

The Impact of LLM-Based Recommendations on Decisions under Uncertainty*

Marty Haoyuan Chen[†] Emel Filiz-Ozbay[‡] Erkut Y. Ozbay[§]

January 16, 2026

Abstract

We study how large language model (LLM) based recommendations affect evaluations of uncertainty. We elicit certainty equivalents for (i) lotteries that share the same reduced-form winning probability but vary in cognitive demands because they are compound, and (ii) an Ellsberg-style ambiguous lottery. Across all lottery types, valuations from participants who receive LLM-based recommendations before each valuation exhibit second-order stochastic dominance relative to valuations from participants who do not. Hence, LLM-based recommendations leave mean valuations unchanged but compress the distribution, reducing the incidence of extreme valuations. Consistent with this pattern, LLM-based recommendations reduce the intensity of risk, complexity, and ambiguity aversion, while leaving the fraction of participants classified as averse unchanged.

JEL Codes: C91, D81, D91

Keywords: Risk, Complexity, Ambiguity, LLM, AI, Recommendation

*We are thankful to the University of Maryland College of Behavioral and Social Sciences for the generous funding used for participant payments. We would like to thank Bakican Ayna and Quang Pham for helpful comments and discussions. All errors are our own.

[†]Department of Economics, University of Maryland (hchen915@umd.edu).

[‡]Department of Economics, University of Maryland (efozbay@umd.edu).

[§]Department of Economics, University of Maryland (ozbay@umd.edu).

1 Introduction

Large language models (LLM) are rapidly becoming a default “second opinion” for everyday economic decisions: from buying insurance to interpreting probabilistic information in medical settings. Investors are also increasingly relying on robo-advisors for portfolio decisions, with the market projected to reach \$33.38 billion by 2030.¹ As LLM-based recommendations diffuse into daily life, a central question for economics has been raised: how do LLM-based recommendations reshape decision making under uncertainty, and how do they affect risk and ambiguity attitudes, which are key measures for economic modeling?

Our experiments introduce standard, incentivized valuation tasks where individuals state their certainty equivalents for risky and ambiguous lotteries while receiving (or not receiving) an LLM-based recommendation. We also manipulate the complexity of the evaluated lotteries using compound lotteries where deriving the objective reduced-form probabilities might be cognitively costly. Our design links two literatures: attitudes toward risk, ambiguity, and complexity, and how decision makers incorporate recommendations into their choices. We show that LLM-based recommendations shift value distributions and reduce variations in them. This result is robust across environments that vary in uncertainty and cognitive complexity, suggesting that LLM-based recommendations may attenuate behaviors driven by extreme risk and ambiguity attitudes.

Our baseline design follows standard preference-elicitation practice using a standard multiple price list method, and the tasks are based on de Clippel et al. (2024). Participants value four prospects that pay \$30 if a “purple” card is drawn from a described deck and \$0 otherwise: (i) a transparent 50/50 lottery; (ii) a compound lottery with two underlying decks that reduces to 50/50; (iii) a more complex compound lottery with four underlying decks that also reduces to 50/50; and (iv) an ambiguous lottery in which the deck composition is unknown. This design varies the informational and computational demands of evaluation while holding fixed the reduced-form probability of winning. Our main intervention randomizes access to LLM-based recommendation generated by ChatGPT (GPT-4o), we call this “AI treatment”. To standardize informational content and rule out a possible impact caused by endoge-

¹Source: absrbd.com, accessed in December 2025.

nous prompting, all participants in the AI treatment view the same pre-generated screenshots for each lottery. The Baseline treatment asks participants to perform the same tasks without LLM-based recommendation.

LLM-based recommendations have economically meaningful effects on valuations in our data, but not through the mean. Across all four lotteries, average certainty equivalents are identical in the Baseline and AI treatments. Instead, the entire distribution shifts: certainty equivalents in AI treatment exhibit second-order stochastic dominance relative to the Baseline treatment. In practical terms, LLM-based recommendations leave mean valuations unchanged while compressing dispersion and hence, reducing heterogeneity. Additionally, while LLM-based recommendations shift the intensity of risk, complexity, and ambiguity aversion, they do not affect the fractions of participants classified as averse, neutral, or loving. After each compound lottery, we also elicit beliefs about the probability of winning the prize and ask participants how certain they feel about those beliefs. LLM-based recommendations improve over mistakes in belief formation caused by additional complexity, and they reduce the additional self-reported cognitive uncertainty for the more complex compound lottery.

Motivated by the evidence that people seek expert advice more when the situation is complex, and uncertainty is enhanced, Engelmann et al. (2009) use fMRI scanning of the participants and show that expert advice corrects understanding of probabilities through “offloading” the calculation burden. While this particular study did not vary the level of difficulty or complexity of the uncertainty the decision makers evaluate, the mechanism identified might be particularly relevant in complex environments where computation is costly and where recommendations may affect confidence and perceived understanding rather than correcting probabilistic reasoning. Related to this intuition, de Clippel et al. (2024) showed that individuals value risk less when they perceive a lottery as complex and ambiguous. Building on these findings, our design can address whether the impact of LLM-based recommendations is robust to variation in the complexity of uncertainty involved.

We also contribute to the recommendation literature. Previous work shows that recommendations, in the form of direct human advice and broader social interactions, systematically shape economic choice. In laboratory settings, consultation with others compresses dispersion in risk taking (Bougeas et al., 2013), observing a peer take a risky action increases risk taking (Cooper and Rege, 2011), and group decision

making can reduce ambiguity sensitivity relative to individual choice (Keck et al., 2014). In financial contexts, social interactions affect retirement plans (Duflo and Saez, 2003; Beshears et al., 2015), and field evidence highlights the role of social learning in portfolio and asset choice (Bursztyn et al., 2014). Related work in strategic environments also documents sizable effects of advice in settings such as ultimatum and coordination games (Schotter, 2003; Çelen et al., 2010). Our study of LLM-based recommendations in decisions under risk and ambiguity builds on these insights: external guidance can change the distribution of choices.

More recently, attention has shifted from human recommendations to algorithmic AI-recommendations, and to natural-language AI recommendations in the early 2020s.² Evidence from large-scale experiments suggests that robo-advisors can reduce investor mistakes (Lambrecht et al., 2024), shift the focus of human investors to sectors where they have advantages over AI (Raymond, 2024), and reduce child hospitalization when the child protective services receive risk scores from algorithms (Grimon and Mills, 2025). However, the results are mixed in different settings, as Kreitmeir and Raschky (2024) showed that better programmers achieve worse outputs when AI assistance is allowed, highlighting an important tension between algorithm aversion (Dietvorst et al., 2015) and algorithm appreciation (Logg et al., 2019). Complementary work shows that algorithmic advice may mitigate behavioral biases such as the disposition effect and trend chasing (D'Acunto et al., 2019), or introduce new biases such as psychological cost of being wrong (Almog et al., 2025). On the theory side, models have begun to formalize how recommendations can enter preferences generating over-responsiveness to recommendations (McLaughlin and Spiess, 2022).

We contribute to this recent literature by asking a basic question: how do LLM-based recommendations affect measured risk, complexity, and ambiguity attitudes in an environment free from social or strategic interactions, and does it shift valuations in one direction or reshape the distribution of valuations across individuals? Moreover, our decision environment should be minimally impacted by some of the behavioral concerns highlighted by the previous research, such as shame, pride, social image, or

²Another branch of this literature has examined the welfare gains and psychological costs that arise when AI overrules human judgments, rather than merely offering recommendations, in settings such as law (Rambachan, 2024), medicine (Raghu et al., 2019), and top-level tennis tournaments (Almog et al., 2025). Note that unlike those studies, the participants of our experiments make the final decisions.

confidence, as there is no right or wrong answer in valuing a lottery. Since LLM-based recommendations in decision making are here to stay, understanding its impact on risky choice is central for economic modeling.

2 Experiment Design

In the experiment, we measure the dollar value that participants assign to multiple lotteries. In each lottery, the computer simulates the draw of a card from a deck of green and purple cards. The bet pays \$30 if the drawn card is purple, and \$0 otherwise. This setting is used in de Clippel et al. (2024). We elicit participants' certainty equivalents for four lotteries using the standard multiple price list (MPL) procedure with enforced single switching (Holt and Laury, 2002; Andersen et al., 2006). The MPL varies the sure amounts from \$1 to \$30 with an increment of \$1 at each row. Participants face several scenarios that differ in the complexity of the lotteries (for a graphical interpretation see of the below lotteries see Figure 1):

- **50/50 Lottery:** The deck contains 50 purple and 50 green cards, making it easy to see that the probability of winning is 50%.
- **Compound Lottery, Two Decks (Compound-2):** There are two decks of cards, each of which contains 100 purple or green cards. While 20% of the cards are Purple in the first deck, the percentage of the Purple cards is 80% in the second deck. Participants are told that the computer selects one of the two decks at random with equal chance, shuffles the selected deck, and draws a card. Hence, this scenario describes a non-degenerate compound lottery that reduces to a 50/50 probability of drawing a purple card.
- **Compound Lottery, Four Decks (Compound-4):** There are four decks of cards, each of which contains 100 purple or green cards. The respective percentage of purple cards in each deck is 10%, 30%, 70%, and 90%. Participants are told that the computer selects one of the four decks at random with equal chance, shuffles the selected deck, and draws a card. Similar to the previous lottery, this is also a non-degenerate compound lottery that reduces to a 50/50 probability of drawing a purple card.

