# Adverse and Advantageous Selection in the Laboratory*

S. Nageeb Ali[†] Maximilian Mihm[‡] Lucas Siga[§] Chloe Tergiman[¶]

March 2, 2020

**Abstract**

We study two-player games where one-sided asymmetric information can lead to either adverse or advantageous selection. We contrast behavior in these games with settings where both players are uninformed. We find stark differences, suggesting that subjects do account for endogenous selection effects. Removing strategic uncertainty increases the fraction of subjects who account for selection. Subjects respond more to adverse than advantageous selection. Using additional treatments where we vary payoff feedback, we show that this difference can be attributed to learning. We also observe a significant fraction of subjects who appear to understand selection effects but do not apply that knowledge.

JEL: C91, C92, D72, D81, D82.
Keywords: Laboratory experiment, adverse selection, advantageous selection, contingent reasoning, strategic uncertainty, voting, social preferences.

*We thank Colin Camerer, Alex Imas, Ryan Oprea, Emanuel Vespa, and Sevgi Yuksel. Robizon Khubulashvili provided excellent research assistance and programming. We are grateful to REF grant #RE099 for funding this experiment.

[†]Department of Economics, Penn State. Email: nageeb@psu.edu.
[‡]Division of Social Science, NYU-Abu Dhabi. Email: max.mihm@nyu.edu.
[§]Division of Social Science, NYU-Abu Dhabi. Email: lucas.siga@nyu.edu.
[¶]Smeal College of Business, Penn State. Email: cjt16@psu.edu.

# Contents

# 1 Introduction

**Motivation:** Asymmetric information is central to many economic and social interactions. When players are asymmetrically informed, it can be rational for the less informed player to be suspicious of the motives of someone who is better informed. This suspicion can significantly alter behavior. For instance, Akerlof (1970) illustrates how buyers should be pessimistic about the quality of products being sold given that better-informed sellers are willing to sell those objects. Rothschild and Stiglitz (1976) argue that insurance providers should set premiums anticipating that those who privately know that they have a higher likelihood of claiming the insurance also have a greater incentive to buy it. Similarly, the "No-Trade Theorem" (Milgrom and Stokey, 1982) articulates how bettors engaged in speculative trading should draw inferences based on the motive that others have for taking opposing bets.

Across these settings, we see a common theme of *adverse selection*. From the perspective of each individual, the payoff of an available option—be it buying used cars, selling insurance, or taking a bet—is determined both by nature and the endogenously chosen actions of other parties. But selection need not always be adverse; when preferences are aligned, then the selection may be *advantageous*. For instance, if potential insurees are better-informed about their risk preferences, those who have a higher demand for insurance may be the risk-averse individuals who are "good risks" for an insurer (De Meza and Webb, 2001; Fang, Keane, and Silverman, 2008). In elections where voters share common preferences, some voters may be willing to abstain on ballot propositions to let better-informed voters cast the decisive votes, and thereby benefit from the selection of outcomes generated by the actions of others (Feddersen and Pesendorfer, 1996).

In this paper, we study how well people account for adverse and advantageous selection. Do people respond differently when they know that their counterpart has better information? Are the effects of asymmetric information uniform across settings where selection is adverse and advantageous?

**Main Design and Findings:** The core of our design is a simple two-player game described in Section 2: Alice and Bob jointly choose between a safe and risky option. The safe option yields identical payoffs for each party. The risky option is a lottery with a high payoff that exceeds that of the safe option and a low payoff below it. In *positively correlated* rounds, Alice's and Bob's payoffs are perfectly aligned so that both always obtain identical payoffs from the risky option; in other words, either both gain or both lose from the risky option being chosen. In *negatively correlated* rounds, Alice and Bob have misaligned interests: exactly one of them gains from the risky option while the other loses (relative to the safe option), but ex ante each is equally likely to be the winner. Each player votes for the safe or risky option, and the risky option is selected only if both vote for it. Importantly, one party—say Alice—privately observes the realized payoffs of the risky option whereas the other party is only told whether it is positively or negatively correlated. That the pair is asymmetrically informed is common knowledge between them.

What do our standard theories of selection predict in this setting? If players are selfish and play weakly undominated strategies, the informed player (Alice) should vote for the risky option if and only if it benefits her. When payoffs are positively correlated, Bob should then always vote for the risky option because he anticipates that the risky option is selected only when it benefits Alice and thus benefits him too. Advantageous selection in this case motivates Bob to vote for the risky option. By contrast, if payoffs are negatively correlated, then Bob should always vote for the safe option: were he to vote for the risky option, it is selected only when he loses from it (and Alice gains from it). In this case, Bob is deterred from voting for the risky option because of adverse selection.[1]

This reasoning suggests a test of whether subjects account for selection effects: when they are in the role of the uninformed player (Bob), are they more likely to choose the risky option when payoffs are positively correlated than when they are negatively correlated? As we illustrate in Table 3, the answer is yes. We consider two payoff variations for each correlation condition: one where the safe option yields a low payoff (below the expected value of the risky option) and one where it yields a high payoff (above the expected value of the risky option).[2] Shifting the payoffs from negative to positive correlation raises the fraction choosing the risky option from 48% to 86% with a low payoff for the safe option, and from 1% to 33% with a high payoff for the safe option. As such, we do see behavior shifting in the direction predicted by theories of asymmetric information.

We control for non-informational confounds[3] by comparing the above behavior with games where both players learn the correlation structure but are symmetrically uninformed about the realized payoffs. In this game, shifting from negative to positive correlation causes a shift in the fraction voting for a risky option from 78% to 89% with a low

[1] These predictions do not require equilibrium reasoning: given selfish preferences, all that is required is two rounds of iterative elimination of weakly dominated strategies. Therefore, a level-$k$ analysis with a random-$L0$ specification generates identical predictions for all types that are $L2$ or above.

[2] The order is randomized across subjects.

[3] For instance, aversion to inequality could be a non-informational confound because the negatively-correlated risky option is ex ante fair but ex post unequal whereas the positively-correlated risky option is both ex ante and ex post equal. As such, if subjects are averse to ex post inequality, there is a confounding rationale for the same behavioral predictions as the standard theory of selection.

payoff for the safe option, and from 4% to 8% with a high payoff for the safe option. Thus, we see shifts in the same direction, but with a much smaller magnitude. Approximately 20% of subjects make every choice in a way that is fully consistent, without error, with predictions of the standard model in all the rounds of the game.

What accounts for inconsistencies with the predictions of the standard model? Using a series of dictator games, we find that only a few of our subjects are guided by preferences for efficiency, suggesting that such motives are not a key explanation. We argue that three potentially important forces are strategic uncertainty, failures of contigent-reasoning, and learning, which we investigate in further treatments described below.

**Role of Strategic Uncertainty:** Even though behavior in our game is pinned down by eliminating two rounds of weakly dominated strategies when players are selfish, subjects may be uncertain about the preferences and hence behavior of others. For example, when information is asymmetric, particularly when the safe option yields high rewards, uninformed subjects are unwilling to vote for the risky option even when payoffs are positively correlated. This finding suggests that subjects are worried about relying on the informed player to make the "right" decision and worry that she may make a mistake.

We conduct a second treatment to assess this issue, discussed in Section 5. In this treatment, subjects are never paired with each other and are instead paired with computerized "robot" players whose strategies are known ahead of time. In the main game, these robot players observe the realized payoff and chooses the risky option if and only if it generates a higher (virtual) payoff for the robot than the safe option; human players never observe the realized payoff but know its correlation. We find that this significantly increases the degree to which subjects account for selection: subjects are far less willing to choose the risky option in negatively-correlated choices and more willing to do so in positively-correlated choices. The corresponding numbers go up from 23% to 82% with a low payoff for the safe option and 2% to 46% with a high payoff for the safe option. In fact, the fraction of subjects who behave according to theory in all rounds almost doubles to 40%. Thus, strategic uncertainty captures (to some degree) a divergence between the selection effects we see in "human-human" interactions and those predicted by theory.

**Role of Contingent-Reasoning:** We also use our second treatment to investigate the degree to which subjects may have difficulties with the contingent-reasoning required to determine selection effects. After subjects play against robots, they are asked several non-leading questions about the inferences they can draw from the robot's choice. These are relatively high stakes questions that deliver a high payment only if subjects answer every question correctly. The questions provide us with a measure of how well subjects understand the relevant contingent-reasoning, but could also provide subjects with a nudge that would alter how they play the game. As such, after answering these questions, the subjects play the asymmetric information game against the robot players once more.

Almost 90% of subjects answer the contingent-reasoning questions perfectly. Moreover, about 30% of the subjects who do not behaving according to theoretical predictions the first time that they played the asymmetric information game do conform to predictions on the second time. This behavior suggests that failures of contingent-reasoning are a reason why some subjects do not respond to asymmetric information as standard theories of strategic selection predict. However, close to 40% of the subjects who answer all of the contingent-reasoning questions correctly continue to depart from theoretical predictions on the second time that they play the asymmetric information game. These subjects appear able to understand each step of contingent reasoning separately but do not piece together that understanding in their subsequent strategic behavior.

**Role of Learning:** Studying positive and negative correlation in a unified framework allows us to compare how subjects respond to adverse versus advantageous selection. We see that subjects seem to respond more to adverse selection. Once strategic uncertainty is removed, about 74% of subjects behave exactly according to our theoretical predictions in every negatively-correlated round, and only approximately 43% do so in every positively correlated round. Even after drawing subjects' attention to the information carried by the computerized player's vote, the fraction of subjects who account for negative selection is substantially higher than the fraction who account for positive selection: 77% and 61%.

One potential reason for this asymmetry is prior learning from day-to-day social and economic interactions with asymmetric information. If an individual were to consistently make the "mistake" of choosing a risky option in settings where payoffs are negatively correlated, she would repeatedly see that she is worse off than were she to vote for the safe option. In everyday life, this is the mistake of "trusting" others when one shouldn't, and this mistake is self-correcting in the long-run. By contrast, if the same individual were to consistently make the mistake of choosing the safe option when payoffs are positively correlated, then she does not observe what would have happened were she to have chosen the risky option instead. Thus, experience does not teach her that this is a mistake. In everyday life, this is the mistake of not trusting others when one should, and this isn't self-correcting because one does not observe the counterfactual. A learning-theoretic concept where players potentially have incorrect beliefs off the equilibrium path, such as a self-confirming equilibria (Fudenberg and Levine, 1993), could in principle rationalize the behavioral differences between how well people learn to account for adverse and advantageous selection.

We test this hypothesis using our third and fourth treatments (described in Section 6) where we vary the level of feedback that subjects obtain after a decision. In the *partial feedback* treatment, in each round, after subjects make their decisions, they observe the payoff they would obtain should that round be selected for payment. By contrast, in the *full feedback* treatment, subjects learn not only the information from the partial feedback treatment but also the realized payoff of the risky option and how the other player voted. Thus, subjects can see the counterfactual outcome of what would have happened if they voted differently. After these feedback rounds, we observe the behavior of subjects in asymmetric information games without feedback.

In support of our hypothesis, we find that in games with adverse selection, about 78% behave exactly according to theoretical predictions in every negatively correlated round after partial feedback, and after full feedback, this percentage is 82%, a difference that is not statistically significant. By contrast, in games with advantageous selection, about 63% do so after partial feedback, and after full feedback, this percentage is 76%, a difference that is large in magnitude and also statistically significantly higher. In addition, after partial feedback, subjects continue to treat adverse and advantageous selection differently whereas after full feedback, there are no longer statistical differences in the degree to which subjects account for adverse and advantageous selection.

We view this finding to be of both theoretical interest and germane to policy. Because in practice, people do not observe counterfactuals, there may be a self-reinforcing cycle whereby individuals learn to distrust those who are better informed (from experiences when preferences are misaligned) and do not learn to rely on others when there are common gains. This finding offers a rationale for the prevalence of "zero-sum bias" (Meegan, 2010; Różycka-Tran, Boski, and Wojciszke, 2015): people may treat strategic interactions as zero-sum games because mistakes made in settings with misaligned preferences are more self-correcting than those in settings with common interests. It also suggests a direct consequence for political and electoral behavior. Given the widespread perception of polarization (Levendusky and Malhotra, 2015), relatively uninformed voters may believe that their interests are misaligned with those of better informed voters. Their suspicion may then lead them to vote in such a way that the election cannot be swung by the choices of better informed voters.[4]

**Related Literature:** A rich literature studies how people respond to asymmetric information in strategic settings including lemons markets (Bazerman and Samuelson, 1983), betting (Sonsino, Erev, and Gilat, 2002; Carrillo and Palfrey, 2011; Magnani and Oprea, 2017), settlements in zero-sum games (Carrillo and Palfrey, 2009), auctions (Kagel and Levin, 1986; Charness and Levin, 2009), elections (Guarnaschelli, McKelvey, and Palfrey, 2000; Battaglini, Morton, and Palfrey, 2010), and many others. Our approach to study how people respond to asymmetric information uses a novel design that varies the correlation in payoffs in order to compare how subjects respond to adverse versus advantageous selection relative to a symmetric information benchmark with no selection. In addition, we use several further treatments to disentangle potential reasons for deviations from our theoretical predictions. Below, we discuss the relationship of our work to the most closely related strands of the prior literature.

One approach roots individual failures to account for asymmetric information to strategic uncertainty or incorrect beliefs about how others play the game. This approach is manifested in, for example, models that use level-$k$, cognitive hierarchy, or cursed equilibrium reasoning; see Crawford, Costa-Gomes, and Iriberri (2013), Eyster (2019) for surveys. Brocas, Carrillo, Wang, and Camerer (2014) distinguish between these models in asymmetric information games by using "mousetracking" to record which payoffs subjects look at, and find support for theories where players are imperfectly attending to relevant information. We contribute to this perspective by seeing the degree to which subjects account for selection in both playing against human players as well as against robot players whose strategies are revealed ahead of time. We find that a greater fraction of subjects account for selection against robot players, but even in that context, a significant fraction do not. Our finding that payoff feedback matters in resolving the discrepancy between how much subjects account for adverse versus advantageous selection suggests that even when human subjects are told the strategies of robot players, experience is essential for them to "trust" the other player to make the right choice.

A more recent literature studies failures in contingent-reasoning and selection-neglect; for example, see Esponda and Vespa (2014, 2018, 2019), Martínez-Marquina, Niederle, and Vespa (2019), Barron, Huck, and Jehiel (2019), and Enke (2019). Relative to this literature, we directly test whether people respond to asymmetric information by com-

[4]This behavior contrasts with that of Feddersen and Pesendorfer (1996) where uninformed voters abstain to let better informed voters swing the election. Ali, Mihm, and Siga (2018) show that negatively correlated payoffs can generally cause failures of information aggregation.

paring choices where players are symmetrically uniformed with those where one player has private information. We see in this comparison that the behavior of an uninformed player changes when he knows that his opponent has private information, and this change is qualitatively in the direction predicted by theory albeit with a smaller magnitude. We also see that the degree to which people account for selection varies between adverse and advantageous selection, suggesting that the ability to reason about contingencies is contextual. Finally, we find that a significant fraction of our subject pool exhibit an understanding of contingent reasoning in principle (when answering understanding questions at high stakes) but do not appear to apply that knowledge in practice (on decisions where the stakes are lower).

Finally, our paper connects to studies of learning in games. We investigate whether learning-based models, like self-confirming equilibria, can explain why subjects account for advantageous selection less than they do for adverse selection. In one case, not accounting for selection is consistent with a self-confirming equilibrium (advantageous selection) but in the other, it is not. To test this hypothesis, we vary the feedback that subjects obtain after the round, where players observe either only the chosen outcome ("partial feedback") or both the chosen outcome and the counterfactual that would have been chosen had the subject voted differently ("full feedback"). We find that the gap between adverse and advantageous selection persists with partial feedback but disappears with full feedback. Our design offers a direct test for self-confirming equilibria in a game with asymmetric information, and provides evidence in support of it, complementing existing studies (e.g. Fudenberg and Levine, 1997; Fudenberg and Vespa, 2019).

# 2 A Conceptual Framework

This section describes the conceptual framework, which also corresponds to the central element of our design. In each round of our experiment, subjects—say Alice and Bob—are matched in pairs and simultaneously choose between two options, $S$ (a safe option) and $R$ (a risky option). Alice and Bob vote simultaneously and $R$ is selected only if both vote for it. The safe option $S$ pays $x > 0$ to each of them. $R$ is a risky option that may offer payoffs of $y$ or $z$ to each player where $0 < y < x < z$, and this lottery is implemented by the toss of a (virtual) fair coin. We denote a vector of payoffs by $(\pi_A, \pi_B)$ where $\pi_A$ is the amount paid to Alice and $\pi_B$ is the amount paid to Bob. We vary whether $R$ is positively or negatively correlated:

1. **Positive Correlation:** If the coin toss is *Heads*, $R$ pays $(y, y)$, and otherwise, $R$

pays $(z, z)$.

2. **Negative Correlation:** If the coin toss is *Heads*, $R$ pays $(y, z)$, and otherwise, $R$ pays $(z, y)$.

Positive correlation reflects a "common-values" environment in which every realization and every choice guarantees that the players' payoffs are identical. By contrast, in the negatively-correlated case, the risky option $R$ benefits one player to the detriment of the other.

In all of our experiments, subjects are told about the correlation of the risky option so they both know the possible payoffs of the risky option. Our setting of interest is one where information is asymmetric: Alice is told the realization of the coin toss, Bob is not, and this is common knowledge. In other words, Bob knows the *potential payoffs* (and the associated probability distribution) of the risky option whereas Alice knows the actual *realized payoffs* of the risky option.

Let us describe the implications of the standard theory (with selfish players) in this setting. Consider equilibra in weakly undominated strategies.[5] In all of these cases, Alice has a unique weakly undominated strategy: vote for $R$ if she is obtaining $z$ and for $S$ if she is obtaining $y$. What does this imply for Bob? In equilibrium, Bob recognizes that his vote affects the outcome only if Alice is voting for $R$ because otherwise $S$ would have been selected regardless of what he does. So in the case where his vote matters, Alice must be obtaining a payoff of $z$ if $R$ wins the election. In the positive-correlation case, this is *advantageous selection* for Bob because he too must be obtaining $z > x$ in this case, which makes voting for $R$ a best-response for him. By contrast, in the negative correlation case, this is *adverse selection* for Bob because then he must be obtaining $y < x$, which makes voting for $R$ a worst-response for him.[6]

Thus, the equilibrium predictions are extremely simple: Alice should be behaving according to her weakly undominated strategy. If the risky option is positively correlated, Bob should vote for $R$ always. By contrast, if the risky option is negatively correlated, Bob should vote for $S$ always. In fact, this behavior is pinned down not only by equilibrium but also by weaker solution-concepts: one round of elimination of weakly dominated strategies guarantees that Alice behaves as above, and a second round guarantees that Bob does so. Therefore, our behavioral predictions are pinned down by two rounds of elimination

[5] There always exist equilibria in which both players choose $S$ with probability 1 because the other is doing so. These equilibria are in weakly dominated strategies.

[6] In the absence of asymmetric information, Bob has no reason to distinguish between positively and negatively correlated rounds.

of weakly dominated strategies.[7]

More broadly, this conceptual framework predicts that we should see the risky option being selected more often in the positively-correlated case than in the negatively-correlated case. One may envision other rationales that motivate behavior in the same direction (e.g., aversion to ex-post inequality). Our design, discussed below, disentangles the selection-story from these other rationales.

# 3 Design and Procedures

This section describes our first treatment, namely the "Human-Human" (HH) treatment, where subjects were matched in pairs. Our second treatment—where subjects were instead matched with robot players—is described in Section 5.

## 3.1 Experimental Design

Our main game, the *Asymmetric Information* (AI) game, is that described in Section 2. In our design, we vary three elements of this game: (1) the payoff of the safe option $S$; (2) whether the risky option $R$ is positively or negatively correlated; and (3) the identity of the player who learns the realized payoffs of the risky option $R$. The payoff of the safe option $S$, denoted by $x$ in Section 2, is either \$12 or \$16 (for both parties). The values for $y$ and $z$ in the risky option $R$ are \$10 and \$20, respectively, and the ex-ante probability that a subject receives either payoff if the risky option is implemented is set to 50%. Subjects played 8 rounds of this game, four where they were uninformed, and four where they perfectly learned the realized payoffs of $R$. These are summarized by Table 1.

Our objective is to assess the degree to which subjects account for selection effects. Following our theoretical predictions in Section 2, do subjects in the role of the uninformed voter vote for $S$ when it is negatively correlated and vote for $R$ when it is positively correlated *because they are strategically accounting for selection*? To answer this question, we have to distinguish the asymmetric-information rationale for this behavior from other rationales for the same behavior. The other parts of the Human-Human treatment are designed with this goal in mind, allowing us to make within-subject comparisons across several games.

An important confounding consideration is *aversion to ex post inequality*: a different reason to favor $S$ when $R$ is negatively correlated is that its payoffs are ex post unequal

[7] For this reason, level-$k$ analyses make the same behavioral predictions for all players that are $L2$ or above (assuming a random-$L0$ specification).

