Rounding the (Non)Bayesian Curve: Unraveling the Effects of Rounding Errors in Belief Updating

October 22nd, 2024

James Bland† Yaroslav Rosokha‡

Abstract

Estimation of belief learning models relies on several important assumptions regarding measurement errors. Whereas existing work has focused on classical measurement errors, the current paper is the first to investigate the impact of a non-classical, behavioral measurement error—rounding bias. In particular, we design and carry out a novel economics experiment in conjunction with simulations and a meta-study of existing papers to show a strong impact of rounding bias on belief updating. In addition, we propose an econometric technique to aid researchers in overcoming challenges posed by the rounded responses in belief elicitation questions.

Keywords: Rounding Bias, Measurement Errors, Bayesian Updating, Belief Updating, Learning, Conservatism, Base-Rate Neglect, Econometrics, Hierarchical Bayesian Models

† The University of Toledo • james.bland@utoledo.edu

‡ Purdue University • yrosokha@purdue.edu

1 Introduction

Belief updating is central to economic models of decision-making in dynamic environments.[1] The prevalent normative paradigm for updating beliefs in settings with subjective probabilities is the Bayes rule (Bayes, 1763). Not surprisingly, behavioral deviations from the Bayesian updating have received much attention in psychology and economics (e.g., Phillips and Edwards, 1966; Tversky and Kahneman, 1974), with many studies finding evidence of conservatism and base-rate neglect (see Benjamin, 2019, for a recent review). Nevertheless, several studies point to Bayesian updating as a good model of human behavior under risk (e.g., El-Gamal and Grether, 1995). In this paper, we investigate a common but overlooked source of behavioral decision error that turns out to impact belief updating estimates—rounding bias. In particular, we show that rounding of reported probabilities to more cognitively simpler numbers can lead to unreliable estimates of the belief updating process. In addition, we uncover a novel link between rounding and overweighting of new information, which may impact the conclusions drawn from previous experiments on (non) Bayesian updating.

Our approach is three-fold. First, we conduct a meta-study of recent experiments that elicit beliefs in the context of belief updating to demonstrate the evidence of rounding and explore the potential impact it has on the estimated parameters. In particular, we split the participant sample of each experiment based on the frequency of rounding to show that instrumental variable regression (henceforth, IV) estimates, which have become a norm in the literature, vary depending on which subset of participants is used. Second, we use simulations to evaluate the accuracy of the traditional IV approach when responses are rounded. Specifically, we show that rounded responses may lead to a substantial bias depending on the prior and the strength or number of signals used in the experiment. In addition, we propose and evaluate an econometric solution to this rounding problem: a model that treats belief data as interval-valued. Simulation results show that our approach does well in recovering learning parameters in rounded data. Finally, we design a new experiment that manipulates the direction and magnitude of the expected IV bias in the scenarios presented to the participants. Our results show that consistent with simulations, there is a bias in estimates of updating both at the individual and aggregate levels. When applying our new estimation approach, however, we show that the magnitude of base-rate neglect and conservatism is reduced substantially. In fact, on average, the behavior is very close to Bayes rule.

Our work contributes to three broad strands of literature in social sciences. First, we contribute to the experimental research in economics and psychology on belief updating. Papers in this stream of research have focused on conservatism (e.g., Phillips and Edwards, 1966), base rate neglect

[1] Examples abound and include investors in financial markets who must update their beliefs and investment decisions in response to unexpectedly high or low quarterly earnings (e.g., Grossman and Stiglitz, 1980); government officials who must update their beliefs and policy when new measurements (e.g., unemployment) become available (e.g., Rogoff, 1985); consumers who must update their beliefs about the quality of a product (e.g., electric cars) based on experience or reviews (e.g., Erdem and Keane, 1996); workers in labor markets who must update their beliefs about the likelihood of receiving better job offers based on past performance and current economic conditions (e.g., Mortensen, 1986).

(e.g., Kahneman and Tversky, 1973), representativeness (e.g., Kahneman and Tversky, 1972), the small sample size biase (e.g., Rabin, 2002), and the large sample size biases (e.g., Samuelson, 1963). Approaches across this literature vary in terms of the type of elicitation method (e.g., Becker-DeGroot-Marschak, vs. quadratic scoring rule), the task type (e.g., inference vs. forecast), the context (e.g., abstract urns and ball vs. own ability), and the estimation procedures used (e.g., structural vs. reduced form). Results and estimates vary but many report conservatism (underweighting of new information) and base rate neglect (underweighting of the prior). Our paper contributes to this body of work by identifying rounding as a source of deviation from Bayesian updating, highlighting instances when the rounding bias could lead to unreliable estimates, and providing potential pathways to mitigate the impact of rounding.