- **Ambiguous Lottery:** Participants are only told that the deck contains 100 purple or green cards, but not the exact composition, and that “they could be all Purple, all Green, or any combination.”

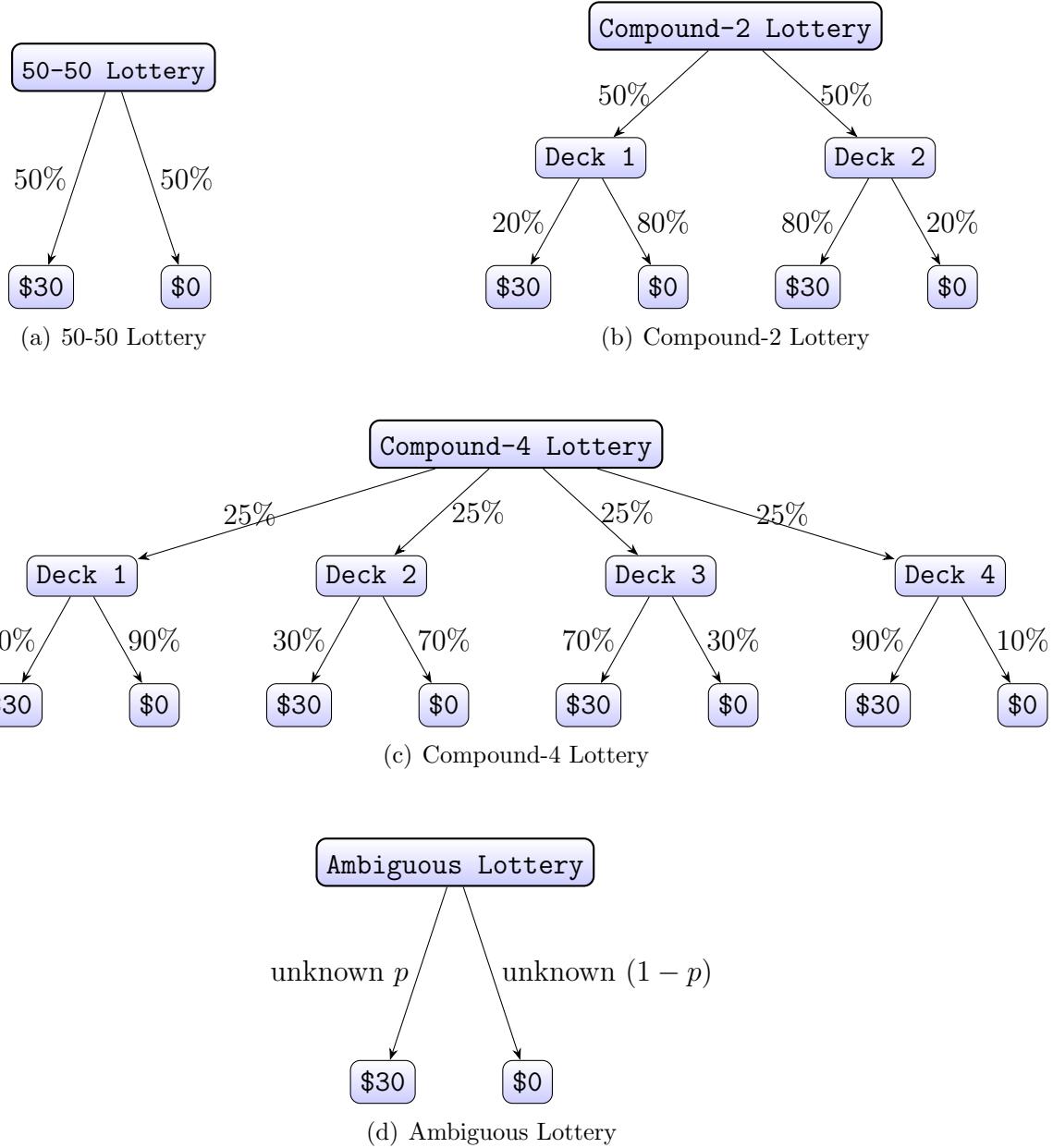


Figure 1: The Lotteries Asked

The experiment proceeds as follows. After training about scenarios and MPLs, and after comprehension quizzes, similar to de Clippel et al. (2024) participants face a 50/50 lottery, the Compound-2 lottery, the Compound-4 lottery, and, finally, the ambiguous lottery. The information for these lotteries is explained verbally (see Figure 2 for the screenshots used in the experiment). After each compound lottery question, the participants' belief about drawing a Purple card is elicited in an incentivized way, and they are asked "how certain" they felt about their beliefs.

A deck contains 50 **Purple** cards and 50 **Green** cards. The computer shuffles the deck and then draws a card.

(a) 50-50 Lottery

There are **2 decks**. Each contains **100** cards:

- In deck 1: 20% of cards are **Purple** and 80% are **Green**
- In deck 2: 80% of cards are **Purple** and 20% are **Green**

The computer **selects one of the two decks at random (each with 50% probability)**, shuffles the selected deck and then draws a card.

(b) Compound-2 Lottery

There are **4 decks**. Each contains **100** cards:

- In deck 1: 10% of cards are **Purple** and 90% are **Green**
- In deck 2: 30% of cards are **Purple** and 70% are **Green**
- In deck 3: 70% of cards are **Purple** and 30% are **Green**
- In deck 4: 90% of cards are **Purple** and 10% are **Green**

The computer **selects one of the four decks at random (each with 25% probability)**, shuffles the selected deck and then draws a card.

(c) Compound-4 Lottery

A deck contains **100** cards. Each card is either **Purple** or **Green**.

You are not told the exact number of **Purple** or **Green** cards. They could be all **Purple**, all **Green**, or any combination.

A card has already been drawn from the deck.

(d) Ambiguous Lottery

Figure 2: The Scenario of the Lotteries

We randomly assigned participants to either the Baseline or the AI treatment. The Baseline treatment replicates the procedures typically used in the existing literature employing the MPL method in eliciting certainty equivalents of lotteries. In the AI treatment, participants were provided with a lottery-specific recommendation from ChatGPT (Model GPT-4o) before they report their certainty equivalents for each lottery. All participants in the AI treatment were given the same screenshots of recommendations for the corresponding lottery.

Each LLM-based recommendation is generated as follows. Prior to the experiment, we take a screenshot of each lottery (one lottery at a time), attach the screenshot to ChatGPT, and ask “What do you recommend for someone making these choices?”. Finally, we take screenshots of both our prompt and ChatGPT’s response to this question. We clear ChatGPT’s memory before we inquire about the next lottery.

Figure 3 is the LLM-based recommendation for the Compound-2 lottery. Note that it first explains the setup and the task. Then, it explicitly calculates the winning probability and the expected value for this compound lottery. It interprets the expected value as the risk neutral switching point. However, in the practical recommendation it provides a classification of what switching at, before, and after \$15 mean in terms of risk attitudes. In Appendix A.1, we provide the screenshots of LLM-based recommendations for all questions.

We ran the experiment on the Prolific platform in October and November, 2025. The experiment was pre-registered on AsPredicted.org.³ Full experimental instructions and screenshots are provided in Appendix A.1. The mean completion time was about 11 minutes. Participants received \$3 for participation, and 10% of participants were eligible for a bonus payment based on their answer to one of the main tasks (drawn at random) following de Clippel et al. (2024). The average bonus was \$1.34. In each of the two treatments, there were 300 participants. None of the 600 participants participated in more than one session.

³The experiment protocol can be found at <https://aspredicted.org/zwmgs8yy.pdf>. The experiment was programmed and implemented using oTree (Chen et al., 2016).



Let's analyze the decision step-by-step.

★ The Setup

There are **2 decks**, each equally likely (50%) to be chosen:

Deck	% Purple	% Green
1	20%	80%
2	80%	20%

The computer picks **one deck at random**, then draws **one card**.

💰 Payoffs

You have two options in each row:

1. **\$30 if the drawn card is Purple**
2. **A certain fixed amount** (\$1 to \$30)

You're deciding when to switch from the risky option to the sure amount.

📊 Expected Value of the Risky Option

Let's compute the overall probability of drawing a Purple card:

$$P(\text{Purple}) = 0.5(0.2) + 0.5(0.8) = 0.5$$

So, there's a **50% chance** of drawing Purple.

Thus, the **expected value (EV)** of the gamble is:

$$EV = 0.5 \times 30 + 0.5 \times 0 = \$15$$

💡 Interpretation

- If you were **risk neutral**, you'd be indifferent between:
 - "\$30 if Purple"
 - "\$15 for sure"
- So, **\$15** is the fair value of the gamble.