*Table 1: Rounds in the Asymmetric Information game.*

| Round | Safe Option $S$ (1 vote) | Risky Option $R$ (2 votes) | Voter Informed | Other Voter Informed |
|---|---|---|---|---|
| 1 | (\$12; \$12) | (\$10, \$20) or (\$20, \$10) | no | yes |
| 2 | (\$12; \$12) | (\$10, \$10) or (\$20, \$20) | no | yes |
| 3 | (\$16; \$16) | (\$10, \$20) or (\$20, \$10) | no | yes |
| 4 | (\$16; \$16) | (\$10, \$10) or (\$20, \$20) | no | yes |
| 5 | (\$12; \$12) | (\$10, \$20) or (\$20, \$10) | yes | no |
| 6 | (\$12; \$12) | (\$10, \$10) or (\$20, \$20) | yes | no |
| 7 | (\$16; \$16) | (\$10, \$20) or (\$20, \$10) | yes | no |
| 8 | (\$16; \$16) | (\$10, \$10) or (\$20, \$20) | yes | no |

but this rationale is absent when $R$ is positively correlated. To assess how much subjects are influenced by this consideration, we precede the AI game with the *Symmetric Information* (SI) game, which uses the same parameters as the AI game, but where players are symmetrically uninformed. That is, in the SI game, neither player is informed about the payoffs of option $R$, other than knowing its correlation structure. Because both players are symmetrically uninformed (and this is common knowledge), there can be no endogenous selection effects in this part. If subjects respond to ex-post inequality in the AI game, the difference in behavior *across* positively and negatively correlated rounds for a given value of the safe option should also be present in the SI game.

To evaluate the strength of social preference considerations (both aversion to ex post inequality and *preferences for efficiency*) without the interference of a voting structure, we had subjects play a series of Dictator games following the AI game. Table 2 shows the rounds that subjects faced in the Dictator games. Rounds 1 through 4 of the Dictator games directly correspond to those in the AI and SI games. Rounds 5 through 8 allow us to evaluate subjects' preferences with respect to efficiency tradeoffs without the presence of uncertainty.[8]

Finally, we also asked subjects a series of 15 questions that tested their understand-

[8]Round 9 is a "sanity check" to evaluate whether subjects paid attention to the values on their screens, and, whether subjects voted for the payoff-maximizing option when inequality and efficiency were the same in both Options.

*Table 2: Rounds in the Dictator Game.*

| Round | Option A | Option B |
|---|---|---|
| 1 | (\$12; \$12) | (\$10, \$20) or (\$20, \$10) |
| 2 | (\$12; \$12) | (\$10, \$10) or (\$20, \$20) |
| 3 | (\$16; \$16) | (\$10, \$20) or (\$20, \$10) |
| 4 | (\$16; \$16) | (\$10, \$10) or (\$20, \$20) |
| 5 | (\$12; \$12) | (\$10, \$20) |
| 6 | (\$12; \$12) | (\$20, \$10) |
| 7 | (\$16; \$16) | (\$10, \$20) |
| 8 | (\$16; \$16) | (\$20, \$10) |
| 9 | (\$12; \$16) | (\$16, \$12) |

ing of the instructions. Six understanding questions focused specifically on how votes translated to outcomes. Four understanding questions focused specifically on the fact that players were symmetrically informed in the SI Game. Five understanding questions focused specifically on the nature of the asymmetric information in the AI game.[9] All the instructions that subjects received are in Appendix A. A series of screen shots showing the understanding questions subjects faced are in Appendix B.

## 3.2 Experimental Procedures

The experiment is comprised of 5 parts. Parts 1 and 2 are not germane to the AI game, but instead allow us to gradually introduce subjects to the different aspects of the experimental procedures.[10] Subjects played the SI game in Part 3, the AI game in Part 4, and ended with the Dictator games in Part 5. The order of rounds within each game was randomly determined at the subject level.

In each session, subjects received printed instructions for each part after they had

[9]We took care not to introduce any elements that might lead subjects to "discover" that the informed player's vote carried information as to the payoffs in the risky option.

[10]Part 1 is a simple decision-making task in which we introduce the notion of uncertainty. Part 2 introduces subjects to the voting structure that exists in the Main Game (i.e. the first option is implemented so long as it receives a single vote, while the second option is implemented only if both voters vote for it) but without uncertainty regarding the second option.

completed the previous part, and those instructions were read aloud each time. Subjects could advance rounds within each part at their own pace, but the experiment advanced from part to part at the pace of the slowest subject. Subjects received no feedback as to their own or anyone else's choices. We conducted four sessions for a total of 86 subjects. Each session lasted about 50 minutes. This experiment took place in the Laboratory for Experimental Management and Auctions (LEMA) at Penn State University in the Spring of 2019.

In terms of payment, at the very start of each session, subjects were told that in addition to their \$7 show-up fee, they would be paid for one part of the experiment only. We divided the understanding questions described above into three groups and attached them to Part 2 (where we introduce the voting structure), Part 3 (where subjects play the SI game) and Part 4 (where subjects play the AI game). Subjects were also told that if Part 2 or Part 3 or Part 4 was randomly chosen to count for payment, then they would be paid either for one randomly selected round in that part or for the understanding questions of that part. If the understanding questions were randomly chosen to count for payment, then they would earn \$10 if they answered *all* questions of that part correctly; otherwise, they earned only \$2.

Because Parts 1 and 2 were primarily included to help subjects understand the AI game, we provide more details on those parts and the choices that subjects made in those parts in Appendix C. The following section will focus solely on the AI game, as well as on behavior in the SI and Dictator games that are directly related.

# 4 Results

We first describe behavior in the Asymmetric Information (AI) game and investigate whether, for a given value of the safe option, subjects in the role of the uninformed voter are more inclined to vote for the risky option when payoffs are positively correlated than when payoffs are negatively correlated. We then compare behavior across games in the HH treatment to distinguish the asymmetric-information rationale for this behavior from other confounds. Unless otherwise stated, all our claims are the results of within-subject analyses and the p-values we report correspond to Wilcoxon matched-pairs signed-rank tests.

Table 3 displays aggregate data of subjects' choices. The fourth column shows the fraction of times subjects voted for the risky option when in the role of the uninformed voter in the AI game. The fifth column shows the same statistic but in the SI game,

*Table 3: Aggregate results: fraction choosing the risky option in the HH treatment.*

| Round | Safe Option | Risky Option | AI Game (uninformed) | SI Game | Dictator Game |
|---|---|---|---|---|---|
| 1 | (\$12; \$12) | (\$10, \$20) or (\$20, \$10) | 47.7% | 77.9% | 72.1% |
| 2 | (\$12; \$12) | (\$10, \$10) or (\$20, \$20) | 86.0% | 88.4% | 82.6% |
| 3 | (\$16; \$16) | (\$10, \$20) or (\$20, \$10) | 1.2% | 3.5% | 0% |
| 4 | (\$16; \$16) | (\$10, \$10) or (\$20, \$20) | 32.6% | 8.1% | 7.0% |

where both subjects are uninformed. The sixth column looks at the same behavior in a Dictator game, where a single uninformed subject chooses between the safe and risky options, knowing that her choice determines outcomes for both her and her partner.

## 4.1 Do Subjects Account for Selection Effects?

At the aggregate level, subjects appear to respond to asymmetric information as predicted by the theoretical framework described in Section 2. In particular, we compare the number reported in the fourth column of Table 3 across Rounds 1 and 2, and then across Rounds 3 and 4. Within each of these pairs of rounds, the value of the safe option is held fixed and the only change is whether the risky option is negatively or positively correlated.

When the safe option is \$12 and the outcomes from the risky option are negatively correlated, subjects choose the risky option 47.7% of the time compared with 86% of the time when they are positively correlated. When the safe option is \$16, these numbers are 1.2% and 32.6%, respectively. For a given value of the safe option, the differences in behavior across positively and negatively correlated risky options are both large in absolute and relative terms, and are statistically significant: whether the safe option is \$12 or \$16, subjects are significantly more likely to choose the risky option over the safe option when payoffs from the risky option are positively correlated than when payoffs are negatively correlated ($p < 0.001$ in both sets of comparisons).

To assess the degree to which subjects are reacting to asymmetric information in the AI game, we compare behavior in the AI with that of the SI games, where both players are symmetrically uninformed. Since the only distinction between these two games is in whether information is asymmetric, a change in subjects' behavior across these games is strong evidence that subjects are reacting to its presence. In particular, comparing the behavior of uninformed players in the AI and SI games, we should observe at least one of the following behaviors for a particular value of the safe option: (1) when the risky option is negatively correlated, a decrease in the fraction that vote for the risky option from the SI game to the AI game; (2) when the risky option is positively correlated, an increase in the fraction that vote for the risky option from the SI game to the AI game. Whether both or only one of these occurs depends on how risk aversion impacts choices in the SI game. Regardless of risk aversion however, the "difference in differences" across correlation structures for a given value of the safe option should be larger in the AI game than in the SI game.

We find substantial differences in behavior across the AI and SI games in line with these predictions. This is the case both at the round level, and when we compare "differences-in-differences" across correlation structures for a given value of the safe option. For example, in Round 1 we see that subjects are far less likely to choose the risky option when information is asymmetric than when it is symmetric (47.7% versus 77.9% — $p < 0.001$). In parallel, in Round 4, subject are far more likely to vote for the risky option when information is asymmetric than symmetric (32.6% versus 8.1% — $p < 0.001$). Both of these patterns are in line with comparative predictions. Also demonstrating the impact of asymmetric information is the differences-in-differences in behavior across Rounds 1 and 2 as well as across Rounds 3 and 4 when we compare both games. Both those differences are much larger in the AI game than in the SI game: 38.3% versus 10.5% when $S = \$12$ ($p < 0.001$) and 31.4% versus 4.6% $S = \$16$ ($p < 0.001$).

While the theory matches qualitative predictions both within the AI game as well as across the AI and SI games, we do see significant departures from the point predictions. If we look across all of the choices, 20.9% of subjects in the AI game behave according to all of the theoretical predictions, voting for the safe option in *both* negatively correlated rounds *and* voting for the risky option in *both* positively correlated rounds.

Among the subjects who do not fully conform to theory, we identify differences in how consistently they conform to theory in the positively and negatively correlated rounds.[11] The fraction of subjects who vote for the safe option in both of the negatively correlated rounds (Rounds 1 and 3) is 52.3%, while the fraction of subjects who vote for the risky option in both of the positively correlated rounds is lower at 30.2% ($p = 0.001$). These findings illustrate that a greater fraction of subjects are sensitive to adverse selection than to advantageous selection. As we described in the introduction, this is consistent with the zero-sum bias in social psychology as well as the possibility that people may

[11]We note no statistical differences in how well subjects conform to theory in the rounds in which $S = \$12$ and those in which $S = \$16$. Indeed, 39.5% of subjects behave according to theory in both rounds where $S = \$12$, and 32.6% do so when $S = \$16$ ($p = 0.239$).

learn to choose the payoff-maximizing action in adverse-selection scenarios but not in advantageous-selection selection scenarios.

What else might be guiding subjects' choices? A poor understanding of our instructions does not appear to be a reason for the departures from theoretical predictions that we observe by some subjects.[12] In Sections 4.2 and 4.3 we discuss the degree to which the behavior that we observe can be explained by social preferences, strategic uncertainty, and failures of contingent reasoning.

## 4.2 The Role of Social Preferences

In this section, we explore the degree to which social preferences can explain behavior. Two leading theories of social preferences that could appear to play a role in our study are *aversion to ex post inequality* (e.g. Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) and *preferences for efficiency* (e.g. Charness and Rabin, 2002; Engelmann and Strobel, 2004). In both cases, our evidence rejects these theories as being good explanations for the behavior that we observe in our experiments.

**Aversion to ex post inequality:** If subjects dislike ex post inequality, then this offers a rationale for them to choose the safe option when the risky option is negatively correlated but not when the risky option is positively correlated. Therefore, it offers an important confound because it can lead to predictions that are identical to those of adverse and advantageous selection in the AI game.

We find little evidence of an aversion to ex-post inequality, both at the aggregate level and in terms of individual level data. To see why, let us turn to the SI and Dictator games where neither player knows the payoffs of the risky option beyond its correlation structure. In both the SI and Dictator games, a large majority of subjects' decisions do not depend on whether the risky option's outcomes are negatively or positively correlated, even controlling for the amount of the safe option. Indeed, at the aggregate level, we see in Columns 5 and 6 of Table 3 that the differences between Rounds 1 and 2, and between Rounds 3 and 4, are not large in magnitude. The fractions of overlap between Rounds 1

[12] Recall that subjects faced a series of 15 questions that tested their understanding of the instructions. These questions were spread over the various Parts of the instructions. The median number of mistakes among our subjects is zero and the mean 0.84 mistakes out of 15 questions. Both Chi Squared and Fisher exact tests show that the distribution of mistakes among subjects who do not behave according to the theory is no different than among those who do those who do ($p = 0.808$ and $p = 0.959$, respectively). Further, A two-sided test of proportions show that there is also no statistical difference in behavior between subjects who make no mistakes at all and those who make at least one ($p = 0.411$). Thus we cannot attribute deviations from theory to confusion.

and 2, and between Rounds 3 and 4 in the SI game are 84.9% and 93.0%, respectively. The corresponding fractions are 80.2% and 93.0% in the Dictator games.[13] Even more informative is behavior at the individual level. If some subjects' choices are guided by aversion to ex-post inequality, then these subjects should behave as the theory predicts in the AI game (though not necessarily due to selection) *and* play identically in the SI game. None of our subjects make choices that follow this pattern. Thus we rule out aversion to ex post inequality as a driver of behavior.

**Preferences for efficiency:** If subjects are motivated by the size of the total surplus, then we should see behavior that differs significantly from the theoretical predictions of Section 2. For example, when the safe option is \$12, then a subject with preferences for efficiency may, depending on how much she values efficiency relative to her own payoff, choose the risky option when it is negatively correlated, even when she is informed that the risky option lowers her own payoff. If the safe option is \$16, then such a subject may never choose the the risky option when it is negatively correlated, even if she is informed that the risky option increases her own payoff. We find that none of our subjects behave in a way that is consistent with preferences for efficiency across all rounds in the AI and Dictator game.[14] Even if we focus on the $s = \$12$ rounds separately, we find that at most 5 of our subjects behave in a way that is consistent with preferences for efficiency, and in the $s = \$16$ rounds, only 6 of our subjects do so. Thus, it appears that the degree to which subjects in our experiment are motivated by efficiency is minimal.[15]

## 4.3 Strategic Uncertainty and Failures of Contingent Reasoning

As described in Section 2, equilibrium in weakly undominated strategies being played by selfish players predicts that in the AI game, uninformed players choose the safe option when payoffs are negatively correlated and choose the risky option when payoffs are pos-

[13] Subjects who do make different decisions across those rounds are more likely to favor the risky option when outcomes are positively correlated than when they aren't (the p-values comparing Rounds 1 and 2 as well as Rounds 3 and 4 in the SI and Dictator games are 0.013, 0.103, 0.029, 0.083).

[14] 14 subjects make decisions consistent with preferences for efficiency when informed (note that subjects do not see all the scenarios when informed, and some subjects only saw "advantageous" risky choices) and 8 subjects make decisions consistent with preferences for efficiency when non-informed. The intersection of those two groups represents 3 subjects. In addition, using behavior in the relevant rounds of the Dictator game, we find that none of those 3 subjects make the same efficient choices (these are rounds 5, 6, 7 and 8 in Table 2).

[15] Note that we do not claim that such preferences do not exist. Rather that the marginal rates of substitution between one's own payoff and the social surplus may be such that, with our parameters, we don't observe such preferences, and thus they can't explain our subjects' behavior.

itively correlated. This prediction is, in fact, pinned down by two rounds of elimination of weakly dominated strategies.

While this may appear straightforward from the perspective of game theory, it involves two cognitive demands. First, it requires subjects to be confident that players behaving as informed voters do not choose weakly dominated actions. An uninformed Bob must attribute sufficiently high probability to the informed Alice choosing what is best for her that it rationalizes the equilibrium choice. This is an issue of strategic uncertainty. Second, it requires subjects to attend to a potentially non-salient feature of the game, namely that one's vote matters only when the other player is voting for the risky option. This is an issue of contingent reasoning. We investigate both of these below.

To assess the issue of strategic uncertainty, we look at whether subjects are best-responding to the empirical distribution of play in the experiment. If it appears that a large fraction of subjects are not doing so, then this behavior suggests that subjects' behavior may be rationalized by them being strategically uncertain, i.e., having incorrect conjectures about the behavior of others. The first two columns in Table 4 show the possible rounds that the informed players saw.[16] In this Table, the informed players' payoffs are listed first. The third column shows the fraction of informed players who choose the option with the payoff in bold. The fourth column shows the (ex ante) expected payoff for the uninformed player of choosing the risky option, given the empirical distribution of the informed players' choices.

We see that subjects who know the realized payoff of the risky option do not necessarily vote for the option that maximize their payoffs. For example, when the safe option is $12 and the risky option has negatively-correlated outcomes, 19.2% of informed subjects choose the safe option when they would have benefited from the risky option, and 9.4% choose the risky option despite it lowering their payoffs relative to the safe option. We see analogous behavior when the safe option is $16 and the risky option has negatively correlated payoffs, but see relatively fewer "mistakes" when the risky option is positively correlated.

Inspecting the expected payoff for a subject in the role of the uninformed voter who votes for the risky option given the empirical distribution of play, we see that if such a subject had correct beliefs about the behavior of informed subjects, her decisions should coincide with the predictions from Section 2. Since we noted that only 20.9% of subjects followed these equilibrium predictions exactly, we do see evidence suggestive of strategic

[16]Recall players did not see all these rounds, but only one in each pair of rows depending on the coin flip.

*Table 4: Rationalizing "Mistakes": Expected Payoffs Given Empirical Distribution.*

| Round | Safe Option | "Risky" Option[a] | Fraction of Informed Players Choosing the "Risky" Option | Expected Payoff of Voting for the Risky Option Given Empirical Distribution[b] |
|---|---|---|---|---|
| 1 | ($12; $12)<br>($12; $12) | **($10, $20)** ~~or ($20, $10)~~<br>~~($10, $20) or~~ **($20, $10)** | 9.4%<br>81.8% | $11.6 |
| 2 | ($12; $12)<br>($12; $12) | **($10, $10)** ~~or ($20, $20)~~<br>~~($10, $10) or~~ **($20, $20)** | 2.4%<br>97.7% | $15.9 |
| 3 | ($16; $16)<br>($16; $16) | **($10, $20)** ~~or ($20, $10)~~<br>~~($10, $20) or~~ **($20, $10)** | 0%<br>82.2% | $13.5 |
| 4 | ($16; $16)<br>($16; $16) | **($10, $10)** ~~or ($20, $20)~~<br>~~($10, $10) or~~ **($20, $20)** | 2.4%<br>97.7% | $17.9 |

[a]The ex-ante probability of either particular outcome was 50% but the informed player knew the outcome.

[b]This is for the uninformed voter given the choices/mistakes the informed voter makes empirically.

uncertainty, which motivates designing a treatment that eliminates strategic uncertainty, which we describe in Section 5. One interesting pattern that we note here is that there are relatively fewer departures from our theoretical predictions at a higher cost of mistakes.

Turning to the other cognitive demand, we investigate the degree to which subjects fail to apply contingent reasoning. Subjects who fail to apply contingent reasoning, should make the same choices in the AI and SI games since they are not thinking about the inference they should draw from being pivotal. Among those subjects who don't play the equilibrium strategies of Section 2, we see that slightly over half (57.4%) behave identically across the AI and SI games.[17]

# 5 The Human-Robot Treatment: Design and Results

To assess the importance of strategic uncertainty and failures of contingent reasoning, we conduct a "Human-Robot" (HR) treatment. Instead of being paired with another human subject, each subject is paired with a robot player whose strategy is revealed ahead of

[17]We note that our understanding questions in this treatment were deliberately designed to focus on the mechanics of the game and to avoid hinting that subjects should think about contingencies. As such, we cannot use the answers to these questions to assess the degree to which subjects fail or succeed in applying contingent reasoning.

time. By pairing subjects with a computerized non-human subject in the SI and AI games, and telling our subjects how it had been programmed, we effectively remove issues of strategic uncertainty that potentially affected behavior in the main treatment.[18] An additional 82 subjects participated in the HR treatment. Below we detail how the HR treatment differs from our earlier HH treatment.

**Symmetric Information Game:** The parameters in the Symmetric Information game of the HR treatment were identical to those in the HH treatment. The instructions closely followed those in the HH treatment, except that subjects were now matched with a robot player that earned "virtual (imaginary) dollars" that "had no impact on you or anyone else at any point, ever." In the SI game, the robot player was programmed to always vote for the risky option. To closely match the understanding questions across treatments, subjects were told how the robot player was programmed only after they answered the understanding questions related to the mechanisms of the SI game. Directly following this information, subjects were asked to confirm they understood how the robot was programmed via one additional understanding question.

**Asymmetric Information Game:** In the Asymmetric Information game in the HR treatment, the robot player was always in the role of the informed voter and our subjects only participated in the role of uninformed voters. The robot player was programmed to always vote for the option that gave it the highest amount of virtual (imaginary) dollars, and this was made known to the human subjects. As in the SI game, the instructions in this treatment closely paralleled those in the HH treatment, as did the understanding questions.

**Contingent Reasoning Questions and Asymmetric Information (2) Game:** To evaluate subjects' ability to do contingent reasoning, we designed a new part following the AI game.[19] Subjects first answered a series of "contingent reasoning" (CR) questions, all of which pertained to the AI game they had just played (examples of these questions are in Section 5.3). These CR questions did not explain contingent reasoning to the subjects, but instead were designed to "nudge" subjects towards paying attention to contingencies. Following the CR questions, subjects again played against the robot players in a repetition of the AI game, which we call the AI(2) game. The CR questions permit us to

[18]It also removes social preferences, but as we concluded in our analysis of the HH treatment, these appear to play only a limited role in our experiment.

[19]This took the place of the Dictator game of the HH treatment.

match behavior in the AI game with subjects' abilities to answer questions on contingent reasoning, and then to see whether such questions have a nudging effect in the AI(2) game.

We begin our analysis by comparing behavior in the AI and SI games in the HR treatment. We then compare behavior in these two games across the HH and HR treatments, and assess the degree to which strategic uncertainty influences behavior. Finally, we explore subjects' potential to reason about contingencies by evaluating their responses to the CR questions as well as behavior in the AI(2) game. Unless otherwise noted, the p-values associated with between-subjects comparisons across treatments are the result of Wilcoxon rank-sum tests, and the p-values associated with within-subject comparisons in the HR treatment are the result of Wilcoxon matched-pairs signed-ranks tests.

