A Simple Test for Decisions under Risk

Jianying Qiu, Jilong Wu, Songfa Zhong*

February 2, 2026

Abstract

We present a simple test for evaluating decisions under risk, where participants assess the value of (almost) degenerate lotteries. This test is model-agnostic, as most preference-based and cognition-based models, which primarily focus on lottery-specific characteristics, predict that valuations should converge to their degenerate values. In an experiment with a Prolific sample, we find that a majority of participants fail the test, and those failures exhibit a more pronounced fourfold pattern of risk attitudes compared to those who pass. Two additional experiments confirm the robustness of these findings, even after controlling for cognition-related measures, namely, deterministic mirrors and cognitive uncertainty. In a separate experiment with a student sample, while participants are less likely to fail the test, the link between test performance and the fourfold pattern persists. Going beyond lottery-specific valuation and cognition, these results highlight the need for more nuanced assessments of cognitive engagement in decisions under risk.

Keyword: preference, non-expected utility, cognition, complexity, risk, experiment

JEL Code: C81, C91

---

*Qiu: Department of Economics, Radboud University, jianying.qiu@ru.nl; Wu: Department of Economics, Ghent University, Jilong.Wu@ugent.be; Zhong: Department of Economics, National University of Singapore, zhongsongfa@gmail.com. Acknowledgment: To be added. This is a preliminary draft. Comments are welcome.

1 Introduction

How individuals evaluate lotteries is central to the study of decisions under risk. While the workhorse model has long been expected utility (EU) theory, a large experimental literature documents behavioral anomalies that systematically depart from EU—notably Allais paradox and the fourfold pattern of risk attitudes. These anomalies have spurred the development of theories and interpretations that can be categorized into two approaches.

The first, classical approach is to have general *preferences over lottery* which generalizes EU.[1] These models typically feature an inverse S-shaped probability weighting function that overweighs small probabilities and underweighs moderate to large probabilities. While models in this approach relax EU assumptions, it continues to maintain behavioral restrictions such as compliance with dominance. As the preference-based approach is "as if" in nature, it does not explicitly model the sources of behavioral anomalies. In particular, it implicitly assumes that individuals do not incur cognitive costs when making decisions.

A second, recent approach focuses on *cognition about lottery* which goes beyond observable choices and considers various aspects of cognitive frictions in processing lottery-specific characteristics.[2] For examples, individual may struggle with contingent reasoning, allocate attention disproportionately to salient payoffs and probabilities, miscalculate expected payoffs, and use imprecise mental representations. As a result of these cognitive frictions, individuals often rely on heuristic rules or make stochastic choices, giving rise to a wide range of behavioral anomalies. As this approach focuses on lottery-specific cognition, it makes sharp predictions when the lottery is cognitively simple.

We propose a simple test of the behavioral implications of these two alternative approaches and their link with four-fold pattern of risk attitudes. Experimental studies of fourfold pattern usually involve the measurement of the certainty equivalents (CEs) of binary lotteries of the form $(x, p; 0)$ that delivers outcome x with probability p and 0 oth-

[1] See, for example, Kahneman and Tversky (1979); Quiggin (1982); Chew (1983); Gul (1991); Cerreia-Vioglio et al. (2015); Masatlioglu and Raymond (2016). See Starmer (2000); Wakker (2010) for reviews.

[2] See, for example, Bordalo et al. (2012); Martínez-Marquina et al. (2019); Khaw et al. (2021); Enke and Graeber (2023); Esponda and Vespa (2024); Oprea (2024); Puri (2025). See reviews on related topics: Gabaix (2019); Maćkowiak et al. (2023); Loewenstein and Wojtowicz (2025) on attention/inattention, Woodford (2020) on imprecision, Bordalo et al. (2012) on salience, Niederle and Vespa (2023) on contingent thinking, de Clippel and Rozen (2024) on bounded rationality, Enke (2024) on cognition, and Oprea (2025) on complexity.

erwise. The outcome x can represent either monetary gains or losses, and the corresponding probability p ranges from small to moderate or high. It is commonly observed that individuals are typically risk averse (seeking) at moderate-to-high probabilities and risk seeking (averse) at low probabilities for gain-(loss-) oriented lotteries. The fourfold pattern of risk attitudes is often referred to as likelihood insensitivity and can be accommodated through models with either preference-over-lottery or cognition-about-lottery approaches. In our simple test, we remove the *lottery-specific characteristics* and elicit the CEs of (almost) degenerate lotteries. More specifically, degenerate lotteries $(px, 1; 0)$ pay the expected value px with certainty, and almost degenerate lotteries $(px + \epsilon, p; px - \epsilon)$ have possible payoffs that are tightly around px. It is important to note that the value px is explicitly provided, which eliminates the need for participants to perform any computation. We say that an individual fails the test when she assigns the CEs of these test lotteries that deviate from px. Our test is model-agnostic because the CE of these (almost) degenerate lotteries ought to be close to px in the above two approaches. Going beyond lottery-specific characteristics, this simple test therefore captures the influences of a wider array of cognitive factors, such as comprehension, incentives, and inattention in the decision environment.

We examine the power of this test in diagnosing the fourfold pattern of risk attitudes in a series of experiments. In the main experiment, we use price lists to elicit CEs of a set of standard binary and test lotteries from a Prolific sample. Lotteries are represented by 100 boxes, with $100p$ containing x and $100(1 - p)$ containing y, and participants' payments are determined by the amount in a randomly selected box. We have three main findings. First, the majority of participants fail the test, and most of them fail frequently. Second, participants who fail the test display a more pronounced fourfold pattern, that is, greater likelihood insensitivity, than those who pass. Third, performance on comprehension questions neither predicts test failure nor helps explain the association between likelihood insensitivity and test failure.

We then conduct three additional experiments to gain insights based on two cognition-related measures and the generalizability of our findings. The first additional experiment incorporates the *deterministic mirrors* proposed in Oprea (2024). The design of deterministic mirrors shares the same structure as the lotteries, with their values determined by the average payoff of all 100 boxes, instead of one randomly chosen box. Notably, such deterministic mirrors are similar to our test lotteries for their objectively defined values, but contain an extra layer of computation complexity: participants need to compute the average of 100 boxes with $100p$ containing x and $100(1-p)$ containing y. We observe that, in both lotteries and deterministic mirrors, the fourfold pattern is stronger for participants who fail the test than for those who pass. Importantly, fourfold pattern is linked to the failure of the simple test but not the extra layer of computation complexity in the deterministic mirrors, when both are jointly considered.

The second additional experiment incorporates the elicitation of *cognitive uncertainty* (Enke and Graeber, 2023), in which participants report how certain they feel about their CEs of the lotteries. We find that the link between likelihood insensitivity and failure of the test holds in both the high- and low-cognitive-uncertainty subsamples and robust to the controlling for cognitive uncertainty. Furthermore, when we examine cognitive uncertainty in both the (almost) degenerate and the standard lotteries simultaneously, we find that the former, but not the latter, is significantly associated with the fourfold pattern. This suggests that the additional layer of lottery-specific cognitive uncertainty plays a limited role, aligning with the finding from the deterministic mirror experiment.

The third additional experiment evaluates the generalizability of our findings by replacing the Prolific sample with a student sample from a Dutch research university. Unlike in the Prolific sample, most participants pass the test, and those who fail typically doing so only once. However, similar to the Prolific sample, the fourfold pattern is stronger among those who fail compared to those who pass, and this difference is also robust when we control for performance on comprehension questions.

In summary, the results from these three additional experiments reinforce our findings from the main experiment: a significant portion of participants fail the test, and this failure is consistently linked to more pronounced fourfold pattern even after controlling for comprehension check performance, although it is weaker in the student sample. Furthermore, this association remains robust after controlling for behavior in deterministic mirrors and self-reported cognitive uncertainty.

These observations add to the literature on decisions under risk. Preference-over-lottery approach is central to the modeling of decisions under risk. By incorporating inverse S-shaped probability weighting function, most such models extend expected utility to rationalize behavioral anomalies. While these models typically incorporate psychological intuition into economic behavior, they do not explicitly model cognitive processes. As a result, they are largely silent on whether and how cognitive measures such as cognitive uncertainty or deterministic mirrors map onto observable behaviors such as probability insensitivity (Wakker, 2025). Yet, these models predict convergence in our test: the valuation of an (almost) degenerate lottery should approach the value of the corresponding degenerate outcome.[3] As a result, they can not rationalize the joint occurrence of fourfold pattern, the failure of test, and the connection between the two, as observed among the majority of Prolific participants and a substantial proportion of student participants.

Cognition-about-lottery approach is increasingly prominent in recent literature. Cognitive frictions in perceiving the lotteries and computing the values can lead to probability insensitivity in lotteries. However, they too imply convergence in our simple test, as evaluation of (almost) degenerate lotteries requires no computation. Thus, like preference-based models, cognition-based accounts cannot explain the joint presence of fourfold patterns in lotteries and the failure of the test as well as their association.

Moreover, our results further suggest limited role of lottery-specific cognition. In the experiment involving deterministic minors, we show that the failure of our simple tests is an important driver of the fourfold pattern across both tasks, and the additional layer of computation complexity of lottery plays is not. In the cognitive uncertainty experiment, we find that cognitive uncertainty beyond lottery valuation contributes to the four-fold pattern, while the additional layer of cognitive uncertainty driven by the lottery appears to be less significant. Taken together, these observations point to the importance of incorporating broader notions of cognition—such as comprehension, incentives, and inattention, salience, memory—into both modeling and measurement of the general choice environment (Bordalo et al., 2012, 2020; Woodford, 2020; Khaw et al., 2021; Frydman and Jin, 2023).

Importantly, this interpretation does not dismiss the relevance of lottery-specific cog-

[3]Since we describe our degenerate lotteries as having two states that yield the same outcome, the test failure may reflect an event splitting effect (Starmer and Sugden, 1993; Bernheim and Sprenger, 2020). In Kahneman and Tversky's original prospect theory (1979), such dominance violation can occur with event splitting, when the probability weighting function exhibits the subcertainty property. However, we observe that the degenerate lottery is valued higher (or lower) than the true value when the true value is low (or high). Therefore, while the event splitting effect may play a role in our observations, it is unlikely to be the primary factor.

nition, as its behavioral significance is inherently linked to the cognitive difficulty of the lottery itself. The limited role observed in our setting likely reflects the simplicity of the binary lotteries $(x, p; 0)$ we employ. In more complex environments, lottery-specific cognition can be much more significant. For instance, Puri (2025) explore lotteries with varying numbers of states, ranging from 2, 4 to 16; Martínez-Marquina et al. (2019) investigate state-contingent payoffs that need to be calculated in bidding tasks; and Khaw et al. (2023) study cognitive imprecision with winning probabilities of 0.58 and payoffs such as 5.55, 15.70, or 31.40, rather than round numbers. We acknowledge the importance of these lottery-specific characteristics and their interplay with the choice environment, along with their behavioral consequences.

Our study also contributes to the design of experiments. Experimental studies typically include detailed instructions, comprehension tests, incentivized choices, attention checks, and cognitive measures. These design features help enhance motivation and reduce misconceptions about the experiments, as well as improve data quality (Plott and Zeiler, 2005, 2007; Cason and Plott, 2014). While these instruments are valuable, they are imperfect proxies for the cognition required to evaluate the lotteries in the experiments. For instance, in our study, the correlation between performance on the simple test and the comprehension check is only moderate. Our test is also closely related to the mirror test proposed in Oprea (2024), but differs in that it does not require any computation, thereby avoiding the computational complexity inherent in lottery valuation. When we combine our test with the measure of cognitive uncertainty, it also allows to separately measure cognitive uncertainty about (almost) degenerate lotteries and standard lotteries. Overall, our test thus offers a practical, task-embedded diagnostic for identifying cognitive factors in the broader decision environment.

The rest of the paper is organized as follows. Section 2 presents experimental design. Section 3 reports results from the main experiment and the three additional experiments. Section 4 concludes and discusses the broader implications of our findings.

2 Experimental design

This section presents experimental design. We first outline the tasks in the main experiment that are common across all experiments and then highlight the differences in three additional experiments.

2.1 Main Experiment

Standard Lotteries and Price-list Elicitation. We study binary lotteries of the form $(\$x, p; \$y)$, in which the decision-maker receives x with probability p and y otherwise. To study the fourfold pattern of risk attitudes (see, e.g., Wakker, 2010), we include both gains and losses with $y = 0$, $x \in \{25, -25\}$, and a range of probabilities $p \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$. For conciseness, we denote these lotteries by GP and LP, respectively, where G (L) indicates a gain of \$25 (a loss of \$25) and P denotes the payout probability expressed as a percentage (e.g., G10 represents $(\$25, 0.10; \$0)$).

We elicit valuations of these lotteries using price lists. Each price list consists of 25 rows of choices between Option A and Option B. Option B is the lottery $(\$x, p; \$0)$ and is fixed across all rows in the list. Option A is a sure payment $(\$c, 1)$ whose amount decreases monotonically in increments of \$1 across rows, from \$25 to \$1 when $x = 25$ and from $-\$1$ to $-\$25$ when $x = -25$. Both payment options are represented by 100 boxes, each containing a specified amount of money, and participants' payments are determined by the amount in a randomly selected box. For example, the lottery $(\$25, 0.10; \$0)$ is implemented by randomly opening one of 100 boxes, where 10 boxes contain \$25 and 90 boxes contain \$0. For a sure payment $\$c$, all 100 boxes contain the same amount $\$c$.

Test lotteries. We include eight test lotteries that are either degenerate, $\pm(\$2.30, 0.10; \$2.30)$ and $\pm(\$22.70, 0.10; \$22.70)$, or almost degenerate, $\pm(\$2.90, 0.90; \$2.10)$ and $\pm(\$22.90, 0.90; \$22.10)$. The expected values of these test lotteries are close to those of the standard lotteries $\pm(\$25, 0.10; \$0)$ and $\pm(\$25, 0.90; \$0)$.

Similar to the standard lotteries, we elicit valuations of these test lotteries using price lists. Because the sure amounts in Option A are integers and the payoffs of these test lotteries lie between two consecutive integers, the row that participants switch from preferring Option A to preferring Option B implied by the first-order stochastic dominance (FOSD) is

Table 1: Option A and Option B in the price lists.

Price lists	Option B	Option A
Standard lotteries	±($25, 0.10; $0), ±($25, 0.25; $0), ±($25, 0.50; $0), ±($25, 0.75; $0), ±($25, 0.90; $0)	From $25 to $1 for gain lists and -$1 to -$25 for loss lists in increments of $1.
simple tests	±($2.3, 0.10; $2.3), ±($22.7, 0.90; $22.7), ±($2.9, 0.10; $2.1), ±($22.9, 0.90; $22.1)	

unique and objectively determined. For example, when the lottery is $\pm(\$2.90, 0.90; \$2.10)$, participants who respect FOSD should switch from preferring Option A to preferring Option B when the sure payment in Option A is $2 or less, irrespective of their risk attitudes or the cognition of valuing lotteries. We refer to price lists involving test lotteries as the *Simple Tests*. We say that a participant passes the simple tests when their choices in these eight price lists respect FOSD. Table 1 summarizes the parameters in Option A and Option B for both the standard lotteries and the simple tests.

2.2 Additional Experiments

Deterministic Mirror. Building on the main experiment, the mirror experiment includes an additional group of Option Bs, in which the lotteries in Table 1 are replaced by their deterministic mirrors as proposed by Oprea (2024). Specifically, whereas payment for lotteries is determined by randomly opening one box from 100 boxes, payment for mirrors is determined by opening all 100 boxes and paying the average amount, thereby eliminating risk. See Figure F.1 for an example screenshot of this task. We also include mirror versions of the test lotteries. In total, the mirror experiment contains 18 price lists for lotteries and 18 price lists for mirrors.

Cognitive Uncertainty. In the cognitive-uncertainty experiment, we additionally elicit participants' cognitive uncertainty about their lottery price-list choices, following Enke and Graeber (2023). After submitting a response to each list, participants are reminded of their choices and then asked to indicate how certain they feel about that choice using a slider. Figure F.2 shows an example screenshot of this interface.

Student Sample. In the student-sample experiment, we implement essentially the same design as in the main experiment, with two differences: (i) the sample consists of students from Radboud University, the Netherlands, rather than Prolific participants; and (ii) participants face stronger monetary incentive (e.g., all instead of only some participants receive a bonus whose amount depends on one randomly selected choice).

2.3 Implementation

For the main, deterministic mirror, and cognitive-uncertainty experiments, we recruited 150 participants from a U.S.-representative sample via Prolific.[4] Participants had a balanced gender composition, were required to be between 18 and 75 years old, and have a historical approval rate of at least 99%. For incentives, in addition to the $5 participation fee, each participant had a 10% chance of being selected to receive payment based on one of their choices made in the experiment. In the student-sample experiment, we recruited 81 participants. They received no participation fee, but all got paid a bonus based on one randomly selected choice.

After reading the experimental instructions and before starting the price lists, participants completed a short training module in which they saw examples illustrating the lotteries and the price lists. They then answered several comprehension questions. In the main experiment, there were six comprehension questions: three about the payment options and three about the price lists (see Appendix F). In the mirror experiment, we added three additional questions about the mirror tasks. In the cognitive-uncertainty experiment, we included one additional comprehension question about cognitive uncertainty. Participants received an additional $0.15 for each correctly answered comprehension question on the first attempt. Participants who failed on the first attempt needed to try again and were required to answer all comprehension questions correctly before proceeding to the price lists.

To reduce fatigue and boredom, we asked participants to indicate the row at which they would switch from preferring Option A to preferring Option B. Once participants indicated

[4]Due to technical issues, we lost two observations for the cognitive-uncertainty experiment.

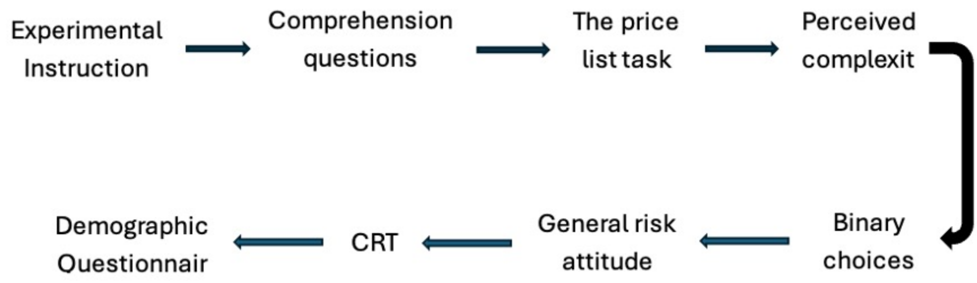


Figure 1: Procedure of the main experiment.

the switching row, the implied choices in each row were highlighted in yellow. Participants could revise their switching row as often as they wished and finalized their decision by clicking the "Submit" button. Appendix F.1 presents an example of the decision screen, along with additional screenshots of the experimental interface.

At the end of the lottery price lists (and of the mirror price lists in the mirror experiment), we repeated two lists. In one list, Option B was either $(\$25, 0.10; \$0)$ or $(\$25, 0.90; \$0)$; in the other list, Option B was either $(-\$25, 0.10; \$0)$ or $(-\$25, 0.90; \$0)$. After completing all price lists, participants reported their perceived complexity of the price-list task. They then faced eight additional binary choices, each on a separate decision screen. Four of these choices were $(\pm\$25, 0.10; \$0)$ *vs.* $\pm\$3$ and $(\pm\$25, 0.90; \$0)$ *vs.* $\pm\$22$, which replicated some of the choices embedded in the price lists. The remaining four binary choices were taken from Kahneman and Tversky (1979): $(\pm\$4000, 0.80; \$0)$ *vs.* $\pm\$3000$ and $(\pm\$5000, 0.001; \$0)$ *vs.* $\pm\$5$. These eight binary choices were not incentivized.

Following these binary choices, we elicited participants' general risk preferences and administered three cognitive-reflection questions. At the end of the experiment, participants completed a short demographic questionnaire (gender, age, education, etc.). Figure 1 illustrates the sequence of tasks in the main experiment. The experimental sequences for the additional experiments are reported in the corresponding appendices. All experimental protocols were approved by the Institutional Review Board at the National University of Singapore.