🧠 Practical Recommendation

Here's what this means for someone making these choices:

Preference Type	What to Choose
Risk neutral	Indifferent at ~\$15 for sure
Risk averse	Switch to the sure option <i>before</i> \$15 (e.g., around \$10-\$14)
Risk seeking	Keep choosing the gamble until the sure amount is <i>above</i> \$15

Figure 3: Screenshot of ChatGPT Recommendation for Compound-2

3 Results

Participants assigned to the Baseline and AI treatments are demographically balanced: gender, race, education distributions and AI use and trust indices are similar across the two treatments (see Table 1). The main body of the paper reports results using all the data. Appendix A.2 repeats the analysis for the participants who pass our comprehension quizzes on the first attempt (82% in the Baseline treatment and 80% in the AI treatment). The results are robust for this filtered subsample.

Table 1: Demographics

	Baseline	AI	t-test (p-value)
Age	43.47	43.15	0.76
Female	0.49	0.48	0.81
White	0.80	0.77	0.32
Black	0.08	0.10	0.38
Asian	0.09	0.08	0.66
Others	0.00	0.00	1.00
Education: Less than High School	0.00	0.01	0.56
Education: High School	0.13	0.13	0.90
Education: Some College	0.24	0.27	0.46
Education: College or Higher	0.62	0.60	0.50
AI Use Index	4.74	5.02	0.25
AI Trust Index	5.16	5.52	0.10
Observations	300	300	

3.1 Treatment Effects

Our main objective is to understand how valuations of risky and ambiguous prospects change when decision makers are assisted by AI. To this end, we compare valuations elicited for each lottery across the two treatments. To convert MPL switching points into point valuations on a discrete grid, we define each participant’s valuation as the sure amount at which they switch to the sure option minus one-half of the grid step size as in de Clippel et al. (2024).

Table 2 summarizes the distribution of valuations by treatment. A clear pattern emerges: *mean valuations are unchanged by LLM-based recommendation, while dis-*

Table 2: Treatment Effect

	Baseline	AI	p-value comparing means	p-value comparing variances
Value (50-50)	13.617 (6.340)	13.590 (4.896)	0.954	0.000
Value (Compound-2)	12.310 (6.209)	12.743 (4.894)	0.343	0.000
Value (Compound-4)	12.357 (6.085)	12.813 (4.910)	0.312	0.000
Value (Ambiguous)	11.290 (6.454)	11.937 (5.243)	0.178	0.000
Observations	300	300		

Notes: In columns 1-2, we report the mean and standard deviation of values participants assign to each lottery in the Baseline and AI treatment, respectively. In columns 3-4, we report the p-values from the t-test for the mean comparisons, and the Levene's robust test for the equality of variances, respectively.

persian falls significantly. Mean differences between the Baseline and AI treatments are small and statistically insignificant for every lottery type (t-tests: all $p \geq 0.178$). In contrast, the equality-of-variance tests reject strongly for every lottery type (Levene's tests: all $p < 0.001$). Quantitatively, the standard deviation of valuations is about 19–23% lower in the AI-treatment across lotteries (50–50: 6.34 vs. 4.90; Compound-2: 6.21 vs. 4.89; Compound-4: 6.09 vs. 4.91; Ambiguous: 6.45 vs. 5.24), corresponding to a roughly 34–40% reduction in variance with LLM-based recommendation. Thus, LLM-based recommendations primarily compress the cross-participant distribution of valuations rather than shifting average certainty equivalents.

Figure 4 visualizes this compression by plotting the cumulative distribution functions (CDF) of elicited values in the two treatments for each lottery type. Across all lotteries, the CDFs in the AI treatment are steeper in the center and has thinner tails, consistent with fewer extreme valuations and a tighter distribution of risk, complexity, and ambiguity attitudes. In other words, all of the Baseline treatment distributions behave like a mean-preserving spread of the corresponding AI treatment distribution.

Applying the Barrett and Donald (2003) second-order stochastic dominance test to all CDF comparisons further validates second-order stochastic dominance between the AI and Baseline treatments distributions (using the code from Lee and Whang, 2023).

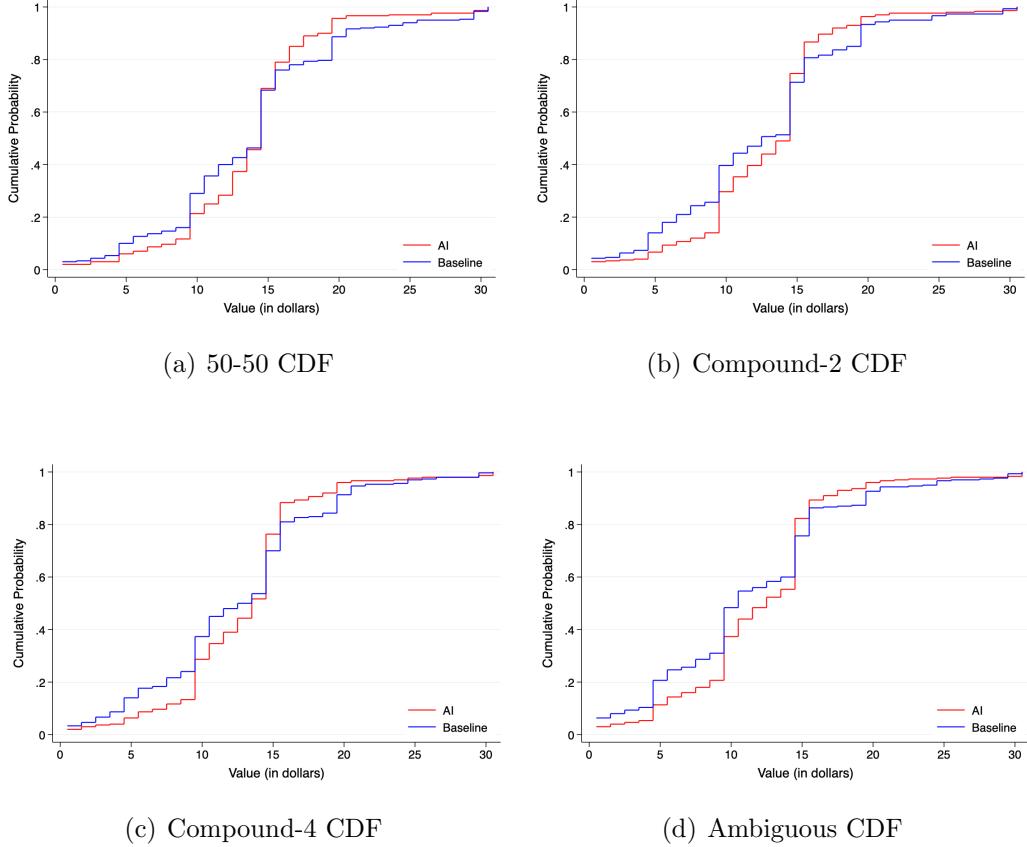


Figure 4: Baseline vs. AI Treatments, CDFs of Values

Figure 5 plots corresponding probability distribution functions (PDFs) of the values of each lottery type. Across all lotteries, the PDFs are narrower in the AI treatment complementing the results based on the CDFs.

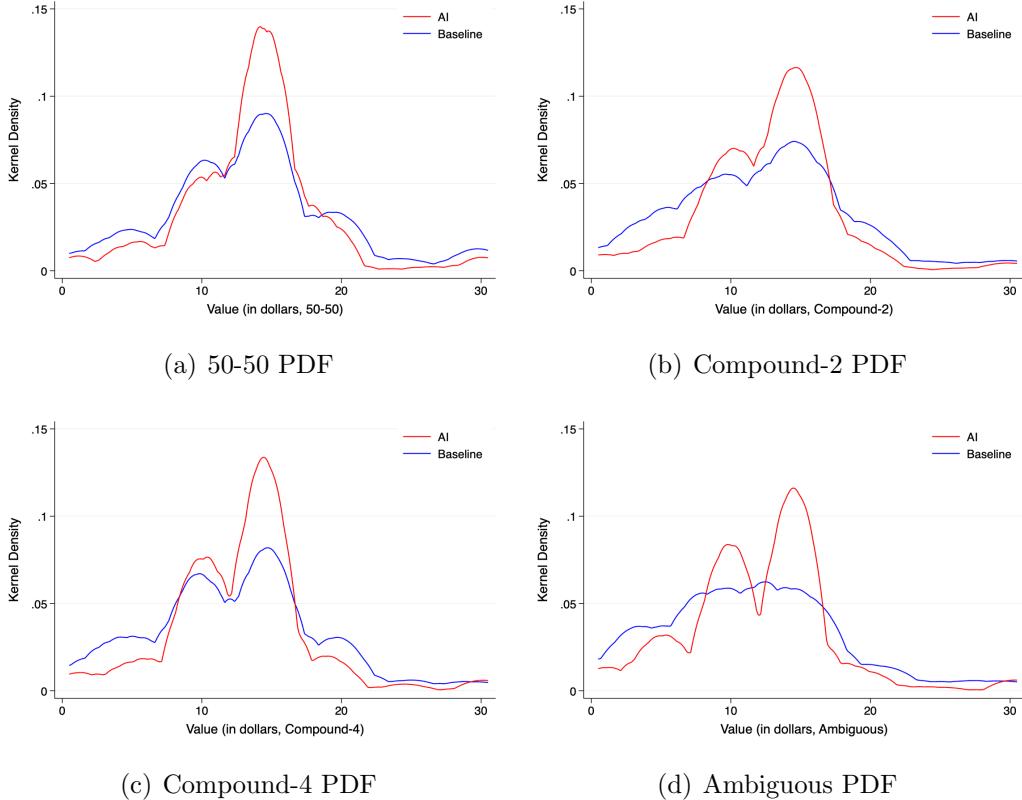


Figure 5: Baseline vs. AI Treatments, PDFs of Values

To quantify the reduction in extreme valuations controlling for the participants' observable heterogeneity, Table 3 reports regressions to measure the treatment effect on the absolute deviation of each valuation from the midpoint (15.5) of the elicitation grid for each lottery type. Consistent with the distributional evidence above, LLM-based recommendation significantly reduces absolute deviations from the neutral benchmark, indicating that the tightening of valuations reflects a reduction in extremes rather than a uniform shift.