## 5.1 Aggregate Results in the HR Treatment

We present the aggregate data of the HR treatment in Table 5.

*Table 5: Aggregate results: fraction choosing the risky option in the HR treatment.*

| Round | Safe Option | Risky Option | Asymmetric Information | Symmetric Information | Asymmetric Information(2)[a] |
|---|---|---|---|---|---|
| 1 | ($12; $12) | ($10, $20) or ($20, $10) | 23.2% | 84.2% | 22.0% |
| 2 | ($12; $12) | ($10, $10) or ($20, $20) | 81.7% | 87.8% | 90.2% |
| 3 | ($16; $16) | ($10, $20) or ($20, $10) | 2.4% | 2.4% | 1.2% |
| 4 | ($16; $16) | ($10, $10) or ($20, $20) | 46.3% | 4.9% | 62.2% |

[a]Restricting attention to subjects who answered all questions correctly would generate fractions of 18.6%, 91.5%, 0% and 71.2% respectively.

We observe sharp differences in behavior when comparing behavior within the AI game across Rounds 1 and 2 as well as across Rounds 3 and 4, consistent with subjects responding to selection effects ($p < 0.001$ in both cases). Overall, 40.2% of the subjects behave in a way that is consistent with standard theoretical predictions across all rounds. We also note a large difference in behavior when comparing the "difference in difference" between Rounds 1 and 2 (as well as Rounds 3 and 4) across the SI and AI games: 58.5% versus 3.6% when $S = \$12$ ($p < 0.001$) and 43.9% versus 2.5% when $S = \$16$ ($p < 0.001$).

Finally, we also see that the difference in behavior in terms of how well subjects respond to adverse and advantageous selection persists in the AI game. In fact, in the HR treatment, almost three quarters of our subjects (74.4%) vote for the safe option in both Rounds 1 and 3, corresponding exactly to our theoretical predictions (from Section 2). In other words, all but a quarter of the subjects account perfectly for adverse selection. The corresponding fraction who vote for the risky option in both Rounds 2 and 4, where payoffs were positively correlated, is 42.7%. Thus, we see evidence both that a substantial fraction of our subjects account perfectly for selection and yet, a gap between adverse and advantageous selection remains among those who do not.

## 5.2 The Impact of Strategic Uncertainty

Our analysis below turns to the effect of eliminating strategic uncertainty, comparing the results of the HH and HR treatments. In Figure 1, we show the fraction of subjects who vote for the risky option across our two treatments. We illustrate the choices in the AI games for both the HH and HR treatment. Since we observe no statistical difference across the HH and HR treatments in the SI games, for simplicity we provide the average choices in the SI games across HH and HR treatments.[20]

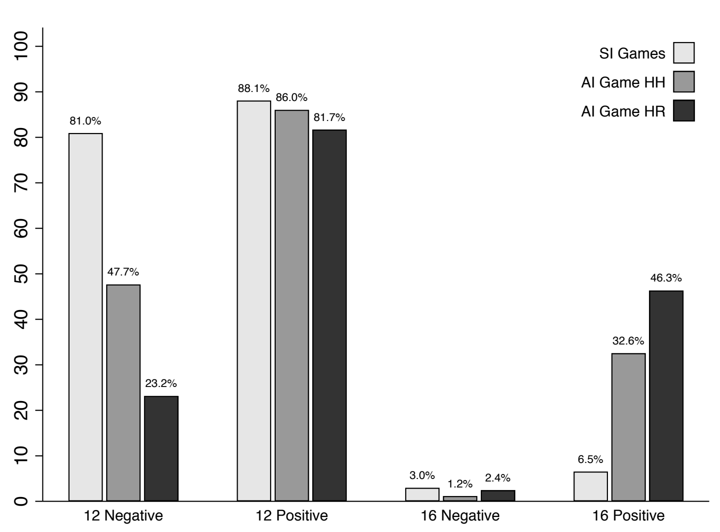



*Figure 1: Fraction choosing the risky option in Human-Human and Human-Robot Treatments.*

We find that subjects respond more to selection effects when strategic uncertainty is removed. This is the case in Rounds 1 and 4, which were the rounds in which a substantial fraction of subjects deviated from the theoretical predictions in the HH treatment. In Round 1, the fraction of subjects who chose the risky option in the HH treatment was

[20]Tests of proportions show that we cannot reject the null that the answers to the four SI rounds come from the same population (the smallest two-sided p-value is 0.303).

47.7%, while it is 23.2% in the HR treatment where strategic uncertainty is eliminated ($p = 0.001$). In parallel, the fraction of subjects who chose the risky option in Round 4 of the HR treatment is now 46.3%, up from 32.6% in the HH treatment ($p = 0.068$).

As mentioned previously, overall, in the HR treatment, 40.2% of subjects behaved according to the theoretical predictions in Section 2 for all rounds of the AI game. This is significantly higher than in the HH treatment, where the fraction was 20.9%.[21]

How much of the difference in behavior across treatments can we attribute to strategic uncertainty, as opposed to difference in subject characteristics across treatments? We note that there are no discernible differences between the subjects in the HH and HR treatments in terms of their demographics or how well they understood the instructions.[22,23] Thus, neither of these explain the increase in the proportion of subjects who play according to theory in the HR treatment. In addition, since removing strategic uncertainty has no impact on how subjects play the SI game (see previous sub-section), it also does not seem that subjects across treatments differed in their social preferences, or beliefs about their pivotality. So, we cautiously attribute treatment differences to strategic uncertainty and estimate that it accounts for about 25% of the deviations from theory that we observed in the HH treatment.

## 5.3 Contingent Reasoning

A plausible conjecture is that subjects who continue to depart from our theoretical predictions even after strategic uncertainty is eliminated are those who simply cannot reason about contingencies. We test this conjecture in Part 5 of the HR treatment. After the AI game, subjects answer a series of questions that draw subjects' attention to the information conveyed in the robot player's vote, and thus contingent reasoning. These contingent reasoning questions (CR) take place before subjects play the AI game a second time. We investigate how responses to these questions correlate with subjects' behavior in the AI game on the first iteration and how answering these questions influences behavior in the AI game on the subsequent play.

The first two CR questions assess whether subjects understand that the vote of the Robot player carried information on the coin flip. The remaining two assess whether they understood that this could impact their own payoff. An example of the former and

[21]The p-value on a 2-sided test of proportions is 0.007).

[22]For demographic data in our two treatments, see Section D.

[23]Restricting attention to the understanding questions in the two treatments that share a common structure, both Chi2 and Fisher exact tests fail to reject that the distribution of mistakes are from the same population (the p-values are 0.797 and 0.940).

latter are below, where the items in the square brackets correspond to the multiple choice answers the subjects faced.[24]

> *Given how the computer player was programmed in Part 4, if the computer player votes for the option requiring 2 votes (the option on the right), what does that tell you about the outcome of the coin flip? [That it landed on HEADS; that it landed on TAILS, it doesn't tell you anything about the outcome of the coin flip]*
>
> *Given how the computer player was programmed in Part 4, if the computer player votes for the option requiring 2 votes and you vote for that option too, how much will you earn? [$15, $17, $20, You will earn $15 or $20 with equal chance of each.]*

We find that 89% of the subjects answer *all* of the CR questions correctly. Correlating these responses with behavior in the preceding asymmetric information game, we find that all subjects who behaved exactly according to our theoretical predictions in the preceding AI game answered each CR question correctly. Of the 11% of players who answered at least one CR question incorrectly, none played according to our theoretical predictions in the preceding AI game.

How does answering these CR questions affect subsequent play? The last column in Table Table 5 displays the fraction of subjects who vote for the risky option for each round of the AI(2) game. The fraction of subjects who behave according to the theoretical predictions increases in both positively correlated rounds ($p = 0.065$ when the safe option is $12, and $p = 0.004$ when the safe option is $16), and is statistically no different in the two negatively correlated rounds ($p > 0.100$ in both cases). Overall, 57.3% of subjects in the AI(2) game make all their choices in a way that is consistent with the theoretical predictions; this fraction is statistically higher than that in the AI game played before the CR questions (40.2%) ($p < 0.001$). Thus, the CR questions help some subjects understand selection effects but a significant fraction of subjects continue to deviate from theoretical predictions.

Interestingly, a large fraction of these subjects appear to understand contingent reasoning when it is broken down into steps: of the 42.7% of subjects whose choices depart from theory in the AI(2) game, 76.5% actually answer all contingent reasoning questions correctly. These subjects show that they understand that the robot player's votes carries information about their own payoff, and yet make choices in the AI(2) game that lead

[24] For reference, in the questions below, "Part 4" refers to the AI game. To explain the nature of the uncertainty, throughout the instructions we used the example of a fair coin flip that determined what the payoffs in the risky option would be if it was to be implemented.

to lower payoffs. Thus, these subjects show that they are able to correctly execute each step of contingent reasoning separately but do not piece together these steps when they subsequently play the AI(2) game.

Furthermore, the data show that the asymmetry in the degree to which subjects account for negative versus positive correlation persists in the AI(2) game. Looking at only decisions in negatively correlated rounds, 76.8% of subjects match our theoretical predictions while the analog for positively correlated rounds is 61.0%, which is significantly less ($p < 0.001$).

# 6 The Feedback Treatments: Design and Results

Across all treatments, we observe a gap between how well subjects reason about contingencies in negatively correlated rounds and the degree to which they do the same in positively correlated rounds. What accounts for this gap?

We hypothesize that it is potentially the result of subjects' real-life experience with asymmetric information. Uninformed individuals who repeatedly choose a risky option when outcomes are negatively correlated would see that they are never better off than if they had chosen the safe option, and in fact may often be strictly worse off. This feedback allows them to learn from their mistakes so that one does not see these mistakes in their long-run behavior. On the other hand, if uninformed individuals repeatedly choose a safe option when payoffs from a risky options are positively correlated, they do not observe what would have happened had they chosen the risky option. Hence, they do not learn from their mistake. Thus, such mistakes can persist in the long-run.

We investigate whether this mechanism can explain the gap between responsiveness to negative and positive correlation by varying the payoff feedback we provide subjects when they play rounds of the AI game. Our two additional treatments—Partial Feedback (PF) and Full Feedback (FF)—are identical to the HR treatment except for Part 4, where subjects play the AI game against a robot player multiple times but with payoff feedback. The PF treatment resembles our description of everyday life: in the PF treatment, after each feedback round, a subject is reminded of how he voted and told what the payoffs would be if that round is selected for payment. Thus, a subject choosing the safe option does not learn the Robot's vote or the coin flip, and cannot deduce what would have happened had he chosen the risky option. By contrast, in the FF treatment, after each feedback round, a subject is reminded of his vote and told the result of the coin flip, how the computer voted, which of the two options was implemented for the round, and what his payoffs would be if that round is selected for payment. Thus, in the FF treatment, regardless of a subject's vote, that subject can deduce what the payoffs would have been had he voted differently.[25]

More specifically, in both the PF and FF treatments, Part 4 comprises 5 repetitions of each of the first four rounds of the AI game described in Table 1 . Within each session, half of the subjects saw the 10 rounds of negatively correlated outcomes first. Within those 10 rounds, the fixed amount was either \$12 or \$16, each happening 5 times, in random order. These subjects then saw the 10 rounds of positively correlated outcomes, again where the fixed amount was \$12 or \$16 in random order. The other half of the subjects saw the positively correlated rounds first.[26] A total of 83 subjects participated in the FF treatment, and a total of 86 subjects participated in the PF treatment.

After these feedback rounds, subjects face positively and negatively correlated payoffs, exactly as in Part 5 of the HR treatment. These Part 5 rounds involve no feedback. We describe how partial and full feedback in Part 4 affects behavior in Part 5 in Section 6.1, interpreting these results through the lens of long-run learning. We describe the evolution of behavior within the feedback rounds in Section 6.2.

## 6.1 Aggregate Results After Partial and Full Feedback

Our first set of results concern behavior in Part 5 of the Partial Feedback and Full Feedback treatments. These are rounds where there is no feedback but that take place after the feedback rounds in Part 4 so that we can see whether the feedback that takes place in Part 4 has an impact on subsequent behavior.

*Table 6: Fraction of subjects following theoretical predictions in Part 5.*

| | Partial Feedback Treatment | Full Feedback Treatment |
|---|---|---|
| Both Negatively Correlated Rounds | 77.9% | 81.9% |
| Both Positively Correlated Rounds | 62.8% | 75.9% |
| All Rounds | 55.8% | 71.1% |

Table 6 presents the fraction of subjects who answered both negatively correlated rounds, both positively correlated rounds and all rounds according to our theoretical

[25]These instructions, as well as screen shot examples showing what the feedback rounds looked like are in Appendix A.

[26]The transition from the negatively to positively correlated rounds (and vice versa) was seamless: subjects simply moved from one type of setting to the next without any announcement.

predictions in Part 5 in the FF and PF treatments. In the negatively correlated rounds, this fraction is not significantly different across the PF and FF treatments (77.9% and 81.9%, $p = 0.516$). However, when outcomes are positively correlated, significantly more subjects behave according to the theoretical predictions in the FF treatment compared to the PF treatment (75.9% versus 62.8%, $p = 0.066$). Additionally, across all rounds, the fraction of subjects who behave according to theoretical predictions increases from 55.8% to 71.1% ($p = 0.040$).

Within the PF treatment, subjects continue to react differently across the positively and negatively correlated rounds ($p = 0.009$).[27] The data from the FF treatment stand in sharp contrast: the difference between the positively and negatively correlated rounds is no longer significant ($p = 0.166$).[28] Thus, full feedback not only increases the fraction of subjects who behave according to the theoretical predictions in the positively correlated rounds but also closes the gap in behavior across positively and negatively correlated rounds. We view this as lending support to our hypothesis that learning and payoff feedback may explain why subjects are better at accounting for adverse selection better than advantageous selection.

## 6.2 Evolution of Subject Behavior in Feedback Rounds

The analysis described above looks at the impact of feedback on learning in subsequent rounds without feedback. In this section, we study the evolution of behavior within the feedback rounds. Our purpose is twofold. First, we document the evidence that the differences across treatments described above are due to the differential impact of partial versus full feedback. Second, we consider how the difference between partial and full feedback can affect the rate of learning across all rounds.

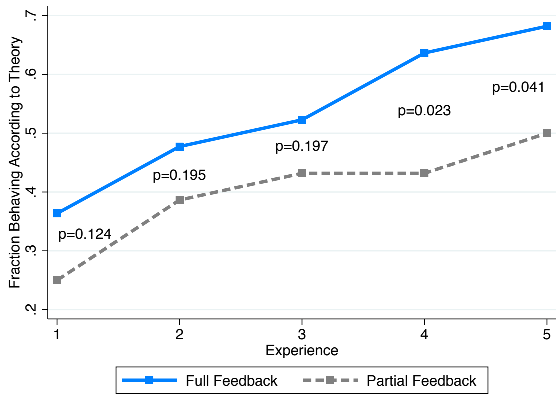
Restricting attention to behavior in the first block of ten rounds of Part 4,[29] Figure 2 depicts the fraction of subjects who behave according to our theoretical predictions across all our settings, by "experience level."[30] We find no treatment differences in how subjects

[27] In fact, we see no statistical difference in behavior between the PF and HR treatments.

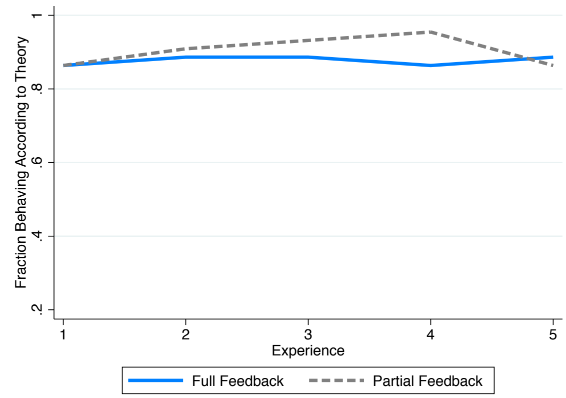
[28] Moreover, compared with the HR treatment, the FF data are significantly higher when looking at the fraction of subjects who behave according to theory in both positively correlated rounds as well as over all rounds ($p = 0.040$ in both cases).

[29] Recall that by design, roughly half of the subjects saw one of the positively correlated rounds first, and the other half saw one of the negatively correlated rounds first. We focus on the first block of ten rounds to more cleanly identify the differential effects of learning from the different types of feedback within type of correlation, before any cross-correlation learning can occur.

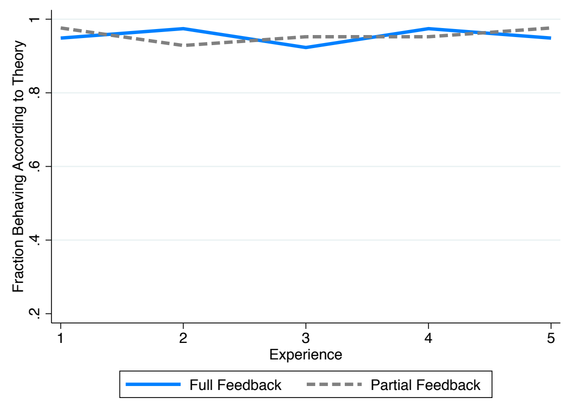
[30] An experience level of $k$ indicates that this is the $k^{th}$ time a subject encounters that specific setting (for example, positively correlated rounds when the safe outcome is 16), though this need not have happened in the $k^{th}$ round of Part 4 since subjects faced settings in a randomized order.



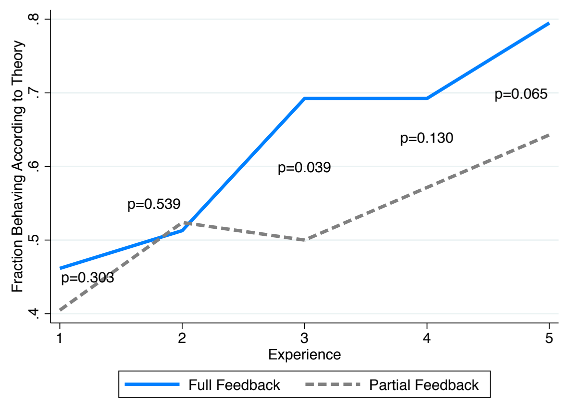

*(a) Positively correlated rounds when the safe option is 16.*




*(b) Positively correlated rounds when the safe option is 12.*




*(c) Negatively correlated rounds when the safe option is 16.*




*(d) Negatively correlated rounds when the safe option is 12.*

*Figure 2: Fraction of subjects behaving according to theory over experience and by feedback. The p-values testing statistical differences are in the graph. The null is that both proportions are equal across treatments and the alternative is that the proportion is higher under full feedback. They are omitted in Panels (b) and (c) where they are all strictly higher than 0.10.*

responded in the very first round in any of our four settings. A between-subject analysis across the two treatments shows that subjects in the FF and PF treatments react similarly in the negatively correlated rounds whether the fixed amount is \$12 or \$16, and that the same is true in the positively correlated rounds, whether the fixed amount is \$12 or \$16 (the p-values for both all pair-wise comparisons are all strictly greater than 0.100).[31] In other words, whether before subjects receive feedback (i.e., in the first round), or after they receive a few rounds of feedback only, we find no differences across behavior in the

[31]These conclusions are robust to looking at first round behavior by correlation structure *pooling both* fixed amounts, and to looking at behavior across multiple early rounds.

positively and negatively correlated rounds in Part 4 when comparing the PF and FF treatments. Thus we attribute the differences observed in Part 5 after all 20 rounds in Part 4 to the difference in feedback subjects received, and not to inherent differences in how subject initially reacted to these rounds before any learning occurred.

We see that feedback has a significant effect in the positively correlated rounds when the safe option is \$16 and in the negatively correlated rounds when the safe option is \$12. In both cases, a non-trivial fraction of subjects deviate from theoretical predictions prior to receiving feedback. In both cases, partial feedback helps but full feedback nevertheless facilitates learning to a greater degree. For the positively correlated rounds, the intuition is clear: subjects learn and observe that if they were to vote for the risky option, it would be selected only in circumstances that benefit them. For the negatively correlated rounds, the intuition for why full feedback is valuable is more subtle: here, a subject observes after choosing the safe option—which is the theoretically predicted choice—that if she were to vote for the risky option, it would be selected only in circumstances that are adverse to her. Thus, learning in full feedback reinforces her vote for the safe option. By contrast, with partial feedback, she would not learn what would happen were she to choose the risky option, and may be tempted to experiment. Thus, full feedback more readily identifies that it would be a mistake to vote for the risky option, regardless of the actual outcome.

# 7 Conclusion

We study how people respond to adverse and advantageous selection effects using a simple two-person game where asymmetrically informed subjects choose between a risky option and a safe option. We vary whether payoffs from the risky option are negatively correlated (inducing adverse selection) or positively correlated (inducing advantageous selection). To isolate the role of asymmetric information from other confounds, these subjects also play a game in which both players are symmetrically uninformed that is otherwise identical.

Consistent with standard theoretical predictions, we find that uninformed subjects are much more likely to choose the risky option when payoffs are positively correlated than when payoffs are negatively correlated. These differences do not arise when players are symmetrically uninformed, indicating that subjects respond to asymmetric information.

We also see departures from theoretical predictions. Strikingly, subjects are better at accounting for adverse rather than advantageous selection. Our subsequent treatments help us diagnose why we see these departures.