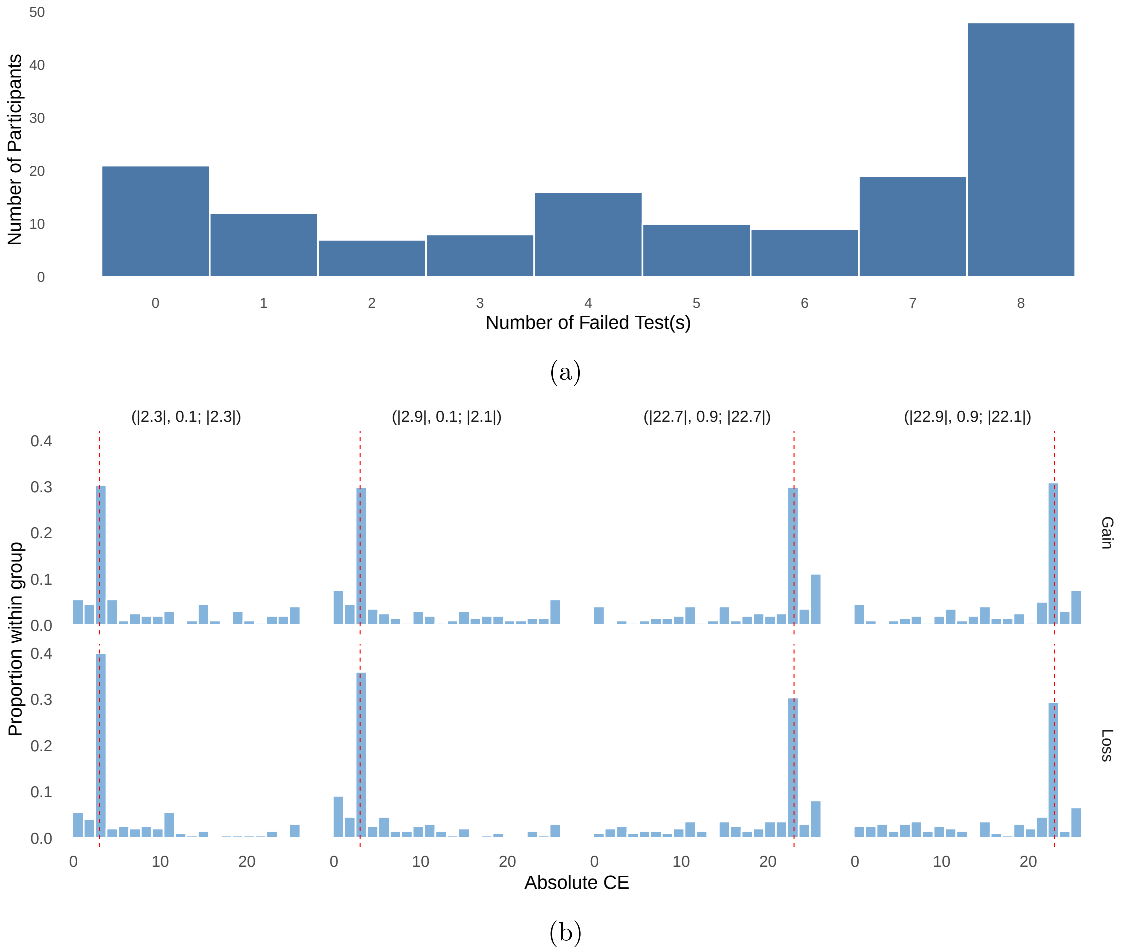


Figure 2: Performance on simple tests. *Notes:* Panel (a) shows the distribution of the number of simple tests failed by each participant. Panel (b) plots the distribution of CEs across participants in each of the eight simple test lists. Red dashed lines denote the theoretically correct CE for each test lottery. The sample consists of 150 participants.

3 Experimental results

3.1 Main Experiment

Observation 1A. *The majority of participants fail at least one of eight simple tests. When participants fail the simple tests, their CEs for the test lotteries are distributed around the correct values.*

Panel (a) of Figure 2 summarizes the number of simple tests each participant fails. For each list, about 30%–40% of participants pass the test. Aggregating across all eight test lists, only 21 out of 150 participants pass all simple tests. Among the remaining 129 participants, 12 fail exactly one simple test, 7 fail exactly two, and 110 fail three or more.

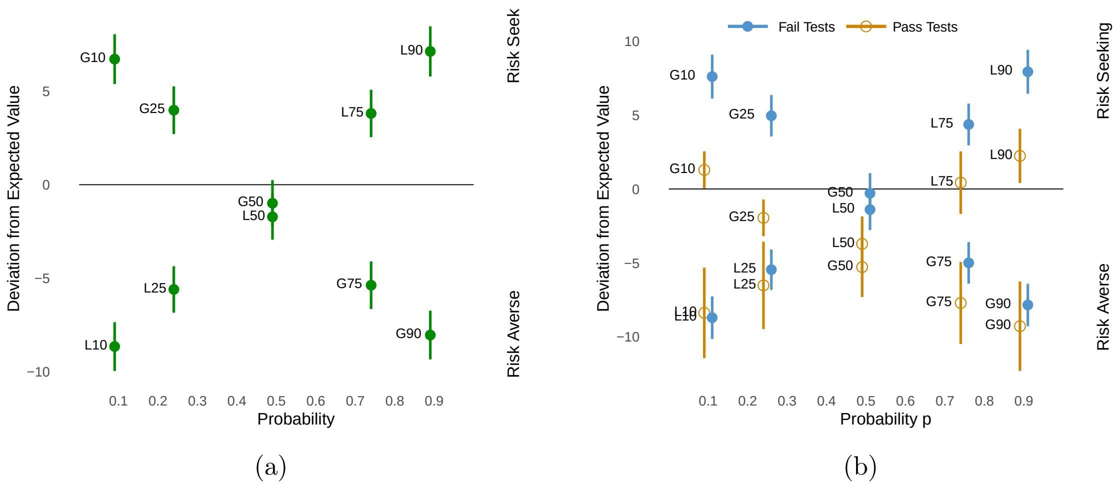


Figure 3: The fourfold pattern of risk attitudes and simple test performance. *Notes:* Panel (a) plots the fourfold pattern of risk attitudes for all participants. Panel (b) plots the same pattern separately for participants who pass versus those who fail the simple tests. In both panels, the x-axis is the payout probability p in the lotteries $(\$25, p; \$0)$ and $(\$-25, p; \$0)$, and the y-axis is the difference between the CE and the expected value, with positive values indicating risk seeking and negative values indicating risk aversion. Error bars correspond to ± 2 standard errors. Appendix Table A.1 reports the corresponding summary statistics of Panel (b).

We set Option B's certainty equivalent (CE) as the average of the sure payments in the two switching rows. For example, if a participant switched between \$9 and \$8, we set Option B's CE as \$8.5. Figure 2b depicts the distribution of CEs across 150 participants in the eight simple test lists. We find that, when participants fail a simple test, most of their CEs are close to the correct value. This pattern indicates that many participants roughly know the correct CE, though not perfectly. Importantly, there is no clear tendency for participants to report the middle value of the price list as their CE (among the incorrect CEs, only 2.04% correspond to the middle value of the list). The proportion of failure of the test is 60.0% in the degenerate lotteries and 62.5% in the almost degenerate lotteries $(p > 0.1)$.

Observation 1B. *Participants exhibit a fourfold pattern of risk attitudes. More importantly, participants who fail the simple tests exhibit a more pronounced fourfold pattern of risk attitudes than those who pass.*

Figure 3a illustrates the fourfold pattern of risk attitudes among all participants. The x-axis is the payout probability p in $(\$25, p; \$0)$ and $(-\$25, p; \$0)$, and the y-axis is the difference between the CE and the expected value, with positive values indicating risk seeking and negative values indicating risk aversion. Participants on average exhibit a clear fourfold pattern: they are risk seeking when lotteries have low-probability gains (G10 and G25) or high-probability losses (L75 and L90), and they are risk averse when lotteries have high-probability gains (G75 and G90) or low-probability losses (L10 and L25). This figure replicates the classical findings in the literature (see, e.g., Kahneman and Tversky, 1979; Tversky and Kahneman, 1992).

Figure 3b plots the fourfold pattern separately for participants who pass versus those who fail the simple tests. The figure reveals a clear pattern: the fourfold pattern is more pronounced among participants who fail the simple tests than among those who pass. In particular, participants who fail the simple tests are significantly more risk seeking toward low-probability gains and high-probability losses than those who pass.

To complement this graphical evidence, we estimate a series of random-effects regressions of lotteries' CEs on the lotteries' payout probability p within each domain. For better comparison across domains, we take the normalized absolute value of the CEs, $\frac{|CE_{ij}|}{25}$. The regression results in Columns (2) and (6) of Table 2 corroborate the patterns in Figure 3b: the coefficient on the interaction between payout probability p and the performance dummy D_{ST} is significantly negative. This implies that participants who fail the simple tests are less sensitive to changes in likelihood, which is consistent with a stronger fourfold pattern.

We conduct several analyses to examine the robustness of these findings. First, we replace the performance dummy variable D_{ST} with the continuous variable F_{ST}, which records the number of failed simple tests (see Table A.2). Second, we estimate Tobit regressions to account for the two-sided censored CEs between 0 and 25 (see Table A.3). Our observations remain essentially unchanged.

Observation 1C. *1) comprehension check performance does not fully capture simple test performance; 2) the association between simple test performance and fourfold risk attitudes remains robust after controlling for comprehension check performance.*

Literature commonly assesses participants' understanding of the experiment through comprehension questions. A natural question is therefore whether participants' performance in the comprehension questions is sufficient to account for the behavioral differences we document. We find that, while performance on the simple test and comprehension check is significantly correlated (Spearman's $\rho = 0.242, p < 0.001$), performance on comprehen-

Table 2: Random-effects regressions of normalized absolute certainty equivalents

	Gain Domain				*Loss Domain*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p	0.247***	0.479***	0.314***	0.495***	0.209***	0.451***	0.224***	0.449***
	(0.030)	(0.079)	(0.040)	(0.079)	(0.030)	(0.078)	(0.040)	(0.079)
D_{ST}		0.314***		0.251***		0.039		0.047
		(0.067)		(0.067)		(0.068)		(0.070)
$p \times D_{\mathrm{ST}}$		−0.270**		−0.230**		−0.281***		−0.286**
		(0.085)		(0.088)		(0.085)		(0.087)
$D_{Compreh}$			0.223***	0.180***			−0.015	−0.023
			(0.046)	(0.047)			(0.048)	(0.049)
$p \times D_{Compreh}$			−0.153*	−0.114			−0.034	0.014
			(0.060)	(0.061)			(0.060)	(0.061)
Intercept	0.347***	0.077	0.249***	0.051	0.436***	0.402***	0.442***	0.406***
	(0.024)	(0.062)	(0.031)	(0.061)	(0.024)	(0.063)	(0.032)	(0.063)
Observations	750	750	750	750	750	750	750	750

Note: The dependent variable is the normalized absolute certainty equivalent, $|CE_{ij}|/25$, for lottery j of participant i. p is the lottery's payout probability; $D_{\mathrm{ST}} = 1$ indicates participants who failed the simplicity tests; $D_{Compreh} = 1$ indicates participants who failed the comprehension questions. Standard errors in parentheses. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. The first four columns (1–4) report results for the gain domain; the last four columns (5–8) report results for the loss domain. All models use random effects with participant-level clustering.

sion questions does not fully account for performance on the simple tests. Among the 65 participants who fail the comprehension questions on their first attempt, 95% (63 participants) fail the simple tests. While this proportion is significantly lower among participants who pass the comprehension questions (79% vs 95%; two-sided proportional test, $p < 0.01$), the majority in this group (66 out of 84) still fail the simple tests.

To examine this more closely, we estimate random-effects regressions of $|CE_{ij}|/25$ on the comprehension dummy $D_{\mathrm{Compreh},i}$ for each payoff domain, which equals 1 if participant i fails to answer all comprehension questions correctly on the first attempt. The results in Columns (3) and (7) of Table 2 suggest that performance on comprehension questions is systematically related to the strength of the fourfold pattern: participants who answer all comprehension questions correctly exhibit greater sensitivity to the changes in payout probability than those who fail at least one question. More importantly, when we include both $D_{\mathrm{ST},i}$ and $D_{\mathrm{Compreh},i}$ in the random regressions, the coefficients associated with D_{ST} remain stable relative to the baseline specification without comprehension controls (compare Columns (2) and (4) for gains and Columns (6) and (8) for losses). By contrast, once D_{ST} is included, the coefficients on D_{Compreh} become weaker and are often statistically insignificant (compare Columns (4) and (3) for gains and Columns (8) and (7) for losses).

These results are robust across alternative specifications (see Tables A.2 and A.3). Overall, these results indicate that comprehension questions do correlate with behavior but are less effective than the simple tests at screening for behavioral heterogeneity in the fourfold risk attitudes.

3.2 Additional Experiment—Deterministic Mirror

We refer to the proportion of boxes containing $25 or -$25 in mirrors as the *payout proportion*, to distinguish it from the payout probability of lotteries. We define CEs as before and define the *simplicity equivalent* (SE) as the value of the mirror, following Oprea (2024). Unless explicitly stated otherwise, we classify participants as passing (failing) the simple tests based on their performance in the lottery task, and use the simple test performance in the mirror task only when necessary. The deterministic mirror experiment replicates the key findings from the main experiment. To avoid duplication, the corresponding figures and tables are reported in Appendix B; here we focus on the new results involving behavioral patterns in the mirrors and its correspondence with those in the lotteries.

Observation 2A. *In both lotteries and mirrors, the fourfold pattern is stronger among participants who fail the simple tests than among those who pass.*

Consistent with Oprea (2024), behavior in the deterministic mirror experiment exhibits a fourfold pattern not only for lotteries but also for mirrors (see Figure B.3 in Appendix B). Figures 4a and 4b plot, respectively, deviations of lotteries' CEs from expected value and mirrors' SEs from the average amount for each payout probability (lotteries) or payout proportion (mirrors), splitting participants by whether they pass or fail the simple tests in the corresponding task. The fourfold pattern is clearly stronger among those who fail the simple tests than among those who pass in respective tasks. This not only reinforces our finding in the main experiment but also provides a complementary perspective: participants who pass the simple tests are significantly more responsive in their SEs to changes in mirror payout proportions.[5]

Observation 2B. *Valuations of lotteries and mirrors differ significantly among participants who pass the simple tests, but not among those who fail.*

[5]These results are robust to the order in which tasks were presented, see Figure B.4 in Appendix B).

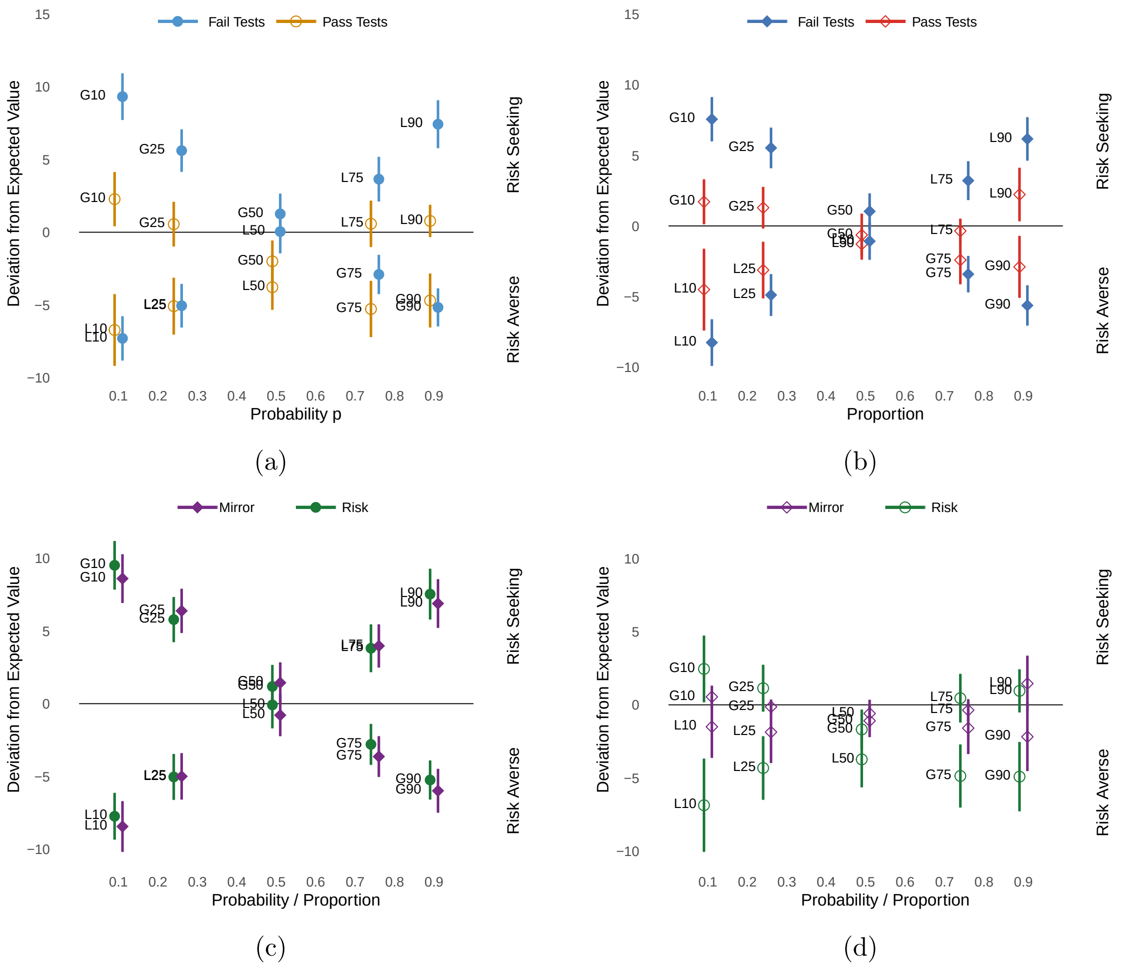


Figure 4: fourfold pattern comparisons across lotteries and mirrors. *Notes:* Panel (a) plots deviations of CEs from expected value for lotteries, separately for participants who pass versus those who fail the lottery simple tests. Panel (b) plots deviations of SEs from the average amount for mirrors, separately for participants who pass versus those who fail the mirror simple tests. Panel (c) plots deviations of CE (lotteries) against deviations of SE (mirrors) for participants who fail the simple tests in both tasks. Panel (d) plots the same comparison for participants who pass the simple tests in both tasks. Error bars correspond to ± 2 standard errors. See Appendix Tables B.1 (Panel a), B.2 (Panel b), B.3 (Panel c), and B.4 (Panel d) for corresponding summary statistics.

For each lottery and its corresponding mirror (i.e., same domain and with lottery probability equal to mirror payout proportion), deviations of CEs from expected value are significantly correlated with the corresponding deviations of SEs from the average amount (Spearman's $\rho = 0.472$, $p < 0.01$), consistent with Oprea (2024). Figure 4c further shows that, among participants who fail the simple tests in both tasks, deviations in lotteries and mirrors move closely together: Spearman's $\rho = 0.522$, $p < 0.01$, and the deviations do not differ significantly between lotteries and mirrors (Wilcoxon signed-rank test, $p > 0.10$). By contrast, among participants who pass the simple tests in both tasks (Figure 4d), there is no clear relationship between the two types of deviations: Spearman's $\rho = 0.138$, $p > 0.10$, and the deviations differ significantly between lotteries and mirrors (Wilcoxon signed-rank test, $p < 0.01$). This observation suggests that the similarity between lotteries and mirrors is not due to the computational complexity of the lotteries. Instead, it is more likely due to the cognition about the general choice environment as measured by the failure of the test.

Observation 2C. *1) Deviations from the correct value in test-lottery CEs are significantly correlated with those in mirror SEs; 2) the additional layer of computation cognition played a limited role in the fourfold pattern of lotteries.*

Like our simple tests, the mirrors also involve an objectively correct valuation. Unlike our simple tests, the mirror task contains an extra layer of computation complexity in that participants need to perform some calculation to arrive at the average. We find a significant correlation in deviations from the correct value between test-lottery CEs and mirror SEs (Spearman's $\rho = 0.472$, $p < 0.01$). Among the 36 participants who pass the simple tests in the lottery task, 24 fail at least one mirror list. Among those who fail the simple tests in the lottery task, all also fail the mirror task.

We estimate a series of random-effects models of CEs on payout probability that includes a dummy variable $D_{\text{mirror},i}$, which equals 1 if participant i fails at least one mirror list. Columns (3) and (7) of Table 3 indicate that participants who fail the mirror task tend to exhibit a stronger fourfold pattern in their CE responses. However, when we extend the model to include the simple test dummy D_{ST}, the coefficients associated with D_{mirror} are no longer statistically significant in either domain (compare Columns (3) and (7) with Columns (4) and (8)). This indicates that, when participants pass the simple tests, passing or failing the mirror task does not lead to qualitatively different behavior. By contrast, the coefficients associated with D_{ST} remain essentially unchanged in both magnitude and significance when D_{mirror} is added to the model. This implies that the association between mirror performance and the fourfold pattern in columns (3) and (7) largely reflects the relationship between simple test performance and the fourfold pattern, with the additional layer of computational complexity playing a limited role. This pattern is robust to several

Table 3: Random effects regressions on normalized absolute certainty equivalents

	Gain Domain				*Loss Domain*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p	0.356***	0.607***	0.697***	0.697***	0.342***	0.592***	0.657***	0.657***
	(0.027)	(0.055)	(0.096)	(0.095)	(0.028)	(0.057)	(0.100)	(0.099)
D_{ST}		0.303***		0.266***		0.060		0.041
		(0.050)		(0.059)		(0.054)		(0.064)
$p \times D_{\mathrm{ST}}$		−0.331***		−0.285***		−0.328***		−0.296***
		(0.063)		(0.074)		(0.065)		(0.077)
D_{mirror}			0.333***	0.113			0.091	0.057
			(0.081)	(0.093)			(0.087)	(0.101)
$p \times D_{\mathrm{mirror}}$			−0.371***	−0.136			−0.342**	−0.098
			(0.100)	(0.116)			(0.104)	(0.121)
Intercept	0.354***	0.123**	0.048	0.048	0.364***	0.318***	0.280***	0.280***
	(0.022)	(0.044)	(0.078)	(0.076)	(0.024)	(0.047)	(0.083)	(0.082)
Observations	750	750	750	750	750	750	750	750

Note: The dependent variable is the normalized absolute certainty equivalent, $|CE_{ij}|/25$, for lottery j of participant i. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. $D_{\mathrm{ST}} = 1$ indicates participants who failed the simplicity tests; $D_{\mathrm{mirror}} = 1$ indicates participants who failed the comprehension questions. Standard errors in parentheses. Columns (1–4) report gain domain results; columns (5–8) report loss domain results. All models use random effects with participant-level clustering.

alternative specifications.[6] Together, these findings suggest that the behavioral tendency captured by our simple tests is an important driver of the fourfold pattern across both tasks, and computation complexity of lottery played a limited role in the fourfold pattern.