Table 3: Regression: Distance to Risk-Neutrality Value

Dep. Variable: Value – 15.5	50-50	Compound-2	Compound-4	Ambiguous
AI	-1.39*** (0.33)	-1.36*** (0.34)	-1.23*** (0.33)	-1.30*** (0.36)
Demographics	Yes	Yes	Yes	Yes
Observations	599	599	599	599

Notes: One participant with missing demographic information is excluded from the regressions. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Extreme Percentiles and IQR

	Baseline	AI
50-50 Lottery		
10th percentile	5.00	8.50
90th percentile	20.50	19.00
IQR	6.00	4.50
Compound-2 Lottery		
10th percentile	4.50	6.50
90th percentile	19.50	17.50
IQR	7.00	6.00
Compound-4 Lottery		
10th percentile	4.50	7.50
90th percentile	19.50	17.50
IQR	6.00	5.00
Ambiguous Lottery		
10th percentile	3.50	4.50
90th percentile	19.50	16.50
IQR	8.00	5.00
Observations	300	300

Furthermore, Table 4 demonstrates 10th and 90th percentile values as well as Interquartile Range (IQR), that is the difference of the values of 75th and 25th percentiles. As can be seen, in all questions, 10th percentile values are higher in AI treatment and 90th percentile values are lower in AI treatment; additionally IQR

statistics are lower in AI treatment, providing additional support to the decrease in the variances of the distributions. Hence, the following result is robustly observed in our data:

Result 1: *The LLM-based recommendations reduce the heterogeneity in participants' lottery valuations toward the Baseline mean without shifting the mean.*

3.2 Attitudes Toward Risk, Complexity, and Ambiguity

Section 3.1 showed that LLM-based recommendations leave average valuations unchanged while compressing the distribution of valuations across participants. A natural next question is whether LLM-based recommendations mitigate risk aversion, complexity aversion, or ambiguity aversion.

In the Baseline treatment, the mean valuations for the 50–50, Compound-2, and Compound-4 lotteries reported in Table 2 are all below 15.5, the risk-neutral benchmark (Wilcoxon sign-rank test: all $p < 0.001$). The participants are ambiguity averse on average in the Baseline, too: the ambiguous lottery is valued less than the 50–50 lottery (Wilcoxon sign-rank test $p < 0.001$). Consistent with the evidence in de Clippele et al. (2024), we also observe aversion to complexity in the Baseline: The average valuations for both Compound-2 and Compound-4 are significantly lower than the average valuation of the 50–50 lottery (Wilcoxon sign-rank tests $p < 0.001$). We observe the same qualitative pattern in the AI treatment: average valuations remain below \$15.5 for the 50–50 lottery and are further discounted for compound risk and ambiguity relative to the 50–50 benchmark, indicating risk aversion, complexity aversion, and ambiguity aversion (Wilcoxon sign-rank tests all $p < 0.001$).

Complexity and ambiguity aversion is robust to controlling the demographics. Table 5 estimates regressions where the 50–50 lottery is the control category. In the Baseline treatment, valuations for Compound-2 and Compound-4 are significantly lower than for the 50–50 lottery, consistent with *complexity aversion*: holding expected value fixed, participants value compound risk less than simple risk. Valuations for the Ambiguous lottery are also lower than for the 50–50 lottery, consistent with *ambiguity aversion* on average. The treatment variable (AI treatment dummy

Table 5: Regression: Lottery Value

	(1)	(2)
Compound-2	-1.08*** (0.19)	-1.08*** (0.19)
Compound-4	-1.02*** (0.19)	-1.02*** (0.20)
Ambiguous	-1.99*** (0.22)	-1.99*** (0.22)
AI	0.38 (0.39)	0.32 (0.39)
Constant	13.41*** (0.33)	13.80*** (2.41)
Demographics		Yes
Observations	2,400	2,396

Notes: The 50-50 lottery value is the comparison group. One participant with missing demographic information is excluded from the regression. Standard errors are clustered at the subject level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

variable) is statistically insignificant, consistent with the results of Section 3.1.

Table 6 classifies participants by their risk, complexity, and ambiguity attitudes in both the Baseline and AI treatments, and the shares in each category are very similar across treatments. For example, 68.3% of participants are classified as risk averse in the Baseline, compared with 69.0% in the AI treatment. These results indicate that AI assistance does not significantly change attitude classifications toward risk, complexity, or ambiguity (Fisher's exact tests $p = 0.485$ in Risk, $p = 0.589$ in Compound-2, $p = 0.813$ in Compound-4, $p = 0.286$ in Ambiguity).

Table 6: Preference Classification, Baseline vs. AI

	Baseline			AI		
	Averse	Neutral	Loving	Averse	Neutral	Loving
Risk	68.3%	7.7%	24.0%	69.0%	10.0%	21.0%
Compound-2	40.0%	37.0%	23.0%	36.0%	38.7%	25.3%
Compound-4	39.7%	37.3%	23.0%	38.0%	40.0%	22.0%
Ambiguity	50.7%	30.7%	18.7%	51.7%	25.7%	22.7%

While classifications are stable, the magnitudes of these attitudes shift under LLM-based recommendations. Figures 6a–d plot CDFs of risk, complexity, and ambiguity aversion measures of the participants who are classified as averse, separately. For instance, Figure 6a shows that, among participants classified as risk averse, the distribution of deviations below the risk neutral benchmark in the AI treatment first-order stochastically dominates the corresponding distribution in the Baseline treatment. In other words, risk averse participants submit systematically higher valuations under LLM-based recommendations (Mann-Whitney test $p = 0.026$). The same qualitative pattern appears for complexity averse participants (Figures 6b and 6c; Mann-Whitney tests $p = 0.029$ and $p = 0.001$, respectively), ambiguity averse participants (Figure 6d; Mann-Whitney test $p = 0.008$). Overall, LLM-based recommendations compress the distribution of attitude magnitudes: it makes risk-, complexity-, and ambiguity-averse participants less extreme, without changing their classification.

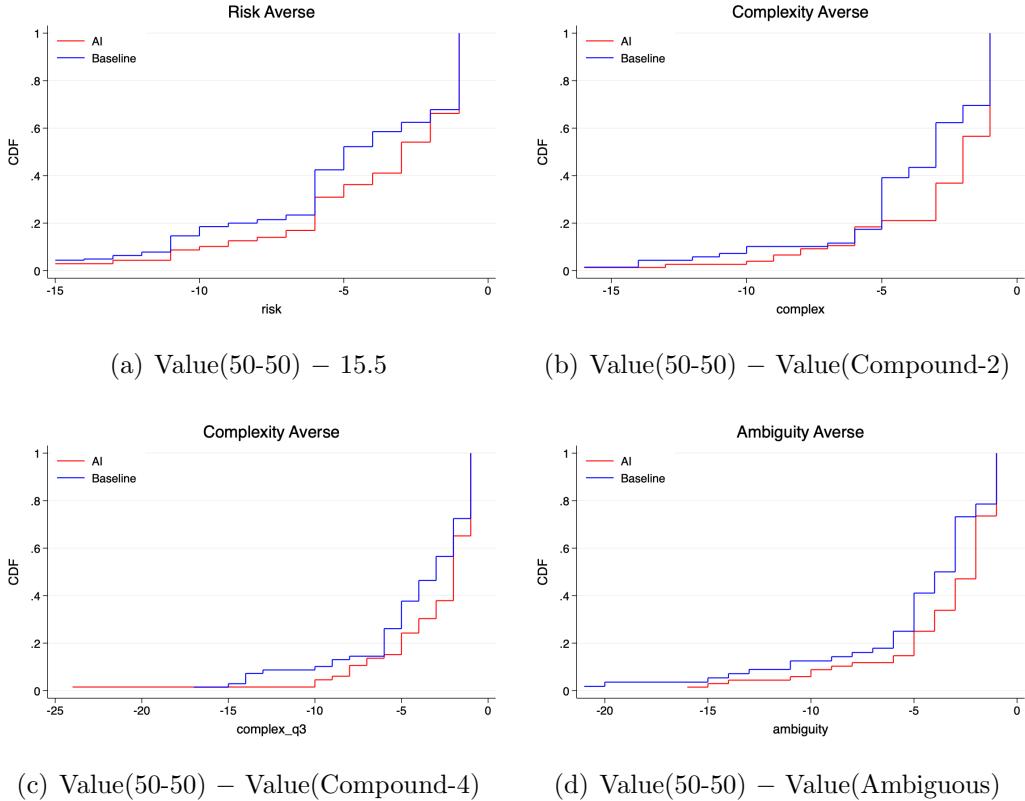


Figure 6: CDFs of the Magnitudes of Risk, Complexity, and Ambiguity Aversion

Result 2: *The LLM-based recommendations reduce the intensity of risk, complexity, and ambiguity aversion while leaving the distributions of classification as averse/neutral/loving unchanged.*