The second treatment studies how strategic uncertainty—uncertainty about the behavior of others—influences choice. We pair subjects with a computerized robot player whose strategy is known. A cross-treatment comparison shows that strategic uncertainty explains up to a quarter of the departures from theoretical predictions in the first treatment (when subjects were paired with each other). In the second treatment, we also ask a number of questions that explore subjects' understanding of contingent reasoning. Answering these questions affects behavior for a significant fraction of our subjects, indicating that contingent reasoning can be taught. However, we also find that a non-trivial fraction of subjects demonstrate an excellent understanding of contingent reasoning when asked questions about it but fail to implement that knowledge in a strategic setting even after those questions. These subjects appear to understand each element of contingent reasoning separately but do not piece them together on their own.

Our third and fourth treatment explore whether the gap in the degree to which subjects account for adverse versus advantageous selection can be explained by everyday learning. We vary payoff feedback in our experimental design and see how much of an effect feedback about counterfactuals can play. We find that it closes the gap: full feedback leads to no significant differences in how well subjects account for adverse versus advantageous selection whereas a significant gap remains with partial feedback.

To summarize, people do account for asymmetric information but the degree to which they do so is contextual. When there is reason for "distrust"—such as in settings of negatively-correlated payoffs—people do not let better informed partners make the final choice. But when there is reason to trust those who are better-informed—because payoffs are positively correlated—people fail to capitalize on these gains.

We view these results to be germane to political and social interactions. They suggest a potential mechanism for the prevalence of "zero-sum thinking" noted in social psychology: people learn to distrust others because mistakes from zero-sum games are self-correcting whereas those from settings with common interests are not. Such behavior may have direct consequences for political behavior and elections: to the degree that voters perceive there to be significant political polarization (Levendusky and Malhotra, 2015; Alesina, Miano, and Stantcheva, 2020), our results suggest that voters are likely to be suspicious of the information possessed by others and unlikely to capitalize on gains that could come from advantageous selection.

# References

Akerlof, G. A. (1970): "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 488–500.

Alesina, A., A. Miano, and S. Stantcheva (2020): "The Polarization of Reality," *American Economic Review Papers and Proceedings*.

Ali, S. N., M. Mihm, and L. Siga (2018): "Adverse selection in distributive politics," Working Paper.

Barron, K., S. Huck, and P. Jehiel (2019): "Everyday econometricians: Selection neglect and overoptimism when learning from others," WZB Discussion Paper.

Battaglini, M., R. B. Morton, and T. R. Palfrey (2010): "The Swing Voter's Curse in the Laboratory," *Review of Economic Studies*, 77, 61–89.

Bazerman, M. H. and W. F. Samuelson (1983): "I Won the Auction But Don't Want the Prize," *Journal of Conflict Resolution*, 27, 618–634.

Bolton, G. E. and A. Ockenfels (2000): "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 90, 166–193.

Brocas, I., J. D. Carrillo, S. W. Wang, and C. F. Camerer (2014): "Imperfect choice or imperfect attention? Understanding strategic thinking in private information games," *Review of Economic Studies*, 81, 944–970.

Carrillo, J. D. and T. R. Palfrey (2009): "The Compromise Game: Two-Sided Adverse Selection in the Laboratory," *American Economic Journal: Microeconomics*, 1, 151–81.

——— (2011): "No Trade," *Games and Economic Behavior*, 71, 66–87.

Charness, G. and D. Levin (2009): "The Origin of the Winner's Curse: A Laboratory Study," *American Economic Journal: Microeconomics*, 1, 207–36.

Charness, G. and M. Rabin (2002): "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 117, 817–869.

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri (2013): "Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications," *Journal of Economic Literature*, 51, 5–62.

De Meza, D. and D. C. Webb (2001): "Advantageous Selection in Insurance Markets," *RAND Journal of Economics*, 249–262.

Engelmann, D. and M. Strobel (2004): "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments," *American Economic Review*, 94, 857–869.

Enke, B. (2019): "What you see is all there is," Working Paper.

Esponda, I. and E. Vespa (2014): "Hypothetical Thinking and Information Extraction in the Laboratory," *American Economic Journal: Microeconomics*, 6, 180–202.

——— (2018): "Endogenous Sample Selection: A Laboratory Study," *Quantitative Economics*, 9, 183–216.

——— (2019): "Contingent Preferences and the Sure-Thing Principle: Revisiting Classic Anomalies in the Laboratory," Working Paper.

Eyster, E. (2019): "Errors in strategic reasoning," in *Handbook of Behavioral Economics - Foundations and Applications 2*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, North-Holland, vol. 2 of *Handbook of Behavioral Economics: Applications and Foundations 1*, 187 – 259.

Fang, H., M. P. Keane, and D. Silverman (2008): "Sources of Advantageous Selection: Evidence from the Medigap Insurance Market," *Journal of Political Economy*, 116, 303–350.

Feddersen, T. and W. Pesendorfer (1996): "The Swing Voter's Curse," *American Economic Review*, 86, 408–424.

Fehr, E. and K. M. Schmidt (1999): "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114, 817–868.

Fudenberg, D. and D. K. Levine (1993): "Self-Confirming Equilibrium," *Econometrica*, 523–545.

——— (1997): "Measuring Players' Losses in Experimental Games*," *The Quarterly Journal of Economics*, 112, 507–536.

Fudenberg, D. and E. Vespa (2019): "Learning Theory and Heterogeneous Play in a Signaling-Game Experiment," *American Economic Journal: Microeconomics*, 11, 186–215.

Guarnaschelli, S., R. D. McKelvey, and T. R. Palfrey (2000): "An Experimental Study of Jury Decision Rules," *American Political Science Review*, 94, 407–423.

Kagel, J. H. and D. Levin (1986): "The winner's curse and public information in common value auctions," *The American economic review*, 894–920.

Levendusky, M. S. and N. Malhotra (2015): "(Mis)perceptions of Partisan Polarization in the American Public," *Public Opinion Quarterly*, 80, 378–391.

Magnani, J. and R. Oprea (2017): "Why Do People Violate No-Trade Theorems? A Diagnostic Test," Working Paper.

Martínez-Marquina, A., M. Niederle, and E. Vespa (2019): "Failures in Contingent Reasoning: The Role of Uncertainty," *American Economic Review*, 109, 3437–74.

Meegan, D. V. (2010): "Zero-Sum Bias: Perceived Competition Despite Unlimited Resources," *Frontiers in Psychology*, 1, 191.

Milgrom, P. and N. Stokey (1982): "Information, Trade and Common Knowledge," *Journal of Economic Theory*, 26, 17–27.

Rothschild, M. and J. Stiglitz (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics*, 90, 629–649.

Różycka-Tran, J., P. Boski, and B. Wojciszke (2015): "Belief in a Zero-Sum Game as a Social Axiom: A 37-Nation Study," *Journal of Cross-Cultural Psychology*, 46, 525–548.

Sonsino, D., I. Erev, and S. Gilat (2002): "On rationality, learning and zero-sum bettingan experimental study of the no-betting conjecture," Working Paper.

# Online Appendices

## A Instructions

Here we present the instructions that subjects saw in the HH and HC treatments. We first show the instructions of the HH treatment (where subjects play against each-other), and then the instructions of the HR treatment (where subjects play against robot players).

### A.1 First treatment: subjects face other subjects

INSTRUCTIONS

This study is in 5 Parts. Only one randomly chosen Part will count for payment. In addition to what you will earn in the study, you will be paid a $7 participation fee if you complete the study.

Importantly, all Rounds and Parts of this study are independent. In other words, nothing you do in any Round or Part of this study will have any impact on your opportunities or payment in any other Round or Part of this study. In addition, since only one randomly chosen Part will be chosen for payment, it is in your best interest to treat each Part as if it was the only one that mattered for payment.

We will now hand out the instructions for Part 1 of the study. We will give you the instructions for Part 2 of the study once you have completed Part 1, for Part 3 after you have completed Part 2, etc.

# Part 1

In this Part of the Study you will make decisions over the course of 9 Rounds. In each Round, you will be asked to choose between two options that determine your payoff.

Below we list exact decision problems that you will all face.

**The List of Rounds in Part 1**

| Decision Problem | Option 1 | | Option 2 |
|---|---|---|---|
| 1 | Fixed amount of \$11 | Versus | Receiving \$10 or \$20 with equal chance of each. |
| 2 | Fixed amount of \$12 | Versus | Receiving \$10 or \$20 with equal chance of each. |
| 3 | Fixed amount of \$13 | Versus | Receiving \$10 or \$20 with equal chance of each. |
| 4 | Fixed amount of \$14 | Versus | Receiving \$10 or \$20 with equal chance of each. |
| 5 | Fixed amount of \$15 | Versus | Receiving \$10 or \$20 with equal chance of each. |
| 6 | Fixed amount of \$16 | Versus | Receiving \$10 or \$20 with equal chance of each. |
| 7 | Fixed amount of \$17 | Versus | Receiving \$10 or \$20 with equal chance of each. |
| 8 | Fixed amount of \$18 | Versus | Receiving \$10 or \$20 with equal chance of each. |
| 9 | Fixed amount of \$19 | Versus | Receiving \$10 or \$20 with equal chance of each. |

These decision problems may appear in different order on your screen. In addition, for any given decision problem, which option appears on the left or the right of your screen may also differ from the examples above.

As you can see above, in each of the Rounds, one choice will be a fixed amount and the other will involve some uncertainty. The uncertainty can be described in the following way. The computer flips a virtual coin that lands either on heads or tails, each with an equal 50% chance. The outcome of the virtual coin flip determines your payment if you chose the uncertain option.

- if the coin lands on tails (which happens with 50% chance) you will receive $10.
- if the coin lands on heads (which happens with 50% chance) you will receive $20.

**Payment:** If this Part is randomly selected to count for payment in this Study, one of the 9 Rounds will be chosen to count for payment. Your earnings would be determined in the following way:

- if you chose the fixed amount, then you will earn that fixed amount;
- if you chose the option with uncertainty, your earnings depend on the result of the virtual coin flip: you receive $10 if the coin lands on tails, and you receive $20 if the coin lands on heads.

# Parts 2, 3 and 4 – Preamble

Each of Parts 2, 3 and 4 consist of 2 Blocks. In each of those Parts, Block 1 consists of a series of questions that test your understanding of the instructions that are relevant to the Part you are in. In each of those Parts, Block 2 consists of several Rounds of the game itself.

In Block 2, in each Round of Parts 2, 3 and 4, you will each be randomly matched into pairs. In each Round you and the person you are matched with will be asked to vote for one of two options that determine payoffs for both you and the person you are matched with. In each Part, who you are matched with will be randomly determined at the start of each Round and nothing you do or anyone else does can influence or impact how this matching occurs. At no point will you find out with whom you were matched, nor will your actions be revealed to anyone else nor will you find out the actions of the person with whom you were matched.

If Part 2 or Part 3 or Part 4 is randomly chosen to count for payment, then you will be paid for Block 1 or Block 2 of that Part.

If Block 1 of a Part is chosen for payment, then if you answered all the questions correctly, you will earn \$10. If you make even one mistake, you will earn \$2.

If Block 2 of a Part is chosen for payment, one of the Rounds in Block 2 will be randomly selected to determine your payment.

# Part 2

Part 2 is in six Rounds. As described in the preamble, in each Round you will be randomly rematched with another person in this room. In each Round you and the person you are matched with will be asked to vote for one of two options that determine payoffs for both you and the voter you are matched with. Here is an example of such a choice you can encounter in one of the Rounds (the choices you face may be different and will vary from Round to Round). Please take a moment to look at this table.

| Votes needed: 1 | Votes needed: 2 |
|---|---|
| Your earnings: \$5 | Your earnings: \$5 |
| Other voter's earnings: \$10 | Other voter's earnings: \$5 |
| vote for this option | vote for this option |

Just as in the example above, each option will differ in terms of the amounts that you and/or the voter you are matched with can earn. The options also differ in how many votes are needed for that option to be the one that is selected for this Round. In each Round, one of the two options will require that both you and the voter you are matched with vote for it in order for it to be selected for this Round. The other option is selected for this Round so long as it receives at least one vote. Which option requires two votes and which option only requires at least one vote will be clearly stated before you and the voter you are matched with make your decisions.

In the example above, for the option on the right to be selected for this Round, both you and the voter you are matched with have to vote for it. On the other hand, for the option on the left to be selected for this Round, only one voter has to vote for it. In other words, if you vote for the option on the left, then it is selected for this Round regardless of what option the voter you are matched with votes for. Similarly, if the voter you are matched with votes for the option on the left it is selected for this Round regardless of which option you vote for.

Note that the option that requires one vote will always be on the left and the option that requires two votes will always be on the right hand side.

**Payment:** If this Part is randomly chosen to count for payment, then one Round will be randomly chosen to count for payment. In the example above, if the option on the left is selected for this Round, then you would receive $5 and the voter you are matched with would receive $10. If the option on the right is selected for this Round, then you would receive $5 and the voter you are matched with would receive $5. Which of the two options is selected for this Round will depend on what happens during the Round.

Also note that you will go through the Rounds of Part 2 without knowing what the voter you are matched with has chosen.

Do you have any questions?

We will now begin Block 1 in which you will be asked questions that test your understanding of this game. If Block 1 of Part 2 is randomly chosen to count for payment, then you earn $10 if you answer ALL the questions correctly. If you make even one mistake you will earn only $2.

After Block 1 is over, you will play the 6 Rounds of the Part 2 game.

# Part 3

Part 3 is in four Rounds. As described in the preamble, in each Round you will be randomly rematched with another person in this room. In each Round you and the person you are matched with will be asked to vote for one of two options that determine payoffs for both you and the voter you are matched with. The game in Part 3 of this study is very similar to the game you played in Part 2. The difference lies in the kinds of options you face. In this part of the study, one of the options involves uncertainty. An example of a choice with an uncertain outcome is shown below. Please take a moment to look at the table below before I describe it.

| **Votes needed: 1** | **Votes needed: 2** |
|---|---|
| Your earnings: $11<br><br>Other voter's earnings: $9 | HEADS: Your earnings: $5<br>Other voter's earnings: $15<br><br>TAILS: Your earnings: $15<br>Other voter's earnings: $5 |
| vote for this option | vote for this option |

In this particular example, there is no uncertainty regarding the option on the left (the one requiring only 1 vote): if this option is the one that is selected for this Round, you would receive $11 and the voter you are matched with would receive $9. However, there is uncertainty regarding the option on the right (the one requiring two votes).

The uncertainty can be described in the following way. The computer throws a fair virtual coin that lands either heads or tails, each with an equal 50% chance. If it lands on heads, then the option on the right is: $5 for you and $15 for the voter you are matched with. If, on the other hand, it lands on tails, the option on the right is: $15 for you and $5 for the voter you are matched with. In other words, there is uncertainty in terms of which of the payoff pairs correspond to the option on the right: you do not know whether the payoff pair will be $5 for you and $15 for the voter you are matched with, or whether it will be $15 for you and $5 for the voter you are matched with. All you know is that the outcomes in the "uncertain" options are equally likely, each having 50% chance. In each Round the computer will flip that virtual coin before you and the voter you are matched with make your choices, but what side the coin landed on and which payoff pair that corresponds to will not be revealed to anyone.

Do you have any questions?

We will now begin Block 1 in which you will be asked questions that test your understanding of this game. If Block 1 of Part 3 is randomly chosen to count for payment, then you earn \$10 if you answer ALL the questions correctly. If you make even one mistake you will earn only \$2.

After Block 1 is over, you will play the 4 Rounds of the Part 3 game.

# Part 4

Part 4 is in eight Rounds. As described in the preamble, in each Round you will be randomly rematched with another person in this room. In each Round you and the person you are matched with will be asked to vote for one of two options that determine payoffs for both you and the voter you are matched with. The game in Part 4 of this study is very similar to the game you played in Part 3. The difference lies in that you OR the voter you are matched with will learn what the outcome of the virtual coin flip is before either of you vote. In other words, either you OR the voter with whom you are matched will observe whether the coin lands on heads or tails **before you have to cast a vote.** Recall the example of the previous Part in which the option requiring two votes (the one on the right) had uncertainty in terms of outcomes:

| **Votes needed: 1** | **Votes needed: 2** |
|---|---|
| Your earnings: \$11<br><br>Other voter's earnings: \$9 | HEADS: Your earnings: \$5<br>Other voter's earnings: \$15<br><br>TAILS: Your earnings: \$15<br>Other voter's earnings: \$5 |
| vote for this option | vote for this option |

Using the example above, this means that one voter in each pair will know whether the option requiring two votes leads to \$5 for you and \$15 for the other voter, or whether that option leads to \$15 for you and \$5 for the other voter.

If you are the one who learns the result of the coin flip, it means the voter you are matched with has not learned the result of the coin flip. That means before you and the other voter vote, you know exactly what happens if the option on the right is selected for this round, but the voter you are matched with does not have this information. If, on the other hand, you don't learn the result of the coin flip, it means that the voter you are matched with does know the result of the coin flip. That means before you and the other voter vote, the voter you are matched with knows exactly what happens if the option on the right is selected for this Round, but you do not have this information.

On the screen for each Round, you will know whether it is you or the voter you are matched with who has learned the result of the coin flip.

If you are the one who learns the result of the coin flip, your screen will display the relevant payoffs in black and the other payoff will be crossed out and in a lighter color. Below is an example of a Round in which you are the one who learned the result of the coin flip. In this example, you learned that the coin landed on Heads. As you can see, the payoff for Tails has been crossed out. Please take a moment to look at this example.

| **Votes needed: 1** | **Votes needed: 2** |
|---|---|
| Your earnings: \$11<br><br>Other voter's earnings: \$9 | HEADS: Your earnings: \$5<br>Other voter's earnings: \$15<br><br>~~TAILS: Your earnings: \$15~~<br>~~Other voter's earnings: \$5~~ |
| vote for this option | vote for this option |

If you do not know the result of the coin flip, your screen will not have anything crossed-out. Instead, you will face a screen like the one below, where you are reminded that the voter you are matched with has learned the result of the coin flip. Your screen would look like the following:

**Recall that the other voter has learned the result of the coin flip.**

| **Votes needed: 1** | **Votes needed: 2** |
|---|---|
| Your earnings: \$11<br><br>Other voter's earnings: \$9 | HEADS: Your earnings: \$5<br>Other voter's earnings: \$15<br><br>TAILS: Your earnings: \$15<br>Other voter's earnings: \$5 |
| vote for this option | vote for this option |

Do you have any questions?

We will now begin Block 1 in which you will be asked questions that test your understanding of this game. If Block 1 of Part 4 is randomly chosen to count for payment, then you earn \$10 if you answer ALL the questions correctly. If you make even one mistake you will earn only \$2.

After Block 1 is over, you will play the 8 Rounds of the Part 4 game.

# Part 5 Preamble

In this final Part of the study you will be assigned a Type. You will be a Type A Player or a Type B Player. Your Type will remained fixed throughout this last Part.

Each Type A Player will be randomly rematched with a Type B Player. You will not know who you are matched with. In this final Part of today's study, only Type A players make decisions that matter for payment, and these decisions affect the payoff of both the Type A Player and the Type B Player he/she is matched with.

Even though your Type will remain fixed for the rest of this study, you will not know which Type of Player you are. Since you do not know which Type of Player you are assigned to be, and since only Type A Players make decisions that matter for payment, we will ask everyone to make decisions as if they were Type A players.

Please note that your Type will remain fixed and at no point will you change roles. Your "true" Types have already been determined by the computer, and your decisions when acting as Player A CANNOT affect you or anyone else in this room if your "true" Type turns out to be Type B. In other words, if it turns out you are a Type B Player, no decision you make here can affect anyone's payoff, including your own. If it turns out your "true" Type is A, there is nothing that anyone else can do that will affect your payoff, and your decisions affect both your payoff and the payoff of the Type B Player you are matched with. Therefore, when making decisions, you should act as Player A. Further, since only "true" Type A Players make decisions that matter for payment in this study, in the remainder of the instructions we will assume you are a Type A Player.

# Part 5

In this Part of the study, you will make decisions over the course of 9 Rounds.

In each Round you will have the choice between two options that determine earnings for you and the Type B player you are matched with. These choices will look like the following:

| | |
|---|---|
| Your earnings: \$6<br>Type B player's earnings: \$12 | Your earnings: \$11<br>Type B player's earnings: \$9 |

Here are how your payments would be determined if this were the Round that mattered for payment:

- If you chose the option on the left, then you would earn \$6 and the Type B player you are matched with would earn \$12.
- If instead you chose the option on the right, then you would earn \$11 and the Type B player you are matched with would earn \$9.

In each Round, you are the one whose decision will matter. That is, it is your choice of option that will be selected for each Round.

Remember that you will not change roles. So as a Type A Player, your payoff will never be determined by someone else in this room. Also remember that only one Part of the study will be chosen to count for payment. If this Part is chosen to count, only *one* Round will matter for payment. So it is in your best interest to treat each Round as if it were the one that mattered for payment.

Do you have any questions?

## A.2 HR treatment: subjects face robot players

INSTRUCTIONS

This study is in 5 Parts. Only one randomly chosen Part will count for payment. In addition to what you will earn in the study, you will be paid a $7 participation fee if you complete the study.

Importantly, all Rounds and Parts of this study are independent. In other words, nothing you do in any Round or Part of this study will have any impact on your opportunities or payment in any other Round or Part of this study. In addition, since only one randomly chosen Part will be chosen for payment, it is in your best interest to treat each Part as if it was the only one that mattered for payment.

We will now hand out the instructions for Part 1 of the study. We will give you the instructions for Part 2 of the study once you have completed Part 1, for Part 3 after you have completed Part 2, etc.

## Part 1

In this Part of the Study you will make decisions over the course of 9 Rounds. In each Round, you will be asked to choose between two options that determine your payoff.

Below we list exact decision problems that you will all face.