Second, we contribute to behavioral research on rounding and grouping of responses to reduce cognitive load. Early research in psychology Rosch (1975) highlighted that round numbers may serve as easy reference points. More recently, studies in economics and management put forth evidence of rounding across a number of domains. For example, Pope and Simonsohn (2011); Allen, Dechow, Pope, and Wu (2017) show that round numbers often serve as goals in sports, standardized testing, and controlled lab experiments. Round numbers are also common in the housing market (Pope, Pope, and Sydnor, 2015; Meng, 2023; Wiltermuth, Gubler, and Pierce, 2022), debt payoff (Isaac, Wang, and Schindler, 2021), in probabilistic expectations surveys (Manski and Molinari, 2010; Giustinelli, Manski, and Molinari, 2022), and even fraudulent elections (Beber and Scacco, 2012). On the experimental side, Ruud, Schunk, and Winter (2014) show that uncertainty causes rounding. However, to the best of our knowledge, rounding has not been studied in the context of belief updating, nor has rounding bias been linked with conservatism or base-rate neglect.

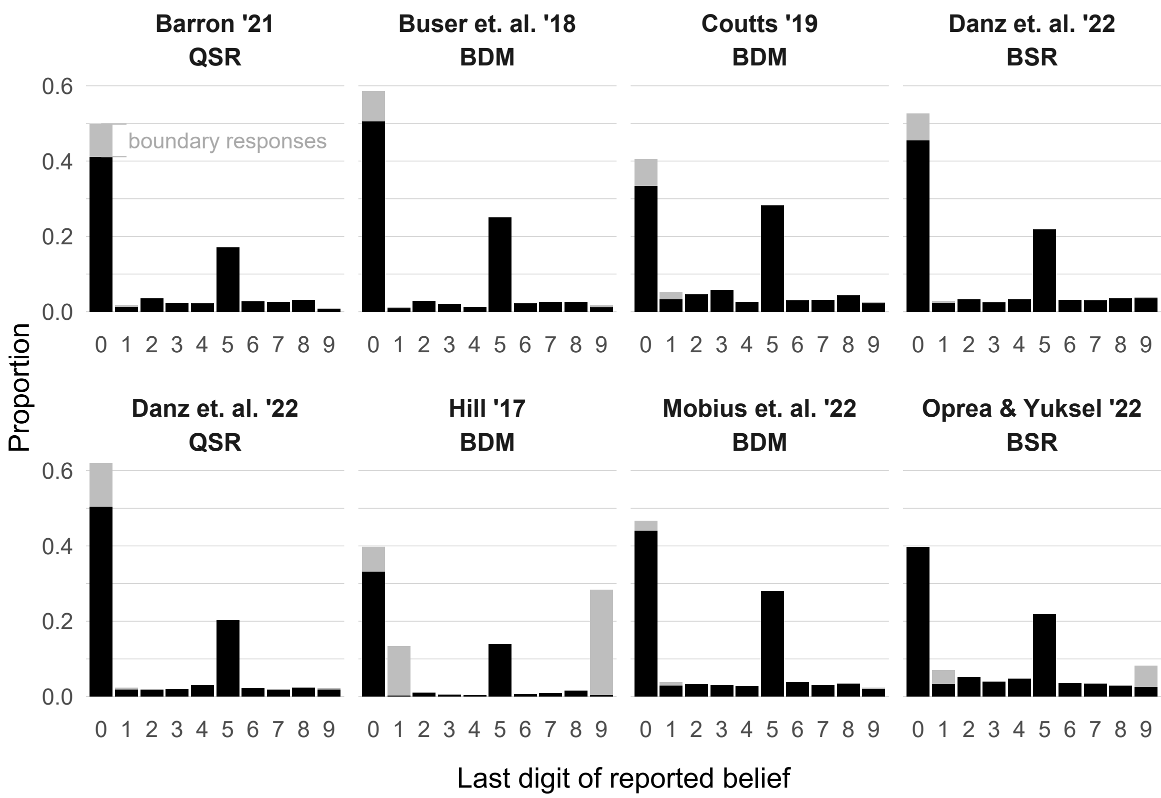
Third, we contribute to the research in statistics and econometrics on rounding as a measurement error. This stream of research has a long history, going back to Sheppard (1897), who introduced a method to correct rounding bias when estimating moments of data. Sheppard's Correction was later expanded by Kendall (1938) and Tricker (1984). More recently, Dempster and Rubin (1983) examined rounding error in the context of the regression model. Our work extends this literature by analyzing the implications of rounding errors on estimates of belief updating models. In particular, we first show how rounding impacts estimates of a common econometric model of belief updating. Next, we augment the simulations with lab experiments to demonstrate how scenarios biased according to our simulations lead to predictable variations in estimated parameters. Finally, we evaluate an econometric for dealing with rounded responses.