Recall that participants also reported how certain they were about their stated beliefs on a 0–100% scale. Following de Clippel et al. (2024), we define cognitive uncertainty as 100 minus the participant’s certainty score. Table 7 shows that, in the Baseline treatment, the percentage of correct beliefs are lower for the more complex Compound-4 lottery than for the Compound-2 lottery (0.70 vs 0.77, Wilcoxon signed-rank test $p < 0.001$). On the other hand, the cognitive uncertainty is significantly higher for the more complex Compound-4 lottery than for the Compound-2 lottery (42.10 vs. 37.13, Wilcoxon signed-rank test $p < 0.001$). This indicates that, without LLM-based recommendations, participants found it harder to assign beliefs to a more complex lottery and were less confident in their answers. In contrast, LLM-based recommendation eliminates this gap: in the AI treatment, uncertainty scores are similar for Compound-2 and Compound-4 (36.43 vs 35.80, Wilcoxon signed-rank test $p = 0.583$). This is consistent with accuracy patterns in the AI treatment, where the rates of correct beliefs are also similar across the two lotteries (0.73 vs. 0.75, Wilcoxon signed-rank test $p = 0.556$).

Table 7: Belief and Cognitive Uncertainty

	Correct Belief		Cognitive Uncertainty	
	Compound-2	Compound-4	Compound-2	Compound-4
Baseline	0.77	0.70	37.13	42.10
AI	0.73	0.75	36.43	35.80

Notes: $N = 300$ in each treatment.

For robustness, for each participant we calculate the difference in cognitive uncertainty between Compound-2 and Compound-4 and regress this difference on AI treatment dummy variable. As can be seen in Table 8, AI dummy variable is significant even when the demographics are controlled, indicating that LLM-based recom-

mendations reduce the perceived complexity.

Table 8: Regression: Change in Cognitive Uncertainty (Compound-4 Minus Compound-2)

	(1)	(2)
AI	-5.60*** (1.51)	-5.68*** (1.52)
Demographics		Yes
Observations	600	599

Notes: One participant with missing demographic information is excluded from the regression. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4 Conclusion

This paper experimentally studies how LLM-based recommendations affect evaluations of uncertainty. Our main result is that providing participants with lottery-specific recommendations from a large language model changes valuations in a second-order stochastic dominance sense: across lotteries, LLM-based recommendation leaves average certainty equivalents unchanged, but it systematically reduces dispersion.

Our findings have two direct implications for economic research and practice. First, they speak directly to the measurement of preferences. Risk and ambiguity preferences are routinely elicited to discipline models, calibrate welfare, and predict behavior in markets and policy settings. Our evidence indicates that when an economically relevant choice environment includes readily available LLM-based recommendations, as it increasingly does, standard elicitation procedures may recover preferences as expressed with AI mediation. This may directly impact risk profiling procedures, which are frequently used by investment platforms.

Second, these results contribute to the economics of complexity in decision making under uncertainty. We compare lotteries that imply the same objective risk but differ in computational burden, and we also consider lotteries that introduce ambiguity. Across these settings, we find that the first result above continues to hold robustly. Since real-world decision environments are often more complex and complexity itself is

one of the key drivers of AI adoption, our finding that LLM-based recommendations reduce heterogeneity in the magnitude of risk and ambiguity attitudes has direct implications for applied contexts, including insurance choice and retirement planning.

Finally, as we implement a tightly controlled LLM-based recommendation treatment in which all participants view the same pre-generated GPT screenshots, produced from a single prompt (“What do you recommend for someone making these choices?”), it strengthens internal validity by eliminating endogenous prompting and heterogeneity in prompt quality. It also enhances replicability by fixing the advice content. Importantly, exposure to identical recommendations does not mechanically anchor participants to the same response. Instead, LLM-based recommendation preserves the qualitative structure of heterogeneity in risk, complexity, and ambiguity attitudes, as participants remain meaningfully dispersed and ordered in their preferences while attenuating the magnitudes of these attitudes by compressing the tails of the valuation distribution.

A natural next step is to endogenize AI use by allowing participants to decide whether to consult the AI and to enter their own prompts. Since take-up and prompt content will vary across individuals, and some may opt out entirely, our current design provides a benchmark. Future work could vary the calibration of advice and allow interactive queries to assess how much of the observed behavioral change reflects the availability of AI per se versus differences in prompting and selection into use. More broadly, as AI systems become embedded in consumer finance, insurance markets, and medical decision support, understanding how recommendations reshape decisions under uncertainty and what aspects of preferences they change will be essential for both modeling and policy design.

References

Almog, D., Gauriot, R., Page, L., and Martin, D. (2025). Human responses to ai oversight: Evidence from centre court. Technical report, Mimeo.

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, 9(4):383–405.

Barrett, G. F. and Donald, S. G. (2003). Consistent tests for stochastic dominance. *Econometrica*, 71(1):71–104.

Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C., and Milkman, K. L. (2015). The effect of providing peer information on retirement savings decisions. *The Journal of Finance*, 70(3):1161–1201.

Bougeas, S., Nieboer, J., and Sefton, M. (2013). Risk-taking in social settings: Group and peer effects. *Journal of Economic Behavior & Organization*, 92:273–283.

Bursztyn, L., Ederer, F., Ferman, B., and Yuchtman, N. (2014). Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica*, 82(4):1273–1301.

Çelen, B., Kariv, S., and Schotter, A. (2010). An experimental test of advice and social learning. *Management Science*, 56(10):1687–1701.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Cooper, D. J. and Rege, M. (2011). Misery loves company: Social regret and social interaction effects in choices under risk and uncertainty. *Games and Economic Behavior*, 73(1):91–110.

de Clippel, G., Moscariello, P., Ortoleva, P., and Rozen, K. (2024). Caution in the face of complexity. Technical report, Mimeo.

Dietvorst, B. J., Simmons, J. P., and Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1):114–126.

Duflo, E. and Saez, E. (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *The Quarterly Journal of Economics*, 118(3):815–842.

D'Acunto, F., Prabhala, N., and Rossi, A. G. (2019). The promises and pitfalls of robo-advising. *The Review of Financial Studies*, 32(5):1983–2020.

Engelmann, J. B., Capra, C. M., Noussair, C., and Berns, G. S. (2009). Expert financial advice neurobiologically “offloads” financial decision-making under risk. *PLOS ONE*, 4(3):e4957.

Grimon, M.-P. and Mills, C. (2025). Better together?: A field experiment on human-algorithm interaction in child protection. Technical report, Working Paper.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655.

Keck, S., Diecidue, E., and Budescu, D. V. (2014). Group decisions under ambiguity: Convergence to neutrality. *Journal of Economic Behavior & Organization*, 103:60–71.

Kreitmeir, D. and Raschky, P. A. (2024). The heterogeneous productivity effects of generative ai. Technical report, arXiv preprint arXiv:2403.01964.

Lambrecht, M., Oechssler, J., and Weidenholzer, S. (2024). On the benefits of robo-advice in financial markets. Conference paper, Verein für Socialpolitik / German Economic Association. Beiträge zur Jahrestagung des Vereins für Socialpolitik 2024; version dated December 20, 2023.

Lee, K. and Whang, Y.-J. (2023). Pysdtest: a python/stata package for stochastic dominance tests. *arXiv preprint arXiv:2307.10694*.

Logg, J. M., Minson, J. A., and Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151:90–103.

McLaughlin, B. and Spiess, J. (2022). Algorithmic assistance with recommendation-dependent preferences. arXiv preprint.

Raghu, M., Blumer, K., Corrado, G., Kleinberg, J., Obermeyer, Z., and Mullainathan, S. (2019). The algorithmic automation problem: Prediction, triage, and human effort. Technical report, Mimeo.

Rambachan, A. (2024). Identifying prediction mistakes in observational data. *The Quarterly Journal of Economics*, 139(3):1665–1711.

Raymond, L. (2024). The market effects of algorithms. Technical report, Working Paper.

Schotter, A. (2003). Decision making with naive advice. *American Economic Review*, 93(2):196–201.

A Supplemental Appendix

A.1 Screenshots

Instructions

Thank you for taking part in this study. You will get \$3 for completing the study, and 10% of participants will be randomly selected to get an **additional bonus**. You must complete the study to receive payments.