**The List of Rounds in Part 1**

| **Decision Problem** | **Option 1** | | **Option 2** |
|---|---|---|---|
| 1 | Fixed amount of \$11 | Versus | Receiving \$10 or \$20 with equal chance of each. |
| 2 | Fixed amount of \$12 | Versus | Receiving \$10 or \$20 with equal chance of each. |
| 3 | Fixed amount of \$13 | Versus | Receiving \$10 or \$20 with equal chance of each. |
| 4 | Fixed amount of \$14 | Versus | Receiving \$10 or \$20 with equal chance of each. |
| 5 | Fixed amount of \$15 | Versus | Receiving \$10 or \$20 with equal chance of each. |
| 6 | Fixed amount of \$16 | Versus | Receiving \$10 or \$20 with equal chance of each. |
| 7 | Fixed amount of \$17 | Versus | Receiving \$10 or \$20 with equal chance of each. |
| 8 | Fixed amount of \$18 | Versus | Receiving \$10 or \$20 with equal chance of each. |
| 9 | Fixed amount of \$19 | Versus | Receiving \$10 or \$20 with equal chance of each. |

These decision problems may appear in different order on your screen. In addition, for any given decision problem, which option appears on the left or the right of your screen may also differ from the examples above.

As you can see above, in each of the Rounds, one choice will be a fixed amount and the other will involve some uncertainty. The uncertainty can be described in the following way. A virtual coin is flipped, that lands either on heads or tails, each with an equal 50% chance. The outcome of the virtual coin flip determines your payment if you chose the uncertain option.

- if the coin lands on tails (which happens with 50% chance) you will receive $10.
- if the coin lands on heads (which happens with 50% chance) you will receive $20.

**Payment:** If this Part is randomly selected to count for payment in this Study, one of the 9 Rounds will be chosen to count for payment. Your earnings would be determined in the following way:

- if you chose the fixed amount, then you will earn that fixed amount;
- if you chose the option with uncertainty, your earnings depend on the result of the virtual coin flip: you receive $10 if the coin lands on tails, and you receive $20 if the coin lands on heads.

# Parts 2, 3 and 4 – Preamble

Each of Parts 2, 3 and 4 consist of 2 Blocks. In each of those Parts, Block 1 consists of a series of questions that test your understanding of the instructions that are relevant to the Part you are in. In each of those Parts, Block 2 consists of several Rounds of the game itself.

In Block 2, in each Round of Parts 2, 3 and 4, you will each be matched with a computer player. In each Round you and the computer player you are matched with will vote for one of two options that determine your payoff. How the computer player has been programmed to vote will be described to you in each Part before you cast your vote. How the computer player has been programmed to vote will vary from Part to Part.

If Part 2 or Part 3 or Part 4 is randomly chosen to count for payment, then you will be paid for Block 1 or Block 2 of that Part.

If Block 1 of a Part is chosen for payment, then if you answered all the questions correctly, you will earn $10. If you make even one mistake, you will earn $2.

If Block 2 of a Part is chosen for payment, one of the Rounds in Block 2 will be randomly selected to determine your payment.

# Part 2

Part 2 is in six Rounds. As described in the preamble, in each Round you will be matched with a computer player. In each Round you will be asked to vote for one of two options that determine how much you will earn. It will also determine how many virtual (imaginary) dollars the computer will earn.

Note that how many virtual (imaginary) dollars the computer player earns will have no impact on you or anyone else at any point, ever. We will describe how the computer player is programmed to cast its vote in this Part of the Study, only after describing the kinds of choices you will face.

We will describe the kinds of choices you will face by using an example. Below is an example of a choice you can encounter in one of the Rounds (the choices you face may be different and will vary from Round to Round). Please take a moment to look at this table.

| Votes needed: 1 | Votes needed: 2 |
|---|---|
| Your earnings: \$5 | Your earnings: \$6 |
| Computer player: \$10 virtual (imaginary) dollars | Computer player: \$5 virtual (imaginary) dollars |
| vote for this option | vote for this option |

Just as in the example above, each option will differ in terms of the amount that you and/or the computer player can earn. The options also differ in how many votes are needed for that option to be the one that is selected for this Round. In each Round, one of the two options will require that both you and the computer player you are matched with vote for it in order for it to be selected for this Round. The other option is selected for this Round so long as it receives at least one vote. Which option requires two votes and which option only requires at least one vote will be clearly stated before you make your decisions.

In the example above, for the option on the right to be selected for this Round, both you and the computer player you are matched with have to vote for it. On the other hand, for the option on the left to be selected for this Round, only one voter has to vote for it. In other words, if you vote for the option on the left, then it is selected for this Round regardless of what option the computer player you are matched with votes for. Similarly, if the computer player you are matched with votes for the option on the left it is selected for this Round regardless of which option you vote for.

Note that the option that requires one vote will always be on the left and the option that requires two votes will always be on the right hand side.

**Payment:** If this Part is randomly chosen to count for payment, then one Round will be randomly chosen to count for payment. In the example above, if the option on the right is selected for this Round, then you would receive \$6 and the computer player would receive 5 virtual (imaginary) dollars. If the option on the left is selected for this Round, then you would receive \$5 and the computer player would receive 10 virtual (imaginary) dollars. Which of the two options is selected for this Round will depend on what happens during the Round.

Recall that the virtual (imaginary) dollars the computer player earns will have no impact on you or anyone else at any point, ever.

Before telling you how the computer player has been programmed to vote, we will ask you 6 questions that test your understanding of these instructions. After you have answered these 6 questions, we will tell you how the computer player has been programmed to vote in Part 2 and then ask you one additional understanding question.

Do you have any questions?

We will now begin Block 1 in which you will be asked questions that test your understanding of this game. If Block 1 of Part 2 is randomly chosen to count for payment, then you earn \$10 if you answer ALL the questions correctly. If you make even one mistake you will earn only \$2.

[*Note: subjects received what follows after the Part 2 Block 1 questions.*]

In Part 2 the computer player has been programmed to **ALWAYS choose the option that gives it the highest number of virtual (imaginary) dollars.** That is, the computer player will look at which option gives it the highest number of virtual (imaginary) dollars, and will vote for that one.

Additional understanding question:

1. Which option will the computer player vote for? [the one that requires 1 vote only; the one that requires 2 votes; it will randomly choose which option to vote for, each option having an equal chance; it will vote for the option that gives it the highest amount of virtual (imaginary) dollars.]

# Part 3

Part 3 is in four Rounds. As described in the preamble, in each Round you will be matched with a computer player. In each Round you will be asked to vote for one of two options that determine how much you will earn. It will also determine how many virtual (imaginary) dollars the computer will earn. Note that how many virtual (imaginary) dollars the computer player earns will have no impact on you or anyone else at any point, ever.

In this Part of the Study, the way the computer player has been programmed to vote is different than in Part 2. Before we describe how the computer player is programmed to cast its vote in this Part of the study, we'll start by showing you an example of what you might see in a Round.

The game in Part 3 of this study is very similar to the game you played in Part 2. The difference lies in the kinds of options you face. In this part of the study, one of the options involves uncertainty. An example of a choice with an uncertain outcome is shown below. Please take a moment to look at the table below before I describe it.

| **Votes needed: 1** | **Votes needed: 2** |
|---|---|
| Your earnings: $11<br><br>Computer player: $9 virtual (imaginary) dollars | HEADS: Your earnings: $5<br>Computer player: $15 virtual (imaginary) dollars<br><br>TAILS: Your earnings: $15<br>Computer player: $5 virtual (imaginary) dollars |
| vote for this option | vote for this option |

In this particular example, there is no uncertainty regarding the option on the left (the one requiring only 1 vote): if this option is the one that is selected for this Round, you would receive $11 and the computer player you are matched with would receive $9 virtual (imaginary) dollars. However, there is uncertainty regarding the option on the right (the one requiring two votes).

The uncertainty can be described in the following way. As before, a coin is flipped and lands either heads or tails, each with an equal 50% chance. If it lands on heads, then the option on the right is: $5 for you and $15 virtual (imaginary) dollars for the computer player you are matched with. If, on the other hand, it lands on tails, the option on the right is: $15 for you and $5 virtual (imaginary) dollars for the computer player you are matched with. In other words, there is uncertainty in terms of which of the payoff pairs correspond to the option on the right: you do not know whether the payoff pair will be $5 for you and $15 virtual (imaginary) dollars for the computer player you are matched with, or whether it will be $15 for you and $5 virtual (imaginary) dollars for the computer player you are matched with. All you know is that the outcomes in the "uncertain" options are equally likely, each having 50% chance. In each Round the coin will be flipped before you and the computer player you are matched with make your choices, but what side the coin landed on and which payoff pair that corresponds to will not be revealed to anyone.

Recall that the virtual (imaginary) dollars the computer player earns will have no impact on you or anyone else at any point, ever.

Do you have any questions?

Before telling you how the computer player has been programmed to vote, we will ask you 4 questions that test your understanding of these instructions. After you have answered these 4 questions, we will tell you how the computer player has been programmed to vote in Part 3, and then ask you one additional understanding question.

We will now begin Block 1 in which you will be asked questions that test your understanding of this game. If Block 1 of Part 3 is randomly chosen to count for payment, then you earn $10 if you answer ALL the questions correctly. If you make even one mistake you will earn only $2.

After Block 1 is over, you will play the 4 Rounds of the Part 3 game.

[*Note: subjects received what follows after the Part 3 Block 1 questions.*]

In Part 3 the computer player has been programmed to **ALWAYS choose the option on the right**. That is, it will vote for the option on the right (the one requiring 2 votes) no matter what.

# Part 4

Part 4 is in four Rounds. As described in the preamble, in each Round you will be matched with a computer player. In each Round you will be asked to vote for one of two options that determine how much you will earn. It will also determine how many virtual (imaginary) dollars the computer will earn. Note that how many virtual (imaginary) dollars the computer player earns will have no impact on you or anyone else at any point, ever.

The game in Part 4 of this study is very similar to the game you played in Part 3 except for two things:

1. The first difference lies in that the computer player you are matched with will learn what the outcome of the coin flip is before it votes. You, however, will not know what the outcome of the coin flip is and will have to cast your vote without knowing the outcome of the coin flip.

2. The second is that the computer player has been programmed to vote for the option that gives it the highest number of virtual (imaginary) dollars. That is, after learning the outcome of the coin flip, it will look at which option gives it the highest amount of virtual (imaginary) dollars and vote for that one.

The kind of screen you face will look very similar to the kind of screen you faced in Part 3, except that you will be reminded that the virtual player knows the outcome of the coin flip AND always votes for the option that gives it the highest amount of virtual (imaginary) dollars:

**Recall that the computer player has learned the result of the coin flip and always votes for the option that gives it the highest amount of virtual (imaginary) dollars.**

| **Votes needed: 1** | **Votes needed: 2** |
|---|---|
| Your earnings: \$11<br><br>Computer player: \$9 virtual (imaginary) dollars | HEADS: Your earnings: \$5<br>Computer player: \$15 virtual (imaginary) dollars<br><br>TAILS: Your earnings: \$15<br>Computer player: \$5 virtual (imaginary) dollars |
| vote for this option | vote for this option |

Recall that the virtual (imaginary) dollars the computer player earns will have no impact on you or anyone else at any point, ever.

Do you have any questions?

We will now begin Block 1 in which you will be asked questions that test your understanding of this game. If Block 1 of Part 4 is randomly chosen to count for payment, then you earn \$10 if you answer ALL the questions correctly. If you make even one mistake, you will earn only \$2.

After Block 1 is over, you will play the four Rounds of the Part 4 game.

# Part 5

Part 5 also consists of two Blocks. In Part 5, Block 1 consists of 4 questions about the game you just played in Part 4. These questions will appear on a series of separate screens. If Block 1 of Part 5 is chosen for payment, you will be paid $10 if you answer all these questions correctly. If you make even one mistake, you will only earn $2.

Recall that in Part 4, in each Round, the computer player learned the outcome of the coin flip before it voted, and then always voted for the option that gave it the highest amount of virtual (imaginary) dollars.

After Block 1 is over, we will hand out instructions for Block 2.

Block 2:

You now are going to play 4 additional Rounds, just like the ones you played in Part 4. Recall that in those Rounds, the computer player you are matched with ALWAYS knows the result of the coin flip before it votes, and then ALWAYS votes for the option that will give it the highest number of of virtual (imaginary) dollars.

## A.3 Full Feedback treatment: subjects face robot players and receive partial feedback in Part 4 over 20 rounds

.

The instructions are identical to the HR treatment except for what happens in Part 4. Below we present the Part 4 instructions for this treatment.

# Part 4

Part 4 is in 20 Rounds. As described in the preamble, in each Round you will be matched with a computer player. In each Round you will be asked to vote for one of two options that determine how much you will earn. It will also determine how many virtual (imaginary) dollars the computer will earn.

Note that how many virtual (imaginary) dollars the computer player earns will have no impact on you or anyone else at any point, ever.

The game in Part 4 of this study is very similar to the game you played in Part 3 except for three things:

1. The first difference lies in that the computer player you are matched with will learn what the outcome of the coin flip is before it votes. You, however, will not know what the outcome of the coin flip is and will have to cast your vote without knowing the outcome of the coin flip.

2. The second is that the computer player has been programmed to vote for the option that gives it the highest number of virtual (imaginary) dollars. That is, after learning the outcome of the coin flip, it will look at which option gives it the highest amount of virtual (imaginary) dollars and vote for that one.

3. The third is that in each round, after you have cast your vote, you will receive information on what the outcome of the virtual coin flip was, and you will be told which option the computer player voted for. In addition, you will be told which option was selected for each round and you will be told how much you would earn if that were the round randomly chosen for payment.

The kind of screen you face will look very similar to the kind of screen you faced in Part 3, except that you will be reminded that the virtual player knows the outcome of the coin flip AND always votes for the option that gives it the highest amount of virtual (imaginary) dollars:

**Recall that the computer player has learned the result of the coin flip and always votes for the option that gives it the highest amount of virtual (imaginary) dollars.**

| **Votes needed: 1** | **Votes needed: 2** |
| --- | --- |
| Your earnings: \$11<br><br>Computer player: \$9 virtual (imaginary) dollars | HEADS: Your earnings: \$5<br>Computer player: \$15 virtual (imaginary) dollars<br><br>TAILS: Your earnings: \$15<br>Computer player: \$5 virtual (imaginary) dollars |
| vote for this option | vote for this option |

Recall that the virtual (imaginary) dollars the computer player earns will have no impact on you or anyone else at any point, ever. So, the only thing that determines your earnings in this study are the number of dollars you earn in the option that is selected for a Round.

Do you have any questions?

We will now begin Block 1 in which you will be asked questions that test your understanding of this game. If Block 1 of Part 4 is randomly chosen to count for payment, then you earn \$10 if you answer ALL the questions correctly. If you make even one mistake, you will earn only \$2.

After Block 1 is over, you will play the 20 Rounds of the Part 4 game.

### A.3.1 Screen Shot - Example of Feedback Screen

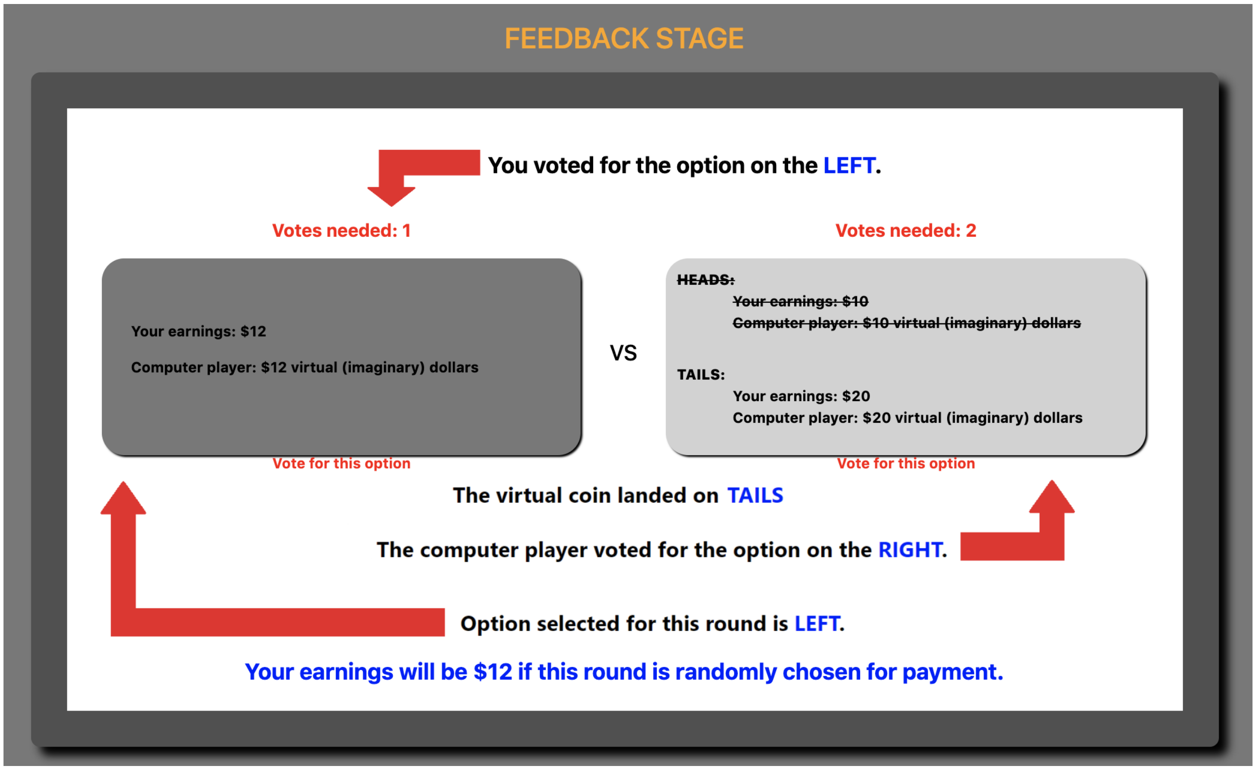



*Figure 3: Full Feedback Example.*

## A.4 Partial Feedback treatment: subjects face robot players and receive partial feedback in Part 4 over 20 rounds

.

The instructions are identical to the HR treatment except for what happens in Part 4. Below we present the Part 4 instructions for this treatment.

# Part 4

Part 4 is in 20 Rounds. As described in the preamble, in each Round you will be matched with a computer player. In each Round you will be asked to vote for one of two options that determine how much you will earn. It will also determine how many virtual (imaginary) dollars the computer will earn.

Note that how many virtual (imaginary) dollars the computer player earns will have no impact on you or anyone else at any point, ever.

The game in Part 4 of this study is very similar to the game you played in Part 3 except for three things:

1. The first difference lies in that the computer player you are matched with will learn what the outcome of the coin flip is before it votes. You, however, will not know what the outcome of the coin flip is and will have to cast your vote without knowing the outcome of the coin flip.

2. The second is that the computer player has been programmed to vote for the option that gives it the highest number of virtual (imaginary) dollars. That is, after learning the outcome of the coin flip, it will look at which option gives it the highest amount of virtual (imaginary) dollars and vote for that one.

3. The third is that in each round, after you have cast your vote, you will be told how much you would earn if that were the round randomly chosen for payment.

The kind of screen you face will look very similar to the kind of screen you faced in Part 3, except that you will be reminded that the virtual player knows the outcome of the coin flip AND always votes for the option that gives it the highest amount of virtual (imaginary) dollars:

**Recall that the computer player has learned the result of the coin flip and always votes for the option that gives it the highest amount of virtual (imaginary) dollars.**

| **Votes needed: 1** | **Votes needed: 2** |
|---|---|
| Your earnings: \$11<br><br>Computer player: \$9 virtual (imaginary) dollars | HEADS: Your earnings: \$5<br>Computer player: \$15 virtual (imaginary) dollars<br><br>TAILS: Your earnings: \$15<br>Computer player: \$5 virtual (imaginary) dollars |
| vote for this option | vote for this option |

Recall that the virtual (imaginary) dollars the computer player earns will have no impact on you or anyone else at any point, ever. So, the only thing that determines your earnings in this study are the number of dollars you earn in the option that is selected for a Round.

Do you have any questions?

We will now begin Block 1 in which you will be asked questions that test your understanding of this game. If Block 1 of Part 4 is randomly chosen to count for payment, then you earn \$10 if you answer ALL the questions correctly. If you make even one mistake, you will earn only \$2.

After Block 1 is over, you will play the 20 Rounds of the Part 4 game.