3.3 Additional Experiment—Cognitive Uncertainty

Observation 3A. *The link between likelihood insensitivity and failure of the simple tests holds in both the high- and low-cognitive-uncertainty subsamples.*

We examine to what extent CU captures the behavioral tendency underlying the failure of the simple tests. Following Enke and Graeber (2023), we divide all risky choices into two groups: those with CU below the overall median (low-CU choices) and those with CU above the median (high-CU choices). Within each subsample, we plot deviations of lotteries' CEs from expected value for participants who pass versus fail the simple tests in Figures 5a and 5b.[7] For high-CU choices (Figure 5a), participants who fail the simple tests display

[6]These results are stable in robustness checks where we replace the dummy variable D_{mirror} with the continuous SE of the corresponding mirror whose payout proportion equals the lottery's probability, and also in Tobit models. See Tables B.5 and B.6.

[7]We again replicate the key findings from the main experiment. To avoid duplication, we report the full set of figures and tables in Appendix C and focus here on how the relationship between simple test

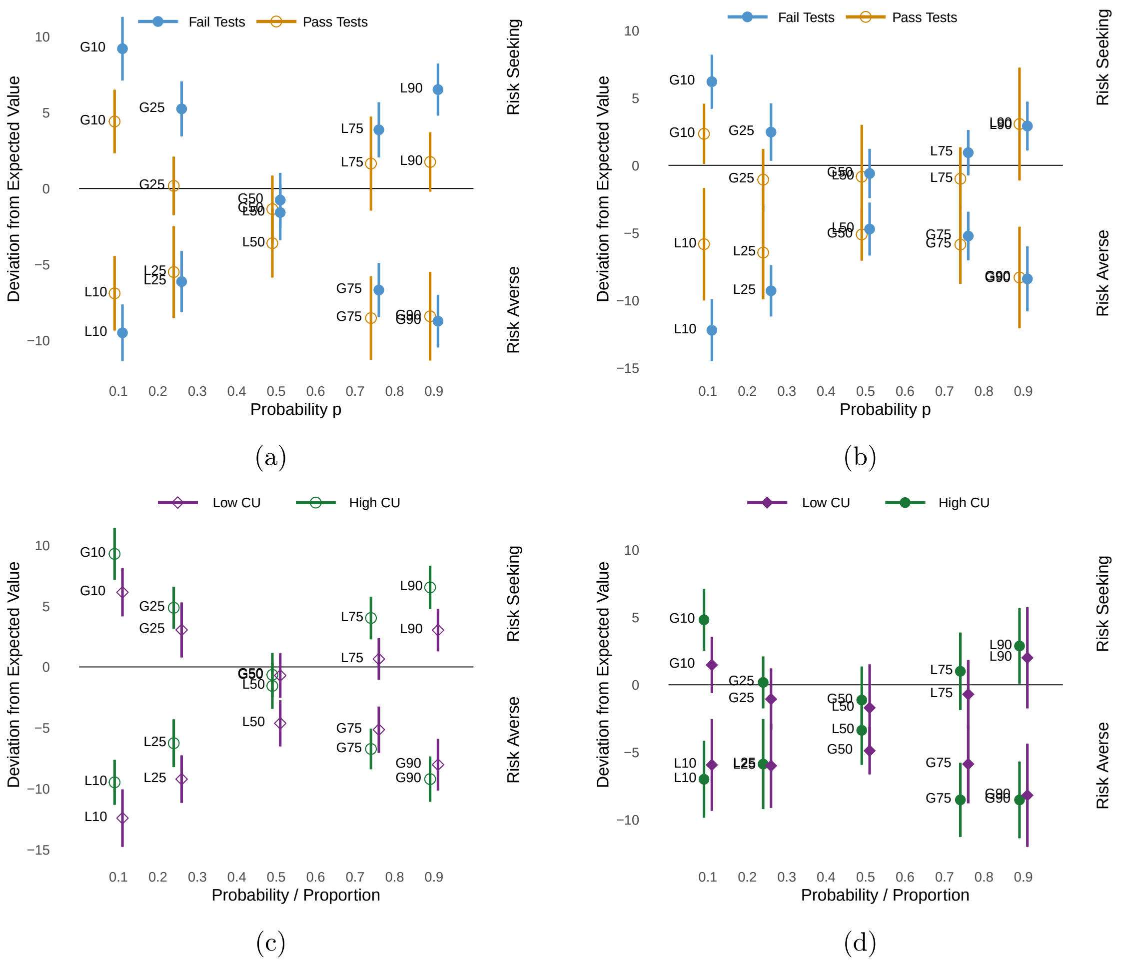


Figure 5: fourfold pattern comparisons by cognitive uncertainty and simple test performance. *Notes:* Panel (a) plots deviations of lotteries' CEs from expected value for high-CU choices, separately for participants who pass and who fail the simple tests. Panel (b) plots the corresponding deviations for low-CU choices. Panel (c) compares high- and low-CU choices among participants who fail the simple tests. Panel (d) makes the same comparison among participants who pass the simple tests. Error bars correspond to ±2 standard errors. See Appendix Tables C.1 (Panel a), C.2 (Panel b), C.3 (Panel c) and C.4 (Panel d) for corresponding summary statistics.

a markedly stronger fourfold pattern than those who pass: they are more risk seeking for low probabilities and more risk averse for high probabilities, reflecting stronger likelihood insensitivity. For low-CU choices (Figure 5b), the difference between the two performance groups is weaker but still visible, especially for the lists involving small probabilities.

We can also flip the comparison and ask how CU moderates the fourfold pattern within each of the two simple test performance groups. As illustrated in Figure 5c, among par-

performance and risky choice patterns varies across CU levels.

ticipants who fail the simple tests there are pronounced "compression effects": high-CU choices exhibit greater risk tolerance for small probabilities and greater risk aversion for large probabilities than low-CU choices in the gain domain, with the effects flipped in the loss domain (Enke and Graeber, 2023; Enke et al., 2024). By contrast, among participants who pass the simple tests (Figure 5d), CU has very limited predictive power for risky choices: risk attitudes are essentially indistinguishable across high- and low-CU choices. Because failing the simple tests reflects limited understanding of the decision environment rather than of the lotteries themselves, this conditional pattern suggests that the cognition of lottery valuation is not the primary driver of the relationship between CU and likelihood insensitivity.

To supplement this graphical evidence, we estimate random-effects regressions analogous to those in the main experiment, but now include a dummy variable $CU_{\text{high},ij}$, which equals 1 for high-CU choices and equal 0 for low-CU choices. Table 4 reports the results. Columns (1)–(2) and (5)–(6) replicate the core patterns in the main experiment and the deterministic mirror experiment: failing the simple tests is associated with greater likelihood insensitivity, which is consistent with stronger fourfold patterns. Columns (3) and (7) show that high-CU choices are less sensitive to changes in payout probability than low-CU choices, consistent with the "compression effects" visible in Figures 5a and 5b. Most importantly, when we include both $D_{\text{ST},i}$ and $CU_{\text{high},ij}$ (and their interactions with p_{ij}) in the regression, the coefficients on the simple test dummy and its interaction change very little relative to the specifications without CU. This indicates that the link between likelihood insensitivity and failure of the simple tests is robust to controlling for CU. These conclusions remain essentially unchanged when we replace the CU dummy with the exact CU measure or when we estimate Tobit models instead of random-effects linear models (see Tables C.7 and C.8).

Observation 3B. *Cognitive uncertainty about lotteries reflects both valuation difficulty and broader environmental factors, with the latter playing a relatively more important role in participants' risky choices.*

We include CU questions for each of the standard and the test lottery, and examine to the extent they are linked to four-fold pattern. We find that cognitive uncertainty (CU) for standard lotteries correlates significantly with CU for the test lotteries (Spearman's $\rho =$

Table 4: Random effects regressions on normalized absolute certainty equivalents

	Gain Domain				*Loss Domain*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p	0.236***	0.424***	0.309***	0.482***	0.271***	0.532***	0.305***	0.566***
	(0.025)	(0.052)	(0.037)	(0.057)	(0.025)	(0.052)	(0.036)	(0.059)
D_{ST}		0.221***		0.213***		0.188***		0.183***
		(0.051)		(0.051)		(0.053)		(0.052)
$p \times D_{\mathrm{ST}}$		−0.241***		−0.230***		−0.337***		−0.328***
		(0.059)		(0.059)		(0.059)		(0.059)
CU_{high}			0.109***	0.101**			−0.032	−0.015
			(0.033)	(0.032)			(0.034)	(0.033)
$p \times CU_{\mathrm{high}}$			−0.148**	−0.136**			−0.070	−0.083
			(0.053)	(0.052)			(0.053)	(0.052)
Intercept	0.331***	0.159***	0.278***	0.116*	0.479***	0.333***	0.495***	0.345***
	(0.022)	(0.045)	(0.027)	(0.047)	(0.022)	(0.047)	(0.027)	(0.050)
Observations	740	740	740	740	740	740	740	740

Note: Dependent variable = normalized absolute certainty equivalent. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. $D_{\mathrm{ST}} = 1$ indicates participants who failed the simplicity tests; $CU_{\mathrm{high}} = 1$ indicates participant i's cognitive uncertainty is higher than the median level of cognitive uncertainty for list j. Standard errors in parentheses. Columns (1–4) report gain domain results; columns (5–8) report loss domain results. All models use random effects with participant-level clustering.

0.576, $p < 0.01$). Interestingly, this correlation is strong among participants who fail the simple tests (Spearman's $\rho = 0.744$, $p < 0.01$) but not among those who pass (Spearman's $\rho = 0.064$, $p > 0.10$). Since the test lotteries are straightforward to evaluate, this result suggests that CU for lotteries can arise from factors other than valuation difficulty. For participants who pass the simple tests, CU reflects the inherent difficulty of valuing the lotteries, whereas for participants who fail the simple tests, CU primarily arises from factors beyond valuation difficulty, consistent with our earlier findings on the deterministic mirrors (see Table C.5 in Appendix C for additional evidence).

To gain deeper insight into the relative influence of the two types of cognition—one stemming from valuation and the other from broader contextual factors—we estimate a series of random effects regressions of normalized absolute certainty equivalents, $\frac{|CE_{ij}|}{25}$, on two individual-level CU dummy variables and their interaction terms with the payout probability. The first dummy indicates whether participant i's average CU on the risk-lottery lists exceeds the aggregate median level (i.e., $CU^{\mathrm{risk}}_{\mathrm{high},ij} = 1$), while the second captures the same condition for the test lists ($CU^{\mathrm{test}}_{\mathrm{high},ij}$). The regression results are reported in Table 5.

First, by comparing Columns (1) and (4) with Columns (2) and (5), we find that the

Table 5: Random effects regressions on normalized absolute certainty equivalents

	Gain Domain			*Loss Domain*		
	(1)	(2)	(3)	(4)	(5)	(6)
p	0.315***	0.269***	0.312***	0.360***	0.359***	0.393***
	(0.032)	(0.036)	(0.037)	(0.032)	(0.036)	(0.037)
$CU^{\text{test}}_{\text{high}}$	0.180***		0.198***	0.044		0.038
	(0.044)		(0.047)	(0.045)		(0.035)
$p \times CU^{\text{test}}_{\text{high}}$	−0.201***		−0.204***	−0.227***		−0.188***
	(0.051)		(0.054)	(0.051)		(0.055)
$CU^{\text{risk}}_{\text{high}}$		0.025	−0.046		0.027	0.013
		(0.043)	(0.046)		(0.044)	(0.047)
$p \times CU^{\text{risk}}_{\text{high}}$		−0.064	0.009		−0.171***	−0.104
		(0.050)	(0.053)		(0.050)	(0.053)
Intercept	0.260***	0.318***	0.277***	0.462***	0.466***	0.458***
	(0.027)	(0.031)	(0.032)	(0.028)	(0.031)	(0.033)
Observations	740	740	740	740	740	740

Note: Dependent variable = normalized absolute certainty equivalent. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. $CU^{\text{test}}_{\text{high}} = 1$ indicates participant i's test-list cognitive uncertainty average is higher than the group median level; $CU^{\text{risk}}_{\text{high}} = 1$ indicates participant i's risk-lottery-list cognitive uncertainty average is higher than the group median level. Standard errors in parentheses. All models use random effects with participant-level clustering.

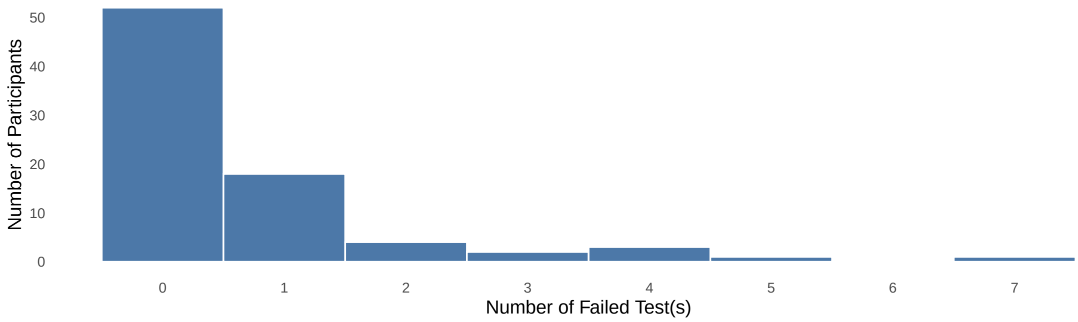
CU dummy for the test lists accounts for more variation in individual risky choices than the dummy for the risk-lottery lists. Notably, Column (2) even suggests a null correlation between CU from the risk-lottery lists and risky choices. Furthermore, when both dummies and their interaction terms with probability are included in the regression (Columns (3) and (6)), the coefficients associated with $CU^{\text{test}}_{\text{high},ij}$ remain largely unchanged compared to specifications (1) and (4), whereas the coefficients related to $CU^{\text{risk}}_{\text{high},ij}$ lose statistical significance. This result indicates that CU beyond lottery valuation contributes to the four-fold pattern, while the additional layer of CU driven by the lottery appears to be less significant.[8]

3.4 Additional Experiment—Student Sample

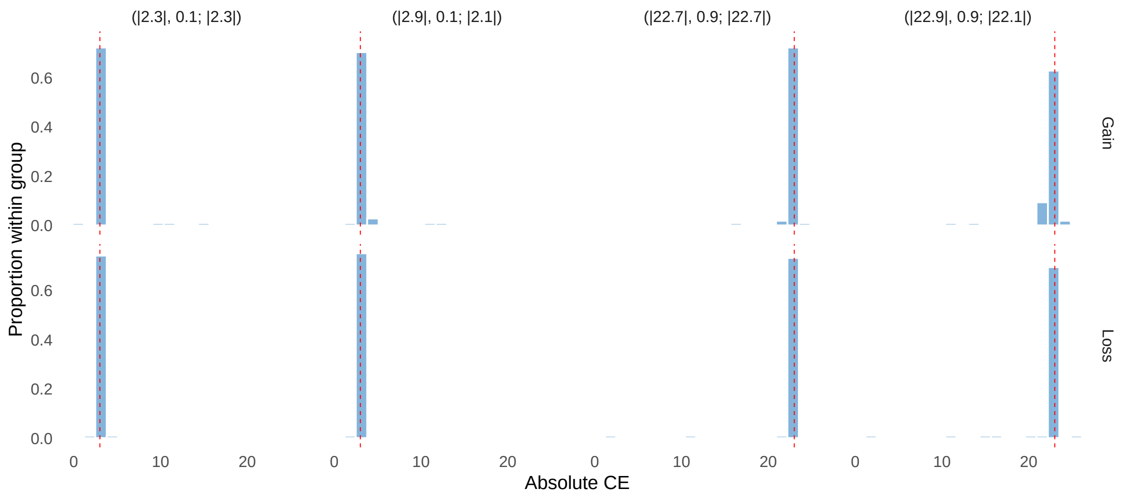
Observation 4A. *More than one third of student participants fail the simple tests.*

Relative to the Prolific sample, the student participants perform substantially better on the simple tests. As shown in Figure 6, a substantially higher proportion of students pass all tests (52 out of 81 in the student sample, versus 21 out of 150 in the Prolific

[8]We also replace the CU dummies with individual median CU values for the two types of lists, and the main findings remain robust. See Table C.6 for further details.



(a)

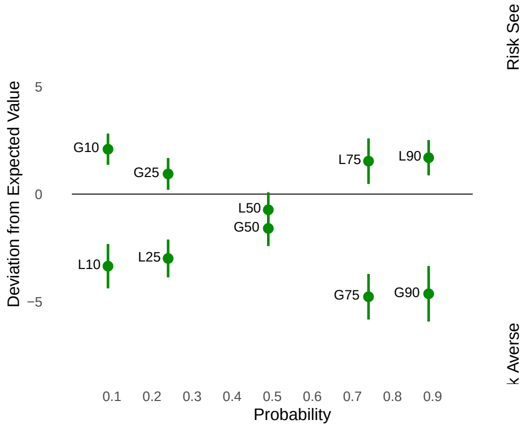


(b)

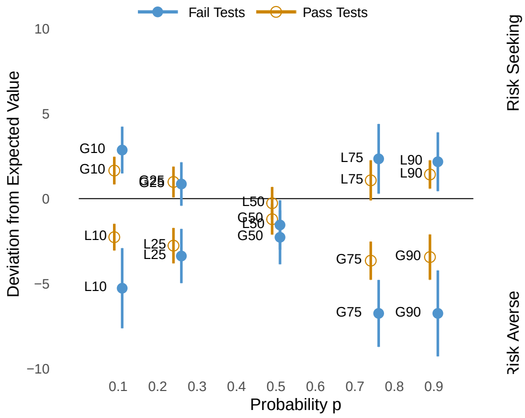
Figure 6: Performance on simple tests. *Notes:* Panel (a) plots the distribution of individual CEs—i.e., switching positions—across the eight simple test lists, separately for gain and loss domains. Red dashed lines indicate the theoretically correct switching position for each list. Panel (b) shows the distribution of the number of simple tests failed by each participant. The sample consists of 81 participants.

sample). Among those who fail at least one test, many students fail only one (18 out of 29 in the student sample, compared to 12 out of 129 in the Prolific sample in the main experiment). Figure D.1 in the Appendix presents histograms of simple test performance for both samples, illustrating this difference. This result is consistent with Snowberg and Yariv (2021), who compare choice behavior across subject pools (a student population, a representative U.S. sample, and Amazon Mechanical Turk subjects) and find lower noise in the student sample.

Observation 4B. *The fourfold pattern is significantly stronger among student participants who fail the simple tests than among those who pass. This pattern is robust with controlling comprehension question performance.*



(a)



(b)

Figure 7: fourfold pattern of risk attitudes and simple test performance (student sample). *Notes:* Panel (a) plots the fourfold pattern of risk attitudes for all participants. Panel (b) plots the same pattern separately for participants who pass and who fail the simple tests. In both panels, the x-axis is the payout probability p in the lotteries (\$25, p; \$0) and (\$ − 25, p; \$0), and the y-axis is the difference between the CE and the expected value, with positive values indicating risk seeking and negative values indicating risk aversion. Two–standard error bars are included for every task. Appendix Table D.1 reports the corresponding summary statistics.

Figure 7 presents the fourfold pattern for the pooled participants and its comparison between two test performance groups. In the left panel, the fourfold pattern is significant, consistent with previous similar studies with a student population (e.g., Bruhin et al., 2010; L'Haridon and Vieider, 2019). In the right panel, we split the sample into two groups based on their simple test performance. Again, student participants who fail the simple tests exhibit a stronger fourfold pattern than those who pass, broadly consistent with those from the main experiment, although the differences between the two groups in student participants are somewhat weaker. This smaller difference likely reflects the fact that many student participants fail only one simple test, as noted above.

Finally, Table 6 presents random-effects regressions. Results suggest that the fourfold pattern is stronger among participants who fail the simple tests, and the relationship between the simple tests performance and the fourfold pattern is robust to the inclusion of comprehension performance, similar to the main experiment (see related robustness checks in Appendix E).

Table 6: Random effects regressions on normalized absolute certainty equivalents

	Gain Domain				*Loss Domain*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p	0.619***	0.701***	0.617***	0.694***	0.706***	0.770***	0.713***	0.770***
	(0.024)	(0.029)	(0.026)	(0.029)	(0.023)	(0.028)	(0.025)	(0.029)
D_{ST}		0.064		0.069		0.112**		0.120***
		(0.036)		(0.036)		(0.035)		(0.036)
$p \times D_{\text{ST}}$		−0.230***		−0.241***		−0.177***		−0.177***
		(0.048)		(0.049)		(0.047)		(0.048)
$D_{Compreh}$			−0.016	−0.034			−0.024	−0.056
			(0.049)	(0.049)			(0.048)	(0.048)
$p \times D_{Compreh}$			0.014	0.077			−0.047	0.000
			(0.067)	(0.066)			(0.064)	(0.064)
Intercept	0.127***	0.104***	0.129***	0.107***	0.178***	0.138***	0.181***	0.143***
	(0.018)	(0.021)	(0.019)	(0.022)	(0.017)	(0.021)	(0.018)	(0.022)
Observations	405	405	405	405	405	405	405	405

Note: Dependent variable = $|CE|/25$ (normalized absolute certainty equivalent). $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. $D_{\text{ST}} = 1$ indicates participants who failed the simplicity tests; $D_{Compreh} = 1$ indicates participants who failed the comprehension questions. Standard errors in parentheses. The first four columns (1–4) report results for the gain domain; the last four columns (5–8) report results for the loss domain. All models use random effects with participant-level clustering.