Answering carefully is in your best interest. If you are selected for the additional bonus, one of the questions you answered will be randomly selected, and your bonus will be **your earnings in that question**.

Next

Figure A.1: Instructions: Part 1

Instructions

Please **read these instructions carefully**. There will be a **short quiz** at the end.

You will face several scenarios. In each scenario, there are **one or more card decks** and **one card is randomly drawn**. The cards in the decks can have different colors.

Each scenario gives you some information about the composition of the decks. For example, you may face the following:

A deck contains **50 Green** cards and **50 Purple** cards.

The computer shuffles the deck and then draws a card.

Or you could face the following:

There are **2 decks**. Each contains **50** cards:

- In deck 1: 50% of cards are **Purple** and 50% are **Green**
- In deck 2: 50% of cards are **Purple** and 50% are **Green**

The computer **selects one of the two decks at random**, shuffles the selected deck and then draws a card.

The decks are **different** in each scenario, so read carefully.

Next

Figure A.2: Instructions: Part 2

Instructions

In some scenarios, you are asked to choose between a fixed amount of money received for sure or winning **\$30** if the drawn card is **Purple**.

For example, you might be asked to choose between:

\$30 if drawn card is Purple	<input type="radio"/> <input type="radio"/>	\$15 for sure
-------------------------------------	---	---------------

- If you select the left option, you win **\$30 if the drawn card is Purple** and **\$0** otherwise.
- If you select the right option, you win **\$15 for sure**, regardless of the color of the drawn card.

You must answer a list of questions like this. For example:

\$30 if drawn card is Purple	<input type="radio"/> <input type="radio"/>	\$11 for sure
\$30 if drawn card is Purple	<input type="radio"/> <input type="radio"/>	\$12 for sure
\$30 if drawn card is Purple	<input type="radio"/> <input type="radio"/>	\$13 for sure
\$30 if drawn card is Purple	<input type="radio"/> <input type="radio"/>	\$14 for sure
\$30 if drawn card is Purple	<input type="radio"/> <input type="radio"/>	\$15 for sure

The option on the left does not change, while the option on the right **gets better as you go down the list**. You must make a **choice in all rows**, but for simplicity, you only have to click once on the row where you want to **switch from left to right**. You can modify your choice as many times as you want, and you can also select only the left or only the right.

Intuitively, you can think about how much you'd value the bet. Then, you switch to the sure amount as soon as it is above the amount you'd value the bet.

Below is an example for you to experience the interface (try to click).

\$30 if drawn card is Purple	<input type="radio"/> <input type="radio"/>	\$0 for sure
\$30 if drawn card is Purple	<input type="radio"/> <input type="radio"/>	\$7 for sure
\$30 if drawn card is Purple	<input type="radio"/> <input type="radio"/>	\$12 for sure
\$30 if drawn card is Purple	<input type="radio"/> <input type="radio"/>	\$18 for sure
\$30 if drawn card is Purple	<input type="radio"/> <input type="radio"/>	\$20 for sure

[Next](#)

Figure A.3: Instructions: Part 3

Quiz 1

There are **2 decks**, each containing **50** cards:

- In deck 1: **90%** of cards are **Green** and **10%** are **Purple**
- In deck 2: **90%** of cards are **Green** and **10%** are **Purple**

The computer **selects one of the two decks at random**, shuffles the selected deck and then draws a card. Which statement is correct?

- The drawn card must be Purple.
- The drawn card must be Green.
- The drawn card can be Purple or Green.

[Next](#)

Figure A.4: Quiz 1

Quiz 2

Which do you choose?

\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$1 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$2 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$3 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$4 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$5 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$6 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$7 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$8 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$9 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$10 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$11 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$12 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$13 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$14 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$15 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$16 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$17 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$18 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$19 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$20 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$21 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$22 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$23 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$24 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$25 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$26 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$27 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$28 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$29 for sure
\$5.5 for sure	<input type="radio"/>	<input type="radio"/>	\$30 for sure

Next

Figure A.5: Quiz 2

\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$1 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$2 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$3 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$4 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$5 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$6 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$7 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$8 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$9 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$10 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$11 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$12 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$13 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$14 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$15 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$16 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$17 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$18 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$19 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$20 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$21 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$22 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$23 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$24 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$25 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$26 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$27 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$28 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$29 for sure
\$30 if drawn card is Purple	<input type="radio"/>	<input type="radio"/>	\$30 for sure

Figure A.6: Choice for Each Question (Omitted in Screenshots Below)

Question 1

A deck contains 50 **Purple** cards and 50 **Green** cards. The computer shuffles the deck and then draws a card.

Which do you choose?

Figure A.7: 50-50 (Baseline)

Question 2

There are **2 decks**. Each contains **100** cards:

- In deck 1: 20% of cards are **Purple** and 80% are **Green**
- In deck 2: 80% of cards are **Purple** and 20% are **Green**

The computer **selects one of the two decks at random (each with 50% probability)**, shuffles the selected deck and then draws a card.

Which do you choose?

Figure A.8: Compound-2 (Baseline)

Question 2 (Continued)

Recall the previous question:

There are **2 decks**. Each contains **100** cards:

- In deck 1: 20% of cards are **Purple** and 80% are **Green**
- In deck 2: 80% of cards are **Purple** and 20% are **Green**

The computer **selects one of the two decks at random (each with 50% probability)**, shuffles the selected deck and then draws a card.

Using the laws of probability, it is possible to calculate the **Exact Chance** that the drawn card is **Purple**. What is this **Exact Chance** (in %)? You can earn a \$5 bonus with your answer. Your probability of winning goes up the more accurate your answer is, using the formula explained here. [Link](#)

How certain are you that your answer above is the **Exact Chance** that the drawn card is **Purple**, calculated using the laws of probability?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

[Next](#)

Figure A.9: Compound-2 Belief Elicitation

Question 3

There are **4 decks**. Each contains **100** cards:

- In deck 1: 10% of cards are **Purple** and 90% are **Green**
- In deck 2: 30% of cards are **Purple** and 70% are **Green**
- In deck 3: 70% of cards are **Purple** and 30% are **Green**
- In deck 4: 90% of cards are **Purple** and 10% are **Green**

The computer **selects one of the four decks at random (each with 25% probability)**, shuffles the selected deck and then draws a card.

Which do you choose?

Figure A.10: Compound-4 (Baseline)

Question 3 (Continued)

Recall the previous question:

There are **4 decks**. Each contains **100** cards:

- In deck 1: 10% of cards are **Purple** and 90% are **Green**
- In deck 2: 30% of cards are **Purple** and 70% are **Green**
- In deck 3: 70% of cards are **Purple** and 30% are **Green**
- In deck 4: 90% of cards are **Purple** and 10% are **Green**

The computer **selects one of the four decks at random (each with 25% probability)**, shuffles the selected deck and then draws a card.

Using the laws of probability, it is possible to calculate the **Exact Chance** that the drawn card is **Purple**. What is this **Exact Chance** (in %)? You can earn a \$5 bonus with your answer. Your probability of winning goes up the more accurate your answer is, using the formula explained here. [Link](#)

How certain are you that your answer above is the **Exact Chance** that the drawn card is **Purple**, calculated using the laws of probability?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

[Next](#)

Figure A.11: Compound-4 Belief Elicitation

Question 4

A deck contains **100** cards. Each card is either **Purple** or **Green**.

You are not told the exact number of **Purple** or **Green** cards. They could be all **Purple**, all **Green**, or any combination.

A card has already been drawn from the deck.

Which do you choose?

Figure A.12: Ambiguous (Baseline)

Question 1

A deck contains 50 **Purple** cards and 50 **Green** cards. The computer shuffles the deck and then draws a card.

For the following question, before you decide what to choose, please see ChatGPT's recommendation for this question:



The screenshot shows a ChatGPT interface with the following content:

What do you recommend for someone making these choices?

💡 The setup
This question is a **decision-making or risk preference** test—similar to those used in behavioral economics or experimental psychology. Here's how to think about it and what I recommend:

💡 How to choose
If you are **risk-neutral**, you should be indifferent between:
- the gamble, and
- a sure \$15.
So, you would switch from choosing the gamble to choosing the sure thing **around \$15**.

🔥 If you're risk-averse
Most people are a little risk-averse. That means they might prefer a sure \$13 or \$14 over a 50-50 chance at \$30. In that case, your "switch point" (where you move from gamble to sure money) might be slightly below \$15.

🔥 If you're risk-seeking
If you like risk, you might still choose the gamble even when the sure amount is \$16, \$17, or higher.

💡 Recommendation
Here's a practical guideline:
- **Below \$13** → Choose "30 if Purple" (good risk/reward).
- **\$13-\$17** → Either could make sense; depends on your comfort with risk.
- **Above \$17** → Take the sure money (statistically better for most people).

Which do you choose?

Figure A.13: 50-50 (AI treatment)

Question 2

There are **2 decks**. Each contains **100** cards:

- In deck 1: 20% of cards are **Purple** and 80% are **Green**
- In deck 2: 80% of cards are **Purple** and 20% are **Green**

The computer **selects one of the two decks at random (each with 50% probability)**, shuffles the selected deck and then draws a card.