The rest of the paper is organized as follows. First, in section 2, we present a meta-study of recent papers on belief updating. Second, in section 3, we carry out simulations to demonstrate that quantitative and qualitative conclusions from the IV approach may be unreliable when dealing with rounded data. We then evaluate an alternative econometric technique to deal with rounded responses. Third, in section 4, we present our main hypotheses, experimental design, and results from our experiment. Finally, in section 5, we conclude and discuss the wide-ranging implications of our research.

2 Stylized Facts

In this section, we consider data from recent papers that study belief updating by eliciting beliefs directly. Among the papers, there are three approaches for eliciting beliefs: the Becket-DeGroot-Marschak (BDM) method, the quadratic scoring rule (QSR), and the binarized scoring rule (BSR). In all three cases, participants are asked to enter their belief as percent. Figure 1 presents histograms of the last digit reported by participants in response to the belief elicitation questions after at least one signal.

Figure 1: Rounding in Past Experiments



Notes: Data from decisions after at least one signal. The figure distinguishes between reports of 0, 1, 99, and 100 as being at the boundary (gray) and reports that are off the boundary (black).

The figure shows that rounding to tens of percent (e.g., .10) is common across all of the studies regardless of the elicitation method. In particular, we find that rounding to .10 is modal across all studies, with the actual proportion varying from 40% to 62%. The second most common form of rounding is to fives of percent (e.g., .15) with approximately 20% of all responses across all studies. We summarize these observations as stylized fact 1:

Stylized Fact 1 *Human decision-makers round their responses to belief elicitation tasks.*

The data from prior studies also reveals that the extent of rounding differs across participants.

In particular, a substantial proportion of human decision-makers always round, whereas there are also participants who rarely round. For example, the proportion of participants who always reported beliefs rounded to .10 was 37% in Danz et al (2022) and 6% in Oprea and Yuksel (2022). One of the goals of the current paper is to understand whether and to what extent the tendency to round responses interacts with belief updating and the estimates thereof. To do this, we estimate the most prevalent model of learning introduced by Grether (1980). The approach relies on the fact that Bayes' rule can be written in log form:

$$\log\left(\frac{p_t}{1-p_t}\right) = \log\left(\frac{p_{t-1}}{1-p_{t-1}}\right) + \log(\lambda_{s_t}) \tag{1}$$

where p_t is the posterior, p_{t-1} is the prior, and λ_{s_t} is the likelihood ratio of observing signal s_t. Setting $y_t \equiv \log(\frac{p_t}{1-p_t})$ and $\lambda_t = \log(\lambda_{s_t})$, leads to the empirical model:

$$y_t = \delta y_{t-1} + \beta\lambda_t + u_t \tag{2}$$

where u_t is the econometric error term.

Comparing Bayes' rule in log form (1) with this empirical model (2) allows us to interpret the parameters as measuring deviations from Bayesian updating. When $\delta = \beta = 1$ and $u_t = 0$, the model collapses to Bayes' rule. Hence, deviations from this restriction imply non-Bayesian updating. Specifically, when $\delta < 1$, the decision-maker does not weigh their prior belief enough. This is called *base rate neglect*. When $\beta < 1$, the decision-maker under-weights the information from the signals and, therefore, updates their belief less aggressively than would a Bayesian. This is called *conservatism*. Both base-rate neglect and conservatism are common findings in the literature (see Benjamin, 2019, for a comprehensive review). Prior literature (e.g., Möbius, Niederle, Niehaus, and Rosenblat, 2022) has also pointed out the dynamic consequences of classical measurement error on belief updating estimates. Namely, the attenuation of the estimates of δ. The proposed solution is to use an instrumental variable approach to address the classical measurement error issue.