4 Conclusion and Discussions

We introduce a test for decisions under risk: asking participants to value (almost) degenerate lotteries. This test is a model-agnostic diagnostic, as valuations should converge to the degenerate value under alternative approaches of decisions under risk. Across experiments, the majority of participants fail this test, and the failure is systematically linked to stronger likelihood insensitivity, even after controlling for performance in comprehension questions, behavior in the deterministic mirror task, and the degree of cognitive uncertainty. This link attenuates but persists in a university student sample relative to the online convenience sample in Prolific. Together, these findings show that our test captures a pervasive, choice-relevant component of cognition. Below, we provide some discussions of our findings.

Cognition in decision environments. Going beyond lottery-specific comprehension and computation, our findings thus point to various sources of cognition in the decision environments.

For example, preference-elicitation mechanisms such as the Becker–DeGroot–Marschak (BDM) mechanism can be cognitively demanding and cause confusion (Plott and Zeiler, 2005, 2007; Cason and Plott, 2014). Much of the methodology literature emphasizes giving clear mechanism descriptions (Gonczarowski et al., 2023), reducing misconceptions through trainings (Plott and Zeiler, 2005, 2007), improving instructions (Freeman et al., 2018), and adopting characterization assessment method (Bernheim et al., 2025). In our experiments, we use price-list elicitation, which is arguably simpler than the BDM mechanism. Moreover, we include both examples and comprehension questions about the operation of the price lists, and we require participants to answer all comprehension questions correctly before proceeding to the price lists. Nevertheless, performance in comprehension question does not appear to be linked to our simple test.

Furthermore, our findings may be partially influenced by inattention of participants. Since the experiments typically involve low stakes, participants may not feel motivated to engage fully, resulting in limited cognitive effort. This is consistent with the different passing rates of the test between the Prolific versus student samples, even though our experimental compensation complies with the hourly wages in both samples. It also echoes the observation in Snowberg and Yariv (2021) that, while behavioral correlations are broadly similar across pools, but noise tends to be higher in online samples than in student samples.

Cognition arising from the choice environment may be mitigated by simplifying elicitation procedures or strengthening participants' incentives and motivation. For example, we would expect fewer test violations under binary-choice elicitation, which is arguably less complex than price-list formats, or when stake sizes are substantially increased (Gneezy et al., 2024). Exploring how different sources of cognition interact remains an important direction for future research.

Heuristics and Noise. When participants confront their cognitive constraints in relation to the overall choice environment, they may adopt heuristics or make random decisions (Simon, 1955; Tversky and Kahneman, 1974; Gigerenzer et al., 2000).[9] A prominent heuristic under risk is middle bias - the tendency to switch in the middle of price lists or to select the

[9]Cognitive constraints may induce the use of heuristics, such as the status quo bias (Masatlioglu and Ok, 2005; Ortoleva, 2010), simplicity seeking (Iyengar and Kamenica, 2010), caution (Cerreia-Vioglio et al., 2015; Gilboa et al., 2021; Cerreia-Vioglio et al., 2024), and mental accounting (Gilboa et al., 2021). The Cognitive difficulty in thinking through uncertainty may lead to suboptimal decisions (see Shafir and Tversky, 1992; Charness and Levin, 2009; Martínez-Marquina et al., 2019).

midpoint in budgetary settings (Read and Loewenstein, 1995; Benartzi and Thaler, 2001; Rubinstein, 2002; Choi et al., 2006; Enke and Graeber, 2023; Halevy and Mayraz, 2024). Individuals may also make noisy decisions by randomizing between options, deliberately or otherwise (Machina, 1985; Hey, 2005; Agranov and Ortoleva, 2017; Cettolin and Riedl, 2019; Halevy et al., 2023). In the context of price lists, both middle bias and random decisions can generate likelihood insensitivity.

When we examine the distribution of CEs for (almost) degenerate lotteries, we find that participants' CEs center around the correct value not bias toward the middle of the price list. This pattern is not fully consistent with the strong form of middle bias in which participants switch right around the middle of the price list. Instead, our finding indicates a milder form of middle bias, where participants anchor on the middle of the price list and adjust insufficiently toward the correct value.Future studies should shed more light on the distinction between heuristics and noise and their implications for likelihood insensitivity and related anomalies.[10]

In summary, we present a model-agnostic, simple test for decisions under risk in which participants value (almost) degenerate lotteries. Our findings emphasize the importance understanding whether and how different cognitive factors interact with decision environments, along with the potential behavioral consequences in relation to preferences, heuristics, and noise.

[10]To explore this issue further, we conduct additional analyses to compare behavioral differences between participants who pass the test and those who fail. First, we compare choices in price lists and the associated binary choices at the end of the experiment and find less preference reversals among participants who pass the test. Second, we assess test–retest reliability between duplicate price lists administered during the experiment and find that participants who pass the test exhibit higher level of consistency in repeated price lists. These results are detailed in Tables E.5, E.6 and E.7.

References

Agranov, M. and Ortoleva, P. (2017). Stochastic choice and preferences for randomization. *Journal of Political Economy*, 125(1):40–68.

Benartzi, S. and Thaler, R. H. (2001). Naive diversification strategies in defined contribution saving plans. *American Economic Review*, 91(1):79–98.

Bernheim, D., Lucia, A., Nielsen, K., and Sprenger, C. (2025). When are decisions improvable? an evaluation of diagnostic methods. *Working Paper*.

Bordalo, P., Gennaioli, N., Ma, Y., and Shleifer, A. (2020). Overreaction in macroeconomic expectations. *American Economic Review*, 110(9):2748–2782.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly journal of economics*, 127(3):1243–1285.

Bruhin, A., Fehr-Duda, H., and Epper, T. (2010). Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica*, 78:1375–1412.

Cason, T. N. and Plott, C. R. (2014). Misconceptions and game form recognition: Challenges to theories of revealed preference and framing. *Journal of Political Economy*, 122(6):1235–1270.

Cerreia-Vioglio, S., Dillenberger, D., and ortoleva, P. (2015). Cautious Expected Utility and the Certainty Effect. *Econometrica*, 83:693–728.

Cerreia-Vioglio, S., Dillenberger, D., and Ortoleva, P. (2024). Caution and Reference Effects. *Econometrica*, 92(6):2069–2103.

Cettolin, E. and Riedl, A. (2019). Revealed preferences under uncertainty: Incomplete preferences and preferences for randomization. *Journal of Economic Theory*, 181:547–585.

Charness, G. and Levin, D. (2009). The origin of the winner's curse: A laboratory study. *American Economic Journal: Microeconomics*, 1(1):207–236.

Chew, S. H. (1983). A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the allais paradox. *Econometrica: Journal of the Econometric Society*, pages 1065–1092.

Choi, S., Fisman, R., Gale, D., and Kariv, S. (2006). Substantive and procedural rationality in decisions under uncertainty.

de Clippel, G. and Rozen, K. (2024). Bounded rationality in choice theory: A survey. *Journal of economic literature*, 62(3):995–1039.

Enke, B. (2024). The cognitive turn in behavioral economics. *Working Paper*.

Enke, B. and Graeber, T. (2023). Cognitive uncertainty. *The Quarterly Journal of Economics*, 138(4):2021–2067.

Enke, B., Graeber, T., Oprea, R., and Yang, J. (2024). Behavioral attenuation. NBER Working Paper 32973, National Bureau of Economic Research.

Esponda, I. and Vespa, E. (2024). Contingent thinking and the sure-thing principle: Revisiting classic anomalies in the laboratory. *Review of Economic Studies*, 91(5):2806–2831.

Freeman, D. J., Kimbrough, E. O., Petersen, G. M., and Tong, H. T. (2018). Instructions. *Journal of the Economic Science Association*, 4(2):165–179.

Frydman, C. and Jin, L. J. (2023). On the source and instability of probability weighting. Working Paper.

Gabaix, X. (2019). Behavioral inattention. In *Handbook of behavioral economics: Applications and foundations 1*, volume 2, pages 261–343. Elsevier.

Gigerenzer, G., Todd, P. M., ABC Research Group, t., et al. (2000). *Simple heuristics that make us smart*. Oxford University Press.

Gilboa, I., Postlewaite, A., and Schmeidler, D. (2021). The complexity of the consumer problem. *Research in Economics*, 75(1):96–103.

Gneezy, U., Halevy, Y., Hall, B., Offerman, T., and van de Ven, J. (2024). How real is hypothetical? a high-stakes test of the allais paradox. Technical report.

Gonczarowski, Y. A., Heffetz, O., and Thomas, C. (2023). Strategyproofness-exposing mechanism descriptions. Technical report, National Bureau of Economic Research.

Gul, F. (1991). A theory of disappointment aversion. *Econometrica*, 59(3):667.

Halevy, Y. and Mayraz, G. (2024). Identifying rule-based rationality. *Review of Economics and Statistics*, 106(5):1369–1380.

Halevy, Y., Walker-Jones, D., and Zrill, L. (2023). Difficult decisions. Technical report, University of Toronto, Department of Economics.

Hey, J. D. (2005). Why we should not be silent about noise. *Experimental Economics*, 8(4):325–345.

Iyengar, S. S. and Kamenica, E. (2010). Choice proliferation, simplicity seeking, and asset allocation. *Journal of Public Economics*, 94(7–8):530–539.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.

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

Khaw, M. W., Li, Z., and Woodford, M. (2023). Cognitive imprecision and stake-dependent risk attitudes. Technical report.

L'Haridon, O. and Vieider, F. M. (2019). All over the map: A worldwide comparison of risk preferences. *Quantitative Economics*, 10(1):185–215.

Loewenstein, G. and Wojtowicz, Z. (2025). The economics of attention. *Journal of Economic Literature*, 63(3):1038–1089.

Machina, M. J. (1985). Stochastic choice functions generated from deterministic preferences over lotteries. *The Economic Journal*, 95(379):575–594.

Maćkowiak, B., Matějka, F., and Wiederholt, M. (2023). Rational inattention: A review. *Journal of Economic Literature*, 61(1):226–273.

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

Masatlioglu, Y. and Ok, E. A. (2005). Rational choice with status quo bias. *Journal of economic theory*, 121(1):1–29.

Masatlioglu, Y. and Raymond, C. (2016). A behavioral analysis of stochastic reference dependence. *American Economic Review*, 106(9):2760–2782.

Niederle, M. and Vespa, E. (2023). Cognitive limitations: Failures of contingent thinking. *Annual Review of Economics*, 15(1):307–328.

Oprea, R. (2024). Decisions under risk are decisions under complexity. *American Economic Review*, pages 3789–3811.

Oprea, R. (2025). Complexity and its measurement.

Ortoleva, P. (2010). Status quo bias, multiple priors and uncertainty aversion. *Games and Economic Behavior*, 69(2):411–424.

Plott, C. R. and Zeiler, K. (2005). The willingness to pay–willingness to accept gap, the "endowment effect," subject misconceptions, and experimental procedures for eliciting valuations. *American Economic Review*, 95(3):530–545.

Plott, C. R. and Zeiler, K. (2007). Exchange asymmetries incorrectly interpreted as evidence of endowment effect theory and prospect theory? *American Economic Review*, 97(4):1449–1466.

Puri, I. (2025). Simplicity and risk. *The Journal of Finance*, 80(2):1029–1080.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of economic behavior & organization*, 3(4):323–343.

Read, D. and Loewenstein, G. (1995). Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. *Journal of Experimental Psychology: Applied*, 1(1):34–49.

Rubinstein, M. (2002). Markowitz's "portfolio selection": A fifty-year retrospective. *Journal of Finance*, 57(3):1041–1045.

Shafir, E. and Tversky, A. (1992). Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology*, 24(4):449–474.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, pages 99–118.

Snowberg, E. and Yariv, L. (2021). Testing the waters: Behavior across participant pools. *American Economic Review*, 111(2):687–719.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of economic literature*, 38(2):332–382.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157):1124–1131.

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

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

Wakker, P. P. (2025). The history of prospect theory for risk and its future for ambiguity: Interactions between economists and psychologists.

Woodford, M. (2020). Modeling imprecision in perception, valuation, and choice. *Annual Review of Economics*, 12(1):579–601.

Appendices

A Appendix: Additional results in the main experiment

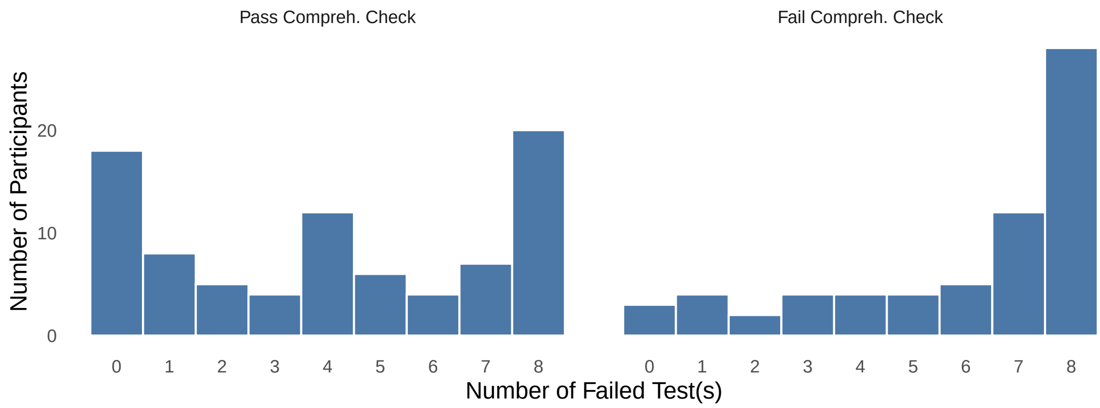


Figure A.1: Participants' performance in the simple tests (the number of failed tests).

Duration and payment. The median duration of the experiment was 35.18 minutes, and participants received \$5 as a participation fee. Participants were additionally rewarded for correctly answering the comprehension questions on their first attempt, with an average reward of \$0.779. Moreover, 17 participants were randomly selected from the total of 150 to receive a bonus based on one of their choices in the experiment, with an average amount of \$18.176. In total, the average payment per participant was \$7.839. This payment rate aligns with Prolific's recommended payment rate of \$12/hour.

Observation A.1. *Most participants fail at least one simple test (129 out of 150), and among those who fail at least one, many (86 out of 129) fail more than 50% of the simple tests. When they fail the simple tests, their implied valuations are distributed around the correct values. Participants' performance in the comprehension checks on their first attempt has a weak correspondence with their performance in the simple tests.*

We compare participants' performance in the simple tests and in the comprehension checks on their first attempt. We find that more than 50% (84 out of 150) participants answered all comprehension questions correctly on their first attempt. However, participants' performance in comprehension checks is a weak predictor of their performance in the simple tests: Among the 84 participants who answered their comprehension questions

Table A.1: Certainty equivalent deviations by gain domain and simplicity test performance.

P	ST	Label	N	Mean	Median	T1	W1	T2	W2
				Panel A: Gain domain					
10	0	G10	21	3.786	2.5	0.054	0.049	0.001	0.002
10	1	G10	129	10.105	8.5	0.000	0.000		
25	0	G25	21	4.548	4.5	0.005	0.012	0.000	0.000
25	1	G25	129	11.453	9.5	0.000	0.000		
50	0	G50	21	7.214	5.5	0.000	0.000	0.005	0.007
50	1	G50	129	12.205	12.5	0.000	0.000		
75	0	G75	21	10.786	9.5	0.019	0.031	0.139	0.103
75	1	G75	129	13.500	14.5	0.000	0.000		
90	0	G90	21	13.214	11.5	0.000	0.002	0.445	0.316
90	1	G90	129	14.655	14.5	0.000	0.000		
				Panel B: Loss domain					
10	0	L10	21	-10.881	-10.5	0.000	0.229	0.864	0.989
10	1	L10	129	-11.205	-10.5	0.000	0.162		
25	0	L25	21	-13.024	-12.5	0.000	0.137	0.553	0.469
25	1	L25	129	-11.958	-10.5	0.000	0.172		
50	0	L50	21	-16.214	-15.5	0.001	0.001	0.189	0.194
50	1	L50	129	-13.896	-14.5	0.046	0.003		
75	0	L75	21	-18.071	-19.5	0.689	0.878	0.031	0.047
75	1	L75	129	-14.136	-15.5	0.000	0.002		
90	0	L90	21	-20.262	-20.5	0.024	0.002	0.003	0.011
90	1	L90	129	-14.570	-15.5	0.000	0.000		

Notes: T1 = one-sample t-test vs. prediction. W1 = Wilcoxon signed-rank test vs. prediction. T2 = two-sample t-test across simplicity groups. W2 = Wilcoxon rank-sum test across simplicity groups. All values rounded to three decimals. See corresponding Figure 3b.

correctly, 66 (79%) fail the simple tests, while among 66 participants who fail some of the comprehension questions on their first attempt, 63 (95%) fail the simple tests. Figure A.1 displays the comparison of the two distributions of number of participants who fail the simple tests 0, 1, 2, ..., 8 times. As we can see, participants who fail the comprehension questions on average fail the simple tests more frequently, and the difference is statistically significant ($p < 0.05$).

Table A.2: Random effects regressions on normalized absolute certainty equivalents

	Gain Domain				*Loss Domain*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p	0.247***	0.488***	0.290***	0.489***	0.209***	0.502***	0.245***	0.501***
	(0.030)	(0.056)	(0.036)	(0.056)	(0.030)	(0.055)	(0.035)	(0.055)
F_{ST}		0.053***		0.047***		0.007		0.008
		(0.008)		(0.008)		(0.008)		(0.008)
$p \times F_{\mathrm{ST}}$		−0.049***		−0.047***		−0.060***		−0.061***
		(0.010)		(0.010)		(0.010)		(0.010)
$F_{Compreh}$			0.084***	0.045*			0.002	−0.005
			(0.019)	(0.020)			(0.019)	(0.020)
$p \times F_{Compreh}$			−0.053*	−0.014			−0.044	0.007
			(0.024)	(0.026)			(0.024)	(0.025)
Intercept	0.347***	0.088*	0.279***	0.082	0.436***	0.400***	0.434***	0.401***
	(0.024)	(0.043)	(0.028)	(0.043)	(0.024)	(0.044)	(0.028)	(0.044)
Observations	750	750	750	750	750	750	750	750

Note: The dependent variable is the normalized absolute certainty equivalent, $|CE_{ij}|/25$, for lottery j of participant i. p is the lottery's payout probability; F_{ST} indicates the number of simplicity tests participants failed; $F_{Compreh}$ indicates the number of comprehension questions participants failed. Standard errors in parentheses. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. The first four columns (1–4) report results for the gain domain; the last four columns (5–8) report results for the loss domain. All models use random effects with participant-level clustering.

Table A.3: Tobit regressions on normalized absolute certainty equivalents

	Gain Domain				*Loss Domain*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p	0.273***	0.506***	0.342***	0.523***	0.234***	0.466***	0.243***	0.461***
	(0.036)	(0.092)	(0.046)	(0.090)	(0.034)	(0.091)	(0.046)	(0.090)
D_{ST}		0.317***		0.242***		0.042		0.049
		(0.082)		(0.083)		(0.077)		(0.081)
$p \times D_{\mathrm{ST}}$		−0.273***		−0.233**		−0.273***		−0.281***
		(0.100)		(0.100)		(0.098)		(0.098)
$D_{Compreh}$			0.255***	0.215***			−0.015	−0.024
			(0.056)	(0.058)			(0.055)	(0.055)
$p \times D_{Compreh}$			−0.156*	−0.118			−0.023	0.027
			(0.070)	(0.071)			(0.070)	(0.069)
Intercept	0.325***	0.052	0.213***	0.024	0.422***	0.388***	0.431***	0.393***
	(0.029)	(0.076)	(0.037)	(0.074)	(0.027)	(0.071)	(0.037)	(0.073)
Observations	750	750	750	750	750	750	750	750

Note: Dependent variable = $|CE|/25$ (normalized absolute certainty equivalent). Reported estimates are posterior means from Bayesian Tobit models with double censoring (left censoring at 0.02 and right censoring at 1.02). Posterior standard errors are reported in parentheses. ***, **, and * indicate that the 99.9%, 99%, and 95% credible intervals, respectively, exclude zero. $D_{\mathrm{ST}} = 1$ indicates participants who failed the simplicity tests; $D_{Compreh} = 1$ indicates participants who failed the comprehension questions. Posterior standard errors are reported in parentheses. The first four columns (1–4) report results for the gain domain; the last four columns (5–8) report results for the loss domain. All models include participant-level random intercepts.