For the following question, before you decide what to choose, please see ChatGPT's recommendation for this question:

Let's analyze the decision step-by-step.

The Setup
There are **2 decks**, each equally likely (50%) to be chosen:

Deck	% Purple	% Green
1	20%	80%
2	80%	20%

The computer picks **one deck at random**, then draws **one card**.

Payoffs
You have two options in each row:
 1. **\$30 if the drawn card is Purple**
 2. **A certain fixed amount (\$1 to \$30)**
 You're deciding when to switch from the risky option to the sure amount.

Expected Value of the Risky Option
Let's compute the overall probability of drawing a Purple card:

$$P(\text{Purple}) = 0.5(0.2) + 0.5(0.8) = 0.5$$

 So, there's a **50% chance** of drawing Purple.
 Thus, the **expected value (EV)** of the gamble is:

$$EV = 0.5 \times 30 + 0.5 \times 0 = \$15$$

Interpretation
 If you were **risk neutral**, you'd be indifferent between:
 - "30 if Purple"
 - "15 for sure"
 So, **\$15** is the fair value of the gamble.

Practical Recommendation
Here's what this means for someone making these choices:

Preference Type	What to Choose
Risk neutral	Indifferent at $\sim \$15$ for sure
Risk averse	Switch to the sure option above \$15 (e.g., around \$10-\$14)
Risk seeking	Keep choosing the gamble until the sure amount is above \$15



What do you recommend for someone making these choices?

Which do you choose?

Figure A.14: Compound-2 (AI treatment)

29

Question 3

There are **4 decks**. Each contains **100** cards:

- In deck 1: 10% of cards are **Purple** and 90% are **Green**
- In deck 2: 30% of cards are **Purple** and 70% are **Green**
- In deck 3: 70% of cards are **Purple** and 30% are **Green**
- In deck 4: 90% of cards are **Purple** and 10% are **Green**

The computer **selects one of the four decks at random (each with 25% probability)**, shuffles the selected deck and then draws a card.

For the following question, before you decide what to choose, please see ChatGPT's recommendation for this question:



Let's analyze this step by step.

The Setup
There are **4 decks**, each equally likely (25% chance each):

Deck	% Purple	% Green
1	10%	90%
2	30%	70%
3	70%	30%
4	90%	10%

The computer:

1. Randomly selects one of the 4 decks (each with 25% chance).
2. Then draws a single card.

Probability that the drawn card is Purple
We can compute the **overall probability** of drawing a Purple card using the law of total probability:

$$P(\text{Purple}) = 0.25(0.10 + 0.30 + 0.70 + 0.90) = 0.25(2.00) = 0.50$$

So, there's a **50% chance** the card will be Purple.

Expected Value (EV) of the gamble
The gamble pays **\$30** if the card is Purple, and **\$0** otherwise.

$$EV = 0.5 \times \$30 + 0.5 \times \$0 = \$15$$

So the gamble's **expected value** = **\$15**.

Recommendation

- If you are **risk neutral**, you should be **indifferent** between the gamble and a sure \$15.
- If you are **risk averse** (prefer certainty), you might prefer a **sure amount slightly below \$15** (e.g., \$10-\$14) instead of the gamble.
- If you are **risk seeking**, you might prefer the **gamble** even if the sure offer is a bit above \$15.

Summary Choice Guidance

Preference Type	Recommended Choice
Risk-neutral	Indifferent at \$15 for sure
Risk-averse	Take the sure option if $\geq \$10-\14
Risk-seeking	Take the gamble unless sure option $\approx \$16-\20

Which do you choose?

Figure A.15: Compound-4 (AI treatment)

Question 4

A deck contains **100** cards. Each card is either **Purple** or **Green**.

You are not told the exact number of **Purple** or **Green** cards. They could be all **Purple**, all **Green**, or any combination.

A card has already been drawn from the deck.

For the following question, before you decide what to choose, please see ChatGPT's recommendation for this question:



This question is designed to probe **decision-making under uncertainty**, specifically exploring **ambiguity aversion**—a well-known concept in behavioral economics and psychology. Here's how to think about it and what to consider when making these choices.

➊ The Setup

- You don't know the **ratio of Purple to Green cards** in the deck of 100.
- You are asked to choose between:
 - A **risky option**, "\$30 if the drawn card is Purple"
 - A **sure payoff**, "X for sure", where X ranges from \$1 to \$30.

➋ What This Measures

This is a classic **Ellsberg-type task**, testing how much **ambiguity aversion** you have:

- **Risk** involves known probabilities (e.g., 50% chance).
- **Ambiguity** involves unknown probabilities (e.g., you don't know if there are any Purple cards at all).
- By choosing the sure option over the ambiguous gamble, you're revealing:
 - Your **subjective belief** about the proportion of Purple cards.
 - Or your **preference for certainty**, even if the gamble might be favorable.

➌ How to Approach It Strategically

Here are a few approaches depending on your mindset:

1. Maximizing Expected Value (EV)

If you're rational and indifferent to ambiguity:

- You might assume the **expected proportion of Purple is 50%**.
- EV of the gamble = $50\% \times \$30 = \15
- So, you'd pick the gamble if the sure amount is **less than \$15**, and pick the sure amount if it's more.

2. Ambiguity Averse

If you're like most people and ambiguity makes you uneasy:

- You might require a **higher sure value** to avoid the gamble.
- For example, you might avoid the gamble even if the sure thing is only \$10 or \$12.

3. Ambiguity Seeking (Less Common)

If you enjoy risk and ambiguity, or believe the chance of Purple is high:

- You might take the gamble unless the sure offer is very close to \$30.

➍ What to Choose?

Here's a simplified recommendation:

Sure Value	Choose...
\$1-\$9	Gamble (too low to avoid)
\$10-\$14	Depends on your ambiguity tolerance
\$15+	Sure thing (expected value or higher)

If you're uncertain about the composition of the deck and want to play it safe, choose the **sure amount once it hits \$15 or more**.

Which do you choose?

Figure A.16: Ambiguous (AI treatment)

A.2 Additional Results

Filtered Sample. In this section, we repeat all the tables and figures in the main text for the participants who pass the comprehension quizzes on the first attempt.

Table A1: Demographics (Filtered Sample)

	Baseline	AI	t-test (p-value)
Age	43.39	42.68	0.52
Female	0.47	0.48	0.87
White	0.84	0.77	0.05
Black	0.05	0.10	0.07
Asian	0.08	0.07	0.93
Others	0.00	0.00	0.99
Education: Less than High School	0.00	0.01	0.55
Education: High School	0.13	0.11	0.47
Education: Some College	0.24	0.27	0.50
Education: College or Higher	0.62	0.61	0.83
AI Use Index	4.74	5.01	0.32
AI Trust Index	5.07	5.45	0.12
Observations	246	240	

Table A2: Treatment Effect (Filtered Sample)

	Baseline	AI	p-value comparing means	p-value comparing variances
Value (50-50)	13.159 (5.487)	13.329 (4.763)	0.715 0.003
Value (Compound-2)	12.354 (5.955)	12.583 (4.836)	0.641 0.000
Value (Compound-4)	12.321 (5.864)	12.725 (5.045)	0.417 0.001
Value (Ambiguous)	11.280 (6.395)	11.542 (5.019)	0.617 0.000
Observations	246	240		

Table A3: Regression: Distance to Risk-Neutrality Value (Filtered Sample)

Dep. Variable: Value – 15.5	50-50	Compound-2	Compound-4	Ambiguous
AI	-0.92*** (0.34)	-1.17*** (0.37)	-0.96*** (0.36)	-1.07*** (0.40)
Demographics	Yes	Yes	Yes	Yes
Observations	485	485	485	485

Table A4: Extreme Percentiles and IQR (Filtered Sample)

	Baseline	AI
50-50 Lottery		
10th percentile	5.50	8.50
90th percentile	19.50	17.50
IQR	6.00	5.00
Compound-2 Lottery		
10th percentile	4.50	6.00
90th percentile	19.50	16.50
IQR	6.00	5.00
Compound-4 Lottery		
10th percentile	4.50	6.50
90th percentile	19.50	17.50
IQR	6.00	5.00
Ambiguous Lottery		
10th percentile	4.50	4.50
90th percentile	19.50	15.50
IQR	9.00	5.00
Observations	246	240

Table A5: Regression: Lottery Value (Filtered Sample)

	(1)	(2)
Compound-2	-0.78*** (0.18)	-0.78*** (0.18)
Compound-4	-0.72*** (0.19)	-0.72*** (0.19)
Ambiguous	-1.83*** (0.22)	-1.84*** (0.22)
AI	0.27 (0.43)	0.18 (0.42)
Constant	13.11*** (0.34)	11.91*** (1.39)
Demographics		Yes
Observations	1,944	1,940

Table A6: Preference Classification, Baseline vs. AI (Filtered Sample)