Table 1: IV Regression Estimates Based on Rounding Frequency

Experiment	Context	Elicit.	N	NS	Sigs	Par.	Data Subset: Freq. round		Data Subset: Infreq. round	p-value
Hill '17	Political facts	BDM	3950	395	5	δ	0.938 (0.085)	$\sim$	0.857 (0.040)	0.271
						β	0.429 (0.065)	$<$	0.746 (0.063)	0.000
Buser et. al. '18	Performance	BDM	6237	297	7	δ	0.924 (0.024)	$\sim$	0.974 (0.015)	0.223
						β	0.253 (0.029)	$\sim$	0.243 (0.015)	0.493
Coutts '19	Objective	BDM	7632	318	4	δ	0.880 (0.025)	$<$	0.934 (0.012)	0.007
						β	0.314 (0.060)	$<$	0.453 (0.041)	0.038
	Weather		3848	326		δ	0.922 (0.056)	$\sim$	1.046 (0.043)	0.051
						β	0.431 (0.065)	$<$	0.676 (0.048)	0.000
	Performance		3816	318		δ	0.991 (0.092)	$\sim$	1.018 (0.040)	0.766
						β	0.331 (0.056)	$<$	0.572 (0.043)	0.000
Barron '21	Objective	QSR	6660	222	6	δ	0.801 (0.041)	$<$	0.893 (0.016)	0.013
						β	0.503 (0.085)	$\sim$	0.568 (0.043)	0.521
Danz et al. '22	Objective	BSR	8370	2790	3	δ	0.632 (0.030)	$<$	0.865 (0.020)	0.000
						β	0.509 (0.023)	$<$	0.679 (0.021)	0.000
		QSR	3480	1160		δ	0.686 (0.038)	$<$	1.055 (0.042)	0.000
						β	0.307 (0.019)	$\sim$	0.383 (0.024)	0.300
Mobius et al. '22	Performance	BDM	4228	1057	4	δ	0.982 (0.054)	$\sim$	0.982 (0.029)	0.929
						β	0.159 (0.019)	$<$	0.233 (0.016)	0.016
Oprea & Yuksel '22	Objective	BSR	880	220	2	δ	0.681 (0.133)	$\sim$	0.420 (0.081)	0.060
						β	0.917 (0.089)	$<$	1.467 (0.080)	0.000

Notes: N denotes number of choices; NS denotes number of subjects; Sigs denotes number of signals; $<$ denotes significant difference at the .05 level.

Table 1 presents a summary of the seven recent studies that elicit beliefs. Six of the seven studies focus on estimating belief updating using the empirical approach specified with equation (2), with several using the IV approach to correct for classical measurement error.[2] The table shows the wide range of setups used, including the various contexts, elicitation methods, number of signals seen by the participants, and the number of decisions made. The table also shows the IV estimates for frequent and infrequent rounders for each experiment, where a frequent rounder is defined as a participant that rounds at least 75% of responses to .10, and an infrequent rounder is defined as a participant that rounds less than 50% of responses to .10.[3] The table shows that eleven out of twenty comparisons lead to a significant difference. What is even more striking is that all eleven of significant comparisons go in the same direction such that estimates from frequent rounders are lower than estimates from infrequent rounders. We summarize these observations as stylized fact 2:

Stylized Fact 2 *There are systematic differences in estimates between frequent and infrequent rounders.*

The difference in estimates between frequent and infrequent rounders may have two distinct reasons. First, the estimates might be different because rounders exhibit different behavioral pat-

[2] Danz et. al. focus on the properties of the elicitation method and do not investigate belief updating.

[3] In the Appendix, we present an alternative split of the data to show that the same general conclusions hold.

terns than non-rounders. That is, rounding might be associated with higher levels of conservatism and base-rate neglect. Second, the estimates might be different due to the dynamic impact of rounding on the estimation procedure. After all, *rounding is not a classical measurement error* and, therefore, deserves more careful attention. Next, we use simulations and new experiments to investigate whether it is the first or the second (or both).

3 Evaluating Econometric Models Using Simulations

In this section, we use simulations to investigate the ability of the commonly used regression approach to recover underlying belief updating parameters of Bayesian agents (i.e., the true values are set to be $\beta = \delta = 1$). In particular, figure 2 presents a scatter plot of estimates of the base rate neglect (β) and the conservatism (δ) from 600 simulations, where each simulation consists of 100 Bayesian decision makers facing a single scenario with prior probability 50% and signal strength of either $\frac{2}{3}$ or $\frac{9}{10}$. The column facets vary in the extent of imposed rounding. The rows differ in the number of decisions made by each agent. Panel (a) of the figure presents estimates from the IV approach commonly used in the literature to solve classical measurement errors. Panel (b) of the figure presents an alternative approach that combines interval-valued regression and Bayesian econometrics.

