complexity" with "cardinality" or "size". Our results, instead, suggest that much of the complexity of lottery choice reflects the difficulty of aggregating relative advantages and disadvantages across payout states ("tradeoff complexity"). While various theoretical papers have proposed that dissimilarity makes decisions difficult or noisy (e.g., Fishburn, 1976, 1978; Natenzon, 2019; He and Natenzon, 2022), only recently have researchers begun to use the idea of tradeoff complexity to explain classic behavioral economics choice anomalies. Shubatt and Yang (2024) develop a model to propose measures of dissimilarity-driven complexity in lottery, intertemporal and multiattribute choice. They document that the logic of dissimilarity explains many famous regularities such as the context-dependent nature of probability weighting, apparent hyperbolic discounting over money, preference reversals and more. We believe that much is to be gained from further studying how complexity arises from the difficulty of aggregating tradeoffs across states or problem dimensions.

References

Agranov, Marina and Pietro Ortoleva, "Stochastic choice and preferences for randomization," *Journal of Political Economy*, 2017, *125* (1), 40–68.

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.

Arrieta, Gonzalo and Kirby Nielsen, "Procedural Decision-Making In The Face Of Complexity," 2023.

Arts, Sara, Qiyan Ong, and Jianying Qiu, "Measuring decision confidence," *Experimental Economics*, 2024, pp. 1–22.

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.

Brown, Alexander L, Taisuke Imai, Ferdinand M Vieider, and Colin F Camerer, "Meta-analysis of empirical estimates of loss aversion," *Journal of Economic Literature*, 2024, *62* (2), 485–516.

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," *The Quarterly Journal of Economics*, 2023, *138* (4), 2021–2067.

_ , _ , **and Ryan Oprea**, "Complexity and Time," *Working Paper*, 2023.

_ , _ , _ , **and Jeffrey Yang**, "Behavioral Attenuation," 2024.

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.

Fishburn, Peter C, "Binary choice probabilities between gambles: Interlocking expected utility models," *Journal of Mathematical Psychology*, 1976, *14* (2), 99–122.

_ , "A probabilistic expected utility theory of risky binary choices," *International Economic Review*, 1978, pp. 633–646.

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 and Thomas Graeber, "The Complexity of Economic Decisions: Theory and Experimental Evidence," *Working Paper*, 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.

He, Junnan and Paulo Natenzon, "Moderate utility," *Available at SSRN 3243657*, 2022.

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

Hu, En Hua, "A Procedural Model of Complexity Under Risk," 2023.

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.

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.

Mononen, Lasse Miikka, "Essays on Rationality: Modeling and Measurement." PhD dissertation, Princeton University 2021.

Natenzon, Paulo, "Random choice and learning," *Journal of Political Economy*, 2019, *127* (1), 419–457.

Oprea, Ryan, "What makes a rule complex?," *American Economic Review*, 2020.

_ , "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.

Shubatt, Cassidy and Jeffrey Yang, "Similarity and Comparison Complexity," *Working Paper*, 2024.

Sonsino, Doron, Uri Benzion, and Galit Mador, "The complexity effects on choice with uncertainty–Experimental evidence," *The Economic Journal*, 2002, *112* (482), 936–965.

Tversky, Amos and Daniel Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and uncertainty*, 1992, *5* (4), 297–323.

Verstyuk, Sergiy, "Log, Stock and Two Simple Lotteries," *Available at SSRN 2763959*, 2016.

Vieider, Ferdinand M, "Noisy coding of time and reward discounting," Technical Report, Ghent University, Faculty of Economics and Business Administration 2021.

_ , "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.

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 4: Experimental literature on lottery complexity

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), Arts et al. (2024)
Number of states	More describable	Arrieta and Nielsen (2023)
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 and Ortoleva (2017)
Payout variance (decisions from experience)	Higher noise	Erev and Barron (2005)

B Derivation and Completeness of Complexity Indices

B.1 Screenshot of Interface

See Figure 7.

B.2 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 $(p_1^j, \ldots, p_{k_j}^j)$ and payoffs $(x_1^j, \ldots, x_{k_j}^j)$, where k_j denotes the number of distinct payout states of lottery j. In the construction of our complexity indices, we include the features listed in Table 5. Whenever a feature is defined for a single lottery rather than a choice set, we include the average feature in

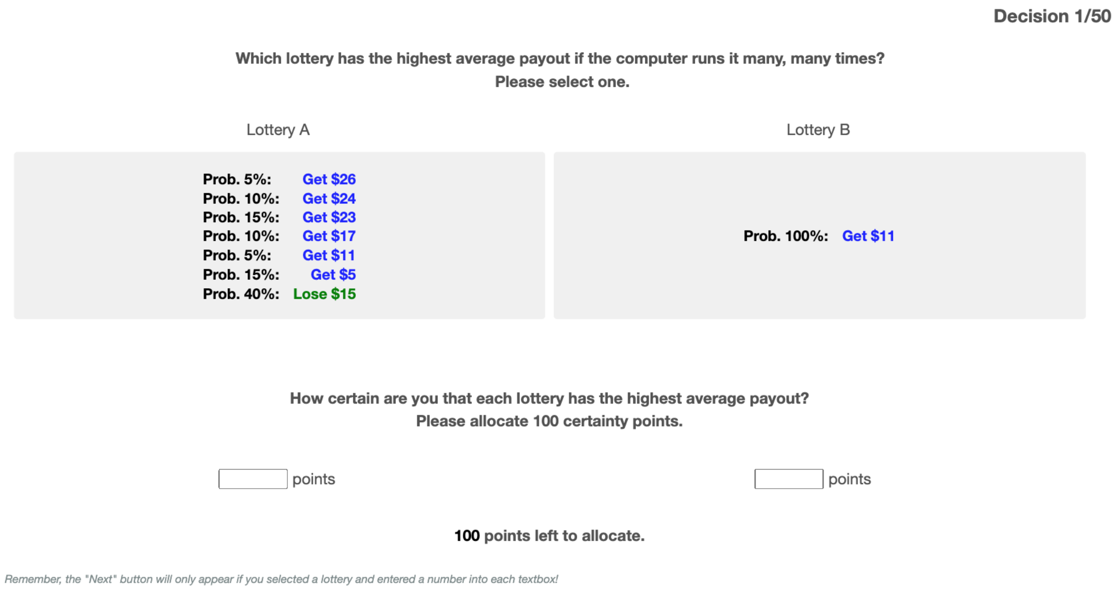


Figure 7: Example decision screen in *EV Tasks*

the set. For continuous features (and "number of states") f, we include the linear term (f), square (f^2), and the natural log ($\ln(f + 0.1)$).

B.3 Derivation of Indices

Table 6 shows the results of the exploratory LASSO regressions.

Table 7 shows the coefficients in each of the handcrafted indices. We winsorize the indices from below at zero to and from above at 0.5 to facilitate their interpretation of predicted error rates / predicted cognitive uncertainty. OAC and SAC capture predicted logit noisiness and are thus only winsorized from below at zero.