### A.4.1 Screen Shot - Example of Feedback Screen

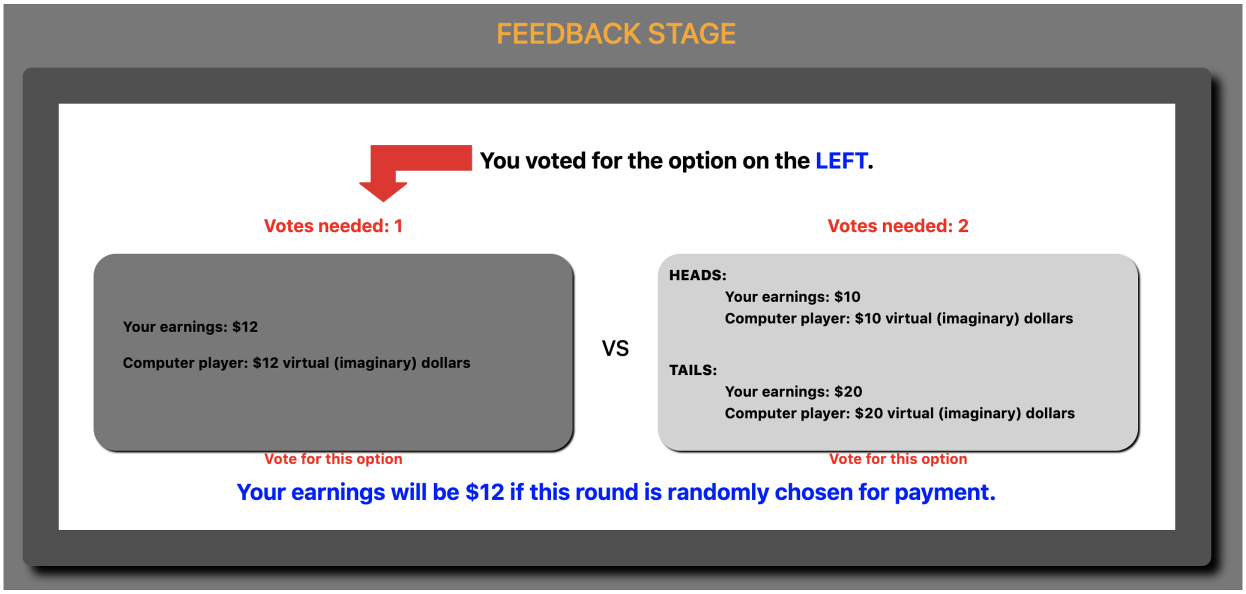
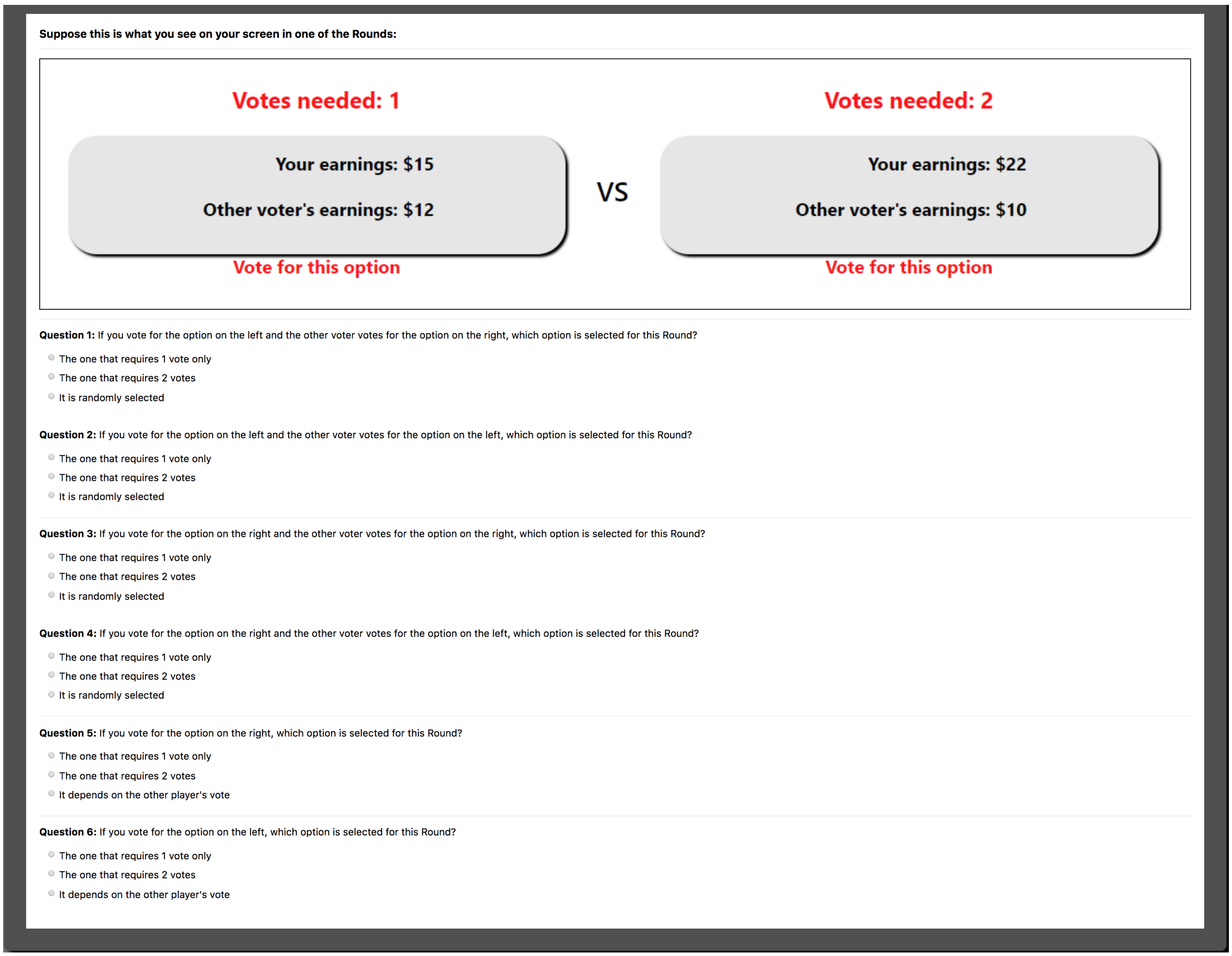



*Figure 4: Partial Feedback Example.*

# B Understanding Questions

Below we present screen shots for each of our understanding questions.

## B.1 Treatment 1: subjects play against each other



**Suppose this is what you see on your screen in one of the Rounds:**

Votes needed: 1

Your earnings: $15

Other voter's earnings: $12

Vote for this option

VS

Votes needed: 2

Your earnings: $22

Other voter's earnings: $10

Vote for this option

**Question 1:** If you vote for the option on the left and the other voter votes for the option on the right, which option is selected for this Round?

- The one that requires 1 vote only
- The one that requires 2 votes
- It is randomly selected

**Question 2:** If you vote for the option on the left and the other voter votes for the option on the left, which option is selected for this Round?

- The one that requires 1 vote only
- The one that requires 2 votes
- It is randomly selected

**Question 3:** If you vote for the option on the right and the other voter votes for the option on the right, which option is selected for this Round?

- The one that requires 1 vote only
- The one that requires 2 votes
- It is randomly selected

**Question 4:** If you vote for the option on the right and the other voter votes for the option on the left, which option is selected for this Round?

- The one that requires 1 vote only
- The one that requires 2 votes
- It is randomly selected

**Question 5:** If you vote for the option on the right, which option is selected for this Round?

- The one that requires 1 vote only
- The one that requires 2 votes
- It depends on the other player's vote

**Question 6:** If you vote for the option on the left, which option is selected for this Round?

- The one that requires 1 vote only
- The one that requires 2 votes
- It depends on the other player's vote

*Figure 5: Part 2 Understanding questions.*

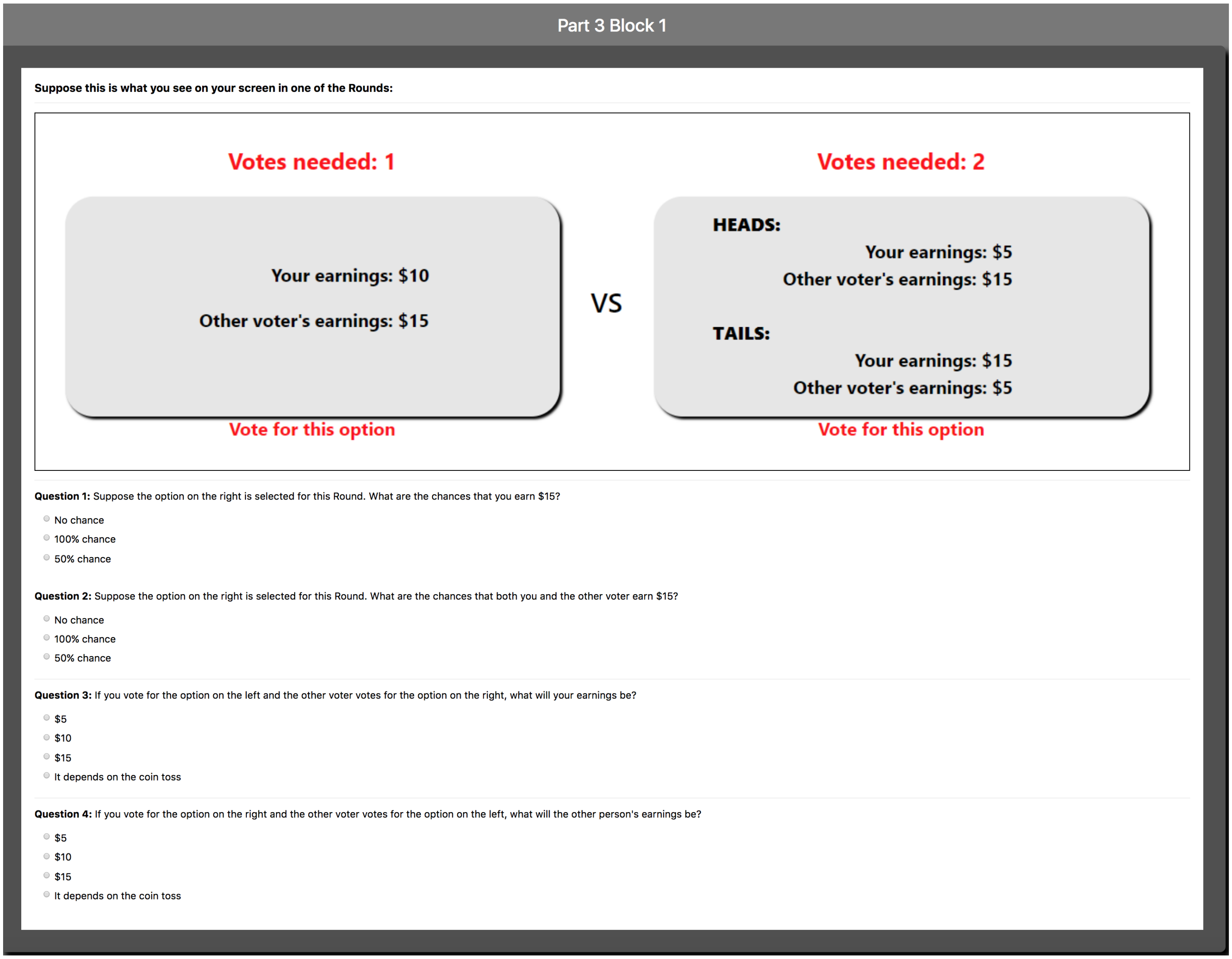
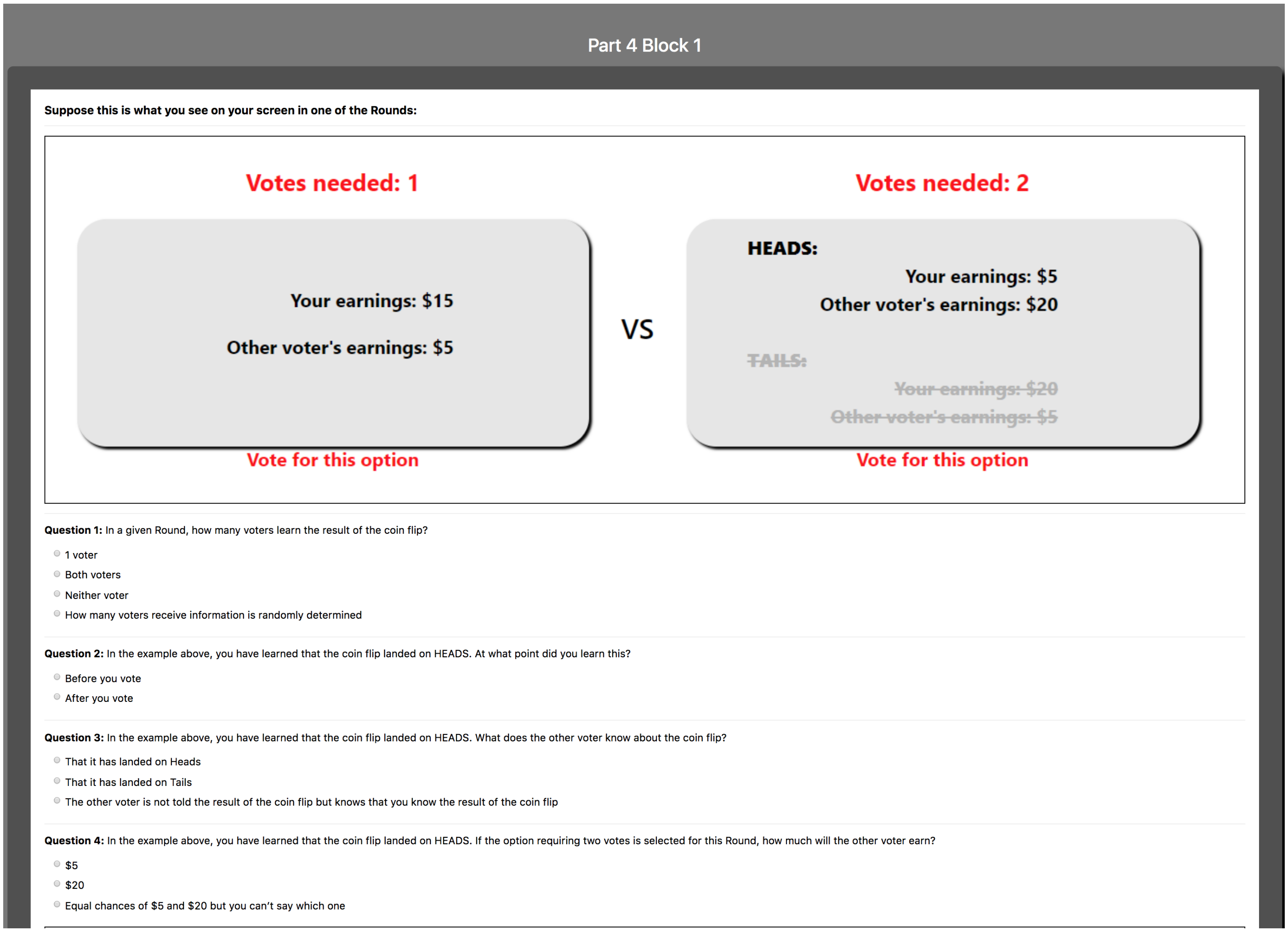



*Figure 6: Part 3 Understanding questions.*


Part 4 Block 1

**Suppose this is what you see on your screen in one of the Rounds:**

| | | |
|---|---|---|
| **Votes needed: 1** | | **Votes needed: 2** |
| **Your earnings: $15**<br>**Other voter's earnings: $5** | **VS** | **HEADS:**<br>**Your earnings: $5**<br>**Other voter's earnings: $20**<br>~~TAILS:~~<br>~~Your earnings: $20~~<br>~~Other voter's earnings: $5~~ |
| **Vote for this option** | | **Vote for this option** |

**Question 1:** In a given Round, how many voters learn the result of the coin flip?

- 1 voter
- Both voters
- Neither voter
- How many voters receive information is randomly determined

**Question 2:** In the example above, you have learned that the coin flip landed on HEADS. At what point did you learn this?

- Before you vote
- After you vote

**Question 3:** In the example above, you have learned that the coin flip landed on HEADS. What does the other voter know about the coin flip?

- That it has landed on Heads
- That it has landed on Tails
- The other voter is not told the result of the coin flip but knows that you know the result of the coin flip

**Question 4:** In the example above, you have learned that the coin flip landed on HEADS. If the option requiring two votes is selected for this Round, how much will the other voter earn?

- $5
- $20
- Equal chances of $5 and $20 but you can't say which one

*Figure 7: Part 4 Understanding questions (questions 1-4).*

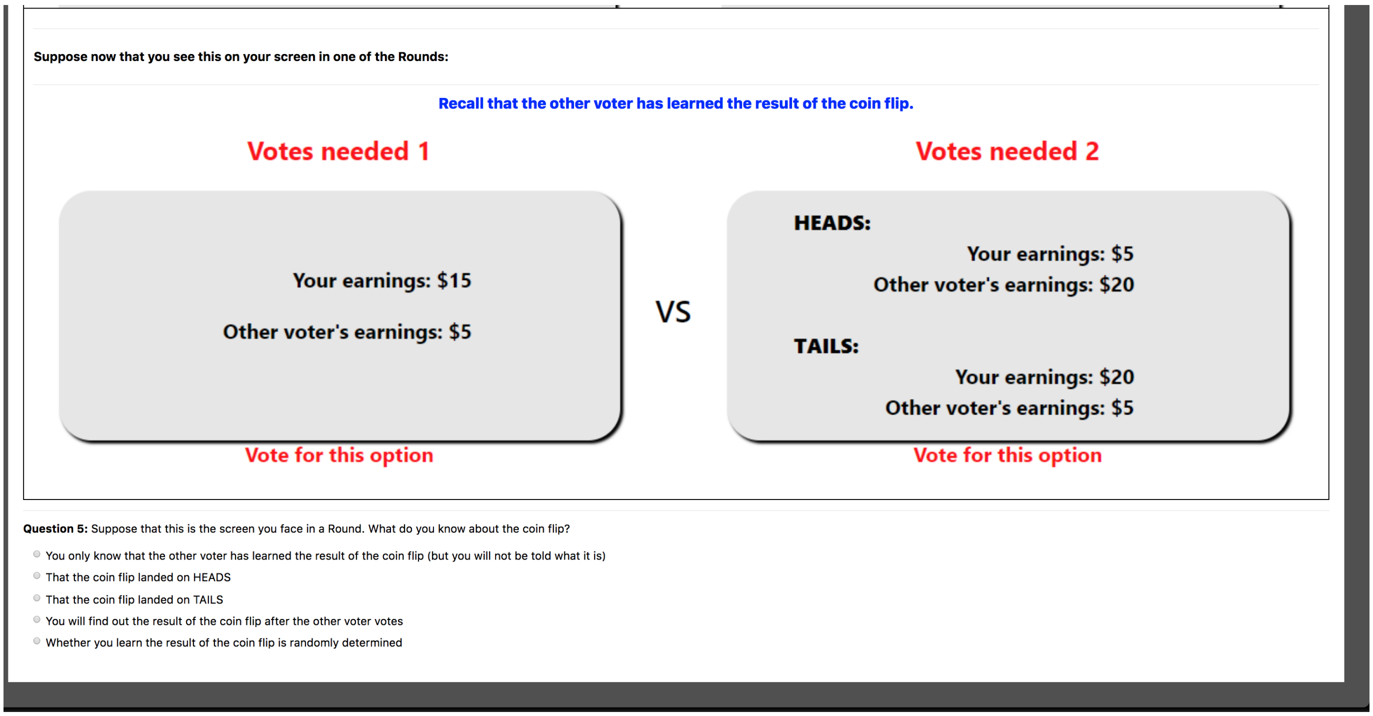
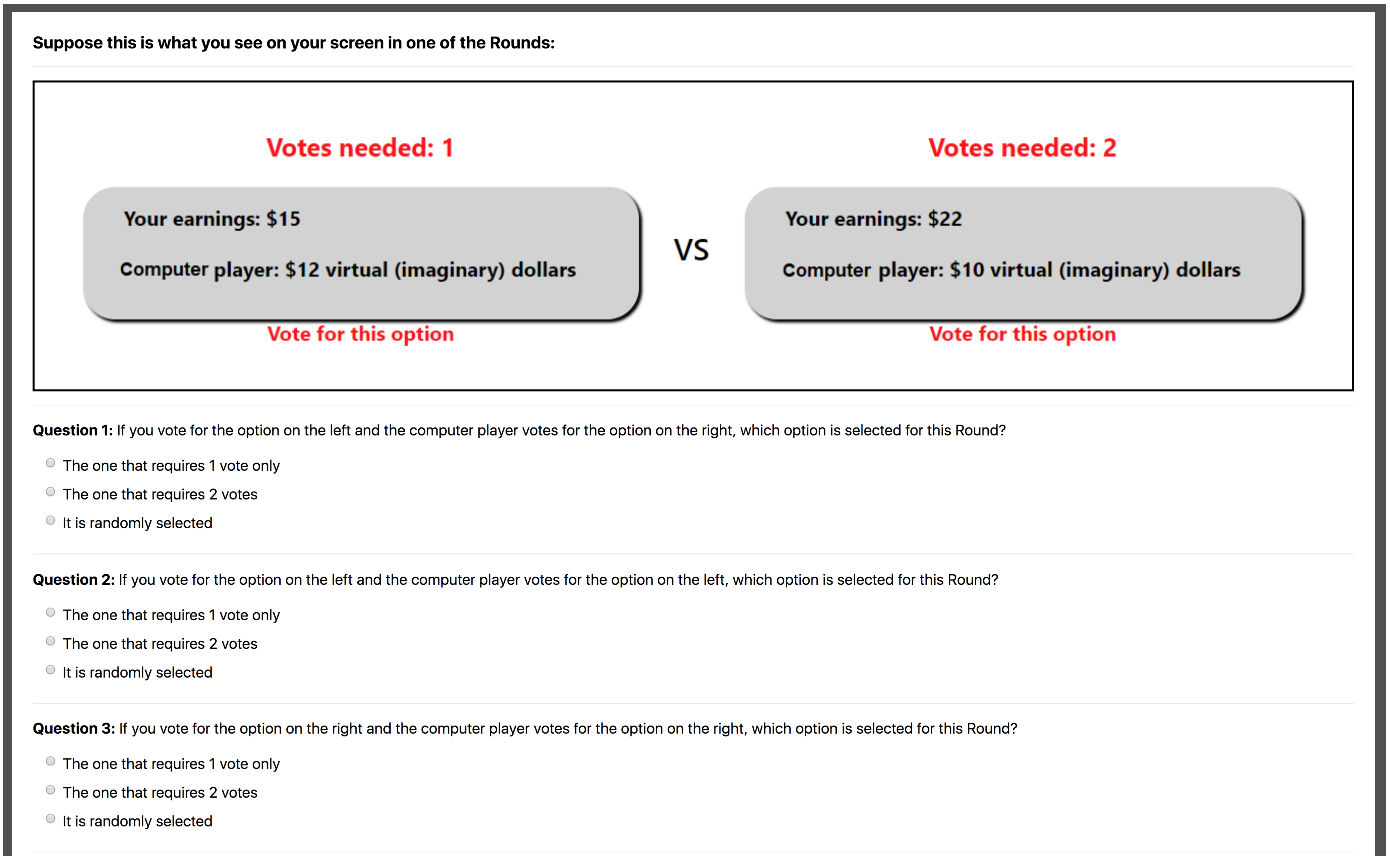



*Figure 8: Part 4 Understanding questions (question 5).*

## B.2 Treatment 2: subjects play against a computerized robot player



**Suppose this is what you see on your screen in one of the Rounds:**

| Votes needed: 1 | | Votes needed: 2 |
|---|---|---|
| Your earnings: $15<br>Computer player: $12 virtual (imaginary) dollars | VS | Your earnings: $22<br>Computer player: $10 virtual (imaginary) dollars |
| Vote for this option | | Vote for this option |

**Question 1:** If you vote for the option on the left and the computer player votes for the option on the right, which option is selected for this Round?