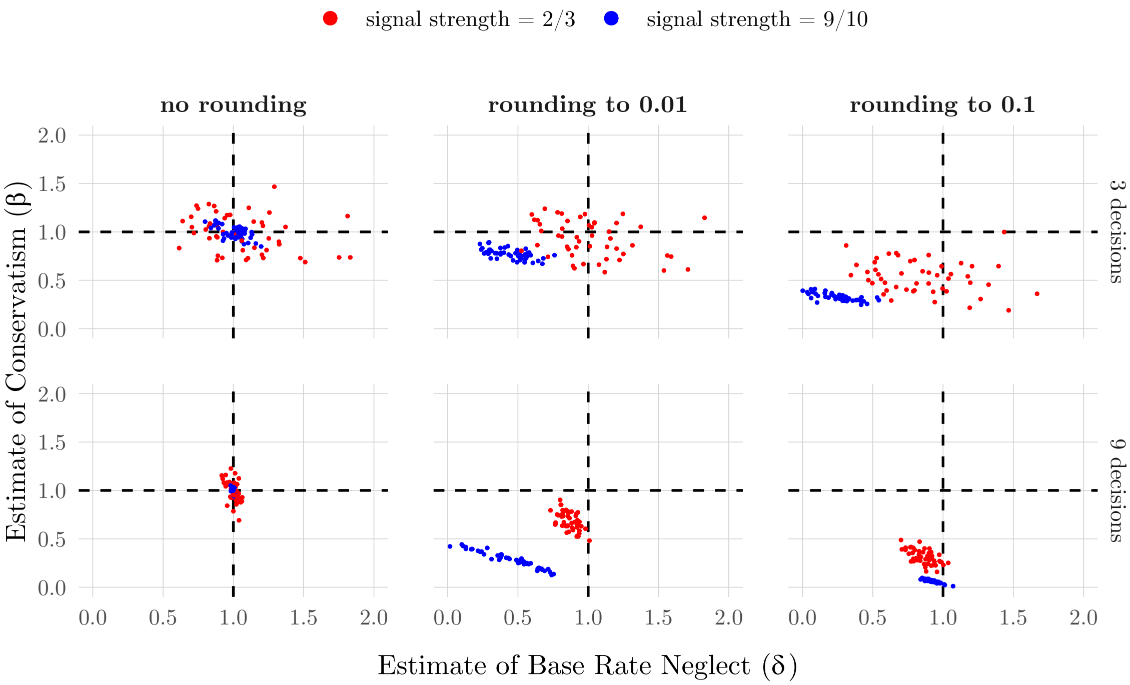
There are three main takeaways from Figure 2. The first takeaway is the IV approach leads to biased estimates of parameters when rounding is present. In fact, our results show that even under seemingly innocuous rounding to .01, which is regularly imposed in surveys and experiments, it is not uncommon to observe an estimate that would be consistent with baser-rate neglect and conservatism. The problem gets exacerbated when rounding increases. For example, in the case of signal strength of $\frac{9}{10}$ when simulated agents make 9 decisions and beliefs are rounded to .01 (blue the bottom center facet of panel (a)), the average estimate of β is .29, and the average estimate of δ is .45. Both of which are substantially lower than the true underlying parameters of 1. We summarize these observations with Simulation Result 1:

Simulation Result 1 *A commonly used IV approach yields unreliable estimates when responses are rounded.*