B.4 Completeness

To assess completeness, we follow Fudenberg et al. (2022) and benchmark the out-of-sample performance of our complexity indices against that of a highly flexible non-parametric model. In particular, we construct an ensemble predictor from the predictions of (1) our objective complexity LASSO, (2) *OCI*, and (3) a fully-flexible convolutional neural network (CNN). The CNN is trained to predict error rates in a non-parametric fashion based on the raw lottery features (payouts and probabilities) and the handcoded features that enter our LASSO regressions. We then form an ensemble predictor by using OLS regression to combine these three predictors in a validation set of 105 problems. We evaluate all models in a test set of 531 problems. Formally, completeness is defined as $C(\hat{y}) = \frac{R^2_{\hat{y}} - R^2_{\text{Prox}}}{R^2_{\text{Ens}} - R^2_{\text{Prox}}}$ where $R^2_{\hat{y}}$ is the variance explained by $\hat{y}$, R^2_{Ens} is the variance ex-

Table 5: 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	$\sum_{s=1}^{k_j} p_s^j \left(x_i^j\right)^2 - \left(\sum_{s=1}^{k_j} p_s^j x_s^j\right)^2$
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	$1/k_j \sum_{s=1}^{k_j} \lvert x_s^j \rvert$
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	$1/k_j \sum_{s=1}^{k_j} \lvert x_s^j \rvert \min\{p_s^j; 1 - p_s^j\}$
Entropy	Option	$\sum_{s=1}^{k_j} p_s^j(-ln(p_s^j))$
Normalized payout dispersion	Option	$1/k_j \sum_{s=1}^{k_j} \frac{\lvert x_s^j - \bar{x}^j \rvert}{\lvert \bar{x}^j \rvert}$
Normalized Variance	Option	$(1/\text{Magn.}^2) \cdot \text{Var}$, with Magn., Var. as defined above
Irregular probabilities	Option	$\mathbb{1}\left(p_s^j \notin \{0.01, 0.05, 0.1, ..., 0.9, 0.95, 0.99\} \text{ for } s = 1, ..., k_j\right)$
CDF self-distance	Option	$\sum_{s=1}^{k_j} \lvert x_s^j - EV(j) \rvert p_s^j$
Compound	Option	
Compound Range	Option	Range of distribution of unknown p
Weak dominance	Choice set	$\mathbb{1}\{F_A(x) \leq F_B(x) \forall x \text{ or } F_B(x) \leq F_A(x) \forall x\}$
Excess dissimilarity	Choice set	$\int_{\mathbb{R}} \lvert F_A(x) - F_B(x) \rvert dx - \lvert EV(A) - EV(B) \rvert$
Average absolute payoff difference	Choice set	$1/k \sum_{s=1}^{k} \lvert x_s^A - x_s^B \rvert$
Probability difference	Choice set	$\sum_{x \in X} \lvert f_A(x) - f_B(x) \rvert$, where $X = \{x_1^A, ... x_{k_A}^A\} \cup \{x_1^B, ..., x_{k_B}\}$

plained by the ensemble predictor $\hat{y}_{\text{Ens}}$, and R^2_{Prox} is the variance explained by a baseline model which depends only on proximity to indifference. Variance explained is always calculated in the test set. Figure 8 shows the variance in mistake rates explained by *OCI*. 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, indices that only consist of the separate aggregation complexity features, such as log excess dissimilarity. Third, our handcrafted *OCI* index. Fourth, the analogous exploratory LASSO indices.

There are three main takeaways. First, proximity to indifference performs substantially worse than our index. By construction, its completeness is 0%. Second, log excess dissimilarity explains three times as much variation as proximity to indifference (66.6% complete). Third, our handcrafted index is almost as complete as the exploratory LASSO index (95% and 98% complete respectively). We conclude that our index captures a large fraction of the predictable component of complexity.

Table 6: LASSO coefficients for *OCI* and *SCI*

	Coefficients			Coefficients	
Feature	**OCI**	**SCI**	**Feature**	**OCI**	**SCI**
Intercept	1.5E-01	4.3E-02	Sq Scale	-2.6E-07	2.4E-06
Log Abs. EV Difference	-9.4E-02	-1.3E-02	Sq Range		1.3E-07
Scale		-1.0E-03	Sq Variance	3.9E-11	-5.0E-10
Range		-3.3E-04	Sq Payout Variance	9.3E-11	2.7E-10
Variance		1.6E-05	Sq Num States		9.1E-04
Pay-wtd DC	-1.7E-03		Sq DC		-1.1E-01
Probability Variance		-4.4E-05	Sq Pay-wtd DC		1.3E-06
Payout Dispersion	6.6E-03		Sq Entropy		-2.4E-02
CDF Self-Dist	-1.2E-03	-2.4E-03	Sq Prob. Variance		1.9E+00
Log Scale	3.8E-02	2.2E-02	Sq Payout Dispersion		6.1E-03
Log Range		1.7E-02	Sq Norm. Variance	6.3E-03	1.8E-02
Log Variance		-2.0E-02	Sq CDF Self Dist		1.3E-05
Log Payout Variance	1.7E-02		Gains	-9.3E-03	-1.9E-02
Log Num States	2.8E-02	-2.4E-02	Irregular Probabilities	4.0E-02	-2.3E-03
Log DC	-1.9E-01		Excess Dissimilarity		-1.2E-03
Log Pay-wtd DC	-7.5E-05	1.6E-02	Log Excess Dissimilarity	5.3E-02	3.5E-02
Log Entropy	-2.4E-02	1.4E-01	Dominance	-5.2E-02	-2.4E-02
Log Prob. Variance		-3.7E-01	Compound	5.5E-02	3.6E-02
Log Payout Dispersion		-8.8E-03	Compound Range	1.2E-01	1.0E-01
Log Norm. Variance		-1.5E-02	Safe Option	7.2E-03	2.0E-02
Log CDF Self-Dist		3.6E-02			

Notes. Coefficients of LASSO regression of problem-level errors rates or cognitive uncertainty on choice set features in the *EV Tasks* experiment. DC = distance to certainty. Features that apply to a single lottery (such as number of states) are averaged across the lotteries in the set. Only features with at least one non-zero coefficient are included.

