ONLINE APPENDIX

Confidence, Self-Selection and Bias in the Aggregate

Benjamin Enke, Thomas Graeber and Ryan Oprea

A More on the Cognitive Tasks

We describe the 15 cognitive tasks in some detail here. See Appendix F for the full task instructions.

Base rate neglect (BRN). An important principle of rational information-processing is to take into account base rates, but a voluminous line of work documents that people tend to neglect base rates (e.g., Bar-Hillel, 1980; Kahneman and Tversky, 1973). We devised a simple variant of Tversky and Kahneman's (1982) well-known taxi-cab problem, which is known to generate responses that neglect the base rate. In our problem, a quality control machine of a bike manufacturer classifies bikes as good or defective but misclassifies any given bike 25% of the time. Subjects are asked to state the percentage chance that a bike is actually defective, given that the base rate for a defective bike is 10% and that the quality control machine classifies the bike as defective. A common incorrect answer is 75%, whereas the statistically correct answer is 25%.

Correlation neglect (CN). Taking into account potential non-independence of data is a core principle of both rational belief updating and econometrics courses. We devised a simplified version of the correlation neglect task developed in Enke and Zimmermann (2019). Subjects are asked to estimate the weight of a bucket. The hypothetical characters Ann and Bob have each examined the bucket and produced an estimate. They share their estimate with Charlie, who computes the average of their two guesses. The subject has access to Ann's estimate of 70 and Charlie's estimate of 40. A common incorrect answer is to compute the average of 40 and 70, i.e., 55. The correct answer is 40.

Balls-and-urns belief updating (BU). A widely used paradigm to study belief updating are so-called balls-and-urns experiments. In our setup, there are two bags. One contains 70 red and 30 blue chips, and one contains 30 red and 70 blue chips. One of them gets selected at random with 50-50 chance, and a balls gets drawn from the selected bag. Subjects are asked to indicate the percentage chance that the selected bag is the one that contains more red chips, given that the drawn chip is red. Subjects commonly exhibit a conservatism bias in this setup and state posteriors strictly between 50% and 70%. The Bayesian answer is 70%.

Gambler's fallacy (GF). Following Dohmen et al. (2009), subjects were asked to predict the next outcome in the following sequence of tosses of a fair coin, where the last three tosses came up Heads: T - T - T - H - T - H - H - H . In this type of task, subjects occasionally believe that Tails is "due." The correct answer is to select that "Both are equally likely."

Sample size neglect (SSN). Adapting the classic "hospital problem" (Kahneman and Tversky, 1972; Bar-Hillel, 1979), subjects were asked whether a factory that produces 45 chairs each day or a factory that produces 15 chairs each day has more days on which more than 20% of chairs are defective. A common incorrect response is that this happens equally often in both factories. The correct answer, however, is that this outcome is more likely in the smaller factory.

Regression to the mean / misattribution (RM). People exhibit a well-known tendency to attribute outcomes to internal factors rather than random noise, which leads them to neglect mean reversion. In a variant of classical work on the failure to appreciate regression (Kahneman and Tversky, 1973), subjects are asked to assess whether the true IQ of a test-taker is more likely to be above or below 140, given that their IQ test score is 140, the average score in the sample is 100, and test scores reflect a combination of true ability and random chance. Subjects commonly believe that the person's true IQ is equally likely to be above or below 140. The correct answer is that the person's true IQ is more likely to be below than above 140. Such failure to account for mean reversion is a special case of a more general class of biases that reflect misattribution.

Acquiring-a-company (AC). This game is one of the most widely studied tasks in experimental economics, both because it reflects a general class of errors in contingent reasoning and because its adverse selection logic has many applications in economics, such as in auctions. Following Charness and Levin (2009), we implement a version of the AC game in which subjects play against a computerized opponent. A seller has a company that is worth either 20 or 120 points to him. The company's value to the buyer is 1.5 times as much as the value to the seller. The subject makes a take-it-or-leave-it-offer, which the seller accepts if and only if the offer is at least as high as the value of the company to him. Subjects frequently bid strictly more than the theoretically optimal bid of 20, which neglects the adverse selection logic of the problem.

Wason selection task (WAS). The Wason task (Wason, 1968) is a widely used task in the social sciences because it captures failures in contingent reasoning and a tendency towards positive hypothesis testing in a simple way. In the problem, there is a deck of four cards that have numbers (odd or even) on one side and a color (brown or green) on the other. Subjects are tasked with turning over (only) those cards that can be useful in assessing whether the statement "All of the cards with an even number on one side are green on the other" is true. In this task, subjects frequently engage in positive testing by turning over the card with the even number and the one that is green. The logically correct choice is to pick the brown card in addition to the card with the even number.

Cognitive Reflection Test (CRT). The CRT is likewise a widely studied task in economics because it effectively captures the intuitive "System 1" responses that the early heuristics and biases program emphasized. Moreover, responses in the CRT have been shown to be correlated with various economic behaviors (Frederick, 2005). We implement the question "It takes 6 machines 6 days to produce 6 cars. How long would it take 12 machines to produce 12 cars?" A tempting, incorrect answer is 12 days. The correct answer is 6 days.

Iterated reasoning / Backward induction (IR). To capture the well-known and widely-studied tendency to iterate the best-response function only a small number of times ("level k reasoning"), we follow Bosch-Rosa and Meissner (2020) in implementing a one-player guessing game that only requires iteration of the best response function but is independent of beliefs about others. Subjects are asked to pick two numbers between 0 and 100, inclusive. Their task is to select numbers whose average is as close as possible to 2/3 of either number. While zero is the correct solution, most subjects state strictly positive values.

Equilibrium Reasoning / Predicting response to incentives (EQ). Many empirical regularities in behavioral economics can be understood as people failing to accurately predict others' behavior from their incentives. One elegant demonstration of this is the experiment in Dal Bó et al. (2018), which we simplify here. Subjects are presented with two similar-looking 2×2 payoff matrices. In Game A, all payoffs are higher than in Game B, but Game B has a cooperative equilibrium, while Game A has a prisoner's dilemma structure. Subjects are asked to predict in which game past participants made more money, on average. A majority of subjects incorrectly believes that people make more in Game A because they fail to anticipate differences in equilibrium play.

Knapsack / identifying constrained optima (KS). Knapsack problems are a simple to explain but canonical instance of constrained optimization, which lies at the heart of a large class of economic consumer and firm maximization problems (Mathews, 1896; Murawski and Bossaerts, 2016). In our implementation, subjects pick from a set of 12 items, each of which has a known value and weight. The objective is to maximize the sum of values chosen, while satisfying a budget constraint on the weights. Experiments typically show that subjects fail to identify the value-maximizing bundle (which, in our instance consists of 4 of the 12 items).

Portfolio choice (PC). Various studies have documented failures to construct efficient financial portfolios. One well-known example of this is failure is the use of the so-called 1/N heuristic (Benartzi and Thaler, 2001), according to which people split their investment uniformly across the available assets. To get at this, we ask subjects to choose between two portfolios that are constructed from four assets each. The portfolios are identical except that one allocates 1/4 of the budget to each asset in a way that makes this “1/N portfolio” strictly dominated by the other available portfolio.

Thinking at the margin (TM). One of the core lessons of economics is to think at the margin rather than the average, yet people have consistently been shown to think in terms of averages. We developed a simplified version of the taxation problem in Rees-Jones and Taubinsky (2020). Subjects are tasked with deciding which of two bank accounts to allocate 20 points to. Both bank accounts already contain 40 points. The trick is that the marginal tax rate for additional 20 points is lower in the bank account that has a higher average tax rate.

Exponential growth bias (EGB). An ability to compute compound interest is essential in numerous economics models and decision contexts, including exponential time discounting, savings and investment. Following previous studies (Levy and Tasoff, 2016), we ask subjects to guess how much a stock that is worth $100 today is worth after 20 years if its value increases 5% each year. People tend to give a response of $200 in this problem, which entirely misses the compounding effect. The correct answer is $265.

B Model Derivations

In the following, we derive our main predictions separately for the model of committee voting or parimutuel betting (Appendix B.1) and for the model of auctions (Appendix B.2).

B.1 Committees and Parimutuel Betting Markets

Note that expanding eq. (6) from Section 3, institutional gain $\mathbb{G}_{bet,com}$ may be expressed as:

$$\mathbb{G}_{bet,com} = \frac{\sum_i p_i \left(N k_i - \sum_j k_j\right)}{N^2 \bar{k}} \tag{10}$$

$$= \omega\beta \frac{\sum_i p_i (p_i - \bar{p})}{N\bar{k}} \tag{11}$$

$$= \frac{(N-1)\beta}{N\bar{c}} s_p^2 \tag{12}$$

where s_p^2 is the sample variance of p. Informally, we may consider $\beta/\bar{k} = \partial \log \bar{k}/\partial \bar{p}$, as the percent increase in the average bid given an increase in the average confidence among the players.

Proof. **(Prediction 1, Part (i))** Stated formally, this part of the prediction states that if $s_p^2 > 0$ and $\beta > 0$, we have $\mathbb{G}_{bet,com} > 0$. Since $\bar{c} > 0$ in our case[21], eq. (12) immediately provides the result. □

Proof. **(Prediction 1, Part (ii))** This part of the prediction says that for $\alpha > 0$ and $s_p^2 > 0$, we have $\frac{\partial \mathbb{G}_{bet,com}}{\partial \beta} > 0$. Using eq. (12) we may compute

$$\frac{\partial}{\partial \beta}\mathbb{G}_{bet,com} = \frac{\partial}{\partial \beta}\frac{(N-1)\beta}{N(\alpha+\beta\bar{p})} s_p^2 \tag{13}$$

$$= \frac{(N-1)\alpha}{N(\alpha+\beta\bar{p})^2} s_p^2 > 0 \tag{14}$$

Given our assumptions the result follows. □

Proof. **(Prediction 2)** This prediction states that the effect of an increase in average overconfidence, d, on institutional improvement, $\mathbb{G}$, is ambiguous. Formally, for the case of betting and committee voting, we predict that for $\beta > 0$ and $s_p^2 > 0$, a change in overconfidence $d \to d + \delta d$ need not imply $\Delta\mathbb{G}_{bet,com} < 0$.

[21]Aside from the empirics, $c_i > 0$ when $\alpha, \beta > 0$.

Firstly, we may compute that: $\frac{\partial \mathbb{G}_{bet,com}}{\partial \alpha} < 0$. Using eq. (12) we may compute

$$\frac{\partial}{\partial \alpha}\mathbb{G}_{bet,com} = \frac{\partial}{\partial \alpha}\frac{(N-1)\beta}{N(\alpha+\beta\bar{p})}s_p^2 \tag{15}$$

$$= -\frac{(N-1)\beta}{N(\alpha+\beta\bar{p})^2}s_p^2 < 0 \tag{16}$$

Secondly, we recall our earlier result that:

$$\frac{\partial}{\partial \beta}\mathbb{G}_{bet,com} > 0 \tag{17}$$

Noting that:

$$\frac{\partial d}{\partial \alpha}, \frac{\partial d}{\partial \beta} > 0 \tag{18}$$

and our assumptions the result follows. □

B.2 Auctions

In the case of a multi-unit auction, eq. (7) showed that:

$$\mathbb{G}_{auc} = \frac{1}{|W|}\sum_{j\in W} p_j - \frac{1}{N}\sum_i p_i. \tag{19}$$

Proof. **(Prediction 1, Part (i))** This part states that when $\beta > 0$, we have $\mathbb{G}_{auc} \geq 0$. Note that when $\beta > 0$, $\frac{1}{|W|}\sum_{i\in W} p_i \geq \bar{p}$. The result then follows. □

Proof. **(Prediction 1, Part (ii))** When $\beta > 0$, we have $\frac{\partial \mathbb{G}_{auc}}{\partial \beta} = 0$. This holds because given a fixed set of probabilities, $\{p_i\}$, the ordering of the bids is invariant under a change, $\beta \to \beta + \delta\beta$ so long as $\delta\beta > -\beta$. □

Proof. **(Prediction 2)** This prediction states that an increase in overconfidence, d, has no effect on institutional improvement $\mathbb{G}$ when $\beta > 0$. In the case of auctions, we have $\frac{\partial \mathbb{G}_{auc}}{\partial \alpha} = 0$. This follows since given a fixed set of probabilities, $\{p_i\}$, the ordering of the bids is invariant under a change, $\alpha \to \alpha + \delta\alpha$. Furthermore, the proof of Pred. 1 part (ii) indicates that a change in β doesn't change $\mathbb{G}$ so long as β remains positive. Accordingly, for any change in average overconfidence, $\delta d = \delta\alpha + \delta\beta\bar{p}$ in which $\beta > 0$, we see that there is no institutional improvement. □

C Additional Figures

C.1 Screenshots of Elicitation Screens

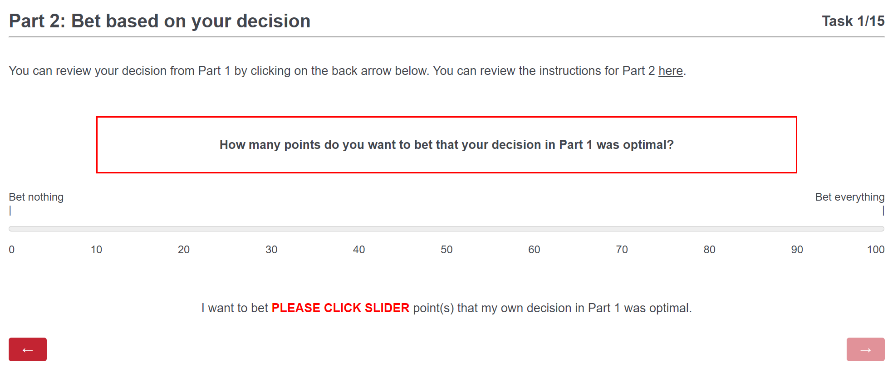


Figure 7: Screenshot of elicitation screen for the institutional decision in treatment *Betting*.

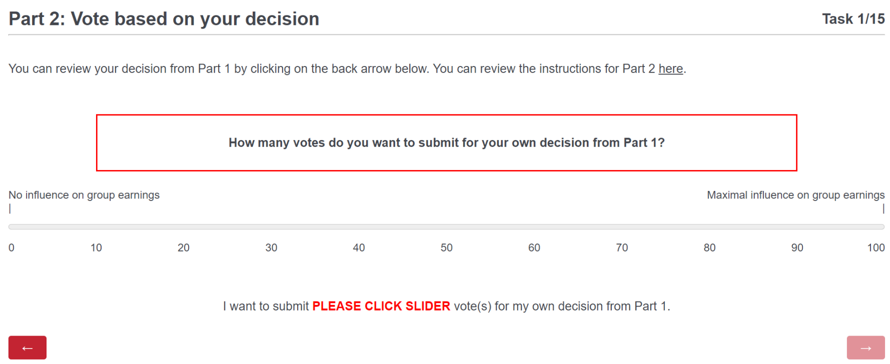


Figure 8: Screenshot of elicitation screen for the institutional decision in treatment *Committee*.