B Appendix: Additional Results in the Mirror Experiment

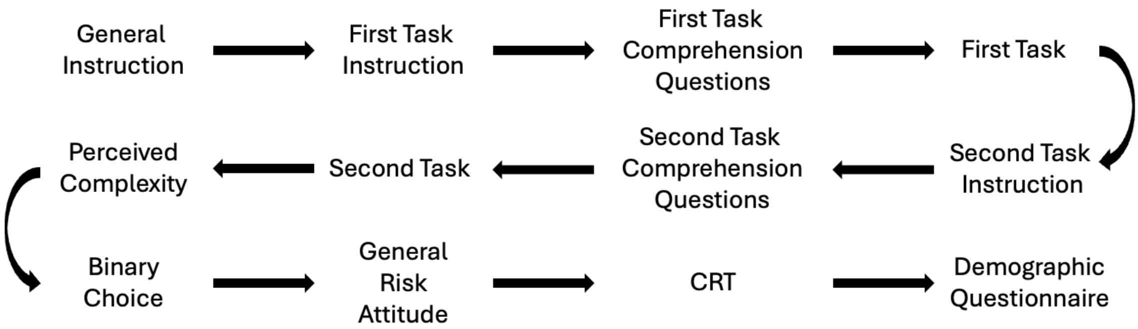
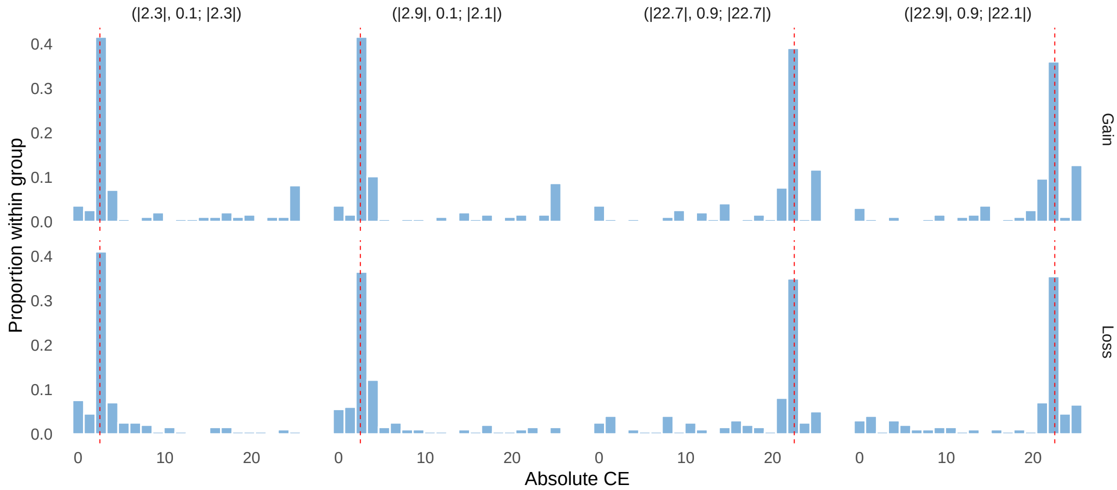


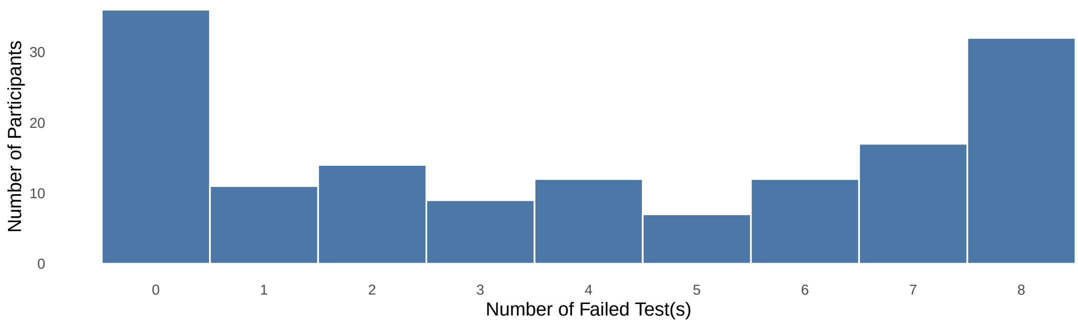
Figure B.1: Procedure of experiment mirror. The lottery task and the mirror task are presented in a randomized order, with one serving as the first task and the other as the second.

Experimental procedure and sample. Figure B.1 depicts the experimental procedure of this additional experiment involving the deterministic mirror task. Participants in this additional experiment ($N = 150$) were on average 38.9 years old (median $= 36.5$, interquartile range $= 28$–48.8), with ages ranging from 18 to 65. Seventy-seven participants (51.3%) were female. In terms of education, 2 participants (1.3%) had completed middle school, 35 (23.3%) had completed high school, 63 (42.0%) held an undergraduate degree, and 50 (33.3%) held a postgraduate degree.

Simplicity performance. Similar to the main experiment, for each simple test, about 30% to 40% of participants switched at the theoretically correct rows, and substantial wrong switches occur around the correct position (See Figure B.2a). Further, as indicated in Figure B.2b, 114 out of 150 participants fail simple tests, and 89 of them fail more than two tests.



(a)



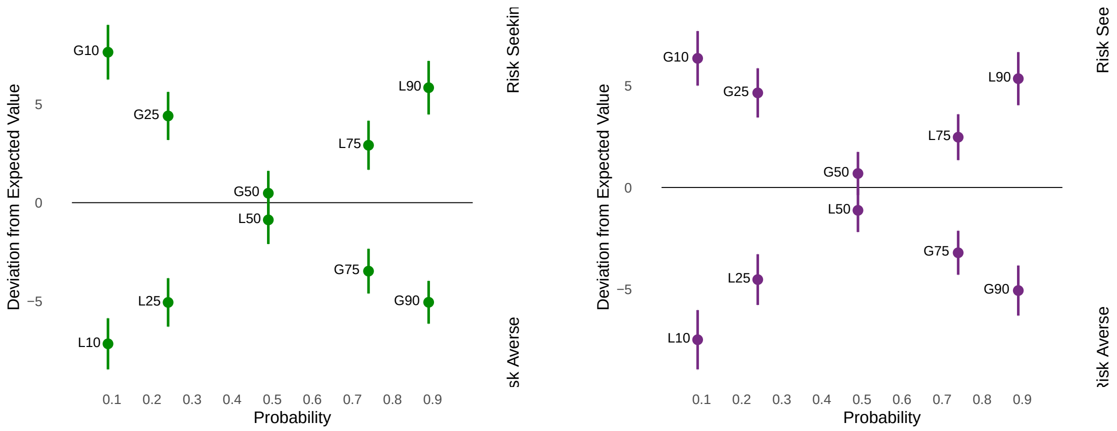
(b)

Figure B.2: Performance on simple tests. *Notes:* Panel (a) plots the distribution of individual CEs—i.e., switching positions—across the eight simple test lists, separately for gain and loss domains. Red dashed lines indicate the theoretically correct switching position for each list. Panel (b) shows the distribution of the number of simple tests failed by each participant. The sample consists of 150 participants.

Table B.1: Certainty equivalent deviations by gain domain and simplicity test performance.

P	ST	Label	N	Mean	Median	T1	W1	T2	W2
Panel A: Gain domain									
10	0	G10	36	4.778	2.5	0.020	0.019	0.000	0.000
10	1	G10	114	11.825	9.5	0.000	0.000		
25	0	G25	36	7.056	6.5	0.474	0.626	0.000	0.000
25	1	G25	114	12.114	10.0	0.000	0.000		
50	0	G50	36	10.500	9.5	0.008	0.001	0.013	0.015
50	1	G50	114	13.763	12.5	0.073	0.101		
75	0	G75	36	13.222	14.5	0.000	0.000	0.074	0.053
75	1	G75	114	15.596	15.5	0.000	0.000		
90	0	G90	36	17.806	19.5	0.000	0.000	0.712	0.901
90	1	G90	114	17.333	19.5	0.000	0.000		
Panel B: Loss domain									
10	0	L10	36	-9.222	-8.0	0.000	0.000	0.706	0.856
10	1	L10	114	-9.798	-6.5	0.000	0.000		
25	0	L25	36	-11.583	-10.5	0.000	0.000	0.983	0.659
25	1	L25	114	-11.553	-10.5	0.000	0.000		
50	0	L50	36	-16.278	-15.5	0.000	0.000	0.007	0.007
50	1	L50	114	-12.456	-12.5	0.953	0.861		
75	0	L75	36	-17.917	-18.5	0.469	0.980	0.035	0.100
75	1	L75	114	-14.851	-15.5	0.000	0.000		
90	0	L90	36	-21.722	-22.5	0.169	0.484	0.000	0.001
90	1	L90	114	-15.070	-15.5	0.000	0.000		

Notes: T1 = one-sample t-test vs. prediction. W1 = Wilcoxon signed-rank test vs. prediction. T2 = two-sample t-test across simplicity groups. W2 = Wilcoxon rank-sum test across simplicity groups. ST = indicator for simple-test failure (1 = fail, 0 = pass). All values rounded to three decimals.



(a) (b)

Figure B.3: fourfold pattern comparison between CE from lottery task (Panel A) and SE from mirror task (Panel B).

Table B.2: Simplicity equivalent deviations by gain domain and mirror simple-test performance.

P	ST	Label	N	Mean	Median	T1	W1	T2	W2
				Panel A: Gain domain					
10	0	G10	31	4.210	2.5	0.040	0.022	0.000	0.000
10	1	G10	119	10.063	6.5	0.000	0.000		
25	0	G25	31	7.790	6.5	0.091	0.091	0.004	0.011
25	1	G25	119	12.029	9.5	0.000	0.000		
50	0	G50	31	11.855	12.5	0.401	0.288	0.202	0.205
50	1	G50	119	13.534	12.5	0.107	0.140		
75	0	G75	31	16.081	18.5	0.008	0.011	0.456	0.674
75	1	G75	119	15.080	17.5	0.000	0.000		
90	0	G90	31	19.597	22.5	0.013	0.012	0.072	0.224
90	1	G90	119	16.861	19.5	0.000	0.000		
				Panel B: Loss domain					
10	0	L10	31	-7.016	-2.5	0.004	0.006	0.037	0.078
10	1	L10	119	-10.769	-8.5	0.000	0.000		
25	0	L25	31	-9.629	-6.5	0.004	0.009	0.252	0.462
25	1	L25	119	-11.399	-8.5	0.000	0.000		
50	0	L50	31	-13.790	-12.5	0.026	0.022	0.876	0.718
50	1	L50	119	-13.584	-12.5	0.102	0.080		
75	0	L75	31	-18.855	-18.5	0.419	0.561	0.010	0.062
75	1	L75	119	-15.290	-18.5	0.000	0.000		
90	0	L90	31	-20.274	-22.5	0.025	0.012	0.014	0.067
90	1	L90	119	-16.332	-20.5	0.000	0.000		

Notes: T1 = one-sample t-test vs. prediction. W1 = Wilcoxon signed-rank test vs. prediction. T2 = two-sample t-test across mirror test groups. W2 = Wilcoxon rank-sum test across mirror test groups. ST = indicator for mirror simple-test failure (1 = fail, 0 = pass). All values rounded to three decimals.

Table B.3: Certainty/simplicity equivalent deviations by gain domain and treatment among who failed simplicity tests.

P	Treatment	Label	N	Mean	Median	T1	W1	T2	W2
				Panel A: Gain domain					
10	mirror	G10	105	11.100	9.5	0.000	0.000	0.441	0.458
10	risk	G10	105	12.014	9.5	0.000	0.000		
25	mirror	G25	105	12.881	11.5	0.000	0.000	0.581	0.666
25	risk	G25	105	12.281	10.5	0.000	0.000		
50	mirror	G50	105	13.938	12.5	0.043	0.056	0.801	0.665
50	risk	G50	105	13.681	12.5	0.115	0.148		
75	mirror	G75	105	14.862	17.5	0.000	0.000	0.400	0.399
75	risk	G75	105	15.700	15.5	0.000	0.001		
90	mirror	G90	105	16.510	17.5	0.000	0.000	0.463	0.641
90	risk	G90	105	17.252	19.5	0.000	0.000		
				Panel B: Loss domain					
10	mirror	L10	105	-10.948	-9.5	0.000	0.000	0.553	0.740
10	risk	L10	105	-10.243	-7.5	0.000	0.000		
25	mirror	L25	105	-11.500	-10.5	0.000	0.000	0.973	0.885
25	risk	L25	105	-11.538	-10.5	0.000	0.000		
50	mirror	L50	105	-13.290	-12.5	0.276	0.239	0.515	0.530
50	risk	L50	105	-12.586	-12.5	0.915	0.715		
75	mirror	L75	105	-14.529	-15.5	0.000	0.000	0.884	0.650
75	risk	L75	105	-14.690	-15.5	0.000	0.000		
90	mirror	L90	105	-15.614	-19.5	0.000	0.000	0.593	0.844
90	risk	L90	105	-14.967	-15.5	0.000	0.000		

Notes: T1 = one-sample t-test vs. prediction. W1 = Wilcoxon signed-rank test vs. prediction. T2 = two-sample t-test across treatments. W2 = Wilcoxon rank-sum test across treatments. Treatment indicates whether the mirror or risk version of the task was used. All values rounded to three decimals.

Table B.4: Certainty/simplicity equivalent deviations by gain domain and treatment among who passed simplicity tests.

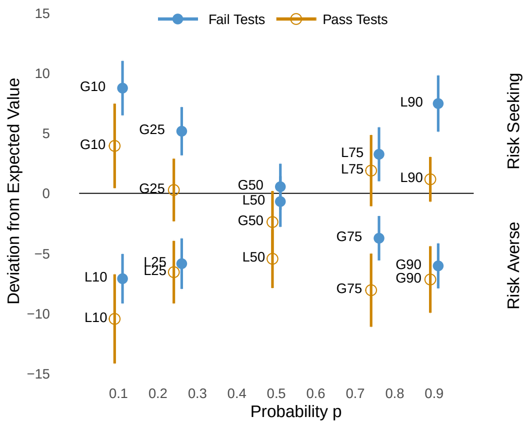
P	Treatment	Label	N	Mean	Median	T1	W1	T2	W2
				Panel A: Gain domain					
10	mirror	G10	22	3.045	2.5	0.168	0.371	0.120	0.081
10	risk	G10	22	4.955	3.0	0.043	0.037		
25	mirror	G25	22	6.364	6.5	0.589	0.789	0.137	0.296
25	risk	G25	22	7.636	6.5	0.171	0.159		
50	mirror	G50	22	11.409	12.5	0.064	0.100	0.505	0.426
50	risk	G50	22	10.818	11.0	0.022	0.008		
75	mirror	G75	22	16.909	18.5	0.087	0.098	0.024	0.011
75	risk	G75	22	13.636	14.5	0.000	0.001		
90	mirror	G90	22	20.318	22.5	0.077	0.057	0.109	0.046
90	risk	G90	22	17.591	18.5	0.000	0.001		
				Panel B: Loss domain					
10	mirror	L10	22	-4.000	-2.5	0.173	0.269	0.008	0.007
10	risk	L10	22	-9.364	-9.0	0.000	0.001		
25	mirror	L25	22	-8.364	-6.5	0.093	0.174	0.113	0.026
25	risk	L25	22	-10.818	-10.5	0.001	0.002		
50	mirror	L50	22	-13.091	-12.5	0.221	0.371	0.005	0.014
50	risk	L50	22	-16.227	-15.5	0.001	0.002		
75	mirror	L75	22	-18.864	-18.5	0.344	0.461	0.375	0.599
75	risk	L75	22	-18.045	-18.5	0.590	0.856		
90	mirror	L90	22	-21.045	-22.5	0.142	0.100	0.680	0.368
90	risk	L90	22	-21.545	-22.5	0.209	0.477		

Notes: T1 = one-sample t-test vs. prediction. W1 = Wilcoxon signed-rank test vs. prediction. T2 = two-sample t-test across treatments. W2 = Wilcoxon rank-sum test across treatments. Treatment indicates whether the mirror or risk version of the task was used. All values rounded to three decimals.

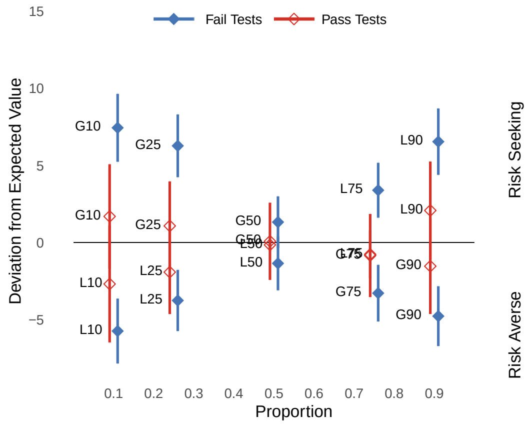
Table B.5: Random effects regressions on normalized absolute certainty equivalents

	Gain Domain				*Loss Domain*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p	0.356***	0.607***	0.352***	0.483***	0.342***	0.592***	0.110	0.323***
	(0.027)	(0.055)	(0.059)	(0.077)	(0.028)	(0.057)	(0.060)	(0.084)
D_{ST}		0.303***		0.212***		0.060		0.028
		(0.050)		(0.051)		(0.054)		(0.053)
$p \times D_{\text{ST}}$		−0.331***		−0.198**		−0.328***		−0.231***
		(0.063)		(0.067)		(0.065)		(0.069)
SE			0.336***	0.256***			0.073	0.050
			(0.057)	(0.060)			(0.056)	(0.056)
$p \times$ SE			−0.181	−0.109			0.282**	0.242**
			(0.094)	(0.095)			(0.092)	(0.091)
Intercept	0.354***	0.123**	0.233***	0.102*	0.364***	0.318***	0.354***	0.339***
	(0.022)	(0.044)	(0.031)	(0.044)	(0.024)	(0.047)	(0.033)	(0.050)
Observations	750	750	750	750	750	750	750	750

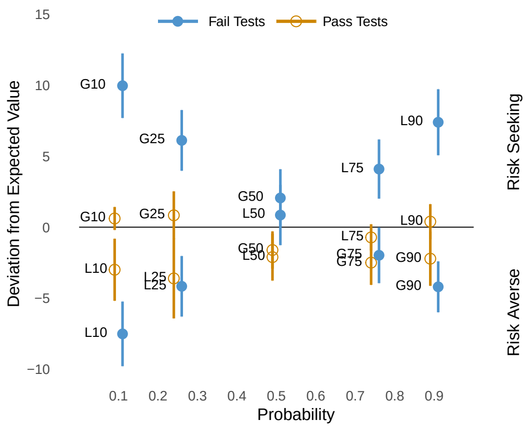
Note: The dependent variable is the normalized absolute certainty equivalent, $|CE_{ij}|/25$, for lottery j of participant i. p is the lottery's payout probability; D_{ST} is a dummy variable (1 = failed simplicity tests, 0 = passed); SE indicates the simplicity equivalents of the corresponding mirror question (percent = proportion). Standard errors in parentheses. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. The first four columns (1–4) report results for the gain domain; the last four columns (5–8) report results for the loss domain. All models use random effects with participant-level clustering.



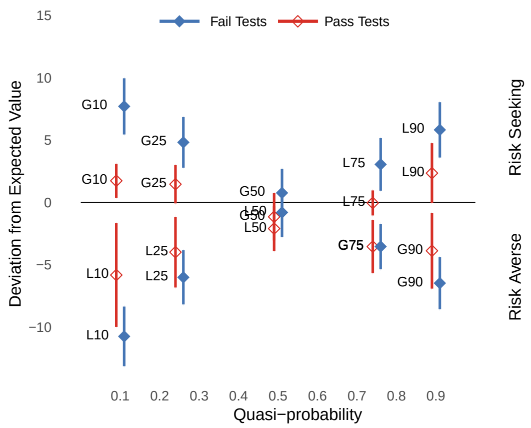
(a) First-answer lottery task response



(b) First-answer mirror task response



(c) Second-answer lottery task response



(d) Second-answer mirror task response

Figure B.4: fourfold pattern comparison in first- and second-answer tasks for participants who pass or fail simple tests. Top row: first-answer tasks; bottom row: second-answer tasks. Left column: lottery task responses; right column: mirror task responses.

Table B.6: Tobit regressions on normalized absolute certainty equivalents

	Gain Domain				*Loss Domain*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p	0.404***	0.766***	0.849***	0.853***	0.379***	0.691***	0.812***	0.816***
	(0.023)	(0.046)	(0.081)	(0.079)	(0.024)	(0.048)	(0.083)	(0.082)
D_{ST}		0.373***		0.357***		0.142***		0.094
		(0.049)		(0.058)		(0.050)		(0.060)
$p \times D_{\mathrm{ST}}$		−0.478***		−0.433***		−0.414***		−0.354***
		(0.053)		(0.063)		(0.055)		(0.064)
D_{mirror}			0.336***	0.045			0.220***	0.147
			(0.080)	(0.091)			(0.081)	(0.094)
$p \times D_{mirror}$			−0.485***	−0.132			−0.473***	−0.186
			(0.085)	(0.098)			(0.087)	(0.099)
Intercept	0.328***	0.044	0.019	0.014	0.351***	0.246***	0.150*	0.146*
	(0.022)	(0.042)	(0.077)	(0.074)	(0.022)	(0.044)	(0.078)	(0.077)
Observations	750	750	750	750	750	750	750	750

Note: Dependent variable = $|CE|/25$ (normalized absolute certainty equivalent). Reported estimates are posterior means from Bayesian Tobit models with left censoring at zero. Posterior standard errors are reported in parentheses. ***, **, and * indicate that the 99.9%, 99%, and 95% credible intervals, respectively, exclude zero. $D_{\mathrm{ST}} = 1$ indicates participants who failed the simplicity tests; $D_{mirror} = 1$ indicates participants who failed the comprehension questions. Columns (1–4) report gain domain results; columns (5–8) report loss domain results. All models include participant-level random intercepts.