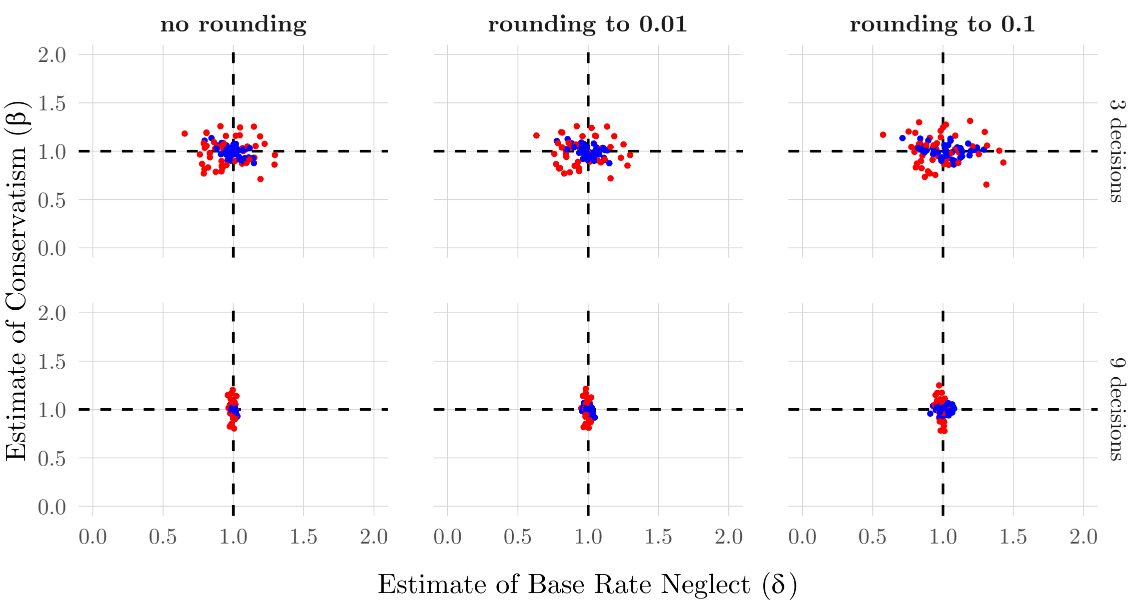
The second takeaway from Figure 2 is regarding the impact of a particular scenario on the estimates. Specifically, the figure presents two scenarios: in red is the scenario with signal strength of $\frac{2}{3}$, whereas in blue is the scenario with signal strength $\frac{9}{10}$. The figure shows a clear difference in estimates depending on the scenario. In general this is problematic, but may explain some of the variation across different studies. In Figure C15 of the Appendix, we provide further evidence of the impact of scenarios on the estimates. We summarize these observations with Simulation Result 2:

Simulation Result 2 *A commonly used IV approach that does not account for rounding yields different estimates depending on the scenarios used.*

Figure 2: Impact of Rounding on Parameter Estimates



(a) Instrumental Variable Regression



(b) Bayesian Interval-valued Regression

Notes: Each facet presents a distribution of 100 estimates (dots). Each dot represents one estimate using data from 100 simulations with Bayesian agents ($\beta = \delta = 1$). Each simulation was carried out for a 50% prior probability and signal strength of either $\frac{2}{3}$ (red) or $\frac{9}{10}$ (blue). Dashed lines show the true values of the parameters. Lagged Bayesian belief was used as an instrument.

As a remedy for this rounding problem, we propose an estimator that treats belief data as interval-valued, while still addressing the classical measurement error problem. Specifically, we assume that participants' beliefs, absent any kind of measurement error, are updated according to the latent process:

$$y_t^* = \delta y_{t-1}^* + \beta \lambda_t + \epsilon_t \tag{3}$$

$$\epsilon_t \sim iidN(0, \sigma_\epsilon^2) \tag{4}$$

where y_t^* is the latent logit belief. To account for classical measurement error, we assume that these latent beliefs receive an additive shock η_t:

$$y_t = y_t^* + \eta_t \tag{5}$$

$$\eta_t \sim iidN(0, \sigma_\eta^2) \tag{6}$$

That is, $\text{logit}^{-1}(y_t)$ is the belief that a participant would report if they did not round their responses. We then adopt the following convention for determining the interval-valued data:

1. If a reported belief is an integer multiple of ten percentage points, then the actual belief was within five percentage points of the reported belief,
2. If a reported belief is an integer multiple of five percentage points (but not ten percentage points), then the actual belief was within 2.5 percentage points of the reported belief, and
3. If a reported belief is an integer multiple of one percentage point (but not a multiple of five or ten percentage points), then the actual belief was within half a percentage point of the reported belief.

The data passed to our estimator therefore includes bounds for y_t which we denote $(\underline{y}_t, \overline{y}_t)$, the prior for each scenario, and the signal information variable λ. We recover the parameters for this model using Bayesian techniques, outlined in Appendix A.[4] Panel (b) of Figure 2 shows that the proposed approach is reliable in recovering true parameters. We summarize observations with Simulation Result 3:

Simulation Result 3 *The proposed econometric approach does well in recovering true parameters.*