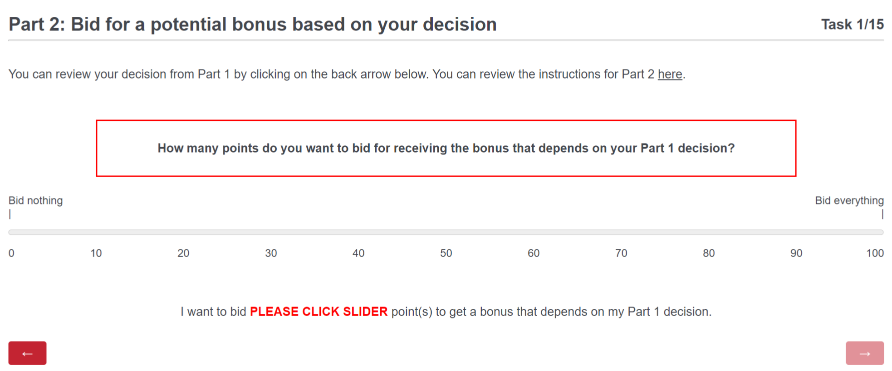


Figure 9: Screenshot of elicitation screen for the institutional decision in treatment *Auction*.

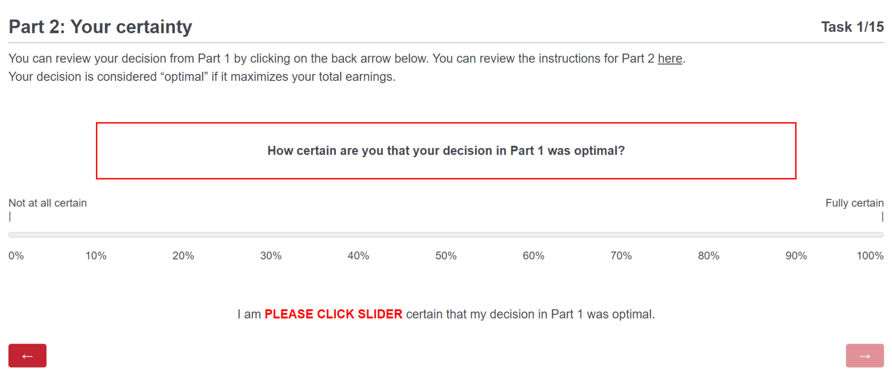


Figure 10: Screenshot of elicitation screen for confidence in the Knapsack task (treatment *Confidence*).

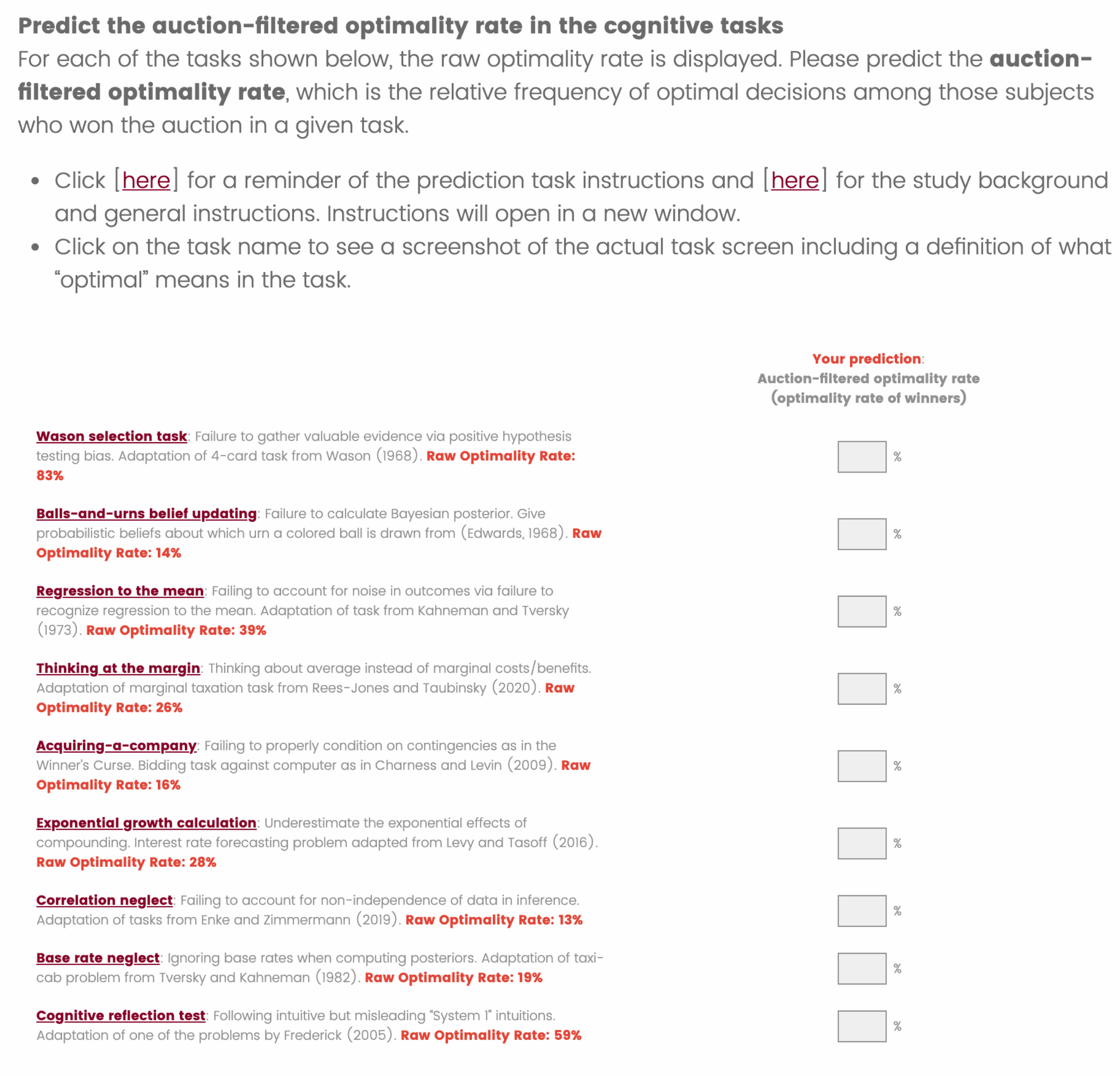
Predict average confidence in the cognitive tasks

For each of the tasks shown below, please predict the average confidence separately for subjects who took the optimal decision and those who did not. Confidence was elicited by asking subjects to provide their assessment of how likely it is that their decision was optimal. 0% means "Not at all certain" and 100% means "Fully certain".

- Click [here] for a reminder of the prediction task instructions and [here] for the study background and general instructions. Instructions will open in a new window.
- Click on the task name to see a screenshot of the actual task screen, including a definition of what "optimal" means in the respective task.

	Your prediction: Average confidence stated for answers that were **optimal**	**Your prediction**: Average confidence stated for answers that were **not optimal**
Wason selection task: Failure to gather valuable evidence via positive hypothesis testing bias. Adaptation of 4-card task from Wason (1968).	%	%
Balls-and-urns belief updating: Failure to calculate Bayesian posterior. Give probabilistic beliefs about which urn a colored ball is drawn from (Edwards, 1968).	%	%
Regression to the mean: Failing to account for noise in outcomes via failure to recognize regression to the mean. Adaptation of task from Kahneman and Tversky (1973).	%	%
Thinking at the margin: Thinking about average instead of marginal costs/benefits. Adaptation of marginal taxation task from Rees-Jones and Taubinsky (2020).	%	%
Acquiring-a-company: Failing to properly condition on contingencies as in the Winner's Curse. Bidding task against computer as in Charness and Levin (2009).	%	%
Exponential growth calculation: Underestimate the exponential effects of compounding.Interest rate forecasting problem adapted from Levy and Tasoff (2016).	%	%
Correlation neglect: Failing to account for non-independence of data in inference. Adaptation of tasks from Enke and Zimmermann (2019)	%	%
Base rate neglect: Ignoring base rates when computing posteriors. Adaptation of taxi-cab problem from Tversky and Kahneman (1982).	%	%
Cognitive reflection test: Following intuitive but misleading "System 1" intuitions. Adaptation of one of the problems by Frederick (2005).	%	%

Figure 11: Screenshot of elicitation screen in the expert survey for forecasts of confidence.



Predict the auction-filtered optimality rate in the cognitive tasks

For each of the tasks shown below, the raw optimality rate is displayed. Please predict the **auction-filtered optimality rate**, which is the relative frequency of optimal decisions among those subjects who won the auction in a given task.

- Click [here] for a reminder of the prediction task instructions and [here] for the study background and general instructions. Instructions will open in a new window.
- Click on the task name to see a screenshot of the actual task screen including a definition of what "optimal" means in the task.

	Your prediction: **Auction-filtered optimality rate (optimality rate of winners)**
Wason selection task: Failure to gather valuable evidence via positive hypothesis testing bias. Adaptation of 4-card task from Wason (1968). **Raw Optimality Rate: 83%**	%
Balls-and-urns belief updating: Failure to calculate Bayesian posterior. Give probabilistic beliefs about which urn a colored ball is drawn from (Edwards, 1968). **Raw Optimality Rate: 14%**	%
Regression to the mean: Failing to account for noise in outcomes via failure to recognize regression to the mean. Adaptation of task from Kahneman and Tversky (1973). **Raw Optimality Rate: 39%**	%
Thinking at the margin: Thinking about average instead of marginal costs/benefits. Adaptation of marginal taxation task from Rees-Jones and Taubinsky (2020). **Raw Optimality Rate: 26%**	%
Acquiring-a-company: Failing to properly condition on contingencies as in the Winner's Curse. Bidding task against computer as in Charness and Levin (2009). **Raw Optimality Rate: 16%**	%
Exponential growth calculation: Underestimate the exponential effects of compounding. Interest rate forecasting problem adapted from Levy and Tasoff (2016). **Raw Optimality Rate: 28%**	%
Correlation neglect: Failing to account for non-independence of data in inference. Adaptation of tasks from Enke and Zimmermann (2019). **Raw Optimality Rate: 13%**	%
Base rate neglect: Ignoring base rates when computing posteriors. Adaptation of taxi-cab problem from Tversky and Kahneman (1982). **Raw Optimality Rate: 19%**	%
Cognitive reflection test: Following intuitive but misleading "System 1" intuitions. Adaptation of one of the problems by Frederick (2005). **Raw Optimality Rate: 59%**	%

Figure 12: Screenshot of elicitation screen in the expert survey for forecasts of the change in optimality rate through the *Auction* institution.

C.2 Additional Figures for Between-Subjects Treatments

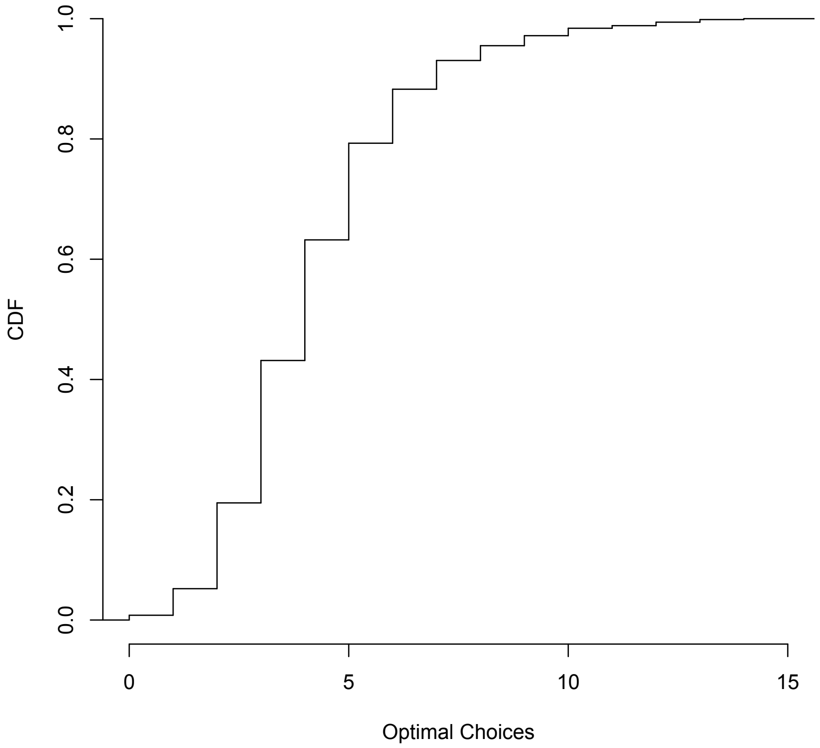


Figure 13: Empirical cumulative distribution of total number of optimal responses to the cognitive tasks (Part 1 decisions), across subjects in treatments *Betting*, *Auction*, *Committee* and *Confidence*. The figure pools data from all 15 tasks across all four between-subject treatments.