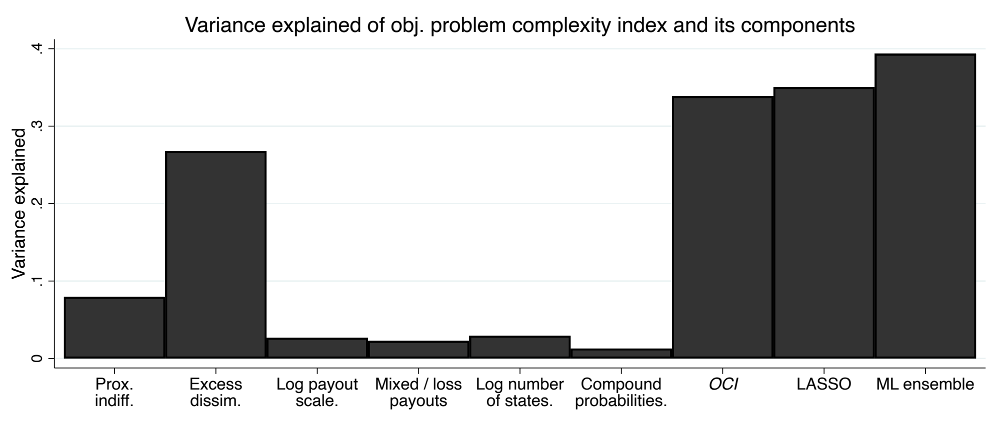


Figure 8: Variance explained of complexity indices and separate features. Each bar reports the variance explained of a regression of mistale rates in the EV task on the respective feature (in the test set). Leftmost model includes linear and squared terms of absolute expected values difference. Second model includes log excess dissimilarity, and so on. Seventh model is *OCI*, eight model the exploratory LASSO index and the ninth model the predictions of a machine learning ensemble.

Table 7: Coefficients of features in complexity indices

	Dependent variable:			
	Error rate	CU	Implied obj. logit imprecision s^{EV}	Implied subj. logit imprecision s^{EV}
Index:	*OCI*	*SCI*	*OAC*	*SAC*
	(1)	(2)	(3)	(4)
Log excess dissimilarity	0.058***	0.022***	1.76***	1.86***
	(0.00)	(0.00)	(0.13)	(0.14)
Average log payout magnitude	0.035***	0.0029	1.11***	1.41***
	(0.00)	(0.00)	(0.18)	(0.20)
Average log number of states	0.044***	0.052***	2.06***	3.09***
	(0.02)	(0.01)	(0.59)	(0.63)
Frac. lotteries involving loss	0.028***	0.027***	1.16***	2.23***
	(0.01)	(0.00)	(0.32)	(0.35)
1 if involves compound prob.	0.072***	0.060***	2.45***	3.39***
	(0.02)	(0.01)	(0.63)	(0.67)
Log absolute EV difference	-0.089***	-0.015***		
	(0.01)	(0.00)		
Constant	0.12***	0.067***	-3.56***	-4.65***
	(0.03)	(0.01)	(0.77)	(0.83)
Observations	1587	1587	1587	1587
R^2	0.31	0.36	0.20	0.24

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 (3) and (4), the dependent variable is the implied logit precision $s^{EV}_{A,B}$ as defined in eq. (11) (either defined through mistakes or through cognitive uncertainty). To calculate $s^{EV}_{A,B}$, we first winsorize the error rates and CU so that they never exceed 0.49. Then, we winsorize the calculated $s^{EV}_{A,B}$ at the 85th percentile. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C Details for and Analyses of Choice Experiments

C.1 Tables

Table 8: Summary statistics for problems across experiments

Experiment		**# options**	**Safe payment**	**# states**	**Var**	**Scale**	**Mixed**	**Dominance**
EV	Mean, median	2.1, 2	58%	3.3, 2	633, 164	34, 26	46%	5%
Tasks	Range	2, 5		2, 7	0, 39440	1, 213		
Choice	Mean	2	80%	3.3	775	25.5	49%	7%
Tasks	IQR	2, 2		2, 5	20, 735	14, 43		
Choice Tasks	Mean	2	59%	3.7	460	30	53%	17%
from PEA	IQR	2, 2		2, 5	20, 553	15, 40		

Notes. *PEA* = Peterson et al. (2021). For the *EV Task*, statistics are limited to problems with menu size two except for *# options*. Scale = absolute average payout. We display information for the lottery with the largest number of distinct payout states.

Table 9: Choice rates and lottery support

	Dependent variable: Fraction of subjects choosing lottery (in %)					
Sample: Payout difference:	Full	≥ 0	< 0	< -10	< -20	< -30
	(1)	(2)	(3)	(4)	(5)	(6)
Support of lottery	0.97***	1.26***	0.34*	-0.26	-1.79***	-2.97***
	(0.08)	(0.10)	(0.18)	(0.30)	(0.44)	(0.55)
CPT value difference	6.67***	5.83***	7.97***	7.50***	6.42***	4.96***
	(0.07)	(0.08)	(0.12)	(0.15)	(0.20)	(0.27)
Constant	48.8***	47.2***	50.8***	51.8***	55.7***	56.6***
	(0.36)	(0.48)	(0.60)	(0.87)	(1.20)	(1.54)
Observations	6476	3725	2751	1682	866	410
R^2	0.59	0.57	0.64	0.60	0.56	0.49

Notes. OLS estimates, robust standard errors in parentheses. An observation is a choice problem. The sample consists of problems in which A is a non-degenerate lottery and B is a safe payment. The sample is split by the difference in average payouts between states that have low probability (defined as smaller than 1/N) and states that have high probability (defined as larger than 1/N), where N is the number of states of A. Positive values mean that low-probability states have relatively high payouts. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C.2 Figures