C Appendix: Additional results in Experiment CU

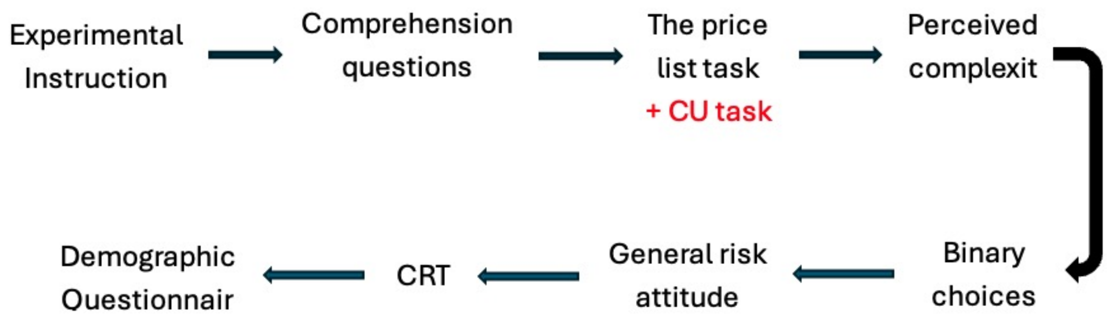


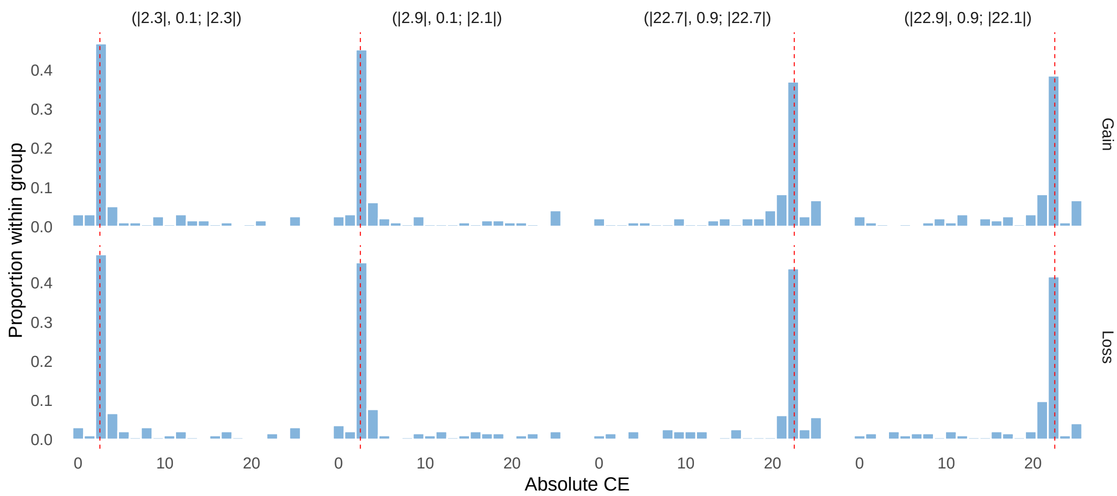
Figure C.1: Procedure of experiment of cognitive uncertainty.

Experimental procedure and sample. Figure C.1 depicts the experimental procedure of this additional experiment involving the cognitive uncertainty task. Participants in this study (N=148) were on average 38.4 years old (median = 34.5, interquartile range = 28–47.3), with ages ranging from 18 to 65. Seventy-three participants (49.3%) were female, and seventy-five (50.7%) were male. In terms of educational attainment, 31 participants (20.9%) had completed high school, 59 (39.9%) held an undergraduate degree, and 58 (39.2%) had completed postgraduate school.

Simplicity performance. We find that 33 out of 148 participants pass all simple tests. Among the remaining 115 participants, 25 fail one test, 18 fail two, and 72 fail three or more. Figure C.2b displays a histogram summarizing participants' performance in the simple tests. All revealed distributions are very close to those in the main experiment and the mirror experiment.

fourfold Pattern. As shown in Figure C.3, the overall risky behavior pattern and the difference across the two groups whether fail or pass the simple tests are essentially same as the main experiment as well as the additional experiment of mirror task.

Cognitive Uncertainty Performance. Figure C.4a displays histograms of each participant's frequency to state positive CU across the ten standard lottery price lists. The vast majority (89.3%) of decisions were associated with strictly positive CU — a rate higher



(a)

(b)

Figure C.2: Performance on simple tests. *Notes:* Panel (a) plots the distribution of individual CEs—i.e., switching positions—across the eight simple test lists, separately for gain and loss domains. Red dashed lines indicate the theoretically correct switching position for each list. Panel (b) shows the distribution of the number of simple tests failed by each participant. The sample consists of 148 participants.

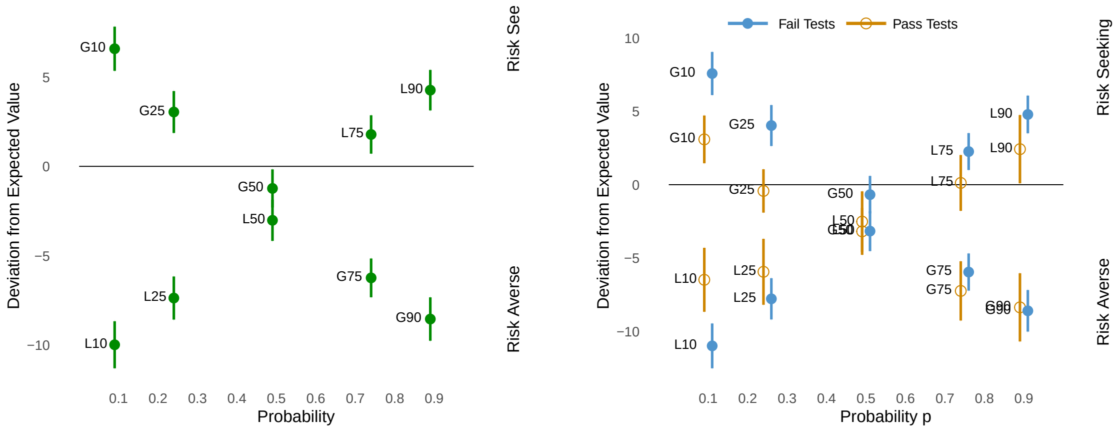
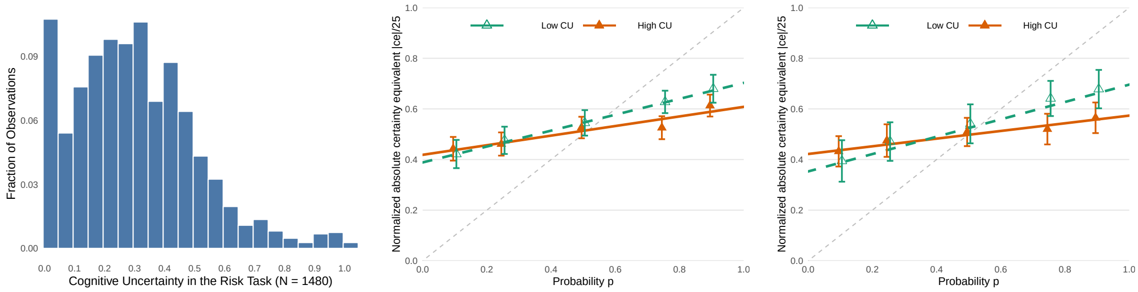


Figure C.3: Panel A: The fourfold pattern in Experiment CU. Panel B: The fourfold pattern among participants who pass and fail the simple tests.

than reported in Enke and Graeber (2023) — and substantial heterogeneity across participants. Figure C.4b replicates the top-left panel of Figure III from Enke and Graeber (2023). Compared to their findings, our participants exhibited substantially weaker differences between low and high CU choices in the association between CE and the probability in the lottery. When we restrict the comparison to the top and bottom 25% of CU scores, the differences become slightly more pronounced (see Figure C.4c), though they remain less clear than those reported by Enke and Graeber (2023). Participants' CU is positively and weakly correlated with the number of failed comprehension questions on the first attempt ($\rho = 0.138$, $p = 0.10$), and positively and strongly correlated with perceived complexity ($\rho = 0.225$, $p = 0.01$).



(a) (b) (c)

Figure C.4: Plot A: Histograms of cognitive uncertainty in the risky lottery tasks (see the left panel of figure II in Enke and Graeber (2023)). Plot B: Cognitive uncertainty (median split) and sensitivity to probabilities in choice under risk (see the left-top panel of figure III in Enke and Graeber (2023)). Plot C: Cognitive uncertainty (quartile split) and sensitivity to probabilities in choice under risk. Two–standard error bars are included for every task.

Table C.1: Certainty equivalent deviations by gain domain and simplicity test performance.

P	ST	Label	N	Mean	Median	T1	W1	T2	W2
				Panel A: Gain domain					
10	0	G10	12	6.917	5.5	0.001	0.004	0.039	0.057
10	1	G10	53	11.700	9.5	0.000	0.000		
25	0	G25	17	6.676	5.5	0.860	0.776	0.008	0.018
25	1	G25	65	11.746	11.5	0.000	0.000		
50	0	G50	17	11.147	9.5	0.239	0.124	0.730	0.941
50	1	G50	54	11.740	10.5	0.401	0.301		
75	0	G75	17	9.971	9.5	0.001	0.000	0.308	0.287
75	1	G75	57	11.816	10.5	0.000	0.000		
90	0	G90	17	14.088	14.5	0.030	0.007	0.865	0.942
90	1	G90	67	13.769	14.5	0.000	0.000		
				Panel B: Loss domain					
10	0	L10	20	-9.400	-10.5	0.000	0.000	0.129	0.177
10	1	L10	52	-12.000	-11.0	0.000	0.000		
25	0	L25	18	-11.200	-10.5	0.000	0.000	0.750	0.662
25	1	L25	55	-12.623	-11.5	0.000	0.000		
50	0	L50	20	-16.100	-15.5	0.005	0.007	0.231	0.229
50	1	L50	57	-14.079	-15.5	0.087	0.047		
75	0	L75	14	-16.857	-18.5	0.309	0.387	0.253	0.321
75	1	L75	52	-14.673	-15.5	0.000	0.000		
90	0	L90	16	-20.750	-21.0	0.093	0.120	0.008	0.010
90	1	L90	60	-15.983	-15.5	0.000	0.000		

Notes: T1 = one-sample t-test vs. prediction. W1 = Wilcoxon signed-rank test vs. prediction. T2 = two-sample t-test across simplicity groups. W2 = Wilcoxon rank-sum test across simplicity groups. All values rounded to three decimals. See corresponding Figure 5a.

Table C.2: Certainty equivalent deviations by gain domain and simplicity test performance.

P	ST	Label	N	Mean	Median	T1	W1	T2	W2
				Panel A: Gain domain					
10	0	G10	21	4.833	3.5	0.005	0.150	0.040	0.028
10	1	G10	62	8.694	6.0	0.000	0.000		
25	0	G25	16	5.438	4.5	0.367	0.163	0.083	0.089
25	1	G25	50	8.960	7.0	0.025	0.164		
50	0	G50	16	7.375	6.5	0.000	0.002	0.017	0.023
50	1	G50	61	11.893	10.5	0.509	0.388		
75	0	G75	16	12.625	14.5	0.001	0.003	0.737	0.812
75	1	G75	58	13.259	14.5	0.000	0.000		
90	0	G90	16	14.188	14.5	0.000	0.001	0.965	0.864
90	1	G90	48	14.083	13.0	0.000	0.000		
				Panel B: Loss domain					
10	0	L10	13	-8.346	-4.5	0.016	0.014	0.021	0.014
10	1	L10	63	-14.722	-15.5	0.000	0.000		
25	0	L25	15	-12.967	-9.5	0.002	0.005	0.181	0.136
25	1	L25	60	-15.800	-16.0	0.000	0.000		
50	0	L50	13	-13.346	-14.5	0.680	0.592	0.092	0.072
50	1	L50	58	-17.224	-20.0	0.000	0.000		
75	0	L75	19	-19.500	-20.5	0.402	0.242	0.248	0.432
75	1	L75	63	-17.563	-19.5	0.271	0.278		
90	0	L90	17	-19.441	-22.5	0.163	0.843	0.940	0.841
90	1	L90	55	-19.591	-22.5	0.002	0.080		

Notes: T1 = one-sample t-test vs. prediction. W1 = Wilcoxon signed-rank test vs. prediction. T2 = two-sample t-test across simplicity groups. W2 = Wilcoxon rank-sum test across simplicity groups. All values rounded to three decimals. See corresponding Figure 5b.

Table C.3: Certainty equivalent deviations by gain domain and cognitive uncertainty (CU) level.

P	CU_{high}	Label	N	Mean	Median	T1	W1	T2	W2
				Panel A: Gain domain					
10	0	G10	62	8.629	6.0	0.000	0.000	0.032	0.019
10	1	G10	53	11.783	9.5	0.000	0.000		
25	0	G25	52	9.538	7.0	0.010	1.000	0.170	0.104
25	1	G25	63	11.357	10.5	0.000	0.000		
50	0	G50	61	11.795	10.5	0.443	0.339	0.965	0.822
50	1	G50	54	11.852	11.5	0.474	0.348		
75	0	G75	57	13.342	14.5	0.000	0.000	0.214	0.031
75	1	G75	58	11.759	10.5	0.000	0.000		
90	0	G90	60	14.467	16.0	0.000	0.000	0.407	0.437
90	1	G90	55	13.282	11.5	0.000	0.000		
				Panel B: Loss domain					
10	0	L10	59	-14.924	-15.5	0.000	0.000	0.054	0.092
10	1	L10	56	-11.982	-10.5	0.000	0.000		
25	0	L25	59	-11.572	-9.5	0.000	0.000	0.036	0.035
25	1	L25	56	-12.769	-12.5	0.000	0.000		
50	0	L50	60	-17.133	-20.0	0.000	0.002	0.024	0.022
50	1	L50	55	-14.064	-15.5	0.103	0.054		
75	0	L75	60	-17.480	-19.5	0.451	0.508	0.007	0.009
75	1	L75	55	-18.250	-15.5	0.000	0.000		
90	0	L90	57	-19.483	-22.5	0.001	0.034	0.005	0.001
90	1	L90	58	-15.966	-15.5	0.000	0.000		

Notes: T1 = one-sample t-test vs. prediction. W1 = Wilcoxon signed-rank test vs. prediction. T2 = two-sample t-test across high/low CU groups. W2 = Wilcoxon rank-sum test across high/low CU groups. All values rounded to three decimals. See corresponding Figure 5c.

Table C.4: Certainty equivalent deviations by gain domain and cognitive uncertainty (CU) level.

P	CU_{high}	Label	N	Mean	Median	T1	W1	T2	W2
Panel A: Gain domain									
10	0	G10	17	3.971	2.5	0.177	0.122	0.038	0.002
10	1	G10	16	7.313	5.5	0.000	0.000		
25	0	G25	16	5.438	4.5	0.367	0.163	0.411	0.127
25	1	G25	17	6.676	5.5	0.860	0.776		
50	0	G50	18	7.611	8.5	0.000	0.001	0.017	0.027
50	1	G50	15	11.367	9.5	0.378	0.235		
75	0	G75	16	12.625	14.5	0.001	0.000	0.194	0.214
75	1	G75	17	9.971	9.5	0.000	0.000		
90	0	G90	16	14.313	14.5	0.001	0.002	0.886	0.701
90	1	G90	17	13.971	14.5	0.000	0.000		
Panel B: Loss domain									
10	0	L10	16	-8.438	-5.5	0.003	0.001	0.634	0.573
10	1	L10	17	-9.500	-10.5	0.000	0.001		
25	0	L25	17	-12.313	-9.5	0.000	0.000	0.957	0.971
25	1	L25	16	-12.375	-10.5	0.003	0.010		
50	0	L50	17	-14.206	-15.5	0.309	0.232	0.428	0.467
50	1	L50	16	-15.875	-15.5	0.018	0.024		
75	0	L75	17	-19.206	-20.5	0.585	0.412	0.379	0.244
75	1	L75	16	-17.500	-19.0	0.000	0.000		
90	0	L90	17	-20.500	-22.5	0.302	0.822	0.714	0.172
90	1	L90	16	-19.625	-20.5	0.058	0.070		

Notes: T1 = one-sample t-test vs. prediction. W1 = Wilcoxon signed-rank test vs. prediction. T2 = two-sample t-test across high/low CU groups. W2 = Wilcoxon rank-sum test across high/low CU groups. All values rounded to three decimals. See corresponding Figure 5d.

Table C.5: Cognitive uncertainty (CU) about the test lotteries and standard lotteries by the frequency of test failures.

	Overall	The frequency of test failures				
		0	1	2–4	5–7	8
Mean CU (Test lotteries)	0.185	0.080	0.104	0.165	0.261	0.330
Mean CU (Standard lotteries)	0.283	0.276	0.229	0.254	0.300	0.355
Difference in CU	0.098	0.196	0.125	0.089	0.039	0.025

Table C.6: Random effects regressions on normalized absolute certainty equivalents

	Gain Domain			*Loss Domain*		
	(1)	(2)	(3)	(4)	(5)	(6)
p	0.296***	0.268***	0.267***	0.367***	0.402***	0.400***
	(0.035)	(0.051)	(0.051)	(0.035)	(0.051)	(0.051)
CU^{test}	0.421***		0.565***	0.017		0.132
	(0.127)		(0.153)	(0.128)		(0.154)
$p \times CU^{\text{test}}$	−0.360*		−0.440*	−0.583***		−0.493**
	(0.147)		(0.178)	(0.147)		(0.178)
CU^{risk}		0.069	−0.283		−0.142	−0.224
		(0.142)	(0.169)		(0.141)	(0.170)
$p \times CU^{\text{risk}}$		−0.117	0.157		−0.484**	−0.177
		(0.164)	(0.197)		(0.164)	(0.198)
Intercept	0.261***	0.312***	0.314***	0.476***	0.518***	0.518***
	(0.030)	(0.044)	(0.044)	(0.030)	(0.044)	(0.044)
Observations	740	740	740	740	740	740

Note: Dependent variable = normalized absolute certainty equivalent. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. CU^{test} represents participant i's median value of test-list cognitive uncertainty; CU^{risk} represents participant i's median value of risk-lottery-list cognitive uncertainty. Standard errors in parentheses. All models use random effects with participant-level clustering.

Table C.7: Random effects regressions on normalized absolute certainty equivalents

	Gain Domain				*Loss Domain*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p	0.236***	0.424***	0.348***	0.512***	0.271***	0.532***	0.348***	0.512***
	(0.025)	(0.052)	(0.044)	(0.061)	(0.025)	(0.052)	(0.044)	(0.061)
D_{ST}		0.221***		0.208***		0.188***		0.208***
		(0.051)		(0.051)		(0.053)		(0.051)
$p \times D_{\text{ST}}$		−0.241***		−0.223***		−0.337***		−0.223***
		(0.059)		(0.059)		(0.059)		(0.059)
CU			0.370***	0.337***			0.370***	0.337***
			(0.091)	(0.090)			(0.091)	(0.090)
$p \times$ CU			−0.450**	−0.411**			−0.450**	−0.411**
			(0.144)	(0.143)			(0.144)	(0.143)
Intercept	0.331***	0.159***	0.239***	0.086	0.479***	0.333***	0.239***	0.086
	(0.022)	(0.045)	(0.031)	(0.049)	(0.022)	(0.047)	(0.031)	(0.049)
Observations	740	740	740	740	740	740	740	740

Note: The dependent variable is the normalized absolute certainty equivalent, $|CE_{ij}|/25$, for lottery j of participant i. p is the lottery's payout probability; D_{ST} is a dummy variable (1 = failed simplicity tests, 0 = passed); CU indicates cognitive uncertainty (derived from mirror task responses). Standard errors in parentheses. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. The first four columns (1–4) report results for the gain domain; the last four columns (5–8) report results for the loss domain. All models use random effects with participant-level clustering.

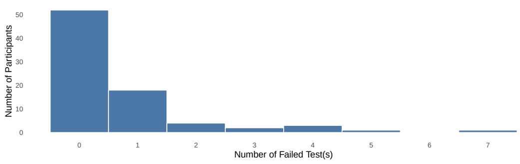
Table C.8: Tobit regressions on normalized absolute certainty equivalents

	Gain Domain				*Loss Domain*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p	0.248***	0.430***	0.329***	0.494***	0.294***	0.557***	0.331***	0.592***
	(0.028)	(0.056)	(0.039)	(0.061)	(0.028)	(0.058)	(0.040)	(0.065)
D_{ST}		0.223***		0.216***		0.200***		0.195***
		(0.055)		(0.057)		(0.060)		(0.057)
$p \times D_{\mathrm{ST}}$		−0.234***		−0.221***		−0.341***		−0.330***
		(0.064)		(0.063)		(0.066)		(0.066)
CU_{high}			0.120***	0.111*			−0.036	−0.020
			(0.036)	(0.035)			(0.037)	(0.037)
$p \times CU_{\mathrm{high}}$			−0.163**	−0.151**			−0.075	−0.087
			(0.057)	(0.056)			(0.058)	(0.057)
Intercept	0.320***	0.146*	0.260***	0.097	0.476***	0.321***	0.494***	0.336***
	(0.024)	(0.049)	(0.029)	(0.053)	(0.025)	(0.053)	(0.030)	(0.056)
Observations	740	740	740	740	740	740	740	740

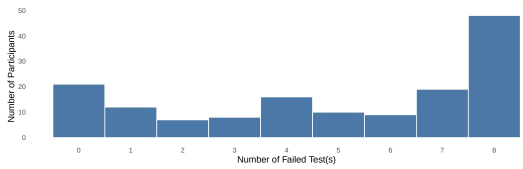
Note: Dependent variable = $|CE|/25$ (normalized absolute certainty equivalent). Reported estimates are posterior means from Bayesian Tobit models with left censoring at zero. Posterior standard errors are reported in parentheses (rounded to 3 decimals). ***, **, and * indicate that the 99.9%, 99%, and 95% credible intervals, respectively, exclude zero. $D_{\mathrm{ST}} = 1$ indicates participants who failed the simplicity tests; $CU_{\mathrm{high}} = 1$ indicates participants whose cognitive uncertainty level is higher than the group median level. Columns (1–4) report gain domain results; columns (5–8) report loss domain results. All models include participant-level random intercepts.