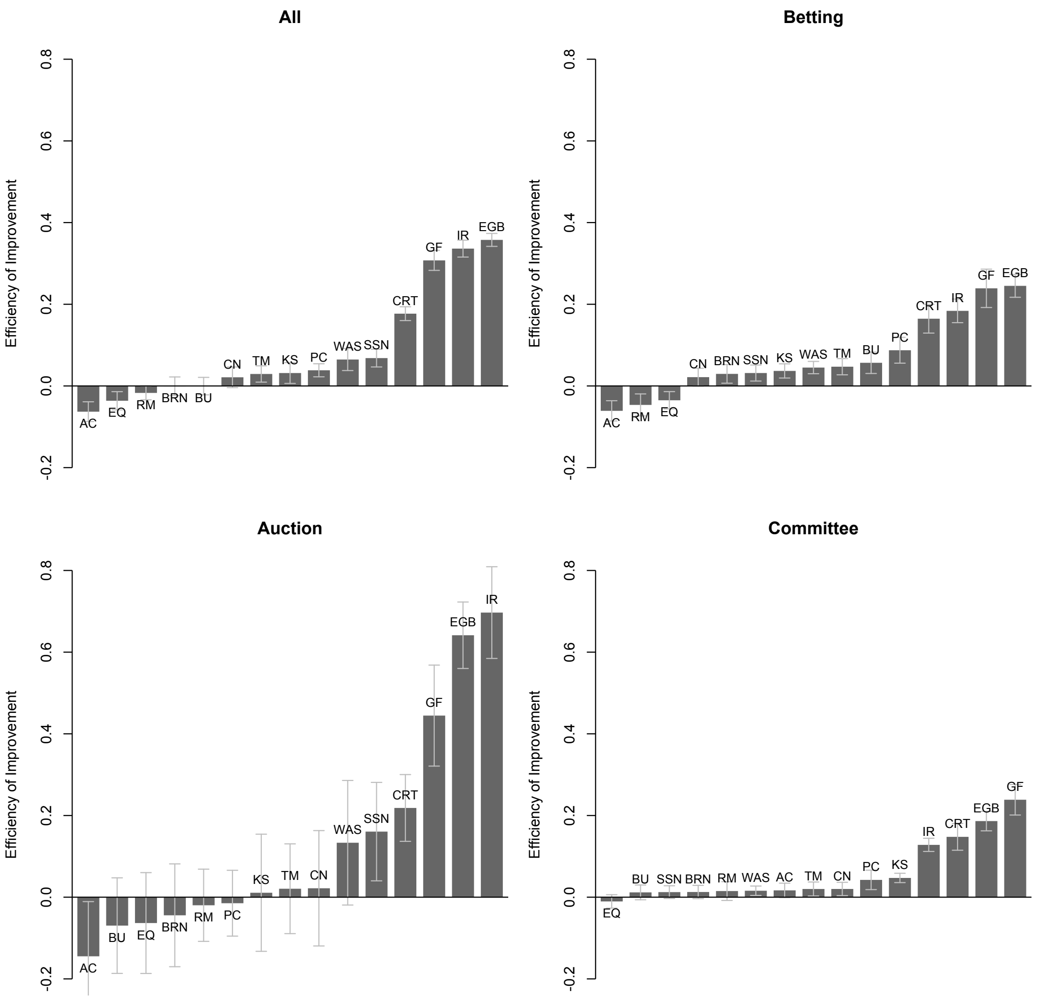


Figure 14: Performance improvement (efficiency measure) through institutions across tasks. Efficiency is computed as the aggregate performance rate after institutional filtering minus the the fraction of optimal Part 1 responses, all divided by the difference between the maximum possible improvement (given responses to the cognitive tasks and the structure of the institution) and the fraction of optimal responses to the cognitive tasks (Part 1 decisions). The aggregate performance rate is based on 10,000 randomly constructed 10-subject cohorts for each institution, taking the mean over all samples. Each participant completed all 15 tasks in random order. Based on $N = 323$ participants in the Auction condition, $N = 387$ in Betting and $N = 337$ in Committee. One-standard error bars are conservatively calculated as the ratio of the standard deviations of efficiencies over these random cohorts divided by the square root of the number of cohorts available in the dataset (e.g., 387/10=38.7 in Betting). Task codes: AC=Acquiring-a-company; BRN=Base rate neglect; BU=Balls-and-urns belief updating; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calc.; EQ=Predict others' play; GF: Predict sequence of draws; IR=Backwards induction; KS=Knapsack; PC=Portfolio choice; RM=Attribution; SSN=Account for sample size; TM=Marginal thinking; WAS=Wason task.

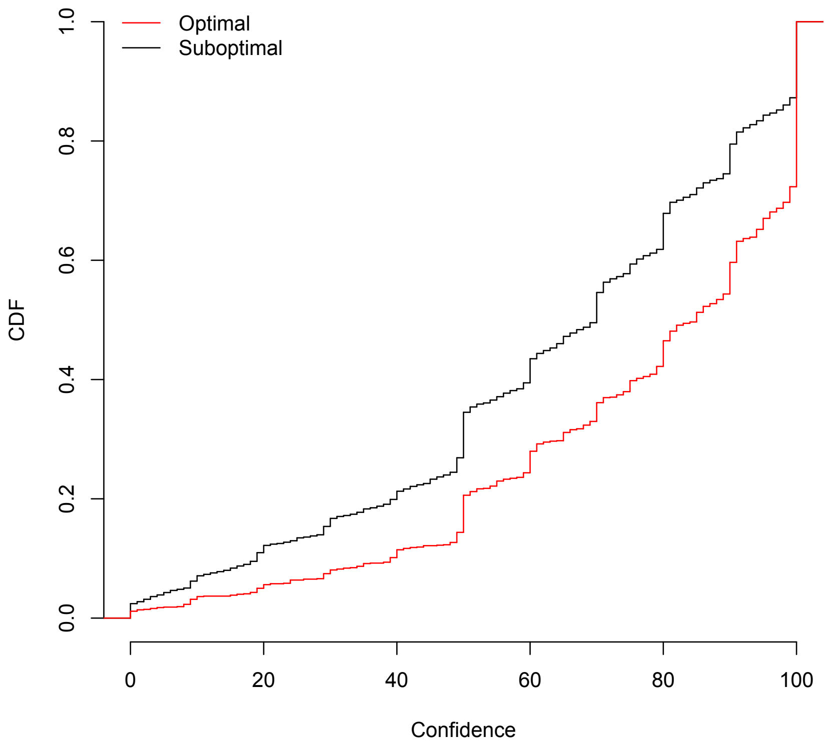


Figure 15: Stated confidence (Part 2 decision) by optimality of the response to a cognitive task (Part 1 decision). Based on $N = 5,010$ Part 2 decisions in the *Confidence* condition, pooled across 15 different cognitive tasks. The sample of Part 2 decisions is split by whether the corresponding Part 1 decision was optimal and empirical distribution functions are displayed.

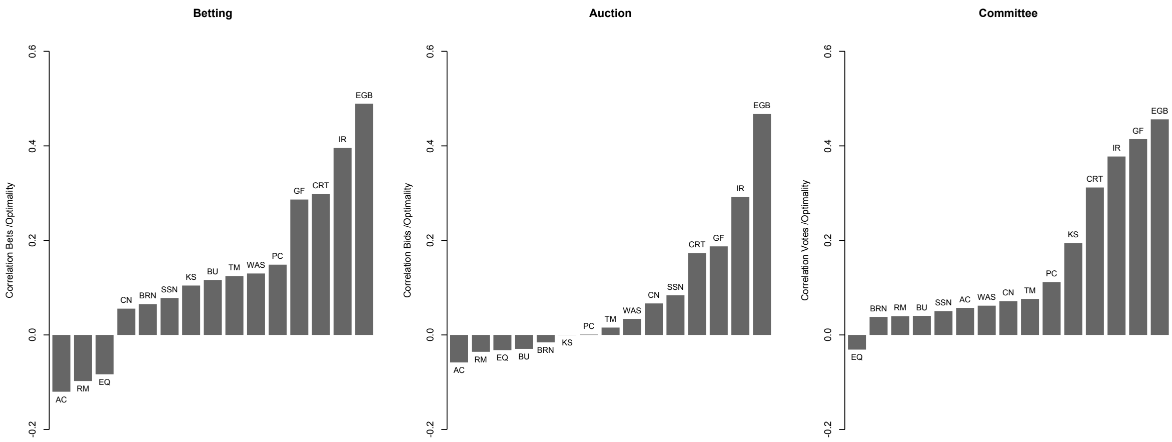


Figure 16: Correlation between optimality of responses to a cognitive task (Part 1 decisions) and institutional choices (Part 2 decisions) by task. Based on $N = 323$ participants in the Auction condition, $N = 387$ in Betting and $N = 337$ in Committee. Task codes: AC=Acquiring-a-company; BRN=Base rate neglect; BU=Balls-and-urns belief updating; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calc.; EQ=Predict others' play; GF: Predict sequence of draws; IR=Backwards induction; KS=Knapsack; PC=Portfolio choice; RM=Attribution; SSN=Account for sample size; TM=Marginal thinking; WAS=Wason task.

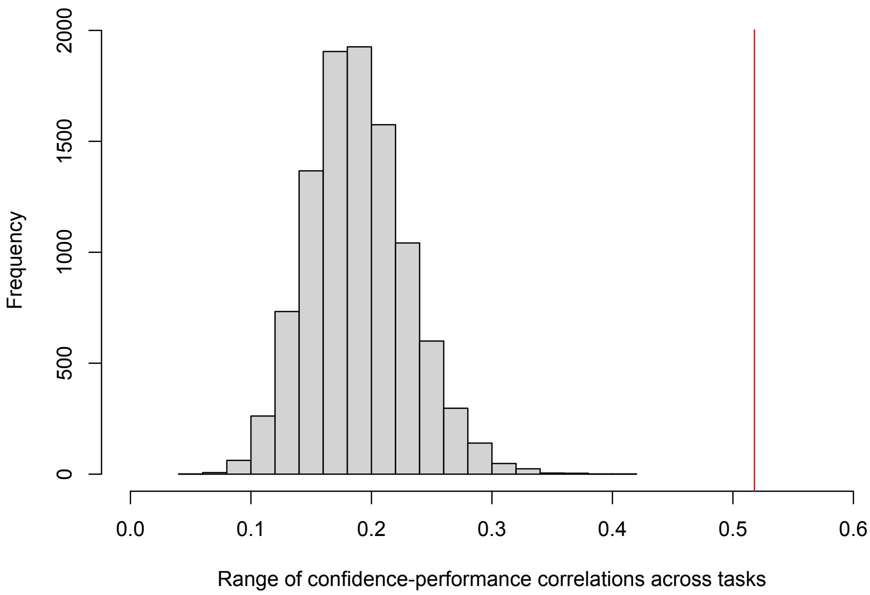


Figure 17: Comparison of empirically observed range of confidence-optimality correlations across tasks with numerical simulations. Red line corresponds to empirical data. The simulations are done as follows. In treatment *Confidence*, we take as given the empirical marginal distributions of both confidence and optimality in each task. We then randomly scramble these two variables such that, in expectation, in each task, the confidence-optimality correlation is identical and equal to zero. In any given simulation the actual confidence-optimality correlation in any given task will usually not be zero (due to the finite sample of 334 subjects). We implement this procedure 10,000 times and, for each iteration, save the range (max minus min) of the within-task confidence-optmality correlation. The figure plots the resulting distribution of simulated ranges along with the empirically observed range.

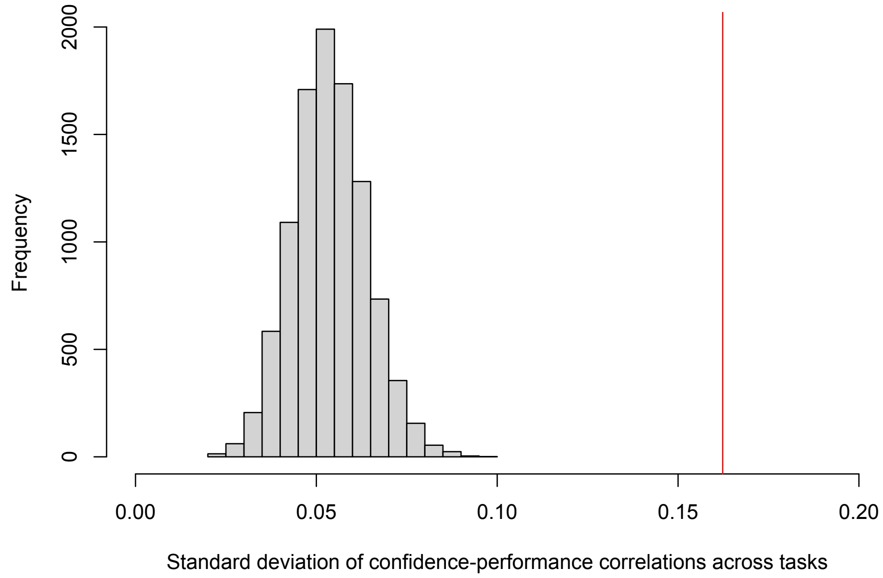


Figure 18: Comparison of empirically observed standard deviation of confidence-optimality correlations across tasks with numerical simulations. Red line corresponds to empirical data. The simulations are done as follows. In treatment *Confidence*, we take as given the empirical marginal distributions of both confidence and optimality in each task. We then randomly scramble these two variables such that, in expectation, in each task, the confidence-optimality correlation is identical and equal to zero. In any given simulation the actual confidence-optimality correlation in any given task will usually not be zero (due to the finite sample of 334 subjects). We implement this procedure 10,000 times and, for each iteration, save the standard deviation of the within-task confidence-optmality correlation. The figure plots the resulting distribution of simulated standard deviations along with the empirically observed standard deviation.

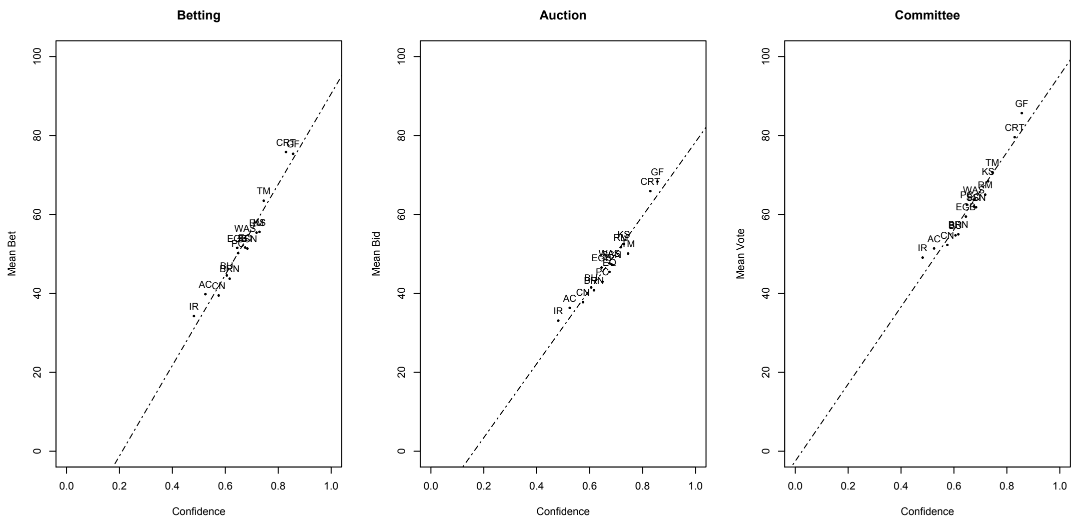


Figure 19: Average confidence and average institutional behavior. For each task, we compute average confidence stated in treatment *Confidence* and plot this against the average Part 2 decision taken in each of the three between-subjects institutional treatments.