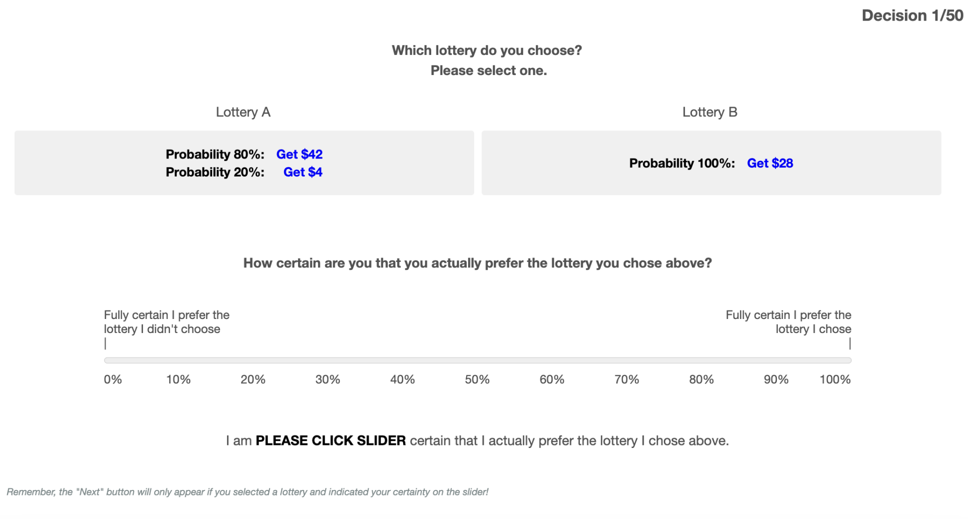


Figure 9: Example decision screen in *Choice Tasks*

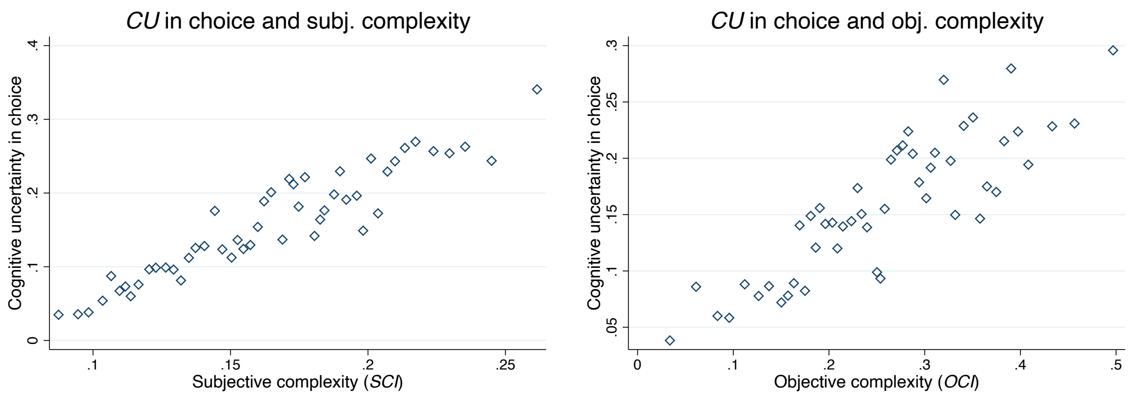


Figure 10: Average cognitive uncertainty in choice and complexity indices in experiment *Choice Tasks*. The figures show binned scatter plots of average cognitive uncertainty in each of 500 lottery choice problems, as a function of the predicted complexity indices. In the figure, an underlying observation is a choice problem. Figure is constructed from subjects who did not complete any EV task. Data from 250 subjects and 500 choice sets.

C.3 Replication of Figures 3 and 4 Excl. Losses in PEA data

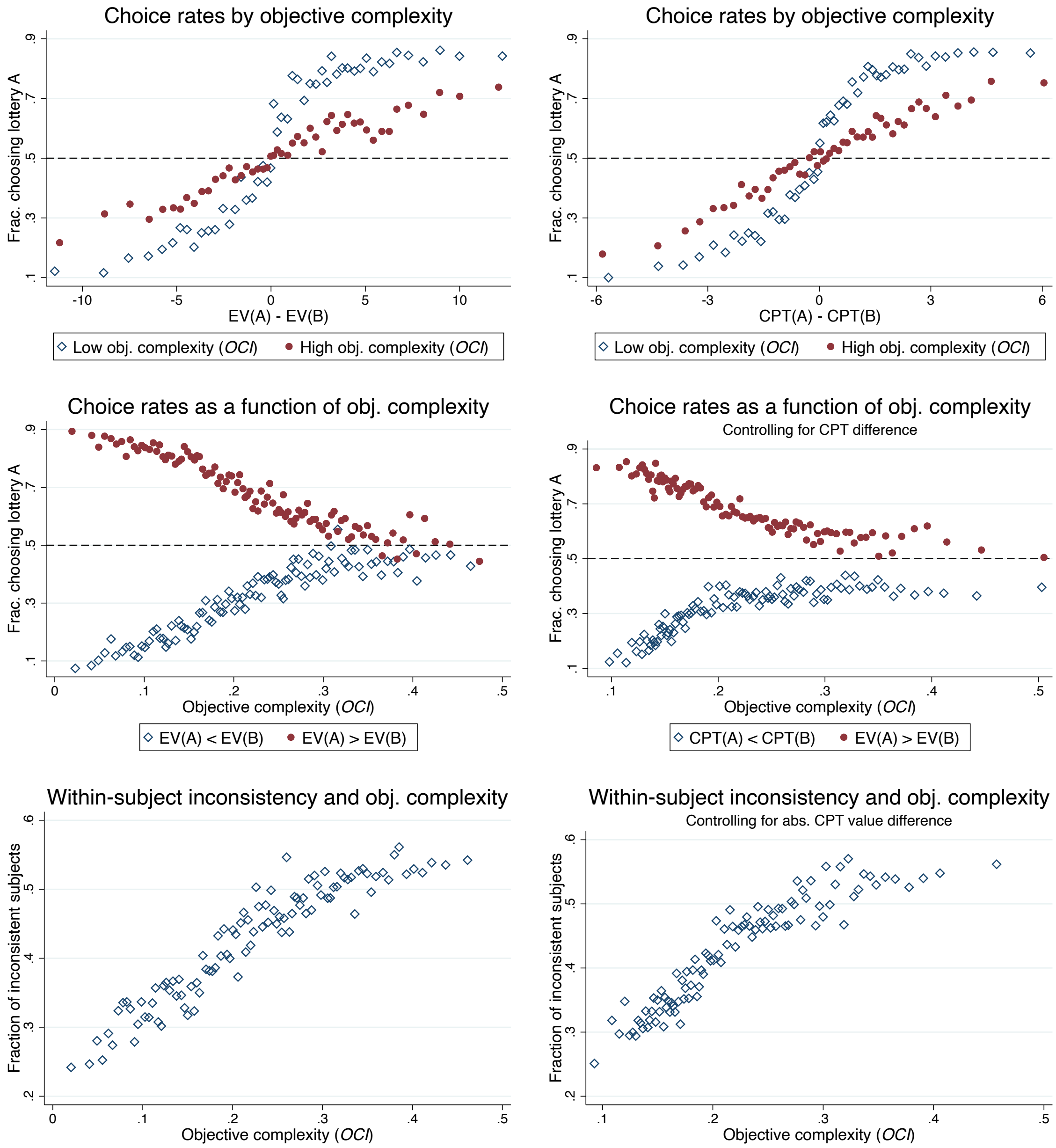


Figure 11: Replication of Figure 3, excluding choice problems in PEA dataset that involve losses.