D Appendix: Additional results in Experiment Student

Experimental procedure and sample. This additional experiment replicates the main experiment but with different university student population rather than representative online population. Participants in this additional experiment ($N = 81$) were on average 22.0 years old (median = 22.0, interquartile range = 21–23), with ages ranging from 19 to 29. Thirty-six participants (44.4%) were female, and forty-five (55.6%) were male. In terms of education, 1 participant (1.2%) had completed middle school, 17 (21.0%) had completed high school, 47 (58.0%) held an undergraduate degree, and 16 (19.8%) held a postgraduate degree. The median payment was €24 for the median duration of 26.05 minutes, significantly higher than the standard in terms of hourly rate.



(a) Individual simple test performance



(b)

Figure D.1: Participants' performance in the simple tests. Panel A is for the student sample. Panel B is for the Prolific sample.

Table D.1: Certainty equivalent deviations by gain domain and simplicity test performance.

P	ST	Label	N	Mean	Median	T1	W1	T2	W2
				Panel A: Gain domain					
10	0	G10	52	4.154	3.5	0.000	0.000	0.111	0.139
10	1	G10	29	5.362	4.5	0.000	0.000		
25	0	G25	52	7.481	6.5	0.035	0.040	0.878	0.580
25	1	G25	29	7.362	6.5	0.189	0.334		
50	0	G50	52	11.288	12.5	0.010	0.002	0.214	0.241
50	1	G50	29	10.224	9.5	0.008	0.006		
75	0	G75	52	14.846	15.5	0.000	0.000	0.004	0.011
75	1	G75	29	11.741	10.5	0.000	0.000		
90	0	G90	52	19.058	20.5	0.000	0.000	0.013	0.025
90	1	G90	29	15.741	16.5	0.000	0.000		
				Panel B: Loss domain					
10	0	L10	52	-4.769	-3.5	0.000	0.000	0.004	0.049
10	1	L10	29	-7.776	-5.5	0.000	0.000		
25	0	L25	52	-9.269	-8.0	0.000	0.000	0.508	0.453
25	1	L25	29	-9.879	-10.5	0.000	0.001		
50	0	L50	52	-12.769	-12.5	0.576	0.288	0.131	0.047
50	1	L50	29	-14.052	-15.5	0.042	0.021		
75	0	L75	52	-17.423	-18.5	0.074	0.229	0.253	0.257
75	1	L75	29	-16.155	-17.5	0.030	0.059		
90	0	L90	52	-21.077	-22.5	0.001	0.000	0.383	0.563
90	1	L90	29	-20.328	-21.5	0.018	0.007		

Notes: T1 = one-sample t-test vs. prediction. W1 = Wilcoxon signed-rank test vs. prediction. T2 = two-sample t-test across simplicity groups. W2 = Wilcoxon rank-sum test across simplicity groups. ST = indicator for simple-test failure (1 = fail, 0 = pass). All values rounded to three decimals.

Table D.2: Random effects regressions on normalized absolute certainty equivalents

	Gain Domain				*Loss Domain*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p	0.619***	0.682***	0.611***	0.668***	0.706***	0.776***	0.710***	0.768***
	(0.024)	(0.026)	(0.026)	(0.026)	(0.023)	(0.025)	(0.024)	(0.025)
F_{ST}		0.019***		0.012***		0.047***		0.059***
		(0.013)		(0.014)		(0.013)		(0.014)
$p \times F_{\text{ST}}$		−0.091***		−0.112***		−0.102***		−0.113***
		(0.018)		(0.019)		(0.017)		(0.018)
$F_{Compreh}$			−0.017	−0.038			−0.031	−0.082*
			(0.033)	(0.033)			(0.031)	(0.038)
$p \times F_{Compreh}$			0.038	0.135**			−0.020	0.242**
			(0.044)	(0.045)			(0.042)	(0.043)
Intercept	0.127***	0.114***	0.130***	0.118***	0.178***	0.145***	0.184***	0.154***
	(0.018)	(0.019)	(0.019)	(0.019)	(0.017)	(0.019)	(0.018)	(0.019)
Observations	405	405	405	405	405	405	405	405

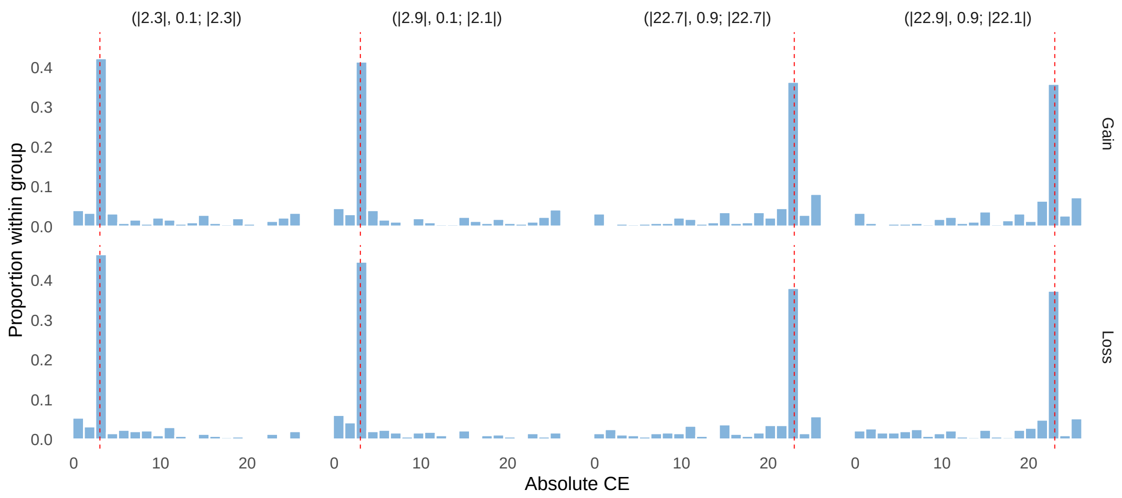
Note: The dependent variable is the normalized absolute certainty equivalent, $|CE_{ij}|/25$, for lottery j of participant i. p is the lottery's payout probability; F_{ST} indicates the number of simplicity tests participants failed; $F_{Compreh}$ indicates the number of comprehension questions participants failed. Standard errors in parentheses. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. The first four columns (1–4) report results for the gain domain; the last four columns (5–8) report results for the loss domain. All models use random effects with participant-level clustering.

Table D.3: Tobit regressions on normalized absolute certainty equivalents

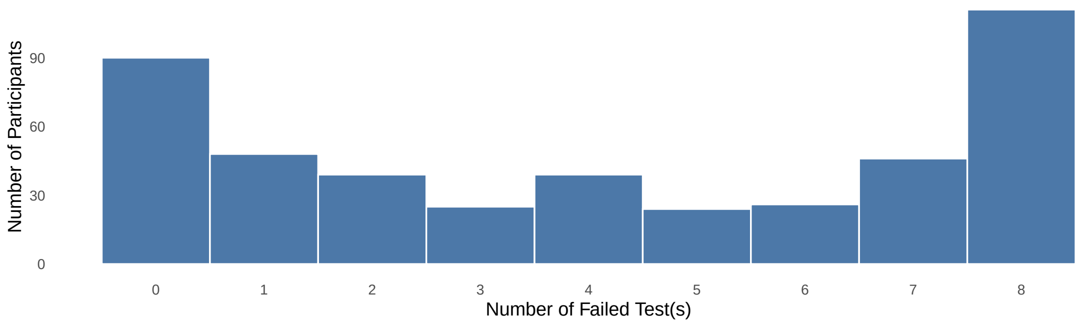
	Gain Domain				*Loss Domain*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p	0.622***	0.702***	0.619***	0.694***	0.710***	0.771***	0.715***	0.771***
	(0.024)	(0.029)	(0.027)	(0.030)	(0.023)	(0.028)	(0.026)	(0.030)
D_{ST}		0.059		0.065		0.110**		0.116**
		(0.036)		(0.036)		(0.037)		(0.038)
$p \times D_{\mathrm{ST}}$		−0.226***		−0.238***		−0.171***		−0.171***
		(0.049)		(0.049)		(0.047)		(0.049)
$D_{Compreh}$			−0.016	−0.035			−0.027	−0.057
			(0.051)	(0.051)			(0.049)	(0.049)
$p \times D_{Compreh}$			0.017	0.080			−0.038	0.005
			(0.067)	(0.068)			(0.066)	(0.065)
Intercept	0.124***	0.103***	0.127***	0.107***	0.175***	0.135***	0.179***	0.141***
	(0.018)	(0.022)	(0.020)	(0.023)	(0.018)	(0.022)	(0.019)	(0.023)
Observations	410	410	410	410	410	410	410	410

Note: Dependent variable = $|CE|/25$ (normalized absolute certainty equivalent). Reported estimates are posterior means from Bayesian Tobit models with double censoring (left censoring at 0.02 and right censoring at 1.02). Posterior standard errors are reported in parentheses. ***, **, and * indicate that the 99.9%, 99%, and 95% credible intervals, respectively, exclude zero. $D_{\mathrm{ST}} = 1$ indicates participants who failed the simplicity tests; $D_{Compreh} = 1$ indicates participants who failed the comprehension questions. Posterior standard errors are reported in parentheses. The first four columns (1–4) report results for the gain domain; the last four columns (5–8) report results for the loss domain. All models include participant-level random intercepts.

E Appendix: Pooling online experiments

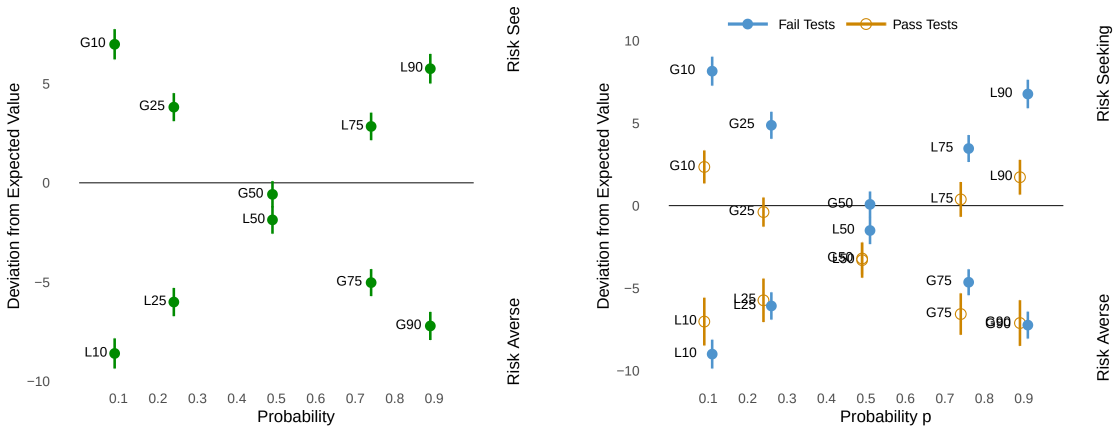


(a)



(b)

Figure E.1: Performance on simple tests. *Notes:* Panel (a) plots the distribution of individual CEs—i.e., switching positions—across the eight simple test lists, separately for gain and loss domains. Red dashed lines indicate the theoretically correct switching position for each list. Panel (b) shows the distribution of the number of simple tests failed by each participant. The sample consists of 448 participants.



(a) (b)

Figure E.2: fourfold pattern of risk attitudes and simple test performance. *Notes:* Panel (a) plots the fourfold pattern of risk attitudes for all participants. Panel (b) plots the same pattern separately for participants who pass and who fail the simple tests. In both panels, the x-axis is the payout probability p in the lotteries $(\$25, p; \$0)$ and $(\$-25, p; \$0)$, and the y-axis is the difference between the CE and the expected value, with positive values indicating risk seeking and negative values indicating risk aversion. Two–standard error bars are included for every task. Table E.1 reports the corresponding summary statistics.

Table E.1: Certainty equivalent deviations by gain domain and simplicity test performance.

P	ST	Label	N	Mean	Median	T1	W1	T2	W2
				Panel A: Gain domain					
10	0	G10	90	4.844	3.5	0.000	0.000	0.000	0.000
10	1	G10	358	10.645	9.5	0.000	0.000		
25	0	G25	90	6.111	4.5	0.381	0.689	0.000	0.000
25	1	G25	358	11.369	9.5	0.000	0.000		
50	0	G50	90	9.300	9.5	0.000	0.000	0.000	0.000
50	1	G50	358	12.578	12.5	0.842	0.932		
75	0	G75	90	11.933	12.0	0.000	0.000	0.023	0.018
75	1	G75	358	13.860	14.5	0.000	0.000		
90	0	G90	90	15.389	15.5	0.000	0.000	0.890	0.721
90	1	G90	358	15.266	15.5	0.000	0.000		
				Panel B: Loss domain					
10	0	L10	90	-9.522	-9.5	0.000	0.000	0.038	0.065
10	1	L10	358	-11.492	-10.5	0.000	0.000		
25	0	L25	90	-12.233	-10.5	0.000	0.000	0.701	0.823
25	1	L25	358	-12.575	-10.5	0.000	0.000		
50	0	L50	90	-15.800	-15.5	0.000	0.000	0.039	0.057
50	1	L50	358	-14.006	-14.5	0.000	0.000		
75	0	L75	90	-18.122	-19.5	0.476	0.908	0.000	0.002
75	1	L75	358	-15.039	-15.5	0.000	0.000		
90	0	L90	90	-20.778	-22.5	0.001	0.014	0.000	0.000
90	1	L90	358	-15.737	-16.5	0.000	0.000		

Notes: T1 = one-sample t-test vs. prediction. W1 = Wilcoxon signed-rank test vs. prediction. T2 = two-sample t-test across simplicity groups. W2 = Wilcoxon rank-sum test across simplicity groups. All values rounded to three decimals. See corresponding Figure E.2b.

Table E.2: Random effects regressions on normalized absolute certainty equivalents

	Gain Domain				*Loss Domain*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p	0.280***	0.510***	0.359***	0.529***	0.274***	0.537***	0.337***	0.546***
	(0.016)	(0.035)	(0.023)	(0.036)	(0.016)	(0.035)	(0.023)	(0.036)
D_{ST}		0.273***		0.224***		0.104**		0.117***
		(0.032)		(0.033)		(0.033)		(0.035)
$p \times D_{\text{ST}}$		−0.288***		−0.255***		−0.329***		−0.313***
		(0.039)		(0.041)		(0.040)		(0.042)
$D_{Compreh}$			0.184***	0.129***			−0.005	−0.034
			(0.026)	(0.027)			(0.027)	(0.028)
$p \times D_{Compreh}$			−0.149***	−0.087**			−0.119***	−0.042
			(0.032)	(0.033)			(0.032)	(0.033)
Intercept	0.344***	0.126***	0.247***	0.097***	0.426***	0.343***	0.429***	0.351***
	(0.013)	(0.029)	(0.019)	(0.029)	(0.013)	(0.030)	(0.019)	(0.030)
Observations	2240	2240	2240	2240	2240	2240	2240	2240

Note: Dependent variable = $|CE|/25$ (normalized absolute certainty equivalent). $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. $D_{\text{ST}} = 1$ indicates participants who failed the simplicity tests; $D_{Compreh} = 1$ indicates participants who failed the comprehension questions. Standard errors in parentheses. The first four columns (1–4) report results for the gain domain; the last four columns (5–8) report results for the loss domain. All models use random effects with participant-level clustering.

Table E.3: Random effects regressions on normalized absolute certainty equivalents

	Gain Domain				*Loss Domain*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p	0.280***	0.506***	0.349***	0.511***	0.274***	0.523***	0.341***	0.526***
	(0.016)	(0.025)	(0.020)	(0.026)	(0.016)	(0.026)	(0.020)	(0.026)
F_{ST}		0.044***		0.038***		0.010*		0.011*
		(0.004)		(0.004)		(0.004)		(0.005)
$p \times F_{\text{ST}}$		−0.055***		−0.051***		−0.061***		−0.059***
		(0.005)		(0.005)		(0.005)		(0.005)
$F_{Compreh}$			0.066***	0.036***			0.004	−0.004
			(0.008)	(0.008)			(0.008)	(0.009)
$p \times F_{Compreh}$			−0.052***	−0.016			−0.054***	−0.008
			(0.009)	(0.010)			(0.010)	(0.010)
Intercept	0.344***	0.155***	0.263***	0.144***	0.426***	0.386***	0.421***	0.387***
	(0.013)	(0.021)	(0.016)	(0.021)	(0.013)	(0.021)	(0.017)	(0.022)
Observations	2240	2240	2240	2240	2240	2240	2240	2240

Note: The dependent variable is the normalized absolute certainty equivalent, $|CE_{ij}|/25$, for lottery j of participant i. p is the lottery's payout probability; F_{ST} indicates the number of simplicity tests participants failed; $F_{Compreh}$ indicates the number of comprehension questions participants failed. Standard errors in parentheses. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. The first four columns (1–4) report results for the gain domain; the last four columns (5–8) report results for the loss domain. All models use random effects with participant-level clustering.

Table E.4: Tobit regressions on normalized absolute certainty equivalents

	Gain Domain				*Loss Domain*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p	0.280***	0.509***	0.359***	0.528***	0.274***	0.537***	0.337***	0.546***
	(0.016)	(0.034)	(0.023)	(0.036)	(0.016)	(0.035)	(0.024)	(0.035)
D_{ST}		0.273***		0.223***		0.104**		0.115**
		(0.033)		(0.034)		(0.033)		(0.035)
$p \times D_{\text{ST}}$		−0.287***		−0.253***		−0.329***		−0.312***
		(0.038)		(0.042)		(0.039)		(0.041)
$D_{Compreh}$			0.183***	0.128***			−0.005	−0.033
			(0.026)	(0.027)			(0.027)	(0.028)
$p \times D_{Compreh}$			−0.149***	−0.087**			−0.118***	−0.042
			(0.031)	(0.033)			(0.032)	(0.034)
Intercept	0.344***	0.126***	0.248***	0.099***	0.426***	0.342***	0.428***	0.352***
	(0.013)	(0.029)	(0.019)	(0.029)	(0.013)	(0.030)	(0.019)	(0.030)
Observations	2240	2240	2240	2240	2240	2240	2240	2240

Note: Dependent variable = $|CE|/25$ (normalized absolute certainty equivalent). Reported estimates are posterior means from Bayesian Tobit models with left censoring at zero. Posterior standard errors are reported in parentheses. ***, **, and * indicate that the 99.9%, 99%, and 95% credible intervals, respectively, exclude zero. $D_{\text{ST}} = 1$ indicates participants who failed the simplicity tests; $D_{Compreh} = 1$ indicates participants who failed the comprehension questions. Posterior standard errors are reported in parentheses. The first four columns (1–4) report results for the gain domain; the last four columns (5–8) report results for the loss domain. All models include participant-level random intercepts.

Table E.5: The proportion of participants choosing the lottery under different conditions.

	($25,0.1;0) *vs* $3		($25,0.9;0) *vs* $22		(-$25,0.1;0) *vs* -$3		(-$25,0.9;0) *vs* -$22	
	PL	BC	PL	BC	PL	BC	PL	BC
Pass (n=90)	0.522	0.344	0.244	0.189	0.267	0.267	0.478	0.422
Fail (n=358)	0.774	0.416	0.268	0.223	0.279	0.190	0.542	0.673
P-value	0.000	0.231	0.690	0.567	0.895	0.111	0.290	0.000

Note: Pass/fail indicates performance on the simple test; PL = price list choice, BC = binary choice. Sample sizes (n) for Pass/Fail groups are shown in parentheses. P-values from Fisher's Exact Test compare lottery choice proportions between subgroups (values rounded to 3 decimal places).

Table E.6: Proportion of consistent choices (retest vs. original) by simple test performance.

	Consistent Choice Proportion			
	($25,0.1;0)	($25,0.9;0)	(-$25,0.1;0)	(-$25,0.9;0)
Pass	21/51	16/39	18/41	28/49
Fail	69/181	61/177	59/180	67/178
p-value	0.746	0.464	0.205	0.021

Note: Values are formatted as "count of consistent choices / valid sample size". Pass/fail refers to performance on the simple test. p-values from Fisher's Exact Test compare consistency proportions between the two subgroups.

To examine what factors affect the performance in the simple tests, we ran an OLS regression with the number of failed simple tests as the dependent variable. Independent variables include the number of correctly answered comprehension questions on the first attempt, participants' perceived complexity about the price list task (a six-point Likert scale: 1 = not complex at all; 6 = very complex), CRT performance (the number of correct answers). Table E.7 summarizes the regression result. Overall, as reported in Column 1, participants' performance in the comprehension checks has a positive relationship with their performance in the simple tests. The self-reported perceived complexity also relates positively to the test performance at 10% significance level. Somewhat surprisingly, CRT performance is positively, though weakly, related to participants' performance in the simple tests. One potential reason is that many participants used google or AI to answer those questions, and thus more correct answers may imply less attention and commitment to the experiment.