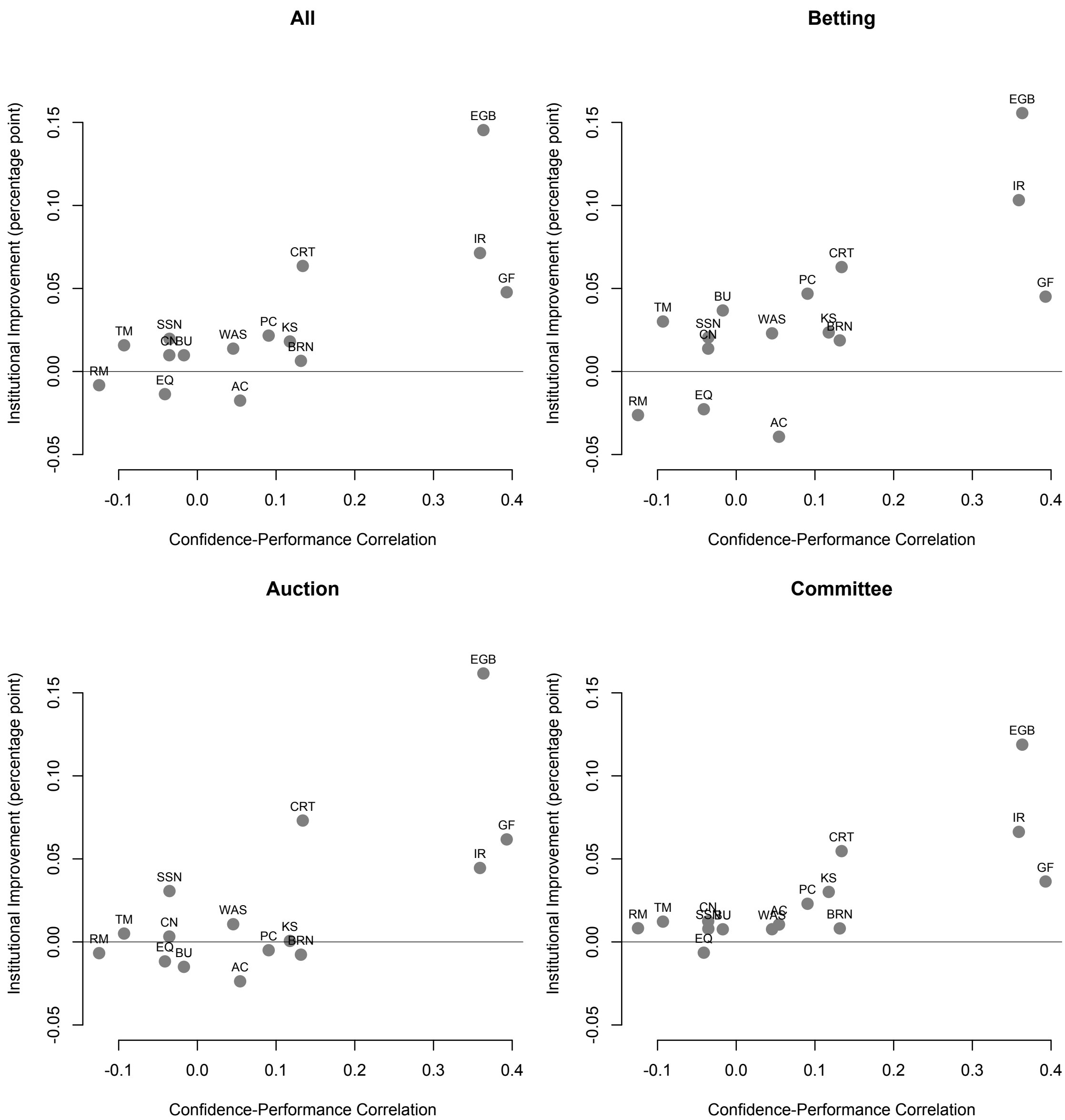


Figure 20: Confidence-optimality correlation and institutional improvement for the separate institutions in the between-subjects treatments. The horizontal axis shows the within-task correlation between confidence and optimality in a given task in treatment *Confidence*. The vertical axis shows the performance improvement that is implied by an institution for the respective cognitive task in treatments *Betting*, *Auction* and *Committee*. Task codes: AC=Acquiring-a-company; BRN=Base rate neglect; BU=Balls-and-urns belief updating; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calc.; EQ=Predict others' play; GF: Predict sequence of draws; IR=Backwards induction; KS=Knapsack; PC=Portfolio choice; RM=Attribution; SSN=Account for sample size; TM=Marginal thinking; WAS=Wason task.

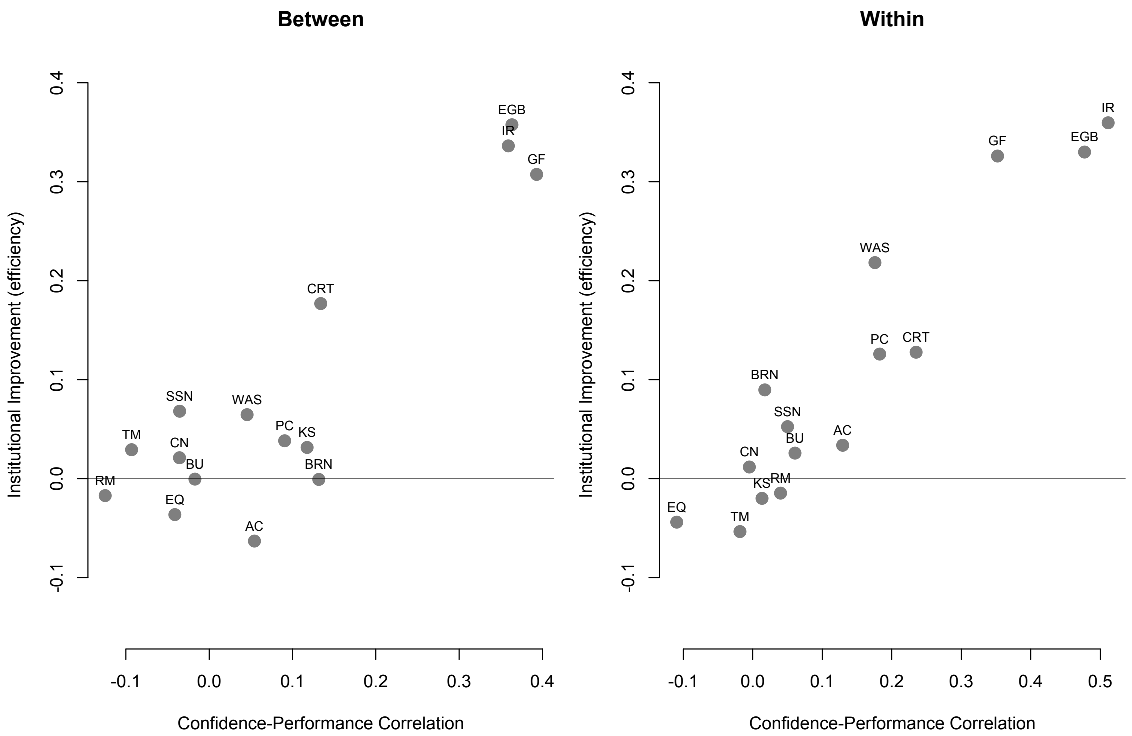


Figure 21: Confidence-optimality correlation and institutional improvement using the efficiency measure. The left panel shows the results for the between-subjects treatments and the right panel those for the within-subjects treatments. In the left panel, the horizontal axis shows the within-task correlation between confidence and optimality in a given task in treatment *Confidence*. The vertical axis shows the efficiency of an institution for the respective cognitive task. Efficiency is computed as the aggregate performance rate after institutional filtering minus the the fraction of optimal responses in a cognitive task, all divided by the difference between the maximum possible improvement (given Part 1 responses and the structure of the institution) and the fraction of optimal Part 1 responses. The data are from treatments *Betting*, *Auction* and *Committee*. In the right panel, we show analogous quantities, except that they are all derived from treatments *Betting Within*, *Auction Within* and *Committee Within*. Task codes: AC=Acquiring-a-company; BRN=Base rate neglect; BU=Balls-and-urns belief updating; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calc.; EQ=Predict others' play; GF: Predict sequence of draws; IR=Backwards induction; KS=Knapsack; PC=Portfolio choice; RM=Attribution; SSN=Account for sample size; TM=Marginal thinking; WAS=Wason task.

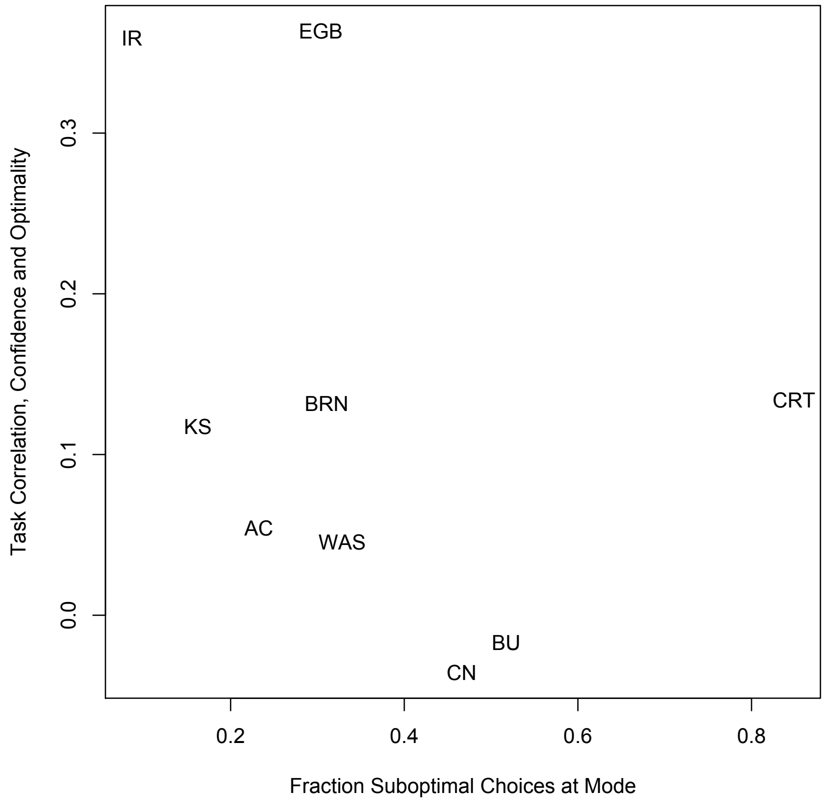


Figure 22: The peakedness of the distribution of suboptimal answers and the confidence-performance correlation in the between-subjects treatments. The horizontal axis displays the fraction of subjects playing the modal suboptimal answer in a cognitive task (conditional on being wrong). The vertical axis displays the confidence-optimality correlation in treatment *Confidence*. Task codes: AC=Acquiring-a-company; BRN=Base rate neglect; BU=Balls-and-urns belief updating; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calc.; IR=Backwards induction; KS=Knapsack; WAS=Wason task.

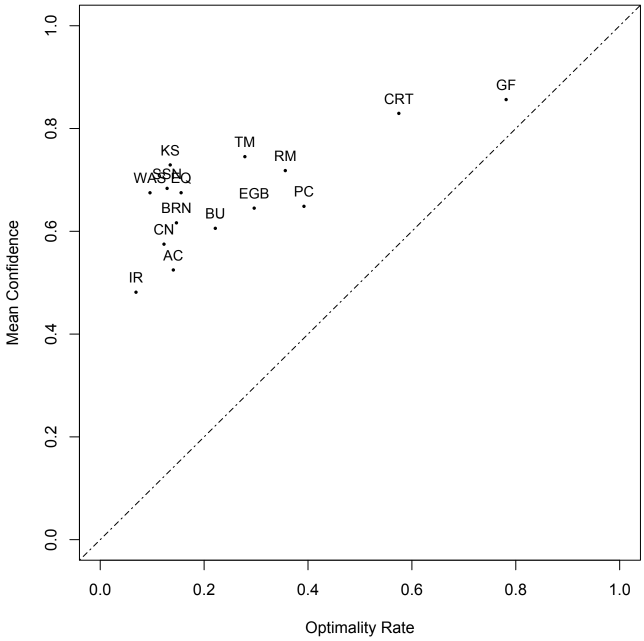


Figure 23: Fraction of optimal responses in a cognitive task (Part 1 decision) and average confidence in treatment *Confidence*, separately for each task. Based on $N = 334$ respondents. Task codes: AC=Acquiring-a-company; BRN=Base rate neglect; BU=Balls-and-urns belief updating; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calc.; EQ=Predict others' play; GF: Predict sequence of draws; IR=Backwards induction; KS=Knapsack; PC=Portfolio choice; RM=Attribution; SSN=Account for sample size; TM=Minimize taxation; WAS=Wason task.

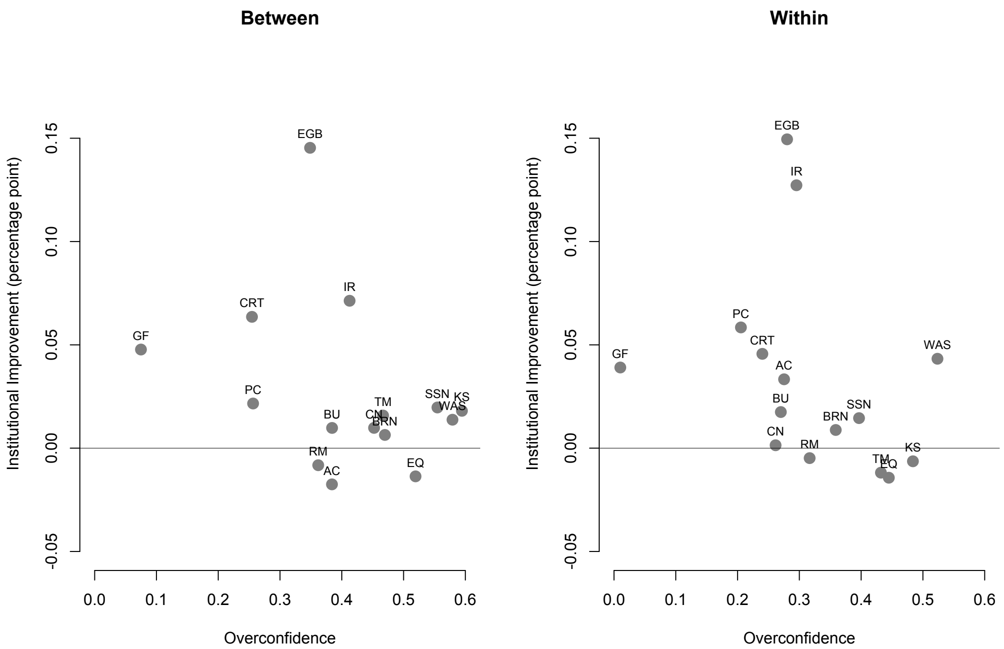


Figure 24: Overconfidence and institutional improvement. The left panel shows the results for the between-subjects treatments and the right panel those for the within-subjects treatments. In the left panel, the horizontal axis shows the average level of overconfidence in treatment *Confidence*. The vertical axis shows the average institutional improvement across treatments *Betting*, *Auction* and *Committee*. In the right panel, we show analogous quantities, except that they are all derived from treatments *Betting Within*, *Auction Within* and *Committee Within*. Percentage point improvement is computed as the aggregate performance of the institutional summary statistic minus the raw fraction of optimal responses in a cognitive task, averaged across institutions. The institutional summary statistics are given by the parimutuel market price in *Betting*, the average rate of bias among the set of winners in the *Auction* and the vote share for the optimal decision in *Committee*. The aggregate performance is based on 10,000 randomly constructed ten-subject cohorts for each institution, taking the mean over all samples. Overconfidence is computed as average confidence minus the optimality rate in a task. Task codes: AC=Acquiring-a-company; BRN=Base rate neglect; BU=Balls-and-urns belief updating; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calc.; EQ=Predict others' play; GF: Predict sequence of draws; IR=Backwards induction; KS=Knapsack; PC=Portfolio choice; RM=Attribution; SSN=Account for sample size; TM=Marginal thinking; WAS=Wason task.