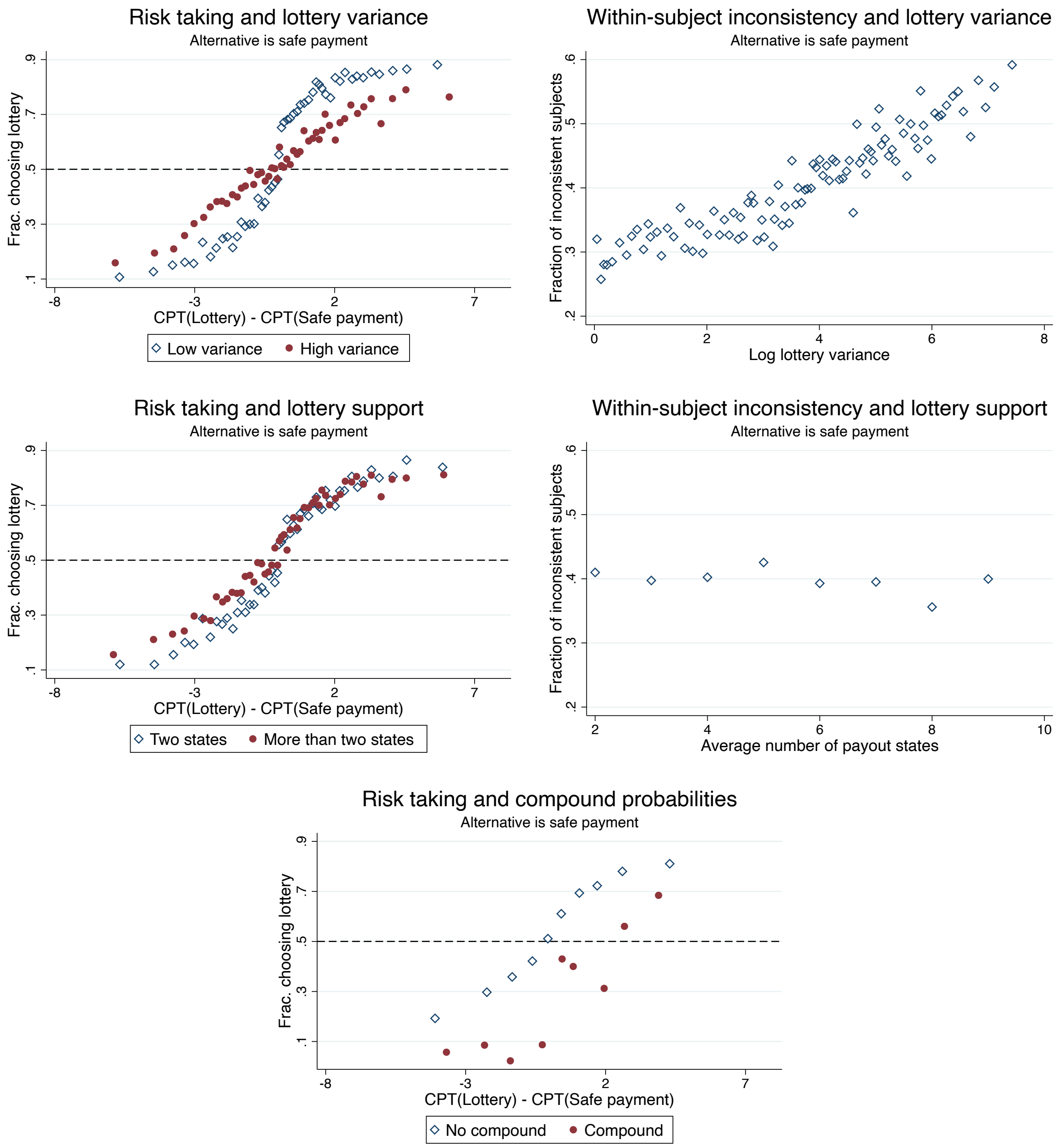


Figure 12: Replication of Figure 4, excluding choice problems in PEA dataset that involve losses.

D Details on Structural Estimations

D.1 Derivation of Structural Complexity Index

Suppose that choice probabilities are given by the logit model:

$$P(A) = F(EU(A) - EU(B); \eta) = \frac{1}{1 + e^{-\eta\,[EU(A) - EU(B)]}}\ , \tag{9}$$

where η is the conventional responsiveness (precision) parameter. We define the aggregation complexity of a choice problem as the inverse of the decision-maker's precision for that problem. Inverting the logit CDF,

$$s_{A,B} := \frac{1}{\eta_{A,B}} = \frac{EU(B) - EU(A)}{\ln\left(\frac{1}{P(A)} - 1\right)}. \tag{10}$$

Again, we cannot observe this object in choice data since we do not know the utility function. However, in the EV task, we can compute the "implied logit precision"[23] because we observe both the EV difference and the empirical rate of choosing A.

$$s^{EV}_{A,B} := \frac{1}{\eta^{EV}_{A,B}} = \frac{EV(B) - EV(A)}{\ln\left(\frac{1}{P(A)} - 1\right)}. \tag{11}$$

Similarly to before, we develop an index of predicted aggregation complexity by imposing the assumption that the problem-specific precision in the choice task is a linear function of the precision in the EV task,

$$\eta_{A,B} = \eta_0 + \eta_1 \frac{1}{s^{EV}_{A,B}} \quad \text{with} \quad \eta_1 > 0. \tag{12}$$

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*

$$s^{EV}_{A,B} = \sum_{i=0}^{N} \alpha_i f_i^{A,B} + \epsilon_{A,B}\ . \tag{13}$$

where $s^{EV}_{A,B}$ is calculated as in (11)*. Thus, the index is* $OAC_{C,D} = \hat{s}^{EV}_{C,D} = \sum_{i=0}^{N} \hat{\alpha}_i f_i^{C,D}$.

We define an analogous index of subjective aggregation complexity (SAC), in which objective mistakes rates in the EV task are replaced with subjective ones (CU).

[23]Empirically, we winsorize selection rates for the higher EV lottery from below at 0.51 (since otherwise (10) is undefined). Next, we winsorize the across-problem distribution of $\hat{s}^{EV}_{A,B}$ at the 85th percentile because $s^{EV}_{A,B}$ can explode when selection rates for the wrong lottery is close to 50% or when the expected values difference is very small.

D.2 Model Estimation

Cumulative prospect theory. We allow up to four additional parameters: loss aversion with respect to a reference point of zero, separate utility curvature for gains and losses, and probability weighting:

$$EU_{CPT}(x) = \sum_i \pi_i u(x_i) \,, \tag{14}$$

where

$$u(x) = \begin{cases} x^\alpha & \text{if } x \geq 0 \\ -\lambda(-x)^\beta & \text{if } x < 0 \end{cases} \tag{15}$$

and π_i is a Tversky and Kahneman (1992) cumulative prospect theory decision weight with capacity function

$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}} \,. \tag{16}$$

Complexity aversion / seeking. In some specifications, we allow the value function to incorporate as-if complexity preferences. Based on earlier results, we allow for aversion (or seeking) responses to the number of states in the lottery and whether the lottery is compound. For these models, we let the value function take the form

$$V(x) = EU(x) + c_0 \cdot \text{ Compound } + c_1 \cdot \log|\text{support(x)}|. \tag{17}$$

Note that negative values of c_0 or c_1 indicate complexity aversion; positive values indicate complexity seeking. In robustness checks (available upon request), we model states-aversion using a three-parameter sigmoid function as proposed by Fudenberg and Puri (2021) and find similar results.

Estimation results. Table 10 summarizes the parameter estimates across the different models. In our likelihood function, we weight each "person-problem" equally. Though the problems from Peterson et. al. were repeated five times, the repetitions were consecutive, so we do not treat them independently. The choice data we collected contains no repetitions. We report standard errors for the maximum likelihood estimation procedure, which we derive from numerically estimated Hessian matrices.