- The one that requires 1 vote only
- The one that requires 2 votes
- It is randomly selected

**Question 2:** If you vote for the option on the left and the computer player votes for the option on the left, which option is selected for this Round?

- The one that requires 1 vote only
- The one that requires 2 votes
- It is randomly selected

**Question 3:** If you vote for the option on the right and the computer player votes for the option on the right, which option is selected for this Round?

- The one that requires 1 vote only
- The one that requires 2 votes
- It is randomly selected

*Figure 9: Part 2 Understanding questions (questions 1-3).*

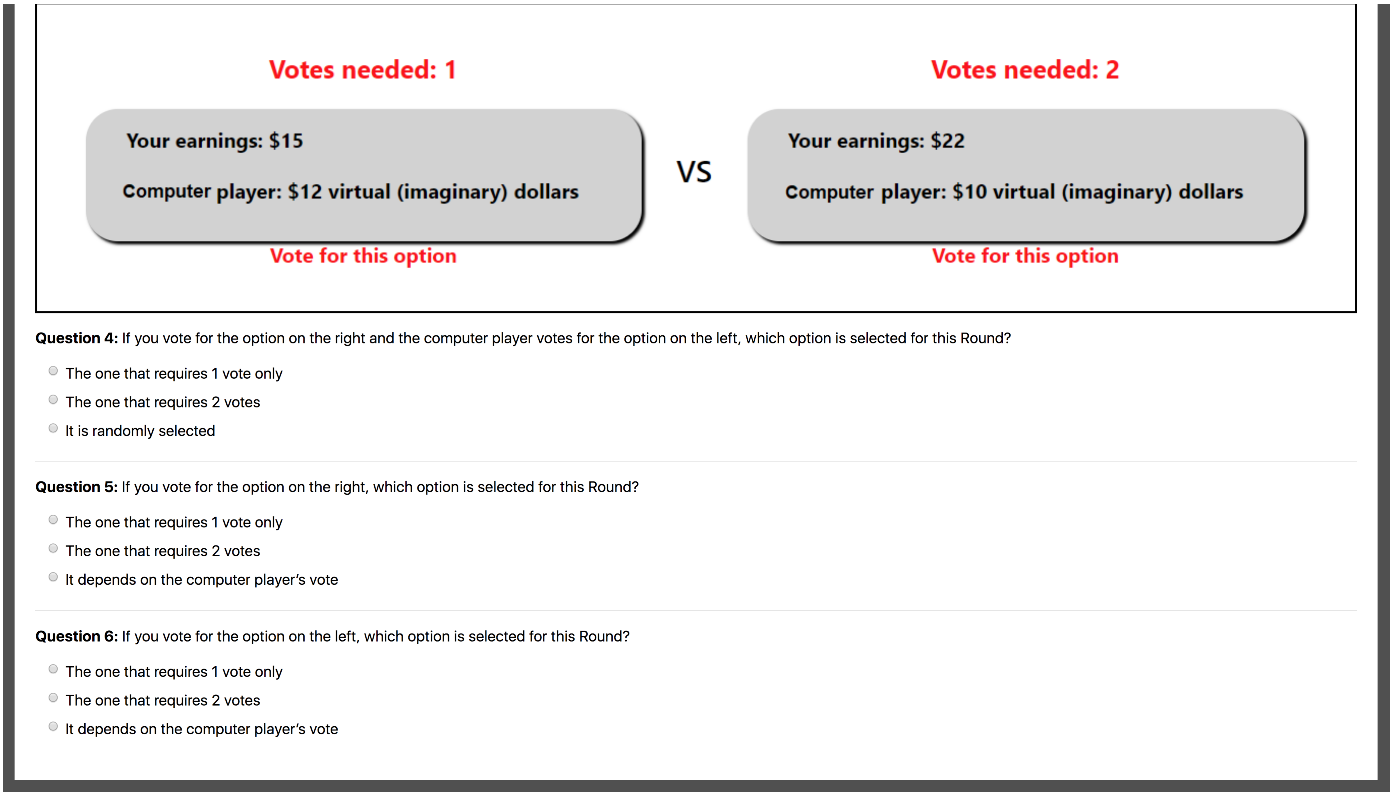



*Figure 10: Part 2 Understanding question (questions 4-6).*

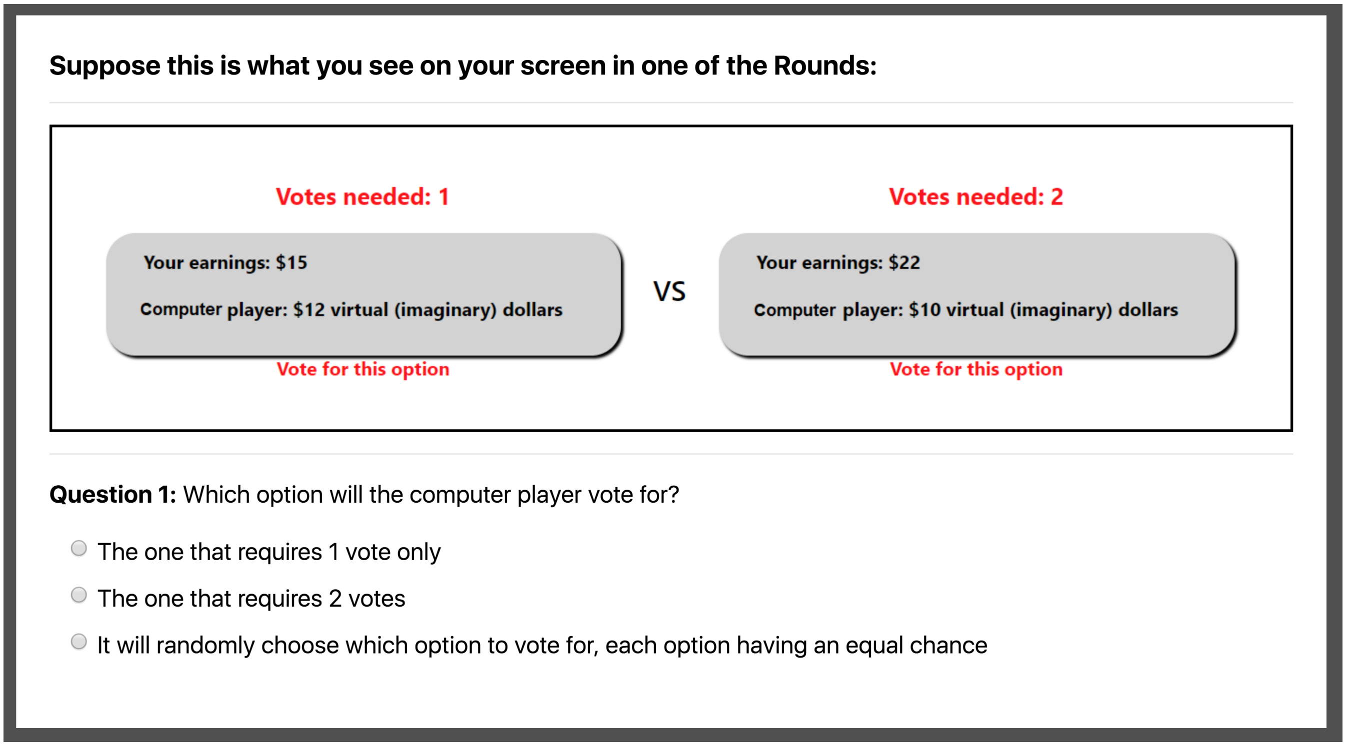



*Figure 11: Part 2 Understanding question (question 7).*

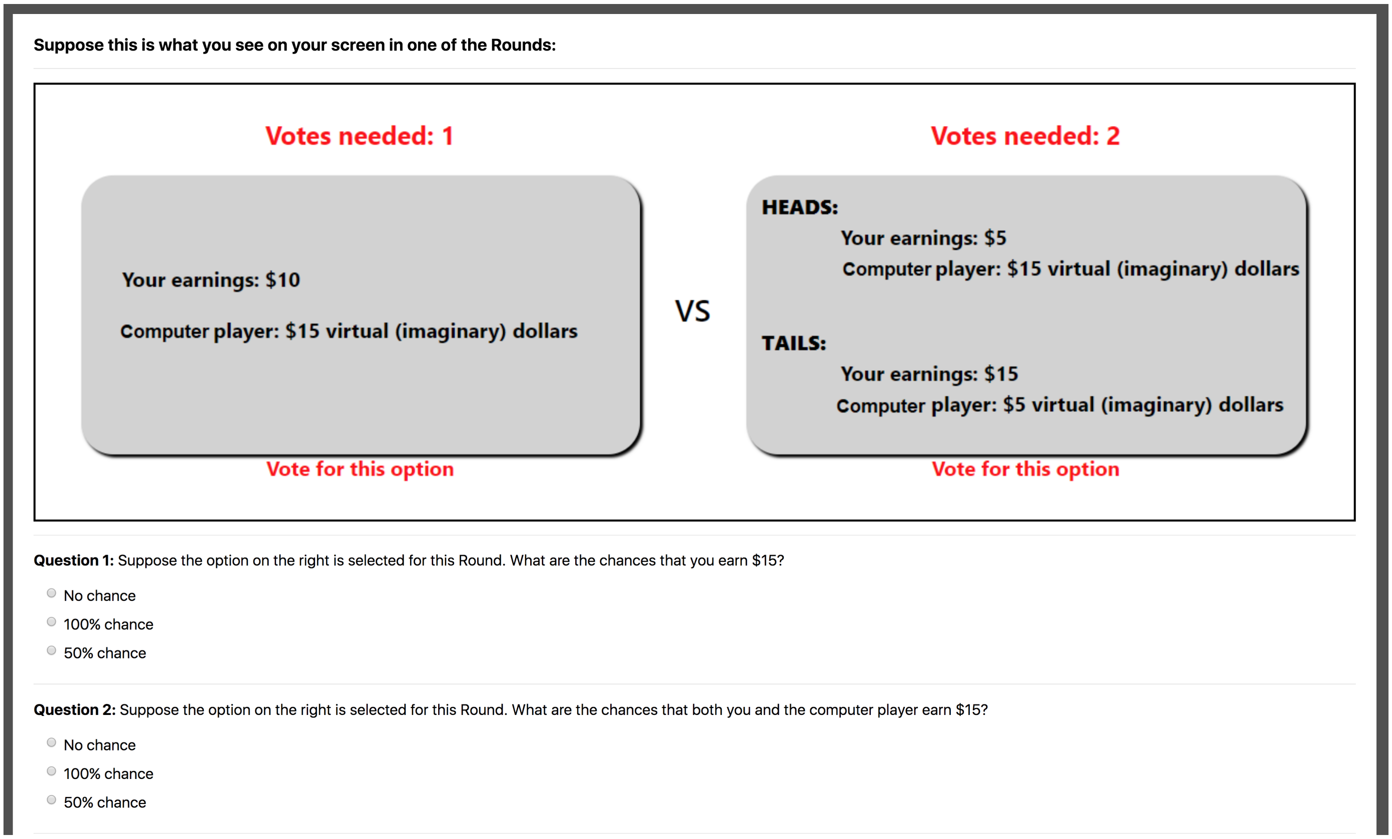



*Figure 12: Part 3 Understanding questions (questions 1-2).*

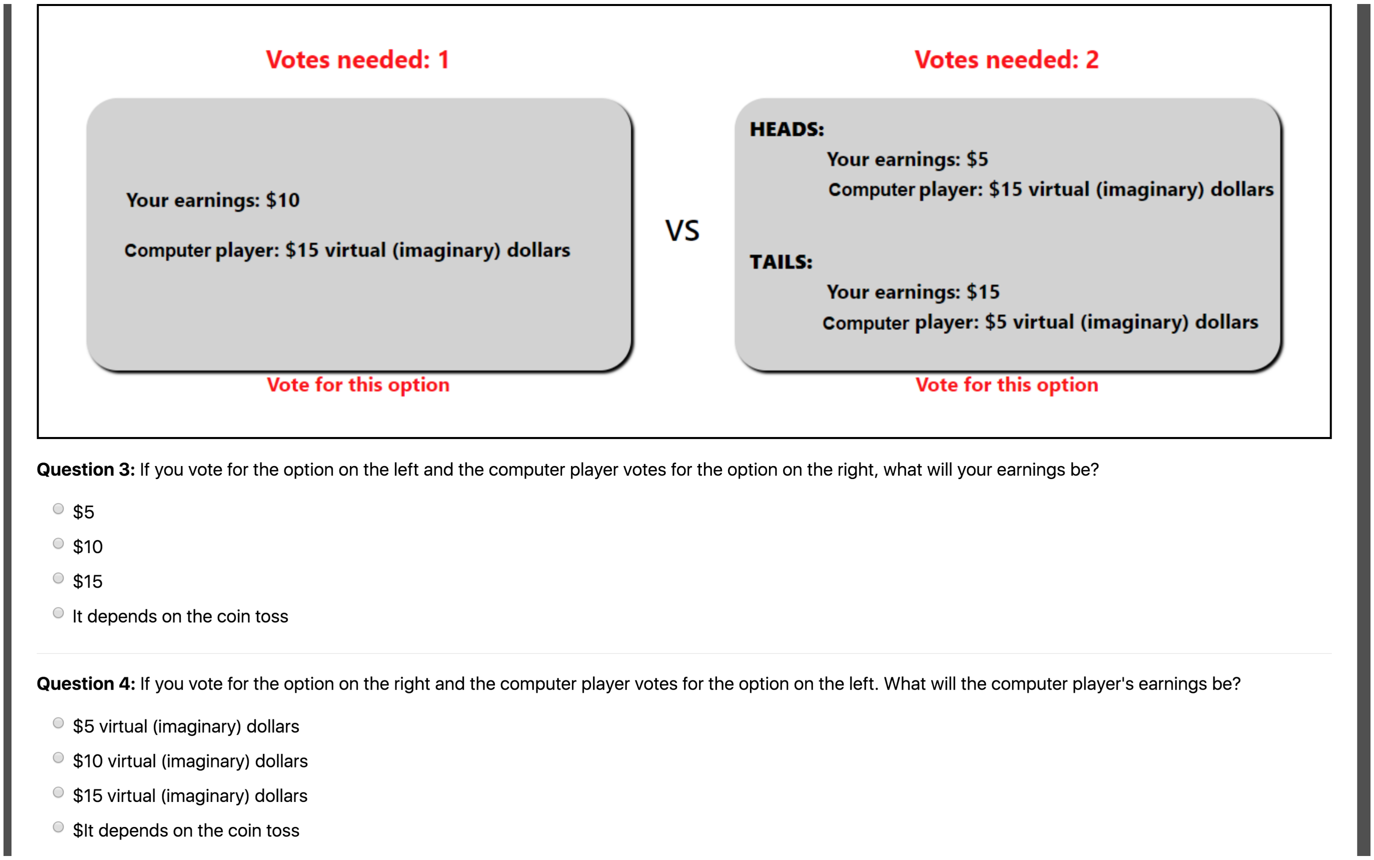



*Figure 13: Part 3 Understanding questions (questions 3-4).*

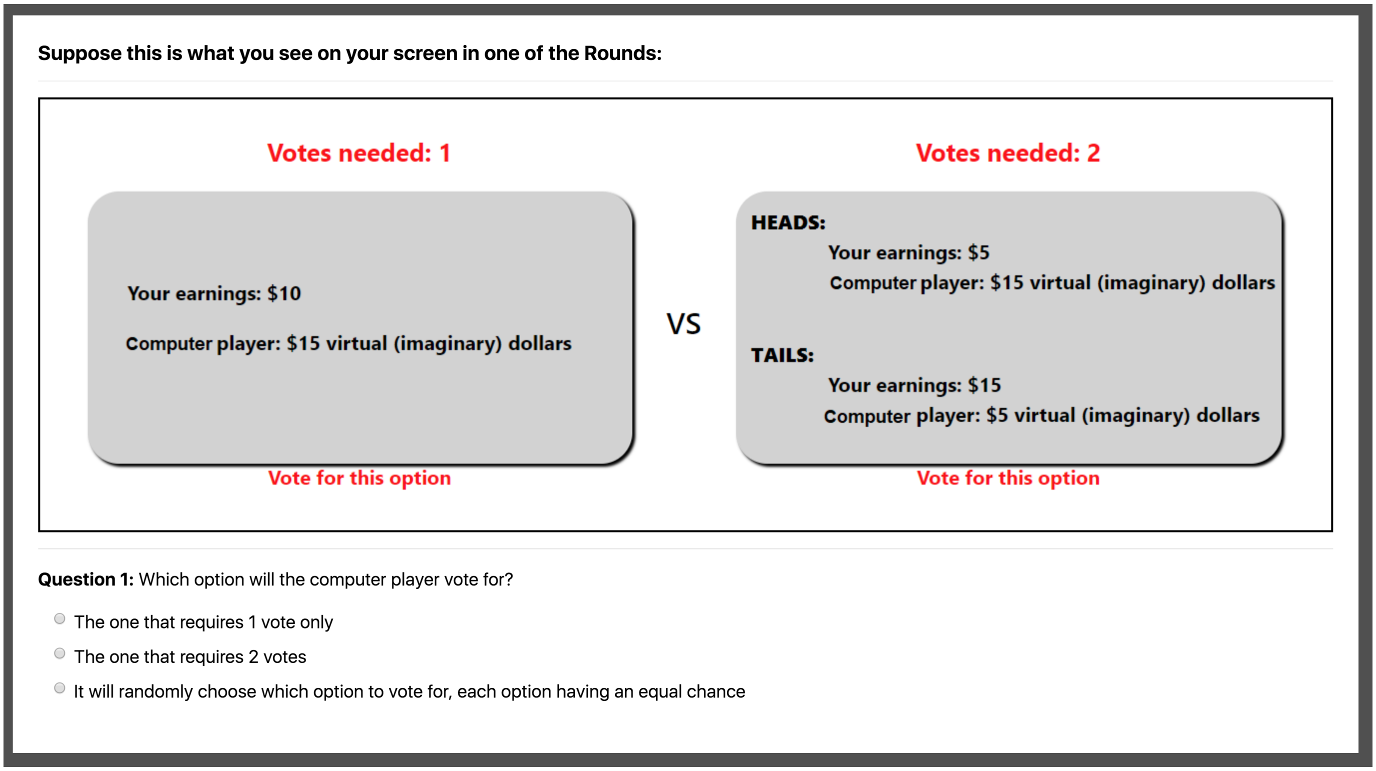
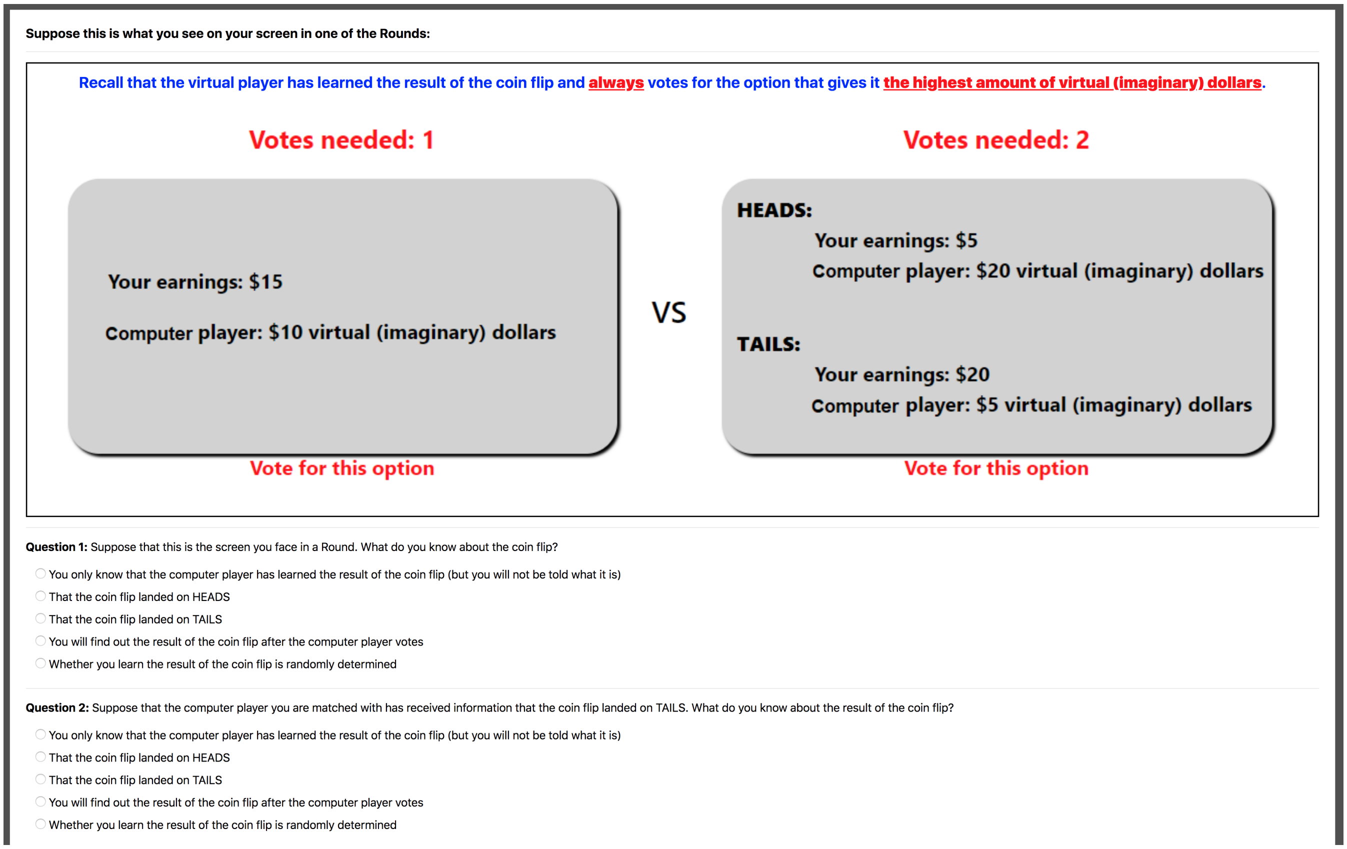