C.3 Additional Figures for Within-Subjects Treatments

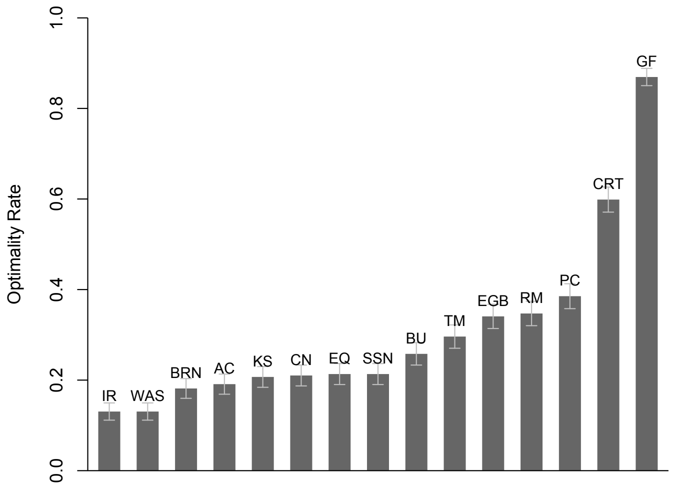


Figure 25: Fraction of optimal responses in a cognitive task (Part 1 decisions) across tasks in treatments *Betting Within*, *Auction Within* and *Committee Within*. The tasks and optimal responses are described in Appendices A and F. Task codes: AC=Acquiring-a-company; BRN=Base rate neglect; BU=Balls-and-urns belief updating; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calc.; EQ=Predict others' play; GF: Predict sequence of draws; IR=Backwards induction; KS=Knapsack; PC=Portfolio choice; RM=Attribution; SSN=Account for sample size; TM=Thinking at the margin; WAS=Wason task. Error bars are the standard error of the binomial mean.

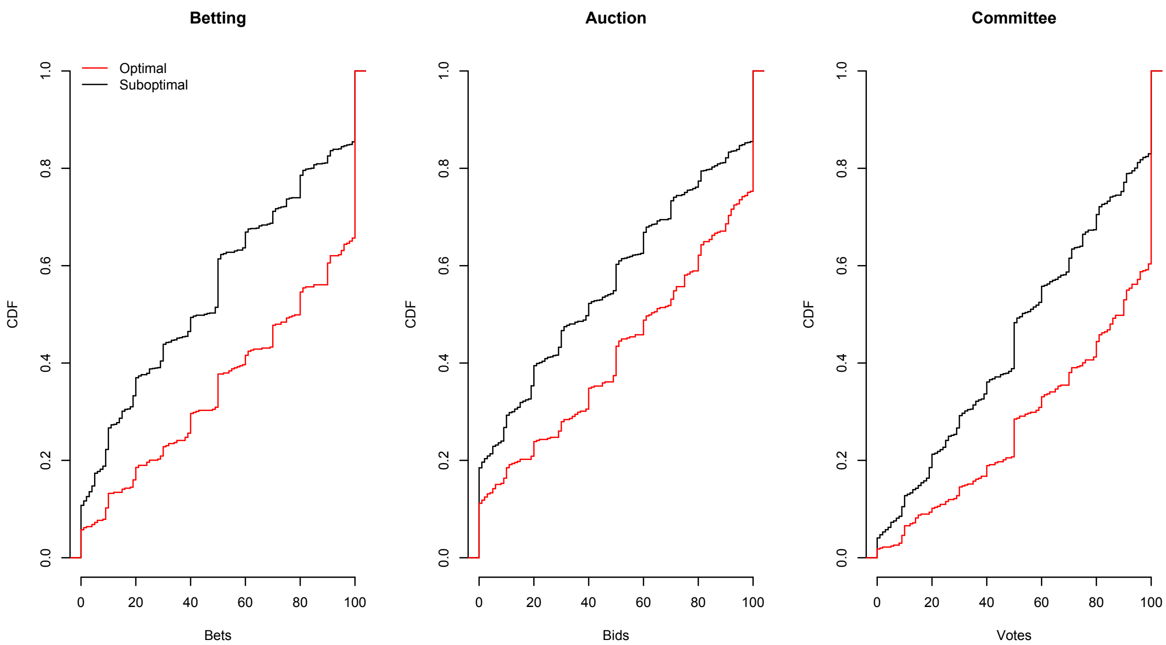


Figure 26: Institutional choices (Part 2 decisions) by optimality of the response to the cognitive task (Part 1 decision). Based on $N = 1575$ Part 2 decisions in the *Auction Within* condition, $N = 1575$ in *Betting Within* and $N = 1560$ in *Committee Within*, pooled across 15 different cognitive tasks. For each institution, the sample of Part 2 decisions is split by whether the corresponding Part 1 decision was optimal and empirical distribution functions are displayed.

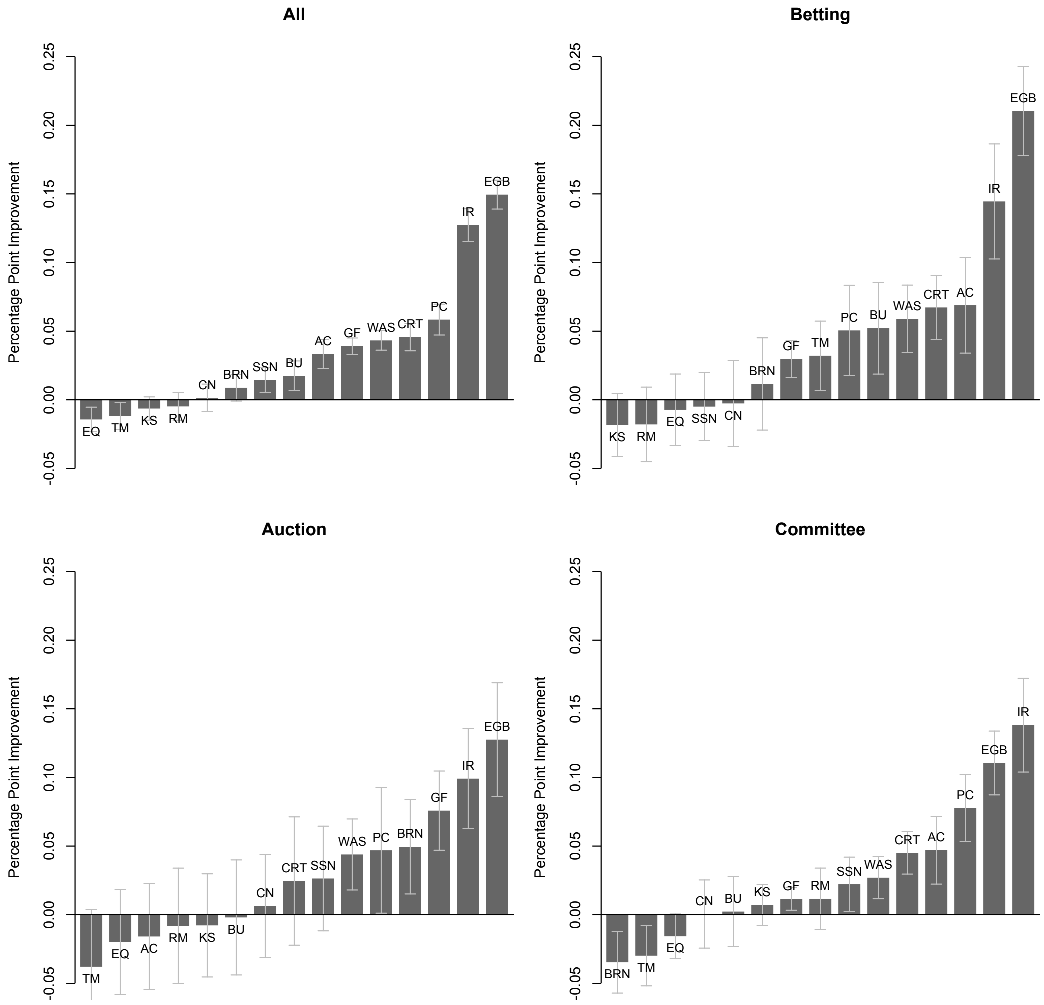


Figure 27: Performance improvement through institutions across tasks in treatments *Betting Within*, *Auction Within* and *Committee Within*. Percentage point improvement is computed as the aggregate performance rate after institutional filtering minus the fraction of optimal responses in a cognitive task. The aggregate performance rate is based on 10,000 randomly constructed 10-subject cohorts for each institution, taking the mean over all samples. Each participant completed all 15 tasks in random order. Based on $N = 105$ participants in the Auction condition, $N = 105$ in Betting and $N = 104$ in Committee. One-standard error bars are conservatively calculated as the ratio of the standard deviations of improvements over these random cohorts divided by the square root of the number of cohorts available in the dataset (e.g., 105/10=10.5 in Betting). Task codes: AC=Acquiring-a-company; BRN=Base rate neglect; BU=Balls-and-urns belief updating; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calc.; EQ=Predict others' play; GF: Predict sequence of draws; IR=Backwards induction; KS=Knapsack; PC=Portfolio choice; RM=Attribution; SSN=Account for sample size; TM=Marginal thinking; WAS=Wason task.

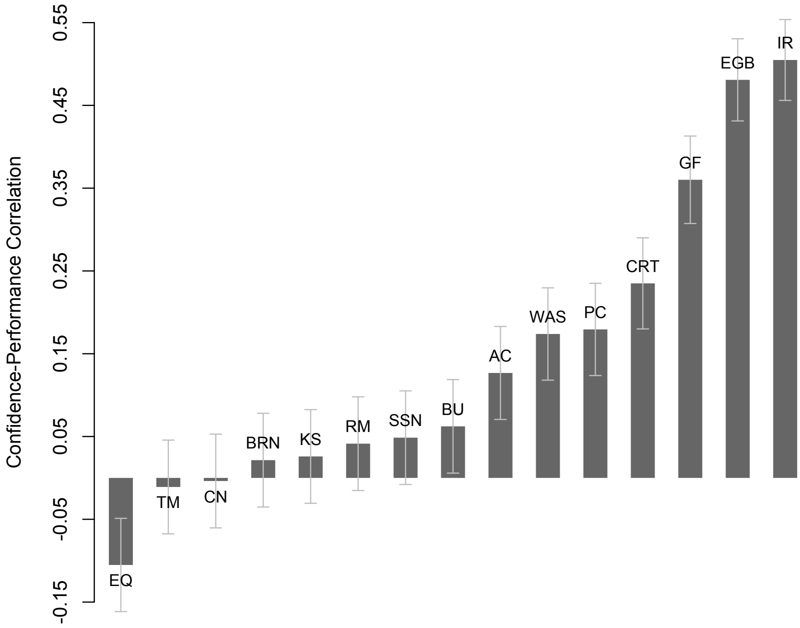


Figure 28: Within-task correlation between confidence and Part 1 optimality across tasks in treatments *Betting Within*, *Auction Within* and *Committee Within*. Displayed are Pearson correlation coefficients, based on $N = 313$ participants. Task codes: AC=Acquiring-a-company; BRN=Base rate neglect; BU=Balls-and-urns belief updating; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calc.; EQ=Predict others' play; GF: Predict sequence of draws; IR=Backwards induction; KS=Knapsack; PC=Portfolio choice; RM=Attribution; SSN=Account for sample size; TM=Marginal thinking; WAS=Wason task.

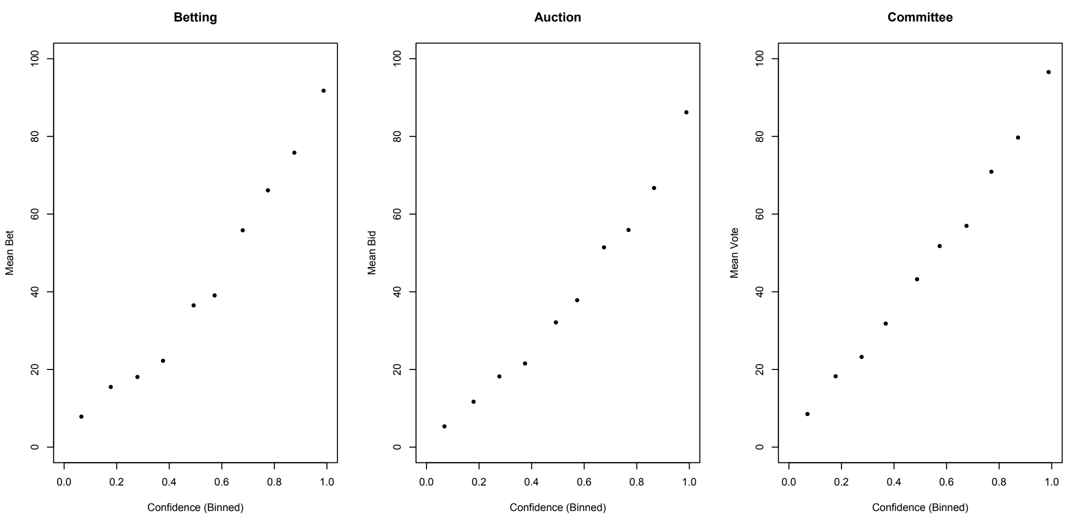


Figure 29: Binned scatterplots of institutional decisions against stated confidence in the within-subjects treatments, separately for each institution. Task codes: AC=Acquiring-a-company; BRN=Base rate neglect; BU=Balls-and-urns belief updating; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calc.; EQ=Predict others' play; GF: Predict sequence of draws; IR=Backwards induction; KS=Knapsack; PC=Portfolio choice; RM=Attribution; SSN=Account for sample size; TM=Marginal thinking; WAS=Wason task.