Alternative specifications. Figure 13 shows the variance explained of a series of alternative models. The first three models are the ones reported in the main paper. The

Table 10: Parameter estimates for model estimations

	Model					
Parameter	EV	CPT	EV + CA	CPT + CA	EV + CN	CPT + CN
α		0.8		0.78		0.745
		(0.005)		(0.01)		(0.006)
β		0.783		0.782		0.74
		(0.01)		(0.022)		(0.01)
λ		0.942		0.922		0.879
		(0.048)		(0.121)		(0.047)
γ		0.829		0.831		0.81
		(0.004)		(0.004)		(0.005)
c_0			-6.295	-0.969		
			(0.622)	(0.228)		
c_1			0.42	0.43		
			(0.039)	(0.025)		
η_0	0.128	0.28			0.044	0.153
	(0.001)	(0.006)			(0.002)	(0.004)
η_1					0.296	0.609
					(0.005)	(0.018)

Notes. Parameter estimates for model estimations described in Section 6. "+ CA" = allows for complexity aversion. "+ CN = allows for complexity-dependent attenuation (η)".

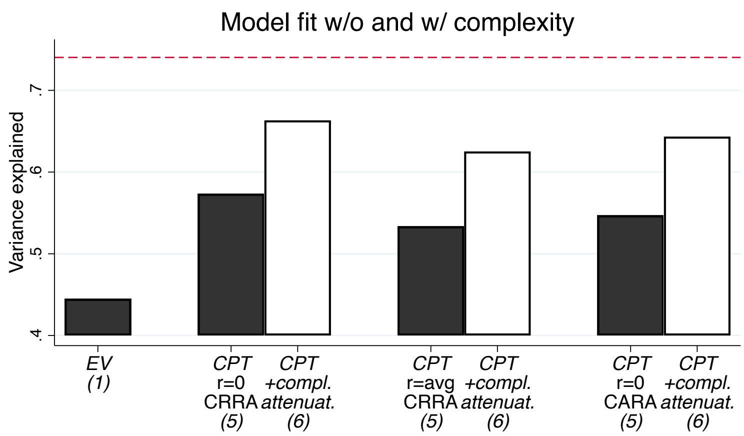


Figure 13: Variance explained of different choice models. Number of estimated model parameters in parentheses. R^2 is computed by first estimating each model, then using estimated model parameters to predict choice rates, and then regressing actual on predicted choice rates in the test set. The first model assumes EV maximization and a constant logit responsiveness term. The second model is the CPT model we estimate throghout the paper, and the third model the extension of CPT in which logit precision depends on complexity. The fourth model is the same CPT model, except that the reference point is assumed to be the average payout across all lotteries in an experiment. The fifth model again augments this CPT model with a complexity-dependent logit precision term. The sixth model is CPT with a reference point of zero and CARA utility, and again the seventh model extends this model with a complexity-dependent logit precision term. For all models, we show the performance in a test set of 2,726 choice problems after the models were estimated in a train sample of 8,172 choice problems. The dashed line corresponds to the performance of a machine learning ensemble.

fourth and fifth are analogous except that the reference point is assumed to be the average payout across all menus in each experiment. In the sixth and seventh models, we estimate CARA rather than CRRA utility (with a reference point of zero).

E Analyses of Within-Subject Experiment

In the *Within Subject* experiment, each participated was confronted with identical problems in a choice frame and an *EV Task* frame. Figure 14 shows the correlation between the elicitations of cognitive uncertainty across these two problem types, controlling for subject fixed effects. These plots are constructed from 300 subjects x 20 problems = 6,000 observations, but we bin them into buckets to make the plot more informative.

Table 11 shows the correlation between cognitive uncertainty in each of the two types of problems with problem features.

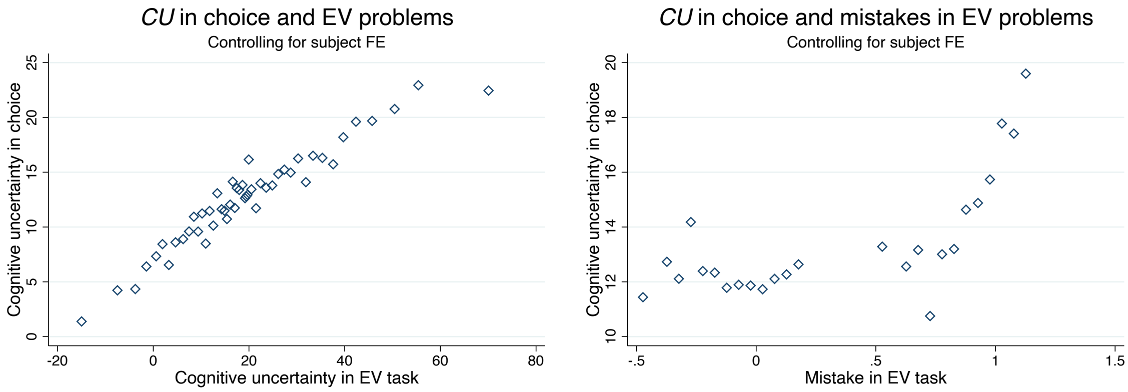


Figure 14: Cognitive uncertainty in choice and in EV task in experiment *Within Subject*. The figures show binned scatter plots of cognitive uncertainty in a choice problem, as a function of cognitive uncertainty in the EV task (left panel) or of a binary mistakes indicator in the EV task. An underlying observation is a subject-decision (not a choice problem). The figure links identical choice sets in the choice and EV task and is constructed controlling for subject fixed effects. Data from 300 subjects and 240 choice sets.

Table 11: Correlations of *CU* in choice and *CU* in EV task with complexity features

	CU in choice	CU in EV	Log excess dissim	Log scale	Log support	Frac. lotteries w/ loss	Compound prob
Avg. CU in choice	1.00	0.72	0.64	-0.05	0.25	0.42	0.05
Avg. CU in EV task	0.72	1.00	0.69	0.13	0.25	0.47	0.29
Log excess dissim.	0.64	0.69	1.00	0.40	0.23	0.47	0.00
Ave log scale	-0.05	0.13	0.40	1.00	0.14	-0.12	0.02
Ave log support	0.25	0.25	0.23	0.14	1.00	0.28	-0.14
Frac lotteries w/ loss	0.42	0.47	0.47	-0.12	0.28	1.00	-0.00
Compound prob	0.05	0.29	0.00	0.02	-0.14	-0.00	1.00

Notes. Pairwise Pearson correlations in the *Within Subject* experiment. The first two columns show the correlations of average cognitive uncertainty in choice and the EV task with the complexity features.