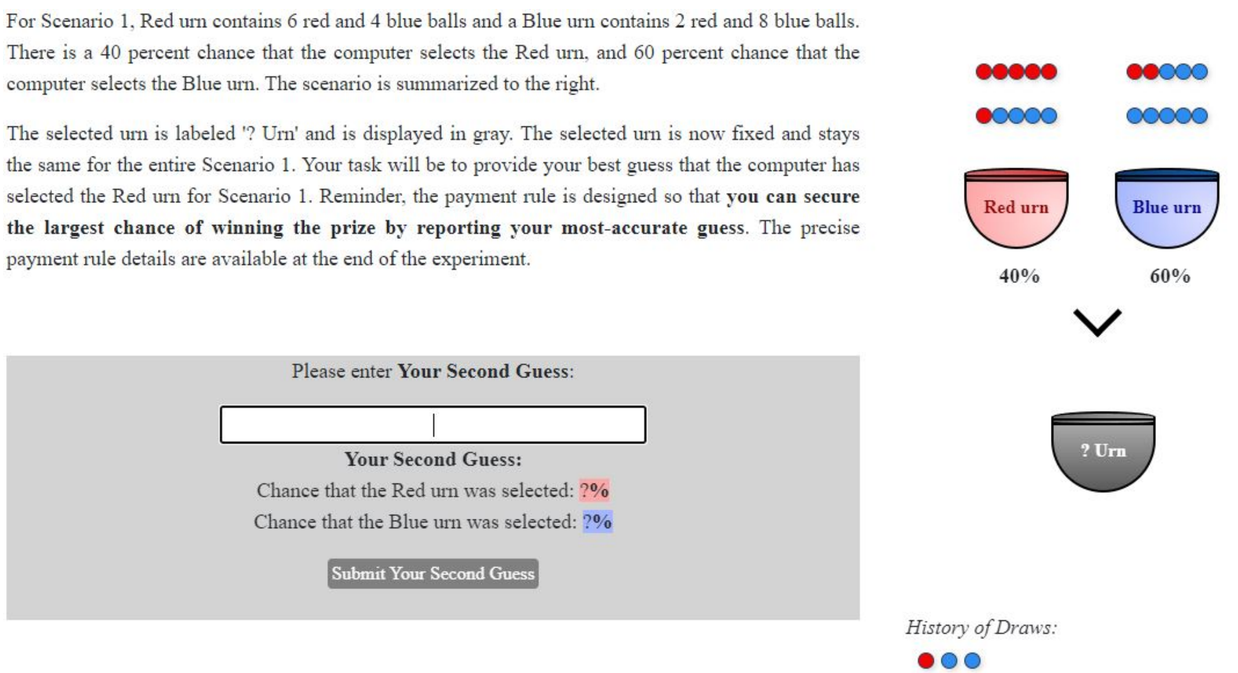
4 The Experiment

For the experiment, we used ORSEE to recruit 139 student participants on the campus of Purdue University in October 2024. Upon arrival, participants were randomly assigned a participant ID.

[4]Posterior simulation is performed in *Stan* (Carpenter, Gelman, Hoffman, Lee, Goodrich, Betancourt, Brubaker, Guo, Li, and Riddell, 2017) and *RStan* (Stan Development Team, 2024).

Each participant faced a unique set of 12 scenarios. Figure 3 presents a screenshot of the interface used for the experiment. The screenshot also presents an example of a scenario. The interface was interactive, such that the draws were visualized to clearly demonstrate how the ball was drawn and placed back into the urn, while the history was recorded and displayed at the bottom. The experiment lasted approximately 45 minutes, with an average payment of 29.06 USD.

Figure 3: Screenshot of Decision Screen



One of the main goals of this paper is to investigate the properties of the existing IV estimation methods. To this end, our experiment contains three features that set it apart from prior studies. First, we consider a large number of decisions faced by each participant, allowing us to reduce variability associated with few decisions and delve dive deeper into individual estimates. Second, our belief elicitations were conducted at random times, allowing us to construct a novel instrumental variable to supplement existing approaches. In particular, the scenarios were pre-drawn such that for each prior, signal strengths and draw realization, there were two participants that faced different timing of decisions (e.g., after 0, 1, 2, and 9 signals vs. after 0, 4, 7, and 9 signals). Third, we chose to randomly generate a large set of possible scenarios to cover the space of possibilities. This is especially relevant in light of the simulation result 2. Notably, a prior, all three three features could be reasons why IV estimates may not lead to reliable estimates. We use simulation results to posit the following hypothesis:

Hypothesis 1 *Learning estimates of the IV approach will predictably respond to the scenario selection such that*

(a) *at the individual level, using a subset of scenarios with higher simulated estimates will result* *in higher estimates of belief updating parameters*

(b) *at the aggregate level, using a subset of participants who faced scenarios with higher simulated estimates will result in higher estimates of belief updating parameters*

To test Hypothesis 1, we carry out two exercises that are based on the simulations with Bayesian agents. Specifically, we rank each scenario according to the estimate that is obtained from running 100 Bayesian agents through that scenario and force-rounding their responses to tens of percent. For the first exercise, we consider different subsets of scenarios that each human-subject participant faced in the experiment. In particular, the subsets differ with respect to the estimate obtained from simulations. If rounding doesn't play a role in the estimates from human-subject data and the difference in estimates is due to random error, then we should not expect any difference in estimates across these subsets. Table 2 presents the results from the IV estimation. The table shows that when scenarios are sorted according to the simulated estimate of base-rate neglect (δ), the estimates of δ from the human-subject experiment predictably vary from .826 to .939, to 1.019. The differences are significant at p-values $< .001$. Similarly, when scenarios are sorted according to the simulated estimate of conservatism (β), the estimates of β from the human-subject experiment vary from .377 to .572, to .656. The differences are also significant at p-values $< .001$, confirming Hypothesis 1a.