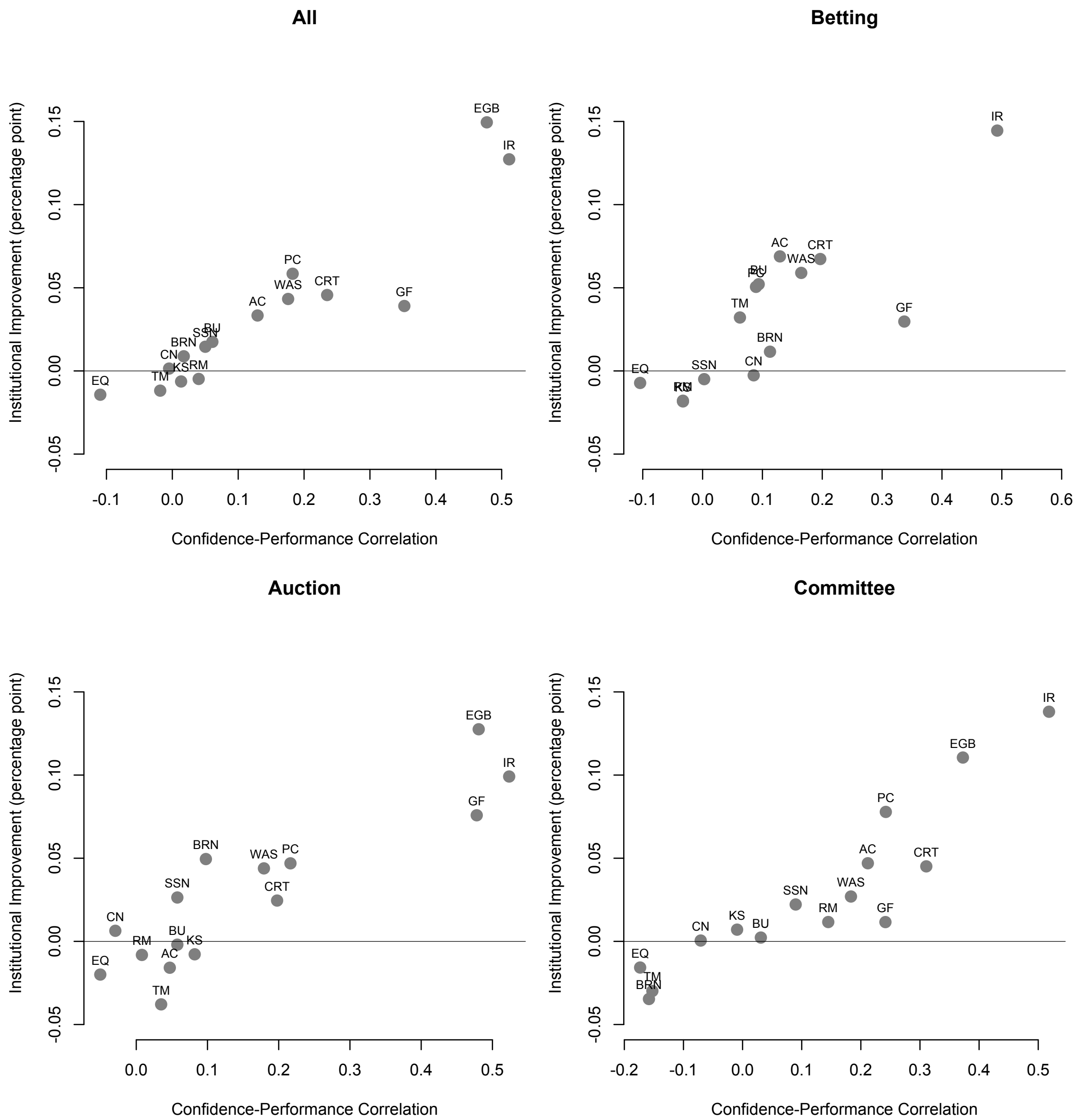


Figure 30: Confidence-optimality correlation and institutional improvement in the within-subjects treatments. The horizontal axis shows the within-task correlation between confidence and optimality in a given task. The vertical axis shows the performance improvement that is implied by an institution for the respective cognitive task. The data are from treatments *Betting Within*, *Auction Within* and *Committee Within*. Task codes: AC=Acquiring-a-company; BRN=Base rate neglect; BU=Balls-and-urns belief updating; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calc.; EQ=Predict others' play; GF: Predict sequence of draws; IR=Backwards induction; KS=Knapsack; PC=Portfolio choice; RM=Attribution; SSN=Account for sample size; TM=Marginal thinking; WAS=Wason task.

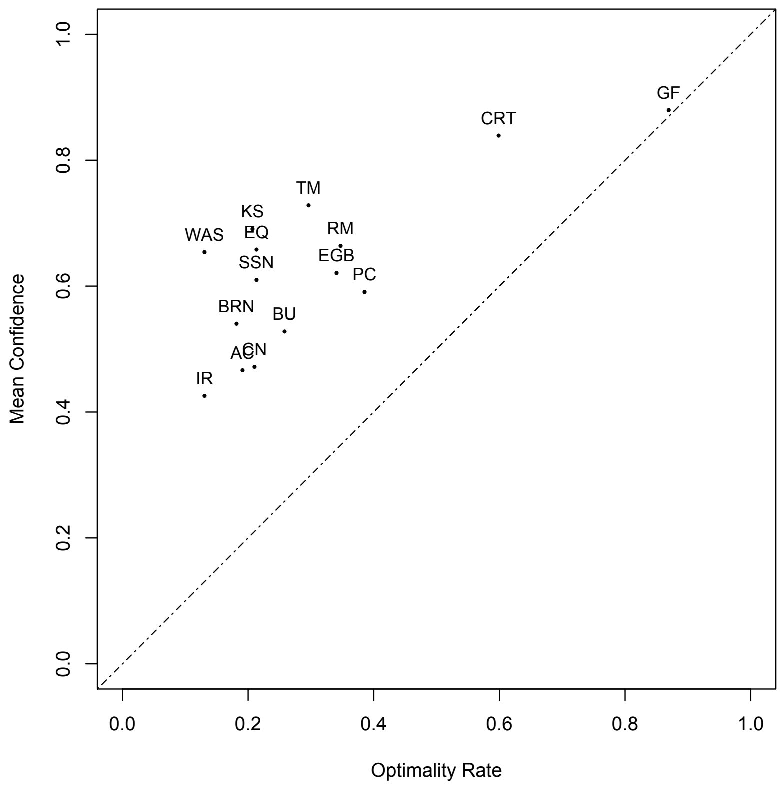


Figure 31: Fraction of optimal Part 1 responses and average confidence in treatments *Betting Within*, *Auction Within* and *Committee Within*, separately for each task. Task codes: AC=Acquiring-a-company; BRN=Base rate neglect; BU=Balls-and-urns belief updating; CN=Correlation neglect; CRT=Cognitive reflection test; EGB=Exponential growth calc.; EQ=Predict others' play; GF: Predict sequence of draws; IR=Backwards induction; KS=Knapsack; PC=Portfolio choice; RM=Attribution; SSN=Account for sample size; TM=Minimize taxation; WAS=Wason task.

D Additional Tables

Table 3: Overview of empirical studies eliciting relative calibration

Study	Tasks in main study	Confidence measurement	Measure of aggregation	Main finding	Sample	Sample size (main study)	Incentivizes for accuracy
Silver et al. (2021)	Numerical estimates of country populations and historical dates	7-point Likert scale	Revised estimates after discussion in groups of 3-6	Quality of revised decisions higher among groups with better initial confidence calibration	Laboratory	676	Yes
Bennett et al. (2018)	Binary choice general-knowledge questions on topics such as world facts and world history	7-point Likert scale judgment of difficulty	In one condition, respondents could self-select which questions to answer	Crowds consisting of participants who could self-selected questions outperformed crowds consisting of participants who were randomly assigned to questions	Online	144	No
Minson et al. (2018)	Numerical estimates of CEO salaries, weights of animals and U.S. demographic statistics	5-point Likert scale	Joint estimates after discussion with one other subject	Individual confidence unrelated to performance. Discussion outperforms average individual decisions in questions where subjects tend to make very large errors.	Laboratory	180	No
Cain et al. (2015)	Mulitple choice trivia quiz questions	Stated belief distribution of obtaining different cumulative scores	Selection into competition for high prize based on hard quiz versus for low prize based on easy quiz.	Selection into competition strongly predicted by overplacement relative to others. Overplacement more likely for easy quiz.	Laboratory	160	Yes
Koriat (2012)	Binary choice in perceptual and general knowledge	Likelihood of being correct (50-100%)	No further choice. Virtual dyads are formed randomly and the response with higher confidence is selected as the group outcome.	Dyads outperform average individual judgments on questions where average accuracy high, which tend to be questions where accuracy and confidence are positively correlated.	Laboratory	40	No
Sniezek and Van Swol (2001)	Multiple choice questions on knowledge about computers	Likelihood of being correct (0-100%)	Advisor recommends an answer to a Judge.	Confidence and accuracy positively correlated on average. Higher Advisor confidence predicts stronger effects of recommendations on Judges.	Laboratory	152	Yes

Notes. This table summarizes a selection of previous experimental work reporting data that permit the calculation of confidence calibration. Specifically, this includes studies that measure objective task performance alongside with respondents' confidence in their performance.

Table 4: Determinants and correlates of subject-level overconfidence

	Overconfidence
Optimality Rate	-84.933^{***}
	(7.111)
Age	0.025
	(0.065)
Male	8.928^{***}
	(1.837)
College?	4.036^{**}
	(1.923)
Income	0.00001
	(0.00002)
Black	3.831
	(4.147)
Hispanic	-14.553^{***}
	(5.197)
Other Race	-1.489
	(5.860)
White	-1.099
	(3.285)
Constant	55.441^{***}
	(4.315)
Observations	334
R^2	0.362
Adjusted R^2	0.344
Residual Std. Error	16.076 (df = 324)
F Statistic	20.413^{***} (df = 9; 324)

Notes. OLS estimates of overconfidence on demographic variables. The unit of the obervation is an individual subject and the dependent variable is the difference between the subject's average confidence and optimality rate in treatment *Confidence*. Independent variables include age in years, an indicator for being male, an indicator for having graduated from college, income and four indicator variables for race (Asian is the excluded variable). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

E Disclosure Statement

The paper reports analyses for all experimental conditions and measures that were collected. All data exclusions are reported. Sample sizes were determined based on a pilot study.

F Experimental Instructions

F.1 Treatment *Confidence*

Instruction screens

Instructions (1/2)

Please read these instructions carefully. There will be comprehension checks. If you fail those, you will not be able to participate in the study and earn a bonus.

This study consists of a total of 15 tasks. Each of these tasks consists of two parts:

Part 1: **You will make a decision by answering a question.** Your decision potentially determines your bonus payment. In each question, there is going to be an optimal decision, by which we mean a decision that maximizes your earnings, on average.

Part 2: **We will ask you how certain you are that your decision in Part 1 was optimal.** Your response to this question does not affect your bonus.

Your bonus
As noted above, there are a total of 15 tasks in this study. In each task, you will make one earnings-relevant decision (Part 1). At the end of the study, the computer will randomly select one of your 15 decisions to determine your bonus. Because we only pay you based on a single decision of yours, there is no point for you in strategizing across decisions or tasks. You should simply always take the decision that you think is best.

→

Instructions (2/2)

Part 1: Your decision

As we described on the previous screen, in each task you will first make a decision (Part 1).

For example, in Part 1 we might ask you a question like "How many cities with more than 2 million people are there in the United States? You will receive 100 points if your decision is correct." Your Part 1 decision is simply your answer to this question, and your decision is "optimal" if it is correct.

Part 2: Your certainty

In Part 2, we ask you "How certain are you that your decision was optimal?" When we ask this question, we are interested in your assessment of how likely it is (in %) that your decision was optimal. You use a slider like the one below to give your answer. If you are completely sure your answer was correct, you should set the slider all the way to the right (100%). If you are certain your answer was not correct, you should set it all the way to the left (0%). In general, the more likely you think it is that you answered the Part 1 question correctly, the further to the right you should set your Part 2 slider.

You need to click on the slider to see the handle.

EXAMPLE:

How certain are you that your decision in Part 1 was optimal?

Not at all certain | Fully certain

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

I am **PLEASE CLICK SLIDER** certain that my decision in Part 1 was optimal.

← →

Example of a task

Here is an example of how a task proceeds. Once the study begins, you will see these two parts on consecutive screens. We just summarize them on one screen here to give you an overview of how things work.

Part 1: Your decision

How many cities with more than 2 million people are there in the United States? You will receive 100 points if your decision is correct.

Part 2: Your certainty

Your decision is considered “optimal” if it is correct.

How certain are you that your decision in Part 1 was optimal?

Not at all certain | Fully certain

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

I am **PLEASE CLICK SLIDER** certain that my decision in Part 1 was optimal.

Once you click the next button, you will not be able to go back to the instructions and the comprehension check questions will start.

← →

Comprehension questions

Comprehension check

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

1. How is your bonus determined?

- I will make 15 decisions in total, and every one of them will get paid. Thus, I can strategize across decisions to hedge my bets.
- I will make 15 decisions in total. The computer will randomly select one of them, and my bonus will depend on my answer to this one question. Thus, there is no point for me in strategizing across decisions.

2. Which of the statements about Part 2 is correct?

- There is no relationship between Part 1 and Part 2.
- My decision in Part 2 builds on my decision in Part 1. Once I get to Part 2, I cannot change my Part 1 decision.
- My decision in Part 2 builds on my decision in Part 1. Once I get to Part 2, I can go back to Part 1 and change my Part 1 decision.

3. Suppose that you DID take the optimal decision in Part 1 of a task. Which Part 2 decision would then be more reflective of your actual performance?

- If I stated that I'm 70% certain I got the task right
- If I stated that I'm 20% certain I got the task right

4. Suppose that you DID **NOT** take the optimal decision in Part 1 of a task. Which Part 2 decision would then be more reflective of your actual performance?

- If I stated that I'm 70% certain I got the task right
- If I stated that I'm 20% certain I got the task right

→

Example screen

Part 2: Your certainty

Task 1/15

You can review your decision from Part 1 by clicking on the back arrow below. You can review the instructions for Part 2 here.
Your decision is considered "optimal" if it maximizes your total earnings.

How certain are you that your decision in Part 1 was optimal?

Not at all certain — Fully certain

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

I am **PLEASE CLICK SLIDER** certain that my decision in Part 1 was optimal.

← →

F.2 Treatment *Betting*

Instructions (1/2)

Please read these instructions carefully. There will be comprehension checks. If you fail those, you will not be able to participate in the study and earn a bonus.

This study consists of a total of 15 tasks. Each of these tasks consists of two parts:

<u>Part 1</u>: **You will make a decision by answering a question.** Your decision potentially determines your bonus payment. In each question, there is going to be an optimal decision, by which we mean a decision that maximizes your earnings, on average.

<u>Part 2</u>: **You will make another decision that relates to the decision you made in Part 1.** This decision will also potentially determine your bonus.

Your bonus

As noted above, there are a total of 15 tasks in this study. In each task, you will make two earnings-relevant decisions (Part 1 and Part 2), for a total of 30 decisions. At the end of the study, the computer will randomly select one of your 30 decisions to determine your bonus. Because we only pay you based on a single decision of yours, there is no point for you in strategizing across decisions or tasks. You should simply always take the decision that you think is best.

→

Instructions (2/2)

Part 1: Your decision

As we described on the previous screen, in each task you will first make a decision (Part 1).

For example, in Part 1 we might ask you a question like "How many cities with more than 2 million people are there in the United States? You will receive 100 points if your decision is correct." Your Part 1 decision is simply your answer to this question, and your decision is "optimal" if it is correct.

Part 2: Bet based on your decision

Once you get to Part 2, you cannot change your Part 1 decision. However, your decision in Part 2 builds on your decision in Part 1. This will work as follows:

- All participants in this study will make a decision in Part 1 by answering the exact same question as you did. In Part 2, 10 participants (including yourself) take part in a betting market that relates to their decision from Part 1.
- Each participant is given a budget of 100 points to participate in the betting market. You can use the slider below to decide **how many points to bet that your decision in Part 1 was optimal. Every point that you don't bet you get to keep**.
- Each of the other 9 participants will also decide how many of their 100 points to bet on the optimality of their decision.
- Based on how much everyone bets, and whether their decisions from Part 1 were actually optimal or not, we determine your bonus as follows:
  - **If your decision in Part 1 was not optimal, every point you bet will be lost.**
  - **If your decision in Part 1 was optimal, every point you bet will yield a positive profit for you**. In this case, your bonus is given by:

    Bonus = Number of points you bet * (Number of points bet by <u>all</u> participants) / (Number of points bet by participants whose Part 1 decision was optimal)

- **While this may sound complicated, what it means is relatively simple: if your decision in Part 1 was optimal, you're guaranteed to earn back at least what you bet, and probably more**.