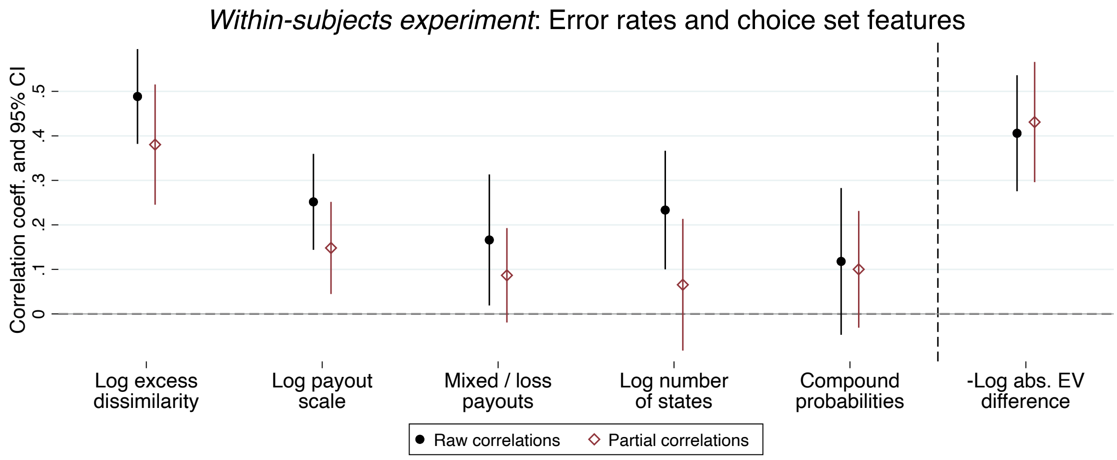


Figure 15: Raw and partial correlation coefficients between task-level error rates and choice set features in the *Within Subject* experiment (240 unique problems). Whiskers show 95% confidence intervals. Partial correlations are calculated controlling for all of the other features in the figure. Log scale, mixed / loss payouts, log number of states and compound probabilities are computed as averages across the lotteries in a set.

F Extension: Larger Menus

In our *EV Tasks* experiment, we also included 102 menus with between three and five options. These menus were excluded from all analyses in the main paper. Figure 16 shows the link between choice set features and errors / cognitive uncertainty in the full sample of 2,220 problems.

With menus larger than two, calculating excess dissimilarity is less trivial. We proceed by calculating 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.

Figure 16 shows that menu size has a large effect on both errors and *CU*, though the effect is still somewhat smaller than that of excess dissimilarity.

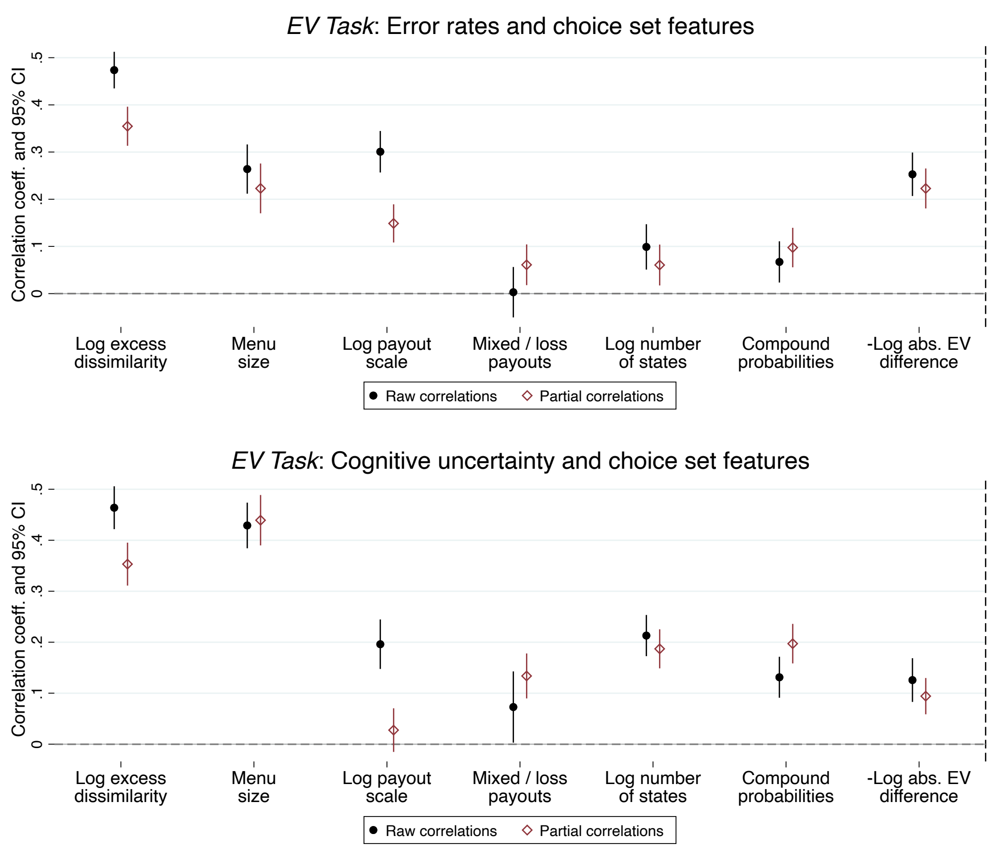


Figure 16: Correlations between choice set features and errors / cognitive uncertainty in the full dataset in the *EV Task*, including menus with more than two options in the train set. Raw and partial correlation coefficients between task-level error rates / average cognitive uncertainty and choice set features. Whiskers show 95% confidence intervals. Excess dissimilarity 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.

G Targeted Problems in *EV Tasks*

The 120 "targeted" *EV Tasks* problems that we manually devised fall into two categories. 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.

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.

Table 12 summarizes the results, which are broadly consistent with those from the

Table 12: Results for targeted *EV tasks*

	Dependent variable:			
	1 if error		Cognitive uncertainty	
	(1)	(2)	(3)	(4)
Higher variance / payout range	0.095***		0.028***	
	(0.02)		(0.01)	
Higher average abs. payout	0.033		-0.0076	
	(0.02)		(0.01)	
Mixed lottery (constant range)	0.033		0.021*	
	(0.04)		(0.01)	
Mixed lottery (larger range)	0.18***		0.038***	
	(0.02)		(0.01)	
Higher number of states	0.069		0.034***	
	(0.04)		(0.01)	
Lower distance to certainty	-0.0088		-0.024*	
	(0.03)		(0.01)	
Compound lottery	0.11***		0.074***	
	(0.03)		(0.01)	
Higher distance between CDFs		0.18***		0.020***
		(0.04)		(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 (twoway-clustered at subject and problem level) in parentheses. An observation is a decision. Each independent variable is a binary dummy for a problem type. The omitted category comprises the base problems. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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.

H Experimental Instructions and Comprehension Checks

H.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.

H.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:

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

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 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.

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.

H.3 Experiment *Within Subject*

The order of the two tasks in our *Within Subject* experiment was randomized. The experiment was structured so that participants first saw a set of general instructions, followed by the instructions for the first task. After completing the first task along with comprehension checks, participants then saw instructions for the second task and associated comprehension checks. The task instructions are virtually independent of order, except that two of the comprehension questions apply to both tasks, and so are only asked for the first. The instructions below have *Choice* tasks first.

General Instructions

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.

There are two parts to this study. In both parts, you will make multiple decisions.