In a second regression, we further included two consistency measures to examine the behavioral implications of participants' performance in the simple tests: the consistency between choices in the repeated price lists and the consistency between price lists and their implied binary choices. Column 2 suggests that participants who pass the simple tests were more consistent between repeated price lists, which is consistent with the interpretation that those participants' choices were more deliberative. However, they were not more consistent between price lists and their implied binary choices. This seems to suggest that binary choices differ from choices in price lists in more fundamental ways.

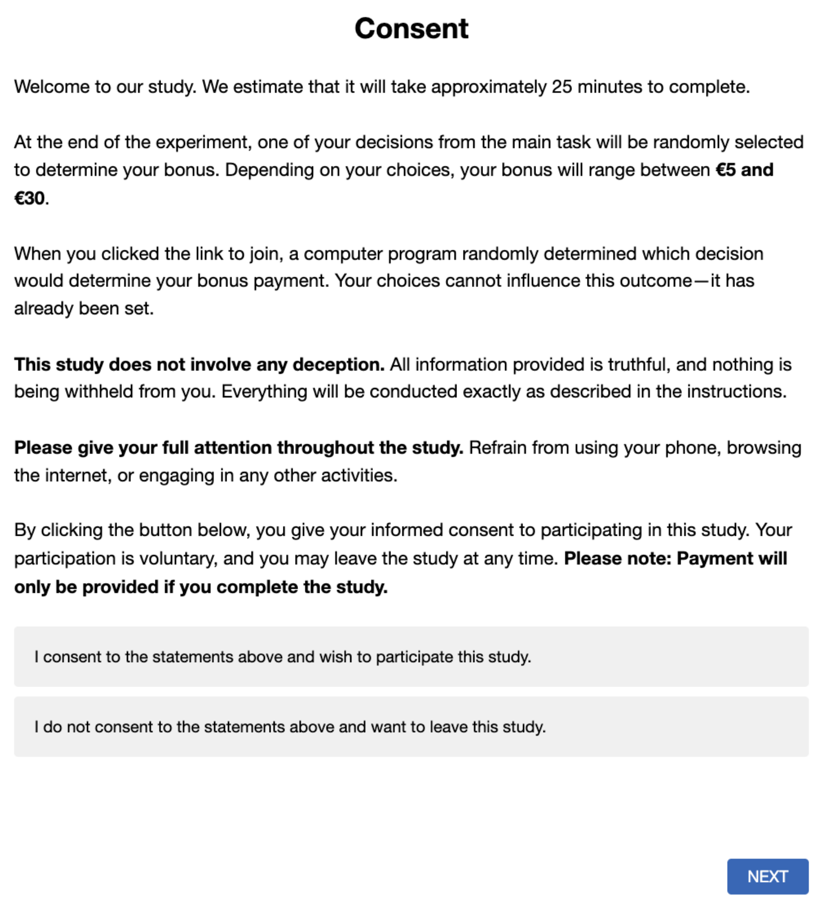
Table E.7: OLS regressions on *Failure frequency in the simple tests*

	Tests failure frequency		
	(1)	(2)	(3)
$F_{\text{Compreh.}}$	0.742***	0.709***	0.702***
	(0.084)	(0.082)	(0.083)
Complexity	0.292*	0.299*	0.306*
	(0.123)	(0.121)	(0.121)
CRT	0.292*	0.262*	0.254*
	(0.130)	(0.127)	(0.128)
PL-Rep		−0.812***	−0.818***
		(0.170)	(0.171)
PL&BC		0.012	0.015
		(0.125)	(0.125)
Female			−0.201
			(0.262)
Age			0.008
			(0.010)
Intercept	1.847***	2.519***	2.334**
	(0.450)	(0.568)	(0.706)
Observations	448	448	448

Note: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. $F_{\text{Compreh.}}$ denotes the number of failed comprehension questions. Complexity refers to participants' self-reported perception of the overall difficulty of the price lists (range: 1–6). CRT indicates correct answers in the Cognitive Reflection Test (range: 0–3). PL-Rep denotes the number of times that choices between the initial and the repeated price list were consistent (range: 0–4). PL&BC represents the number of times that choices implied in the price lists were consistent with their corresponding binary choices (range: 0–4). Standard errors are in parentheses.

F Appendix: Experimental Design

F.1 Screenshots of the Main Experimental Interface



Consent

Welcome to our study. We estimate that it will take approximately 25 minutes to complete.

At the end of the experiment, one of your decisions from the main task will be randomly selected to determine your bonus. Depending on your choices, your bonus will range between **€5 and €30**.

When you clicked the link to join, a computer program randomly determined which decision would determine your bonus payment. Your choices cannot influence this outcome—it has already been set.

This study does not involve any deception. All information provided is truthful, and nothing is being withheld from you. Everything will be conducted exactly as described in the instructions.

Please give your full attention throughout the study. Refrain from using your phone, browsing the internet, or engaging in any other activities.

By clicking the button below, you give your informed consent to participating in this study. Your participation is voluntary, and you may leave the study at any time. **Please note: Payment will only be provided if you complete the study.**

I consent to the statements above and wish to participate this study.

I do not consent to the statements above and want to leave this study.

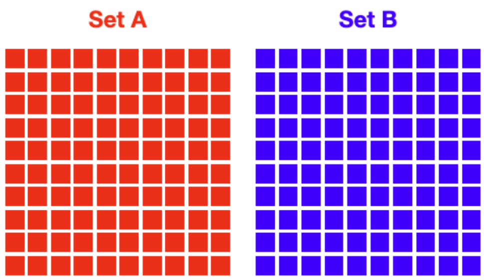
NEXT

General Instructions

Boxes With Money

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

- You will then choose which set of BOXES – Set A or Set B – the computer should open.



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

Continue Instructions

General Instructions

The Decision Table

- The two sets of boxes will be shown in a TABLE as illustrated below. For each set, the number of boxes (e.g., 100, 75, or 25 boxes) is listed at the top, and the corresponding monetary value (e.g., $7, $20, or $0) for that count is displayed in the table row.

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

- In the example above, Set A consists of 100 boxes, each containing $7; Set B consists of 75 boxes with $20 in each box and 25 boxes with $0 in each box.

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

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

- Your task is to click on the table to choose which set of boxes (A or B) you want the computer to use to determine your payment. Clicking on the table will highlight one set in yellow. The set highlighted in yellow will be used to determine your FINAL EARNINGS.

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

- In the example above, you have highlighted Set A, and your payment will be based on that set.

Continue Instructions

Instruction – A Random Box

- In the upcoming tasks, the computer will RANDOMLY select one of the 100 boxes from the set you choose. Every box in your chosen set is EQUALLY likely to be selected. If the selected box contains a positive amount, that amount will be ADDED to your initial money; if it is negative, the amount will be SUBTRACTED from your initial money.

- Example: In the example below, each set contains 100 boxes. For Set A, every one of the 100 boxes contains $4.00. Therefore, if you choose Set A, you have a 100% chance of having $4.00 added to your initial money. In Set B, 50 boxes contain $16.00 and the remaining 50 contain $0.00. Thus, if you choose Set B, there is a 50% chance that $16.00 will be added and a 50% chance that $0.00 will be added.

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

- Example: In the example below, each set also contains 100 boxes. For Set A, all 100 boxes contain -$6.00. If you choose Set A, you have a 100% chance of having $6.00 subtracted from your initial money. In Set B, 50 boxes contain -$8.00 and the other 50 contain $0.00. Therefore, if you choose Set B, there is a 50% chance that $8.00 will be subtracted and a 50% chance that $0.00 will be subtracted.

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

Continue Instructions

Here are some examples to help you answer the comprehension questions in next page correctly.

Suppose in the following two sets, you chose Set B.

Set A	Set B	
100 Boxes	60 Boxes	40 Boxes
$2.00	$5.00	$0.00

Q1: What is the chance that $5 is added to your earnings?

- ○ 0 in 100 (0%)
- ○ 40 in 100 (40%)
- ◉ 60 in 100 (60%)
- ○ 100 in 100 (100%)

The correct answer is C for this question. This is because you selected Set B. In Set B, the chance that $5 is added to your earnings is 60 in 100 (60%).

Q2: What is the chance that $2 is added to your earnings?

- ◉ 0 in 100 (0%)
- ○ 40 in 100 (40%)
- ○ 60 in 100 (60%)
- ○ 100 in 100 (100%)

The correct answer is A for this question. This is because you selected Set B, not Set A. Thus, the chance that $2 (Set A) is added to your earnings is 0 in 100 (0%).

Q3: What is the chance that $0 is added to your earnings?

- ○ 0 in 100 (0%)
- ◉ 40 in 100 (40%)
- ○ 60 in 100 (60%)
- ○ 100 in 100 (100%)

The correct answer is B for this question. This is because you selected Set B. In Set B, the chance that $0 is added to your earnings is 40 in 100 (40%).

By clicking "Continue," you will begin the real comprehension questions. Please note that you can earn an extra **reward of $0.15** if you answer a comprehension question correctly on your first try.

Continue

Comprehension Questions - A Random Box

Suppose in the following two sets, you chose Set B. Please answer the three questions below:

Set A	Set B	
100 Boxes	40 Boxes	60 Boxes
$5.00	$10.00	$0.00

Q1: What is the chance that $10 is added to your earnings?

- ○ 0 in 100 (0%)
- ◉ 40 in 100 (40%)
- ○ 60 in 100 (60%)
- ○ 100 in 100 (100%)

Correct!

Q2: What is the chance that $4 is added to your earnings?

- ◉ 0 in 100 (0%)
- ○ 40 in 100 (40%)
- ○ 60 in 100 (60%)
- ○ 100 in 100 (100%)

Correct!

Q3: What is the chance that $5 is added to your earnings?

- ◉ 0 in 100 (0%)
- ○ 40 in 100 (40%)
- ○ 60 in 100 (60%)
- ○ 100 in 100 (100%)

Submit Quiz

General Instructions

Decision Screen: Choosing a Set of Boxes

- In the actual experiment, you will face a sequence of questions to choose between Set A and Set B. Each question is presented as a different row in the table.

- Example: In the first row (Question 1) in the example below, Set A has 100 boxes containing $10 while Set B has 40 boxes containing $10 and 60 boxes containing $0. However, in the second row (Question 2), Set A has 100 boxes containing $9, while Set B has 40 boxes containing $10 and 60 boxes containing $0. Additional rows display other Question of Set A and Set B.

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

SUBMIT

- You will make a choice between Set A and Set B for all questions by clicking on the table and highlighting your preferred set in each row. Please try clicking the table above.

- When making your choices, the computer will limit you to switching from Set A to Set B only once across the table (although you may choose exclusively Set A or exclusively Set B for all rows). You may click on the table as many times as needed until you are satisfied with your selections. Then, press the SUBMIT button to finalize your decision.

- At the end of the experiment, the computer will randomly select one row from the table (each row is equally likely) and pay you based on your choice in that row. This means you should carefully consider your decision for every row (every question), as any row could determine your final payment.

Continue Instructions

Here are some examples to help you answer the comprehension questions in next page correctly.

Suppose you made the decisions shown in the table below.

	Set A	Set B	
Question	100 Boxes	60 Boxes	40 Boxes
1	$5.00	$5.00	$0.00
2	$4.00	$5.00	$0.00
3	$3.00	$5.00	$0.00
4	$2.00	$5.00	$0.00
5	$1.00	$5.00	$0.00

Q1: If Question 2 of this table is selected to determine your payment, which statement is correct?

- ○ You will receive $5.
- ◉ You will receive $4.
- ○ You will receive $3.
- ○ You will receive $0.

The correct answer is B. This is because you chose Set A for Question 2, and Set A adds $4 to your earnings for sure (100%).

Q2: If Question 4 of this table is selected to determine your payment, which statement is correct?

- ○ You will receive $4.
- ○ You will receive $2.
- ◉ You will have 60% chance of receiving $5 and 40% chance of receiving $0.
- ○ You will have 50% chance of receiving $5 and 50% chance of receiving $0.

The correct answer is C. This is because you chose Set B for Question 4. Its Set B could add $5 to your earnings with a chance of 60 in 100 (60%), and add $0 with a chance of 40 in 100 (40%).

Q3: Which statement best describes your choices?

- ○ You chose Set A over Set B for Question 5.
- ○ You chose Set A over Set B for Question 4.
- ◉ You chose Set A over Set B for Question 3.
- ○ You chose Set B over Set A for Question 2.

The correct answer is C. This is because you chose Set A for Questions 1 to 3, and then chose Set B for remaining Questions 4 and 5.

By clicking "Continue," you will begin the real comprehension questions. Please note that you can earn an extra **reward of $0.15** if you answer a comprehension question correctly on your first try.

Continue

Comprehension Questions – Decision Table

Suppose you made the decisions shown in the table below. Please answer the three questions below:

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

Q1: If Question 3 of this table is selected to determine your payment, which statement is correct?

- ○ You will receive $4.
- ○ You will receive $10.
- ◉ You will receive $8.
- ○ You will receive $0.

Correct!

Q2: If Question 6 of this table is selected to determine your payment, which statement is correct?

- ○ You will receive $4.
- ○ You will receive $5.
- ○ You will have 50% chance of receiving $10 and 50% chance of receiving $0.
- ◉ You will have 40% chance of receiving $10 and 60% chance of receiving $0.

Correct!

Q3: Which statement best describes your choices?

- ○ You chose Set A over Set B for Question 9.
- ○ You chose Set A over Set B for Question 5.
- ○ You chose Set A over Set B for Question 4.
- ○ You chose Set B over Set A for Question 2.

Submit Quiz

General Instructions

Multiple Tables

- Throughout the experiment, you will be presented with several tables. Each table displays a different initial amount of money along with different monetary amounts in the boxes arranged in rows. You must make a decision for every question (row) of each table.
- You will receive a BONUS based on your choices. Specifically, the computer will randomly choose one table and then randomly select one question (row) from that table to determine your bonus payment.
- Since you will not know which decision will be used to calculate your payment, please treat each choice as if it alone determines your final payment.

Continue Instructions

Choice task screenshot. In this example, a participant switch from Option A to Option B between Row 17 and Row 18.

Initial Money: $5.00

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

	Set A	Set B	
Question	100 Boxes	50 Boxes	50 Boxes
1	$25.00	$25.00	$0.00
2	$24.00	$25.00	$0.00
3	$23.00	$25.00	$0.00
4	$22.00	$25.00	$0.00
5	$21.00	$25.00	$0.00
6	$20.00	$25.00	$0.00
7	$19.00	$25.00	$0.00
8	$18.00	$25.00	$0.00
9	$17.00	$25.00	$0.00
10	$16.00	$25.00	$0.00
11	$15.00	$25.00	$0.00
12	$14.00	$25.00	$0.00
13	$13.00	$25.00	$0.00
14	$12.00	$25.00	$0.00
15	$11.00	$25.00	$0.00
16	$10.00	$25.00	$0.00
17	$9.00	$25.00	$0.00
18	$8.00	$25.00	$0.00
19	$7.00	$25.00	$0.00
20	$6.00	$25.00	$0.00
21	$5.00	$25.00	$0.00
22	$4.00	$25.00	$0.00
23	$3.00	$25.00	$0.00
24	$2.00	$25.00	$0.00
25	$1.00	$25.00	$0.00

SUBMIT

You have finished the first part of the experiment.

How would you rate the complexity of the choice task you completed?

- Very easy to understand
- Esay to understand
- Somewhat easy to understand
- Somewhat difficult to understand
- Difficult to understand
- Very difficult to understand

NEXT

Binary Choice Example

Imagine you have been given $30. Which of the following would you prefer?

Lose $25 with probability 0.1; 0 with probability 0.9

Lose $3 with certainty

NEXT

How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?

Please drag the slider, where the **value 0 means: 'not at all willing to take risks'** and the **value 10 means: 'very willing to take risks'**.

0	1	2	3	4	5	6	7	8	9	10

If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together?

Answer: [ ] day(s).

A man buys a pig for $60, sells it for $70, buys it back for $80, and sells it finally for $90. How much has he made?

Answer: [ ] dollar(s).

Simon decided to invest $8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%.

At this point, Simon has:

Answer: [ ] dollar(s).

This concludes the MAIN portion of the experiment.
Thank you for your participation!

To finish the study, please answer the following questions about yourself.

Q1: Please indicate your gender

Male

Female

Q2: Please drag the slider to indicate your age

You are **"Use the slider!"** years old.

18 **65**

Q3: What level of education do you complete?

No formal education

Elementary school

Middle school

High school

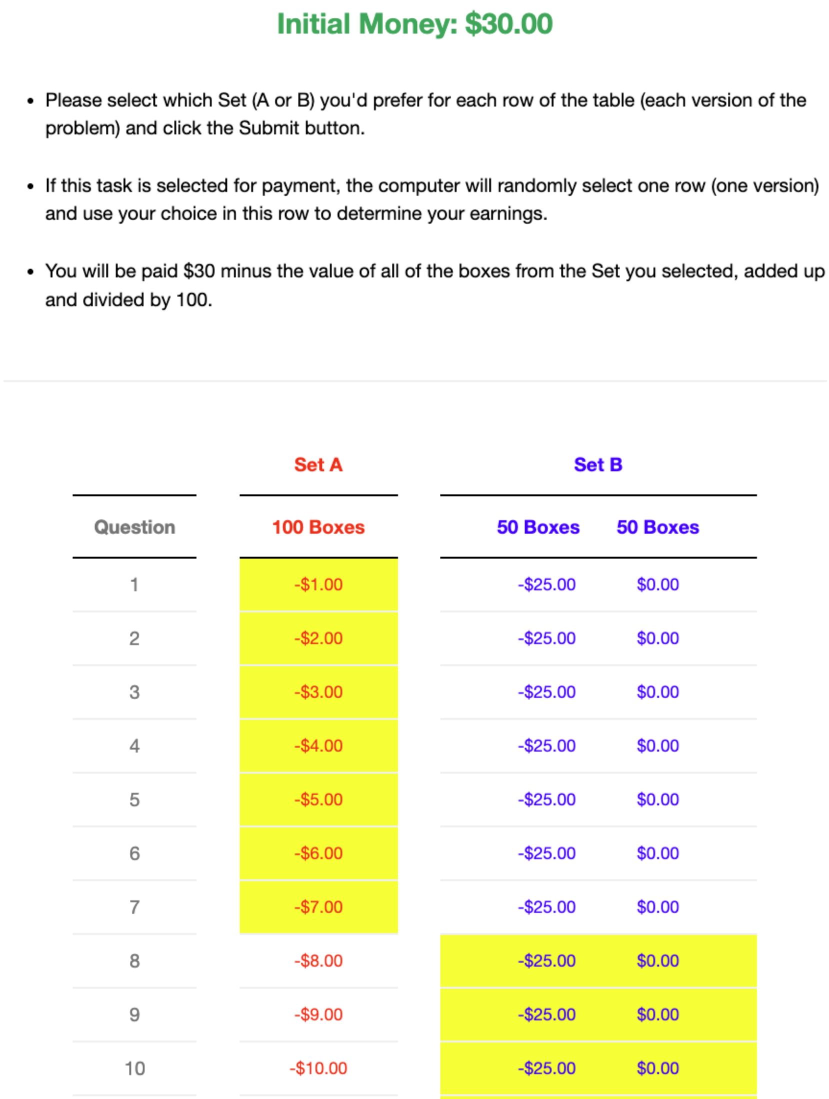
Undergraduate school

Postgraduate school

Q4: Do you have any comments, suggestions, or feedback about your participation, or anything else you'd like to share about this experiment? (Optional)

NEXT

F.2 Screenshots of the Mirror Task



Initial Money: $30.00

- Please select which Set (A or B) you'd prefer for each row of the table (each version of the problem) and click the Submit button.
- If this task is selected for payment, the computer will randomly select one row (one version) and use your choice in this row to determine your earnings.
- You will be paid $30 minus the value of all of the boxes from the Set you selected, added up and divided by 100.

	Set A	Set B	
Question	100 Boxes	50 Boxes	50 Boxes
1	-$1.00	-$25.00	$0.00
2	-$2.00	-$25.00	$0.00
3	-$3.00	-$25.00	$0.00
4	-$4.00	-$25.00	$0.00
5	-$5.00	-$25.00	$0.00
6	-$6.00	-$25.00	$0.00
7	-$7.00	-$25.00	$0.00
8	-$8.00	-$25.00	$0.00
9	-$9.00	-$25.00	$0.00
10	-$10.00	-$25.00	$0.00

Figure F.1: (Partial) Screenshot of the deterministic mirror task. *Notes*: This is a loss-domain mirror list from the mirror task, where the participant switches from the sure-amount option to the mirror option between Question 7 and Question 8.

F.3 Screenshots of the Cognitive Uncertainty Question

13	$13.00	$25.00	$0.00
14	$12.00	$25.00	$0.00
15	$11.00	$25.00	$0.00
16	$10.00	$25.00	$0.00
17	$9.00	$25.00	$0.00
18	$8.00	$25.00	$0.00
19	$7.00	$25.00	$0.00
20	$6.00	$25.00	$0.00
21	$5.00	$25.00	$0.00
22	$4.00	$25.00	$0.00
23	$3.00	$25.00	$0.00
24	$2.00	$25.00	$0.00
25	$1.00	$25.00	$0.00

You chose Set A over Set B in Versions 1 to 17 and Set B over Set A in Versions 18 to 25.
How certain do you feel about your decision?

very uncertain **completely certain**

I am **60%** certain about my choices in the Table.

Figure F.2: Screenshot of the cognitive uncertainty task.