You need to click on the slider to see the handle.

EXAMPLE:

Bet nothing | Bet everything

0 10 20 30 40 50 60 70 80 90 100

I want to bet **PLEASE CLICK SLIDER** point(s) that my own decision in Part 1 was optimal.

← →

Example of a task

Here is an example of how a task proceeds. Once the study begins, you will see these two parts on consecutive screens. We just summarize them on one screen here to give you an overview of how things work.

Part 1: Your decision

How many cities with more than 2 million people are there in the United States? You will receive 100 points if your decision is correct.

Part 2: Bet based on your decision

How many points do you want to bet that your decision in Part 1 was optimal?

Bet nothing | Bet everything

0 10 20 30 40 50 60 70 80 90 100

I want to bet **PLEASE CLICK SLIDER** point(s) that my own Part 1 decision was optimal.

Once you click the next button, you will not be able to go back to the instructions and the comprehension check questions will start.

← →

Comprehension questions

Comprehension check

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

1. How is your bonus determined?

- I will make 30 decisions in total, and every one of them will get paid. Thus, I can strategize across decisions to hedge my bets.
- I will make 30 decisions in total. The computer will randomly select one of them, and my bonus will depend on my answer to this one question. Thus, there is no point for me in strategizing across decisions.

2. Which of the statements about Part 2 is correct?

- There is no relationship between Part 1 and Part 2.
- My decision in Part 2 builds on my decision in Part 1. Once I get to Part 2, I cannot change my Part 1 decision.
- My decision in Part 2 builds on my decision in Part 1. Once I get to Part 2, I can go back to Part 1 and change my Part 1 decision.

3. Suppose that you DID take the optimal decision in Part 1 of a task. Which Part 2 decision would then lead to higher Part 2 earnings for you?

- If I bet 70 points
- If I bet 20 points

4. Suppose that you DID **NOT** take the optimal decision in Part 1 of a task. Which Part 2 decision would then lead to higher Part 2 earnings for you?

- If I bet 70 points
- If I bet 20 points

→

Example screen

Part 2: Bet based on your decision — **Task 1/15**

You can review your decision from Part 1 by clicking on the back arrow below. You can review the instructions for Part 2 here.

How many points do you want to bet that your decision in Part 1 was optimal?

Bet nothing — Bet everything

0 10 20 30 40 50 60 70 80 90 100

I want to bet **PLEASE CLICK SLIDER** point(s) that my own decision in Part 1 was optimal.

← →

F.3 Treatment *Auction*

Instructions (1/2)

Please read these instructions carefully. There will be comprehension checks. If you fail those, you will not be able to participate in the study and earn a bonus.

This study consists of a total of 15 tasks. Each of these tasks consists of two parts:

Part 1: **You will make a decision by answering a question.** Your decision potentially determines your bonus payment. In each question, there is going to be an optimal decision, by which we mean a decision that maximizes your earnings, on average.

Part 2: **You will make another decision that relates to the decision you made in Part 1.** This decision will also potentially determine your bonus.

Your bonus

As noted above, there are a total of 15 tasks in this study. In each task, you will make two earnings-relevant decisions (Part 1 and Part 2), for a total of 30 decisions. At the end of the study, the computer will randomly select one of your 30 decisions to determine your bonus. Because we only pay you based on a single decision of yours, there is no point for you in strategizing across decisions or tasks. You should simply always take the decision that you think is best.

→

Instructions (2/2)

Part 1: Your decision

As we described on the previous screen, in each task you will first make a decision (Part 1).

For example, in Part 1 we might ask you a question like "How many cities with more than 2 million people are there in the United States? You will receive 100 points if your decision is correct." Your Part 1 decision is simply your answer to this question, and your decision is "optimal" if it is correct.

Part 2: Bid for a potential bonus based on your decision

Once you get to Part 2, you cannot change your Part 1 decision. However, your decision in Part 2 builds on your decision in Part 1. This will work as follows:

- All participants in this study will make a decision in Part 1 by answering the exact same question as you did. In Part 2, 10 participants (including yourself) will take part in an **auction** that relates to your decisions in Part 1.
- The winners of this auction will be eligible for an additional bonus. However, as explained below, the winners of the auction only receives a bonus from winning the auction if they **also** made the optimal decision in Part 1.
- You and the other participants are each given a budget of 100 points to participate in the auction. You can use the slider below to decide **how many points (maximum 100) to bid. Every point that you don't bid you get to keep no matter what**.
- The other participants will also decide how many of their 100 points to bid to get the potential bonus.
- Your Part 2 bonus is then determined according to the outcome of the auction:
  - The five highest bids win the auction. That is, the auction will have FIVE winners. If multiple participants make exactly the same fifth-highest bid, the winner will be chosen randomly from among those bidders.
  - If you are NOT among the five highest bidders, you simply keep your entire initial budget of 100 points, and you won't have to pay the bid you make.
  - If you ARE among the five highest bidders, you will have to pay the amount you bid out of your 100-point budget. In addition, you will receive a bonus:
    - If your Part 1 decision was optimal, your bonus from being a winner in the auction is 100 points.
    - If your Part 1 decision was not optimal, your bonus from being a winner in the auction is 0 points.
- **While this may sound complicated, what it means is relatively simple: if your decision in Part 1 was optimal, you receive a bonus of 100 points from being one of the winners in the auction in Part 2, but you will also have to pay your winning bid. If, on the other hand, your decision in Part 1 was not optimal, you don't receive a bonus from being one of the winners in the auction, but you will still have to pay your winning bid.**

You need to click on the slider to see the handle.

EXAMPLE:

Bid nothing | Bid everything

0 10 20 30 40 50 60 70 80 90 100

I want to bid **PLEASE CLICK SLIDER** point(s) to get a bonus that depends on my Part 1 decision.

← →

Example of a task

Here is an example of how a task proceeds. Once the study begins, you will see these two parts on consecutive screens. We just summarize them on one screen here to give you an overview of how things work.

Part 1: Your decision

How many cities with more than 2 million people are there in the United States? You will receive 100 points if your decision is correct.

Part 2: Bid for a potential bonus based on your decision

How many points do you want to bid for receiving the bonus that depends on your Part 1 decision?

Bid nothing | Bid everything

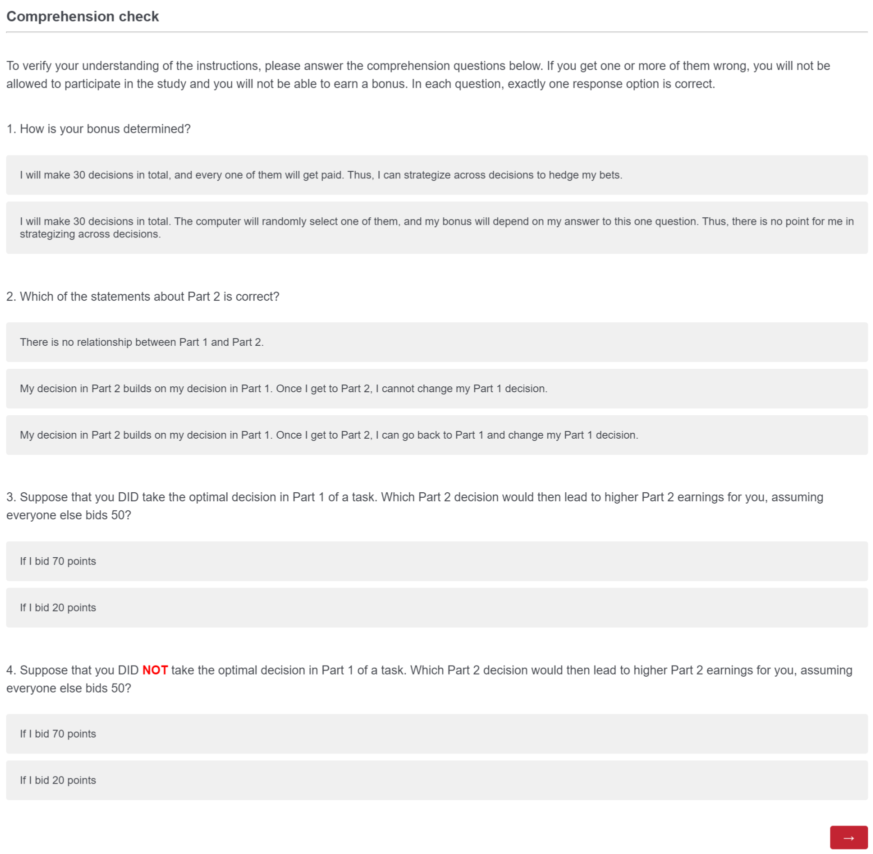
0 10 20 30 40 50 60 70 80 90 100

I want to bid **PLEASE CLICK SLIDER** point(s) to get a bonus that depends on my Part 1 decision.

Once you click the next button, you will not be able to go back to the instructions and the comprehension check questions will start.

← →

Comprehension questions



Comprehension check

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

1. How is your bonus determined?

I will make 30 decisions in total, and every one of them will get paid. Thus, I can strategize across decisions to hedge my bets.

I will make 30 decisions in total. The computer will randomly select one of them, and my bonus will depend on my answer to this one question. Thus, there is no point for me in strategizing across decisions.

2. Which of the statements about Part 2 is correct?

There is no relationship between Part 1 and Part 2.

My decision in Part 2 builds on my decision in Part 1. Once I get to Part 2, I cannot change my Part 1 decision.

My decision in Part 2 builds on my decision in Part 1. Once I get to Part 2, I can go back to Part 1 and change my Part 1 decision.

3. Suppose that you DID take the optimal decision in Part 1 of a task. Which Part 2 decision would then lead to higher Part 2 earnings for you, assuming everyone else bids 50?

If I bid 70 points

If I bid 20 points

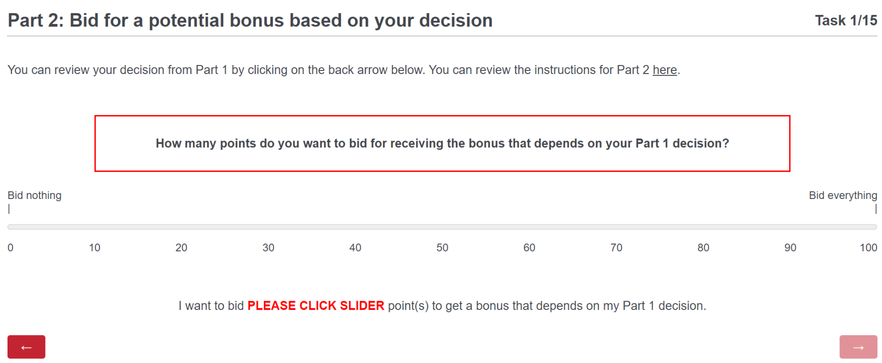
4. Suppose that you DID **NOT** take the optimal decision in Part 1 of a task. Which Part 2 decision would then lead to higher Part 2 earnings for you, assuming everyone else bids 50?

If I bid 70 points

If I bid 20 points

→

Example screen



Part 2: Bid for a potential bonus based on your decision **Task 1/15**

You can review your decision from Part 1 by clicking on the back arrow below. You can review the instructions for Part 2 here.

How many points do you want to bid for receiving the bonus that depends on your Part 1 decision?

Bid nothing | Bid everything

0 10 20 30 40 50 60 70 80 90 100

I want to bid **PLEASE CLICK SLIDER** point(s) to get a bonus that depends on my Part 1 decision.

← →

F.4 Treatment *Committee*

Instructions (1/2)

Please read these instructions carefully. There will be comprehension checks. If you fail those, you will not be able to participate in the study and earn a bonus.

This study consists of a total of 15 tasks. Each of these tasks consists of two parts:

Part 1: **You will make a decision by answering a question.** Your decision potentially determines your bonus payment. In each question, there is going to be an optimal decision, by which we mean a decision that maximizes your earnings, on average.

Part 2: **You will make another decision that relates to the decision you made in Part 1.** This decision will also potentially determine your bonus.

Your bonus

As noted above, there are a total of 15 tasks in this study. In each task, you will make two earnings-relevant decisions (Part 1 and Part 2), for a total of 30 decisions. At the end of the study, the computer will randomly select one of your 30 decisions to determine your bonus. Because we only pay you based on a single decision of yours, there is no point for you in strategizing across decisions or tasks. You should simply always take the decision that you think is best.

→

Instructions (2/2)

Part 1: Your decision

As we described on the previous screen, in each task you will first make a decision (Part 1).

For example, in Part 1 we might ask you a question like "How many cities with more than 2 million people are there in the United States? You will receive 100 points if your decision is correct." Your Part 1 decision is simply your answer to this question, and your decision is "optimal" if it is correct.

Part 2: Vote based on your decision

Once you get to Part 2, you cannot change your Part 1 decision. However, your decision in Part 2 builds on your decision in Part 1. This will work as follows:

- All participants in this study will make a decision in Part 1 by answering the exact same question as you did. In Part 2, we will combine the decisions of 10 participants (including yourself) through a voting procedure to determine the group's Part 2 earnings.
- You can submit **up to 100 votes for your own decision from Part 1**. The more votes you choose to submit, the more your decision will influence the group's earnings.
  - For instance, if you choose to submit 0 votes, you are choosing that your decision from Part 1 does not have any influence on the group's earnings. If you choose to submit 100 votes, you are choosing to have as much influence as possible. You can choose how many votes you'd like to submit using a slider like the one below
- Each of the other 9 participants will also decide how many of their 100 votes to submit for their own decision from Part 1. **The group's earnings are higher the more total votes (across all participants in your group) get submitted for the optimal decision, and lower the more total votes get submitted for decisions that are not optimal.**
- **Important: You and all other participants will all earn an identical amount if Part 2 is selected to count for payment, regardless of how many votes you each individually choose to submit**. Everyone's bonus from Part 2 is only determined by the **fraction of total submitted votes** that are for the optimal decision. Therefore, from the perspective of your earnings, it doesn't matter whether you or other people submit votes – **all that matters is that the votes that do get submitted (whoever they are from) are for the optimal decision**. In case you're interested, the specific formula we use to determine everyone's bonus is given by

  Bonus = 100 * (Number of votes for the optimal decision) / (Total number of votes)

You need to click on the slider to see the handle.

EXAMPLE:

No influence on group earnings | Maximal influence on group earnings

0 10 20 30 40 50 60 70 80 90 100

I want to submit **PLEASE CLICK SLIDER** vote(s) for my own Part 1 decision.

← →

Example of a task

Here is an example of how a task proceeds. Once the study begins, you will see these two parts on consecutive screens. We just summarize them on one screen here to give you an overview of how things work.

Part 1: Your decision

How many cities with more than 2 million people are there in the United States? You will receive 100 points if your decision is correct.

Part 2: Vote based on your decision

How many votes do you want to submit for your own decision from Part 1?

No influence on group earnings | Maximal influence on group earnings

0 10 20 30 40 50 60 70 80 90 100

I want to submit **PLEASE CLICK SLIDER** vote(s) for my own Part 1 decision.