*Figure 14: Part 3 Understanding questions (question 5).*



**Suppose this is what you see on your screen in one of the Rounds:**

**Recall that the virtual player has learned the result of the coin flip and always votes for the option that gives it the highest amount of virtual (imaginary) dollars.**

| Votes needed: 1 | | Votes needed: 2 |
| --- | --- | --- |
| Your earnings: $15<br>Computer player: $10 virtual (imaginary) dollars | VS | HEADS:<br>Your earnings: $5<br>Computer player: $20 virtual (imaginary) dollars<br>TAILS:<br>Your earnings: $20<br>Computer player: $5 virtual (imaginary) dollars |
| Vote for this option | | Vote for this option |

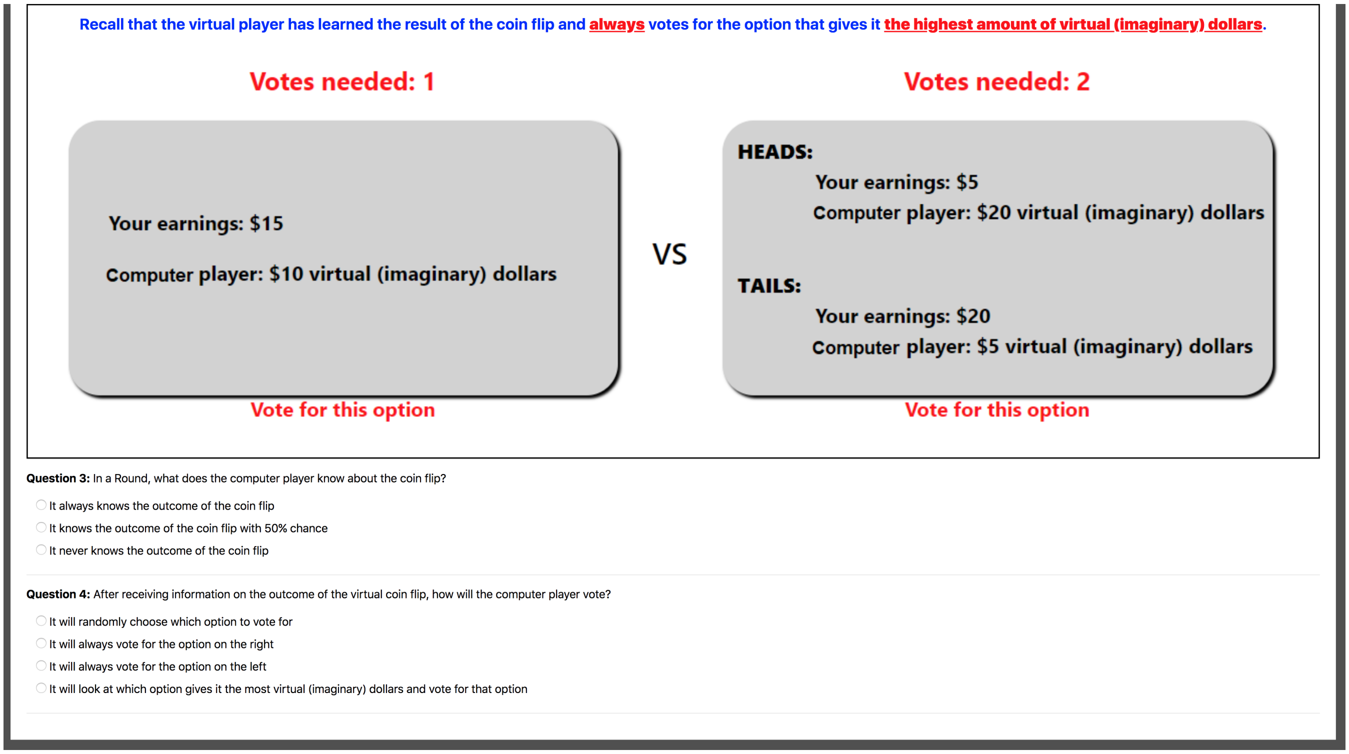
**Question 1:** Suppose that this is the screen you face in a Round. What do you know about the coin flip?

- You only know that the computer player has learned the result of the coin flip (but you will not be told what it is)
- That the coin flip landed on HEADS
- That the coin flip landed on TAILS
- You will find out the result of the coin flip after the computer player votes
- Whether you learn the result of the coin flip is randomly determined

**Question 2:** Suppose that the computer player you are matched with has received information that the coin flip landed on TAILS. What do you know about the result of the coin flip?

- You only know that the computer player has learned the result of the coin flip (but you will not be told what it is)
- That the coin flip landed on HEADS
- That the coin flip landed on TAILS
- You will find out the result of the coin flip after the computer player votes
- Whether you learn the result of the coin flip is randomly determined

*Figure 15: Part 4 Understanding questions (questions 1-2).*


Recall that the virtual player has learned the result of the coin flip and **always** votes for the option that gives it **the highest amount of virtual (imaginary) dollars**.

**Votes needed: 1**

Your earnings: $15

Computer player: $10 virtual (imaginary) dollars

Vote for this option

VS

**Votes needed: 2**

HEADS:

Your earnings: $5

Computer player: $20 virtual (imaginary) dollars

TAILS:

Your earnings: $20

Computer player: $5 virtual (imaginary) dollars

Vote for this option

**Question 3:** In a Round, what does the computer player know about the coin flip?

- It always knows the outcome of the coin flip
- It knows the outcome of the coin flip with 50% chance
- It never knows the outcome of the coin flip

**Question 4:** After receiving information on the outcome of the virtual coin flip, how will the computer player vote?

- It will randomly choose which option to vote for
- It will always vote for the option on the right
- It will always vote for the option on the left
- It will look at which option gives it the most virtual (imaginary) dollars and vote for that option

*Figure 16: Part 4 Understanding questions (questions 3-4).*

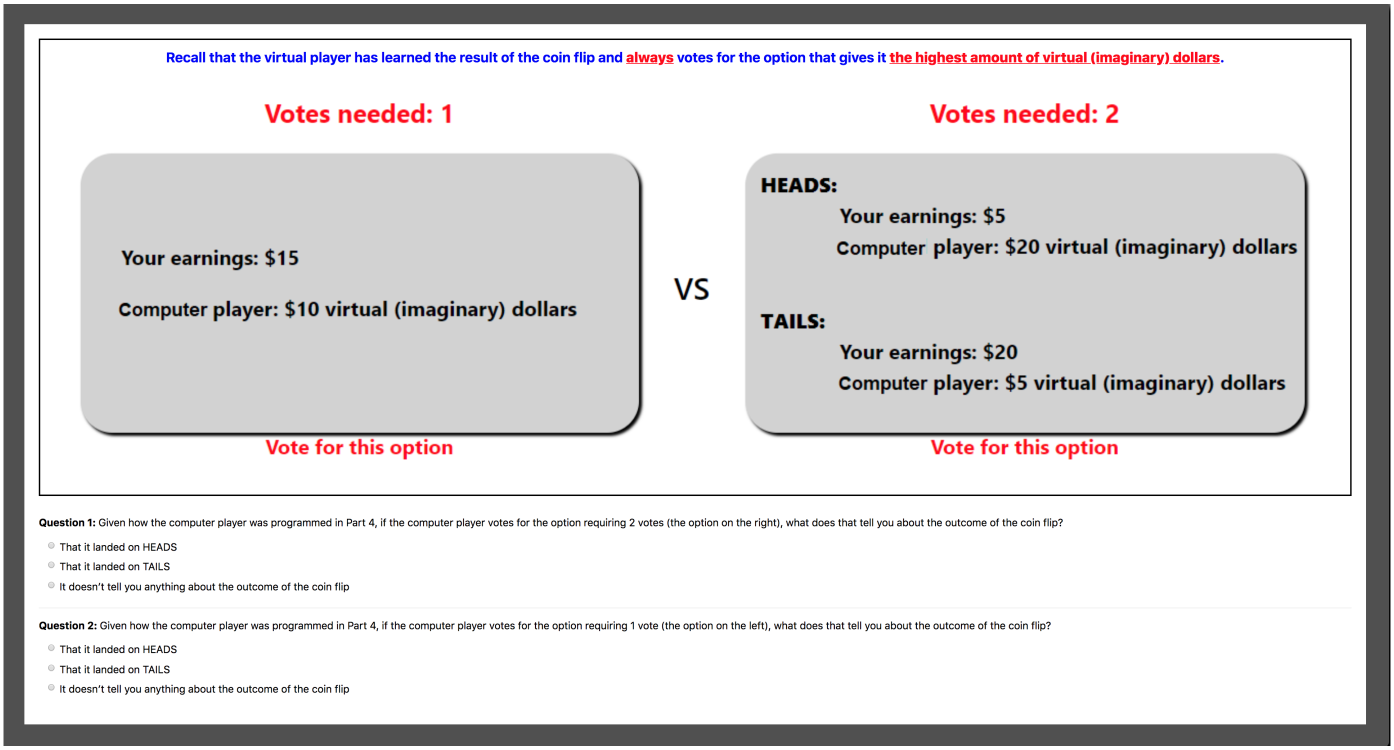



*Figure 17: Part 5, Contingent-Reasoning Understanding questions (questions 1-2).*

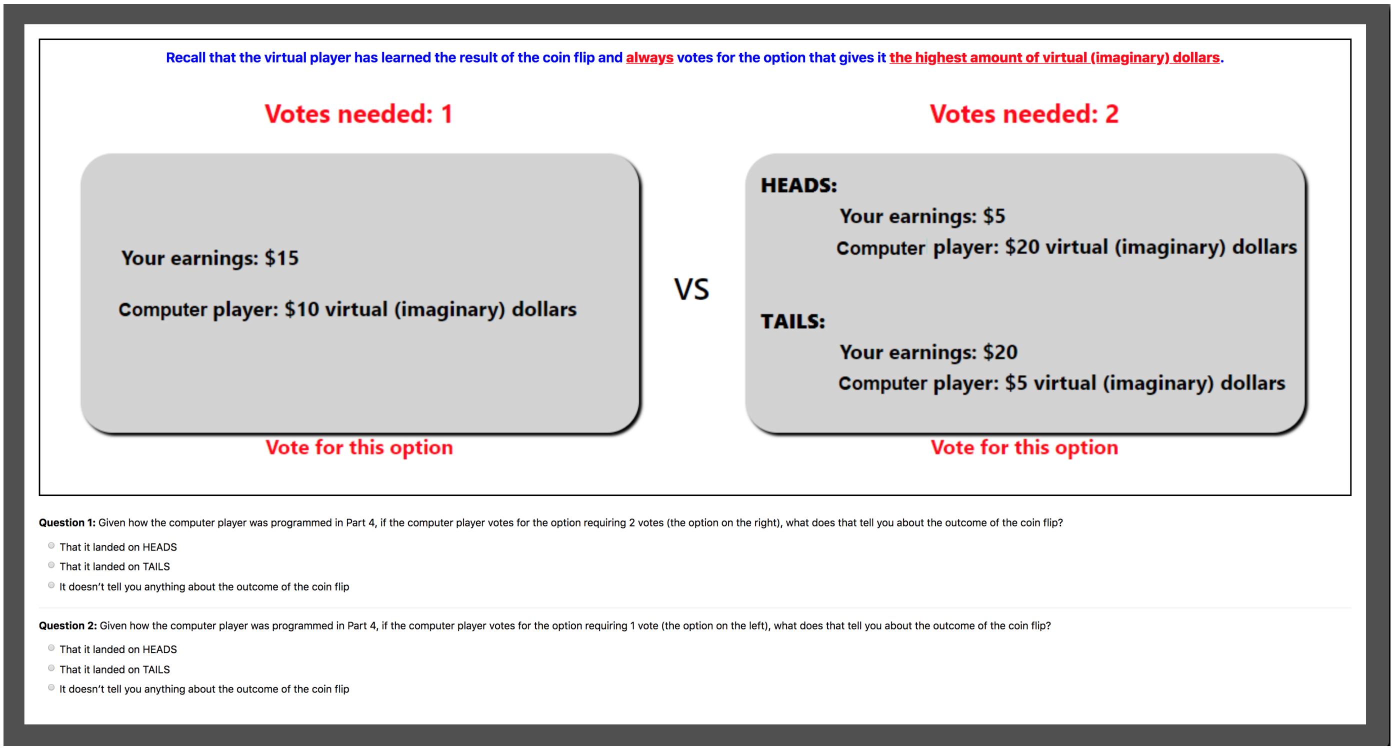



*Figure 18: Part 5, Contingent-Reasoning Understanding questions (question 3).*

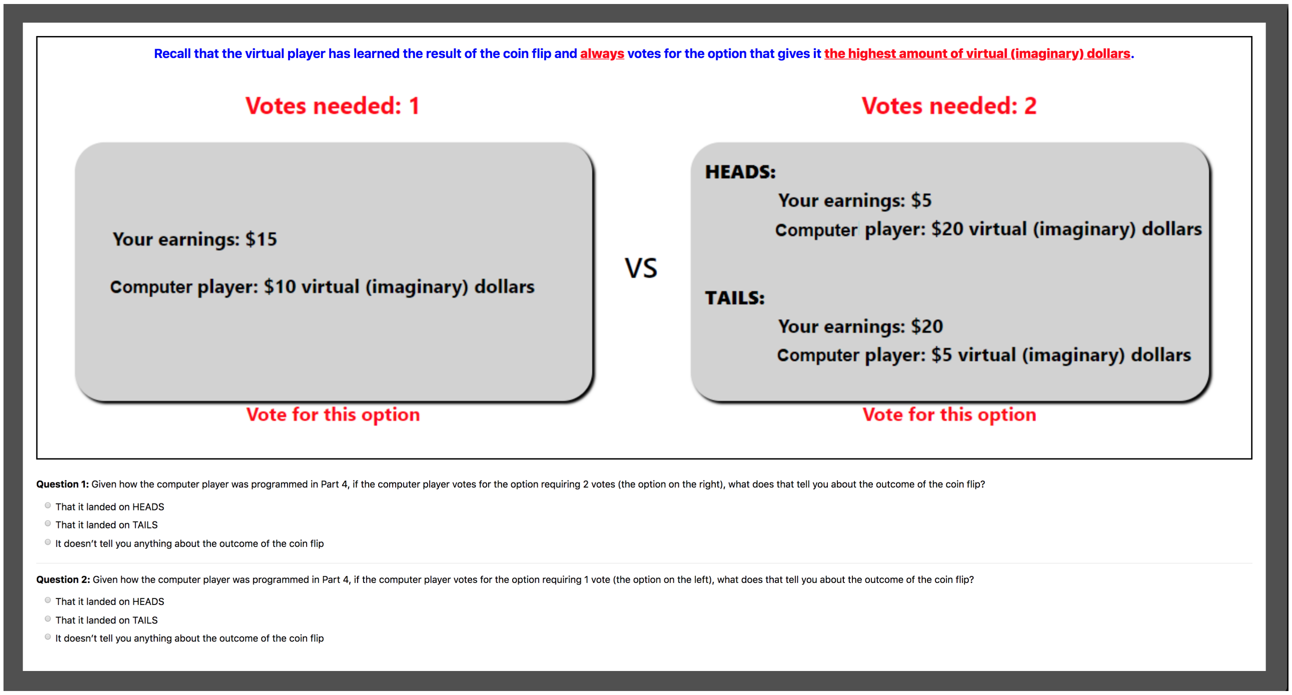



*Figure 19: Part 5, Contingent-Reasoning Understanding questions (question 4).*

# C Parts 1 and 2 analysis

## C.1 Part 1 - lottery questions

In Part 1, subjects faces a series of 9 rounds of an individual decision-making task. In each round subjects had the choice between two Options. Option A was a fixed amount (that varied from round to round). Option B was a lottery that paid \$10 with 50% chance, and \$20 with 50% chance. For each subject, the fixed amounts were drawn randomly without replacement from the following list $\$11, \$12, ..., \$18, \$19$. After Part 1 was over, subjects were informed that from that point onwards, they would be matched into pairs with another player in the room.

Table 7 shows the fraction of subjects choosing the lottery for each round. A large majority of subjects (88.4%) have a single switching point. That is, there is a fixed amount above which they always forgo the lottery and below which they always choose the lottery. Almost all subjects prefer the lottery over the fixed amount of \$12, and almost all prefer \$16 to the lottery.

*Table 7: Part 1 choices.*

| Round[a] | Option A | Option B[b] | Fraction choosing the lottery | Switching point (assuming single SP) |
|---|---|---|---|---|
| 1 | 11 | Lottery | 97.7% | 0% |
| 2 | 12 | Lottery | 94.2% | 5.3% |
| 3 | 13 | Lottery | 90.7% | 4% |
| 4 | 14 | Lottery | 62.8% | 26.3% |
| 5 | 15 | Lottery | 15.1% | 51.3% |
| 6 | 16 | Lottery | 4.7% | 11.8% |
| 7 | 17 | Lottery | 1.2% | 1.3% |
| 8 | 18 | Lottery | 2.3% | 0% |
| 9 | 19 | Lottery | 0% | 0% |
| Fraction (number) of subjects with a single switch point: 88.4% (76) | | | | |

[a]The order of rounds was randomly determined for each subject in each session, so was which option appeared on the left or right of the screen.

[b]In all rounds, the lottery paid \$10 with 50% chance and \$20 with 50% chance.

## C.2 Part 2 - voting, no uncertainty

In Part 2, subjects play 6 rounds of a game in which each player in a pair has to vote for one of two options that determine outcomes for both players. Just as in the Main Game, Option A is implemented so long as it receives at least one vote, while Option B requires two votes to be implemented. Unlike in the Main Game, Option B here consists of a fixed and known allocation. Table 8 presents the parameters that subjects faced in each round as well as the fraction of subjects who chose Option B in each of those rounds. The order in which rounds were presented to the subjects was randomly determined and thus varied from subject to subject.

*Table 8: Part 2 choices.*

| Round[a] | A (1 vote) | B (2 votes) | Fraction Choosing Option B |
|---|---|---|---|
| 1 | (\$12 ; \$12) | (\$10 ; \$20) | 0.198 |
| 2 | (\$12 ; \$12) | (\$20 ; \$10) | 0.686 |
| 3 | (\$16 ; \$16) | (\$10 ; \$20) | 0.023 |
| 4 | (\$16 ; \$16) | (\$20 ; \$10) | 0.419 |
| 5 | (\$12 ; \$16) | (\$16 ; \$12) | 0.872 |
| 6 | (\$16 ; \$12) | (\$12 ; \$16) | 0.058 |

[a]The order of rounds was randomly determined for each subject in each session.

Each of rounds 1−4 correspond to choices that parallel the Main Game. Round 5 was included to see whether subjects voted as if they were pivotal. Round 6 was included as a "sanity check" to see whether subjects voted for Option A, since the allocations in Options A and B are symmetric. More often than not, subjects vote for the option that gives them the highest payoff, and act as if they are pivotal.[32] This is especially salient in round 5, where inequality and efficiency are held constant across Options A and B. There, over 87% of our subjects vote for Option B, even though they may anticipate that their pair member will vote for Option A, and that their own vote will not be relevant. Our round 6 "sanity check" shows that all but a few subjects make choices that are payoff-maximizing in the absence of inequality and efficiency concerns.

[32] We do not expect that this number be 100% since subjects may have social preferences, and may hold beliefs on their partner's voting that would render their own vote irrelevant.

# D Demographic Information

Table 9 presents statistics on the demographic information that we collected from subjects via a questionnaire.

*Table 9: Average demographic information.*

| | Human-Human Treatment | Human-Robot Treatment | Partial Feedback Treatment | Full Feedback Treatment |
|---|---|---|---|---|
| Female | 55.8% | 64.3% | 64.0% | 63.9% |
| Age | 21.1 | 21.0 | 22.2 | 22.0 |
| GPA | 3.4 | 3.5 | 3.4 | 3.5 |
| Nb Years at PSU | 3.2 | 2.9 | 3.0 | 2.6 |
| Nb. of subjects | 86 | 82 | 86 | 83 |

A series of Fisher exact and Chi-squared tests (age, school/major at PSU), test of probability (female) and ranksum test (GPA) reject the hypotheses that subjects in the treatments come from different populations.[33] We also find that behaving according to theory is no different across the different groups. For example, women aren't more likely to behave according to theory relative to men (in either treatment). Likewise, age, GPA or other observables don't appear to have influence on making choices according to predictions.

[33] In all treatments, the most represented major/school was Economics or Business followed by Science and Engineering. The fraction of subjects who are in those fields in each treatment represents 74.4%, 63.4%, 67.4% and 63.9%, respectively.