Your payment will consist of two components:

- Completion fee: If you pass all our **comprehension checks and complete both parts of the study**, you will receive a **completion fee of $3.50.**
- Additional bonus: In each part of the study, you will face 20 decision screens where you will make a decision. One of these decision screens will be selected at random by the computer, with equal probability, and will determine your **bonus**. Your bonus will actually be paid out to you with probability 1/5.

Budget

- For some of the decisions which are eligible for a bonus payment, it is possible to make a "loss." If such a decision is selected to determine your bonus, you will receive a budget.
- This budget will always be equal to the largest loss you can potentially incur for that decision.
  - If your decision happens to earn a loss, then this loss will be subtracted from the budget in that particular task to give your final bonus.
  - If you don't incur a loss, the budget will simply be added to whatever you earn from your decision.
- The budget cannot be transferred across decisions.

Part 1 Instructions

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.

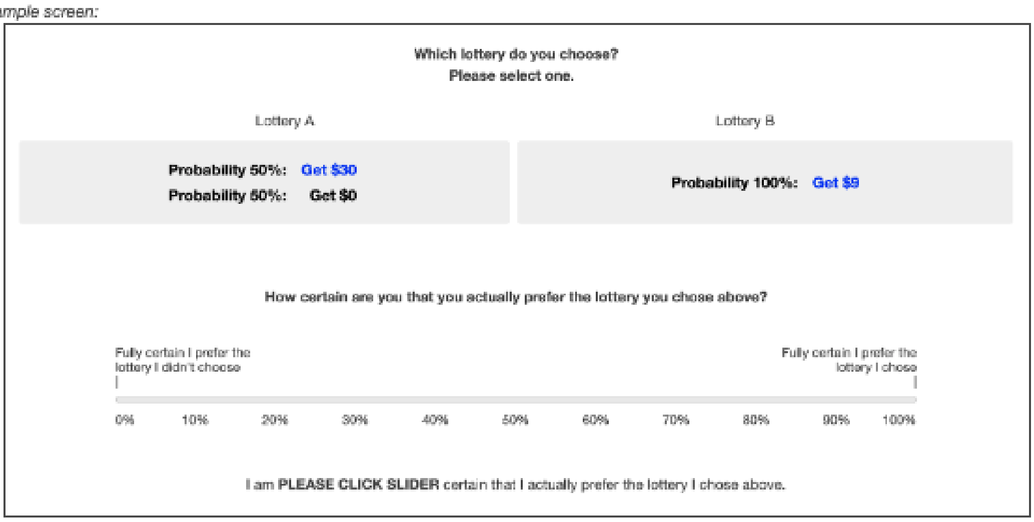
Step 2: Indicate your certainty about your choice

- 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 80% likely that you actually prefer the lottery that you chose, you should set the slider to 80%.**
  - If you are certain that you prefer the lottery you chose, you should set the slider to 100%.

Unknown Probabilities

For some lotteries, you will not know the precise probabilities. Instead, the probabilities will be randomly drawn, and we will explain the random process in the instructions for that choice.

Example screen:



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 (Part 1)

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. 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.

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

→

Part 2 Instructions

Guessing task: which lottery has the highest average payout?

On each decision screen in the second part of the study, you will again be presented with two lotteries that pay out different amounts of money with different probabilities.

This time, your task is **NOT** to indicate which lottery you would personally prefer to receive if it was only played out once.

Instead, you will be **paid according to the average payout of each lottery,** 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.**
  - We know from statistics that when a lottery is run so many times, we can trust that the probabilities will match the true shares of payouts. Put simply, after running a lottery that pays $4 with probability 30% many, many times, statistics tells us that 30% of the payouts will *actually* be $4.
- **On each decision screen, you will make two decisions that reflect which lottery you think has the highest average payout.**

Step 1: Select lottery with 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 **the average payout of the lottery you selected.**
- Your payment is based only on the **average payout** of the lotteries. This means it is always in your best interest to choose the lottery which pays out more on average, even if that lottery is riskier.

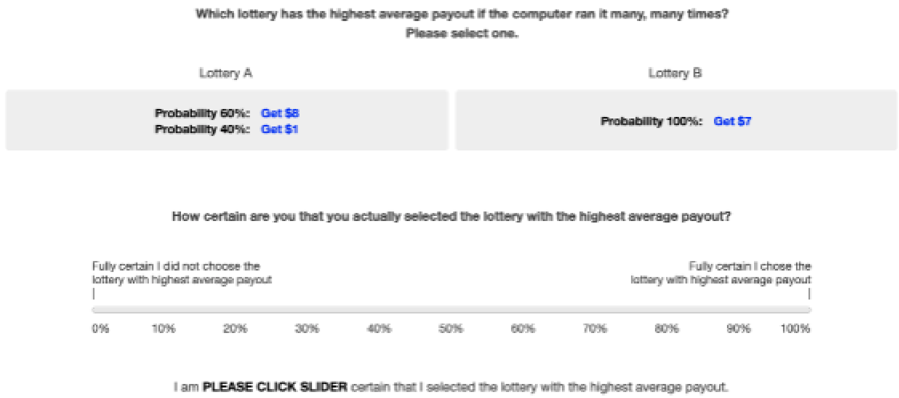
Step 2: Indicate your certainty about your choice

- You might feel uncertain about which lottery has the highest average payout. Like in the previous part, we will ask you to indicate how certain you are (in percent) that you actually chose the lottery with the highest average payout.

Unknown Probabilities

For some lotteries, you will not know the precise probabilities. Instead, the probabilities will be randomly drawn, and we will explain the random process in the instructions for that choice. To determine how much each lottery pays on average, the probabilities are randomly drawn many, many times. Each time, the lottery is run with the chosen probabilities and the payout is recorded. We then compute the average payout across all of these draws.

Example screen:



In this example, if Lottery A was run many, many times, it would have a lower average payout ($5.20) than Lottery B if it was run many, many times ($7).

Here is how your bonus would be determined in this example:
- If you selected lottery A, your bonus would be $5.20
- If you selected lottery B, your bonus would be $7.

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 (Part 2)

To verify your understanding of the instructions, please answer the comprehension questions below. You must answer each question correctly 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. Suppose that you picked Lottery B in one of the tasks and this decision is selected to earn a bonus payment.

Lottery B

Probability 85%:	**Get $9**
Probability 15%:	**Get $4**

Which of the following statements is correct?

- I know that my payoff from this lottery will be a profit of $9 AND a profit of $4, for a total of $13.
- I know that my payoff from this lottery will be either a profit of $9 OR a profit of $4, but not both.
- I know that my payoff from this lottery will be equal to the average payoff of the lottery, which is $8.25.

2. Which of the following statements is correct?

- On each decision screen I will be asked to indicate how certain I am that I selected the lottery with the highest average payout.
- On each decision screen I will be asked to indicate how certain I am that I actually prefer the lottery that I selected.
- On each decision screen I will be asked to indicate how certain I am that my bonus will be determined by each lottery.