Once you click the next button, you will not be able to go back to the instructions and the comprehension check questions will start.

← →

Comprehension questions

Comprehension check

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

1. How is your bonus determined?

- I will make 30 decisions in total, and every one of them will get paid. Thus, I can strategize across decisions to hedge my bets.
- I will make 30 decisions in total. The computer will randomly select one of them, and my bonus will depend on my answer to this one question. Thus, there is no point for me in strategizing across decisions.

2. Which of the statements about Part 2 is correct?

- There is no relationship between Part 1 and Part 2.
- My decision in Part 2 builds on my decision in Part 1. Once I get to Part 2, I cannot change my Part 1 decision.
- My decision in Part 2 builds on my decision in Part 1. Once I get to Part 2, I can go back to Part 1 and change my Part 1 decision.

3. Suppose that you DID take the optimal decision in Part 1 of a task. Which Part 2 decision would then lead to higher Part 2 earnings for you, on average?

- If I submit 70 votes
- If I submit 20 votes

4. Suppose that you DID **NOT** take the optimal decision in Part 1 of a task. Which Part 2 decision would then lead to higher Part 2 earnings for you, on average?

- If I submit 70 votes
- If I submit 20 votes

→

Example screen

Part 2: Vote based on your decision **Task 1/15**

You can review your decision from Part 1 by clicking on the back arrow below. You can review the instructions for Part 2 here.

How many votes do you want to submit for your own decision from Part 1?

No influence on group earnings | Maximal influence on group earnings

0 10 20 30 40 50 60 70 80 90 100

I want to submit **PLEASE CLICK SLIDER** vote(s) for my own decision from Part 1.

← →

F.5 Task Descriptions

Acquiring a compnay

Part 1: Your decision **Task 1/15**

- You have a budget of 180 points. You can either keep it or use it to buy a company.
- Bob is selling his company. The **VALUE** of Bob's company to him is either **20** or **120 points**, but you do not know which. There is a **50% chance** it is worth 20 points to him and a **50% chance** it is worth 120 points to him.
- **Bob's company has a higher value to you than to Bob**. If you acquire his company, it will pay you 1.5 times its value to Bob. Therefore, if the value of the company turns out to be 20 points for Bob, it would be worth 30 points for you. If the value of the company turns out to be 120 points for Bob, it would be worth 180 points for you.
- The realized value is determined randomly by the computer, and you will not know the value until after you've made your decision.
- You can make a **PRICE** offer to Bob of up to 180 points.
- Your earnings will be determined as follows:
  - If you offer a PRICE **that is at least as high** as Bob's realized VALUE, Bob will accept your offer, and your earnings will be
    *Earnings = (Your budget) + 1.5 * (Bob's **VALUE**) – (the **PRICE** you offered)*
  - If you offer a PRICE **less** than Bob's realized VALUE, you will not acquire his company and your profits will be
    Earnings = Your budget

How much do you bid for Bob's company?

[ ] point(s)

→

Knapsack

Part 1: Your decision **Task 1/15**

- There are 12 **ITEMS** shown in the Table below. Your task is to choose one or more of these items.
- Each item has a **VALUE** in points to you. Your earnings for this task are given by the **SUM OF VALUES** of the items you choose.
- However, each item also has a **WEIGHT**. The total **SUM OF WEIGHTS** of the items you choose **CANNOT EXCEED 14**. If your selection exceeds this weight limit, you will earn nothing.

Which items do you choose? (Please click on the columns)

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
Value	2	3	4	5	6	9	8	7	6	5	8	9
Weight	3	4	6	3	5	13	6	9	2	4	7	7

Current sum of weights chosen **(cannot exceed 14)**: **0**

→

Iterated reasoning / backward induction

Part 1: Your decision

Task 1/15

- Your task is to **pick two numbers between 0 and 100 each**. Let's call these numbers A and B.
- You will earn more points the **closer** A and B each are to 2/3 of the **average** of A and B.
- Specifically, we will pay you 100 base points and subtract from this:
  - The absolute difference between A and 2/3 of the average of both numbers AND
  - The absolute difference between B and 2/3 of the average of both numbers.
- You cannot make losses, meaning you always earn at least 0 points.
- The rules are:
  - All numbers between 0 and 100 are acceptable, including 0 and 100.
  - You are welcome to pick the same number for A and B, or different numbers for each.
- **While all this may sound complicated, all it means is that you will receive more points the closer your chosen numbers A and B are to 2/3 of the average of the two numbers.**

Which numbers do you choose?

A

B

→

Exponential growth bias

Part 1: Your decision

Task 1/15

- **Suppose a stock starts at a value of $100.**
- **It grows by 5% each year relative to its beginning-of-year value.**
- **How much is it worth after 20 years?** If necessary, round your decision to the nearest dollar value.
- We will pay you more points the closer your decision is to the correct answer.
  - Specifically, we will pay you 100 base points and subtract from this the absolute difference between your decision and the correct stock value. For example, if the absolute difference between the true stock value and your decision is $10, we will subtract 10 points.
  - You cannot make losses, meaning you always earn at least 0 points.

How much is the stock worth after 20 years? (round to the nearest integer)

$

→

Correlation neglect

Part 1: Your decision

Task 1/15

- There are three people: Ann, Bob and Charlie. Each of them is interested in estimating the weight of a water bucket in pounds.
- Ann and Bob both **get to take a peek at the bucket**. They are equally good at estimating weight. Each of them gets weight estimates right, on average, but sometimes makes **random mistakes**. Ann and Bob are equally likely to make mistakes in any given estimate they make.
- Ann and Bob both share their estimates with Charlie, who has never seen the bucket. Because he has never seen the bucket, Charlie computes his best estimate of the weight of the bucket as the average of the estimates of Ann and Bob.
- **You have never seen the bucket either, but you're asked to produce an estimate of its weight.** You now talk to Ann and Charlie. They share the following estimates with you:
  - Ann's estimate: 70
  - Charlie's estimate: 40
- **Your task is to estimate the weight of the bucket.**
- We will pay you more points the closer your decision is to the statistically-correct estimate given the information you are provided.
  - Specifically, we will pay you 100 points if your decision corresponds to this correct answer. We subtract 3 points for every number you are away from the correct answer.
  - You cannot make losses, meaning you always earn at least 0 points.

What is your best estimate of the weight of the bucket? (round to the nearest integer)

→

CRT

Part 1: Your decision

Task 1/15

It takes 6 machines 6 days to produce 6 cars. **How long would it take 12 machines to produce 12 cars?** (round to the nearest integer)

day(s)

We will pay you 100 points if you get it right, and nothing otherwise.

Wason

Part 1: Your decision

Task 1/15

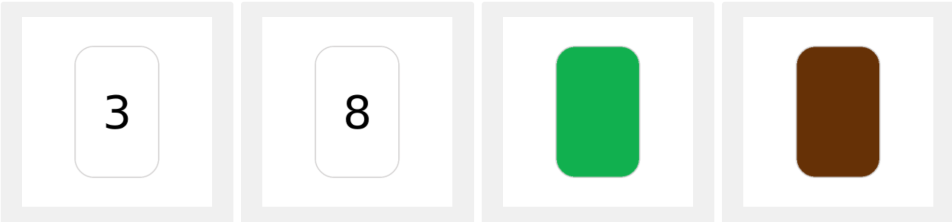
- Suppose your friend has a special deck of cards.
- His special deck of cards all have **numbers** (odd or even) on one side and **colors** (brown or green) on the other side. Suppose that the **4 cards** below are from his deck.
- Your friend claims that: "**In my deck of cards, all of the cards with an even number on one side are green on the other.**"

We will pay you 100 points if:

- you turn over **ALL** of the card(s) that can be helpful in determining whether your friend's statement is true AND
- you do not turn over **ANY** of the cards that **CANNOT** be helpful in assessing whether his statement is true.

In other words, you won't earn 100 points if you turned over a card that isn't actually helpful in determining whether your friend's statement is true. Likewise, you won't earn 100 points if you failed to turn over a card that is actually helpful.

Which card(s) do you want to turn over?



→

Thinking at margin

Part 1: Your decision

Task 1/15

- You are given 100 points (money) to store in two different **BANK ACCOUNTS**, A and B. Points stored in each account are **TAXED** by the government in different ways. You can store your points in 20-unit increments.
- Here is how much you pay in taxes for account A based on the total amount stored:

Investment in account A (in points)	20 total points stored	40 total points stored	60 total points stored
Total taxes to be paid for account A (in points)	4	12	24

- For instance, if you store 20 points in account A, you pay 4 points, leaving you with 16 points in account A after taxes. If you store 40 points, you pay 12 points in taxes, leaving you with 28 points in account A after taxes, etc.
- Here is how much you pay in taxes for account B based on the total amount stored:

Investment in account B (in points)	20 total points stored	40 total points stored	60 total points stored
Total taxes to be paid for account B (in points)	10	20	30

- For instance, if you store 20 points in account B, you pay 10 points in taxes, leaving you with 10 points in account B after taxes. If you store 40 points, you pay 20 points in taxes, leaving you with 20 points in account B after taxes, etc.
- In total, you can put 100 points into the bank.
  - We already put 40 points into bank account A for you.
  - We also put another 40 points into account B for you.
  - You now have an **ADDITIONAL 20** points to put into the bank. You must now decide into which account you would like to put these last 20 points.
- We will pay you the 100 points in the bank, minus total taxes from accounts A and B.

Into which account do you put your additional 20 points?

Account A

Account B

→

We noticed a typo in the instructions of this task after beginning the data collection: In the first sentence, instead of "100 points", the instructions read "60 points". Despite this typo, we believe that it was still possible to follow the task description and arrive at the correct decision. The optimization rate in this task did not change after fixing this typo: it was 28.28% (N=1,372) before and 28.24% (N=170) after correcting the mistake. In all our analysis, we thus pool these data.

Portfolio choice and 1/N

Part 1: Your decision

Task 1/15

- **In this task, you'll be asked to choose an investment portfolio that consists of different stocks.**
- There are **four stocks** that pay you different amounts of money depending on the color of a ball that a computer will randomly draw. Each of the colors red, blue, and green is equally likely to get selected by the computer.
- The table below shows you the **payment rate** of each stock, depending on which ball the computer randomly draws. For example, a realized return of 10% means that if you invest 20 points, you end up with 22 points. Likewise, a realized return of -10% means that if you invest 20 points, you end up with 18 points.

Color of ball drawn	Return of **Stock A**	Return of **Stock B**	Return of **Stock C**	Return of **Stock D**
Red	13%	-2%	-9%	17%
Blue	-8%	8%	12%	-9%
Green	8%	6%	7%	7%

- In total, you need to invest **100 points** across these stocks. You can select one of the portfolios (combinations of stock purchases) below.
- The computer will randomly draw a ball and pay you the total amount earned across the stocks in your portfolio.

Portfolio	Points in **Stock A**	Points in **Stock B**	Points in **Stock C**	Points in **Stock D**
I	50	25	0	25
II	25	25	25	25

Which investment portfolio do you choose?

Portfolio I

Portfolio II

→

Balls and urns

Part 1: Your decision

Task 1/15

- There are **two bags**. One bag contains 70 red chips and 30 blue chips. The other one contains 30 red chips and 70 blue chips.
- We secretly flipped a (fair) coin. If it came up **HEADS**, we chose the bag with more red chips. If the coin came up **TAILS**, we chose the bag with more blue chips. **You do not observe which bag was selected.**
- Next, **we drew one chip at random from the bag selected by the coin toss**. You will learn the color of this randomly-drawn chip below. Then, you need to guess (in percent) which bag was selected.
- We will pay you more points the closer your decision is to the statistically-correct percentage chance given the information you are provided.
  - Specifically, we will pay you 100 points if your decision corresponds to this correct answer. We subtract 3 points for every percentage point you are away from the correct answer.
  - You cannot make losses, meaning you always earn at least 0 points.

You are told that one red chip has randomly been drawn from the secretly selected bag. What do you think is the likelihood (percentage chance) that the selected bag is the one with more red chips? (round to the nearest integer)

%

→

Sample size neglect

Part 1: Your decision

Task 1/15

- There are **two factories** that make office chairs. The **larger** factory produces **45 chairs** each day, and the smaller factory produces **15 chairs** each day.
- For both factories, there is a **10% random chance** that any given chair is **defective**. However, since this is random, the exact percentage varies from day to day. Sometimes it may be higher than 10%, sometimes lower.
- For a period of 1 year, each factory recorded the days on which **MORE THAN 20%** of the chairs were defective.

Which factory do you think recorded more days on which more than 20% of the chairs were defective?

The larger factory

The smaller factory

About the same (that is, within 2% of each other)

We will pay you 100 points if your decision corresponds to the statistically-correct option given the information you are provided, and nothing otherwise.

→

Base rate neglect

Part 1: Your decision

Task 1/15

- Assume that, on average, out of every 100 bicycles produced by a bike manufacturer, **90 are good** and **10 are defective**.
- There is a quality control machine that classifies bicycles as either good or defective at the end of the production process. This quality control machine makes classification mistakes from time to time. On average, the machine **correctly** classifies a bicycle (as good or defective) **75 out of 100 times**, but **incorrectly** classifies it **25 out of 100 times**.
- Now a bicycle produced by the manufacturer has randomly been selected. Next, this specific bicycle was run though the quality control machine, and you will learn about the machine's classification below. Based on this classification, your task is to state the likelihood (percentage chance) that this specific bicycle is actually defective.
- We will pay you more points the closer your decision is to the statistically-correct percentage chance given the information you are provided.
    - Specifically, we will pay you 100 points if your decision corresponds to this correct answer. We subtract 3 points for every percentage point you are away from the correct answer.
    - You cannot make losses, meaning you always earn at least 0 points.

You learn that the randomly selected bicycle has been **classified as defective** by the quality control machine. **What do you think is the likelihood (percentage chance) that it is actually defective?** (round to the nearest integer)

%

→

Gambler's fallacy

Part 1: Your decision

Task 1/15

Imagine you are tossing a **fair coin**. After eight tosses you observe the following result (where **T** stands for **TAILS** and **H** stands for **HEADS**):

T – T – T – H – T – H – H – H

Which event is more likely to happen on the next coin toss?

Heads is more likely

Tails is more likely

Both are equally likely

We will pay you 100 points if your decision corresponds to the statistically-correct option given the information you are provided, and nothing otherwise.

→

Regression to the mean / misattribution

Part 1: Your decision

Task 1/15

- The **average** score on a standard **IQ test** is **100**. Suppose a **randomly** selected individual has obtained a score of **140**.
- Suppose further that an IQ score is the sum of both **true ability** and **random good or bad luck**. The luck component can be positive or negative but equals zero on average (over all people).

Which of the following statements is correct?

This person's true IQ is more likely to be above than below 140.

This person's true IQ is more likely to be below than above 140.

This person's true IQ is equally likely to be above or below 140.

We will pay you 100 points if your decision corresponds to the statistically-correct statement given the information you are provided, and nothing otherwise.

→