	Baseline			AI		
	Averse	Neutral	Loving	Averse	Neutral	Loving
Risk	69.5%	8.9%	21.5%	72.1%	9.6%	18.3%
Complexity-2	38.6%	37.4%	24.0%	35.4%	42.9%	21.7%
Complexity-4	37.4%	39.4%	23.2%	36.2%	42.1%	21.7%
Ambiguity	48.8%	32.9%	18.3%	54.2%	25.8%	20.0%

Table A7: Belief and Cognitive Uncertainty (Filtered Sample)

	Correct Belief		Cognitive Uncertainty	
	Compound-2	Compound-4	Compound-2	Compound-4
Baseline	0.79	0.72	37.32	42.80
AI	0.75	0.76	35.54	35.58

Table A8: Regression: Change in Cognitive Uncertainty (Compound-4 Minus Compound-2) (Filtered Sample)

	(1)	(2)
AI	-5.45*** (1.67)	-5.54*** (1.67)
Demographics		Yes
Observations	486	485

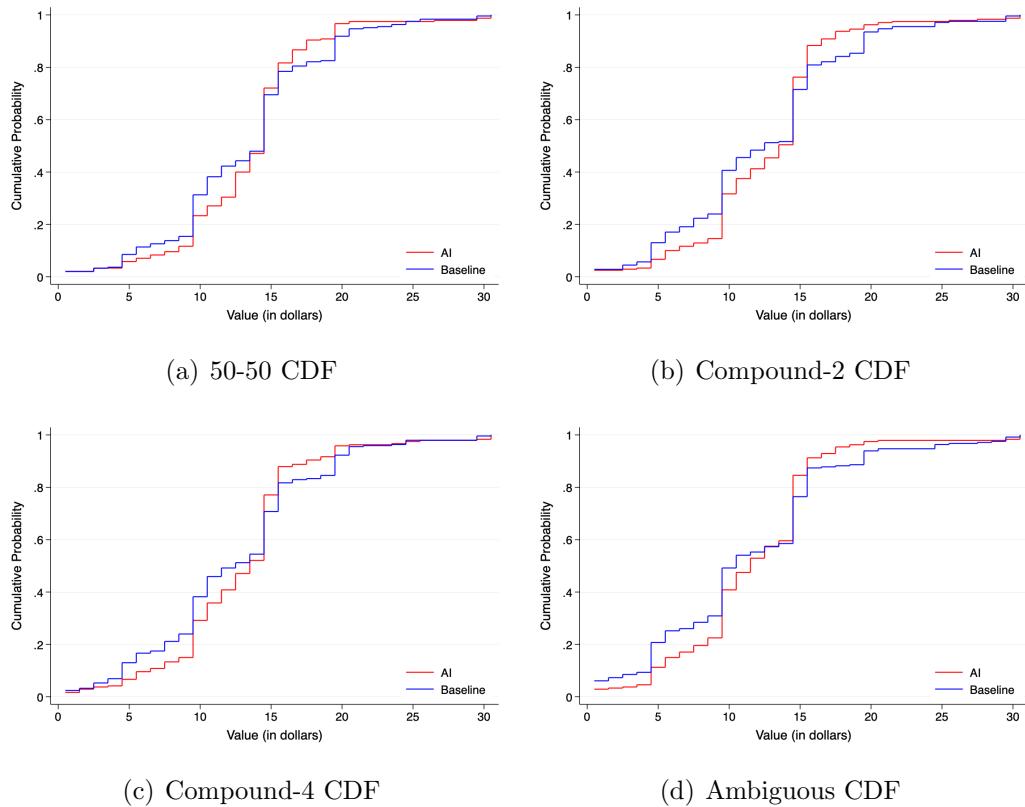


Figure A.17: Baseline vs. AI Treatments, CDFs of Values (Filtered Sample)

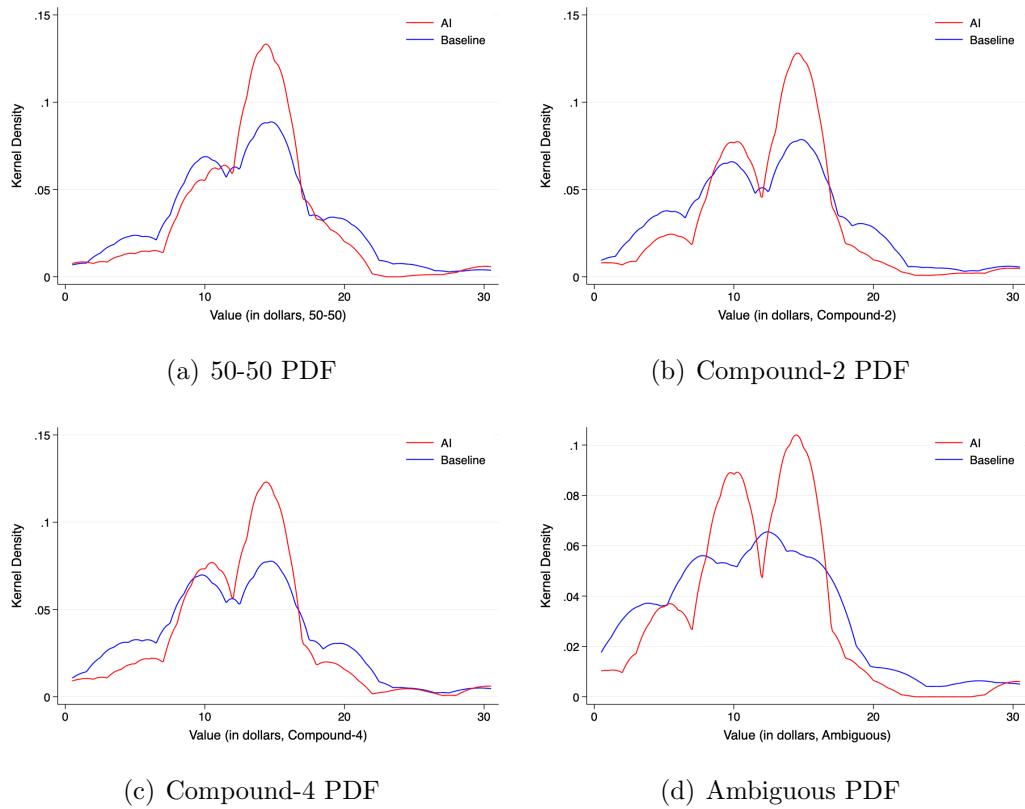


Figure A.18: Baseline vs. AI Treatments, PDFs of Values (Filtered Sample)

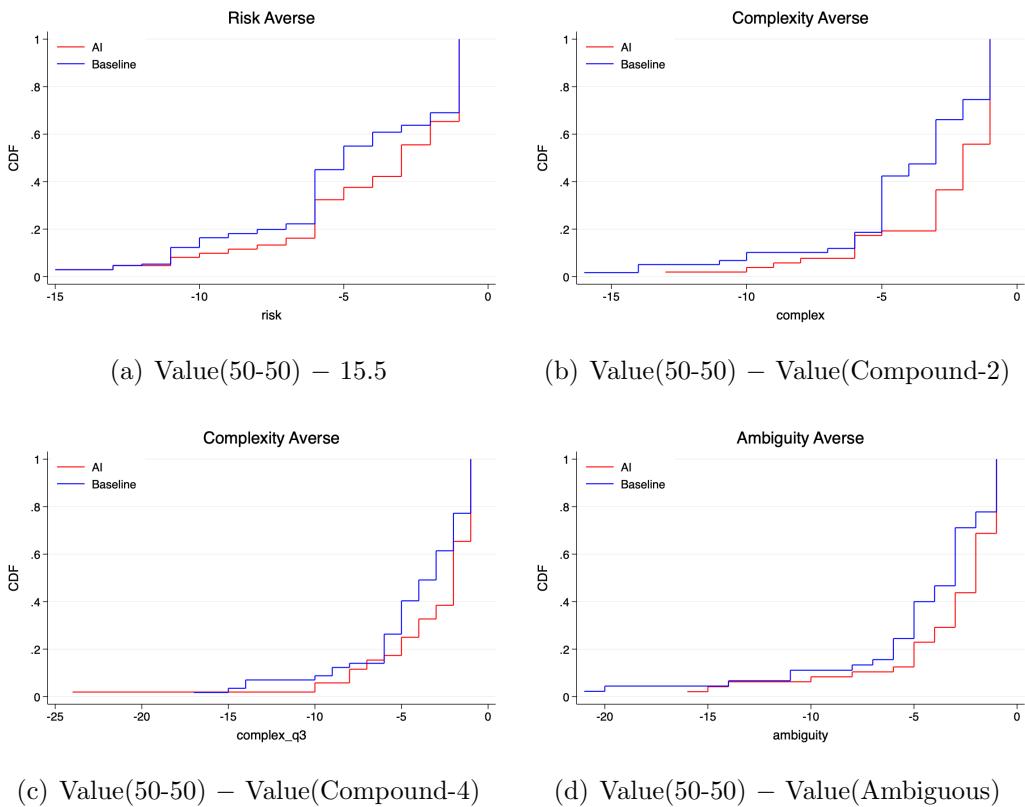


Figure A.19: CDF of the Magnitudes of Risk, Complexity, and Ambiguity Aversion (Filtered Sample)