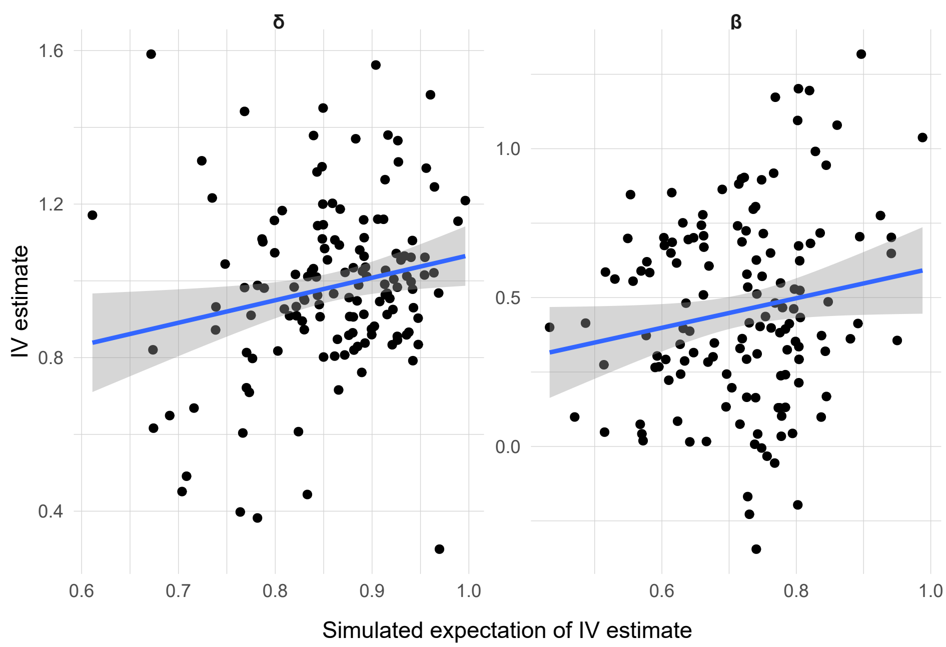
Table 2: IV Regression: Human-subject estimates vs simulated estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
δ	0.928	0.826	0.939	1.019	0.946	0.987	0.793
	(0.017)	(0.031)	(0.026)	(0.026)	(0.024)	(0.030)	(0.041)
β	0.465	0.466	0.543	0.434	0.377	0.572	0.656
	(0.025)	(0.031)	(0.038)	(0.033)	(0.023)	(0.043)	(0.052)
Sorted by	-	δ			β		
Data	All	smallest	middle	largest	smallest	middle	largest
Observations	4,995	1,665	1,653	1,668	1,665	1,653	1,668
Subjects	139	139	139	139	139	139	139

Notes: Column (1) includes all data from the experiment. Columns (2)–(4) split individual scenarios based on the estimates of δ in simulations with Bayesian agents. Columns (5)–(7) split individual scenarios based on the estimates of β in simulations with Bayesian agents. Standard errors clustered at the participant level.

For the second exercise we consider the difference across scenarios faced by participants. That is, due to random chance embedded in our experimental design, some participants faced a set of scenarios that pulled estimates higher than a set of scenarios faced by other participants. Thus, for this exercise, we run Bayesian agents through each set of 12 scenarios, round their responses, and estimate belief updating parameters. This serves as our expectation of IV estimate. We then estimate the same IV model on the responses by human subjects. Figure 4 presents our results. In particular, the figure presents the expectation of the IV estimate from simulations on the x-axis and the IV estimate from individual participants on the y-axis. We find that there is a significant correlation between the estimate of individual parameters and the simulated value confirming Hypothesis 1b (p-values of $< .10$ for δ and $< .05$ for β). We summarize the two exercises with Result 1.

Figure 4: Simulated Rounding Bias vs. Actual Estimate of IV Approach



Notes: Each dot represents a participant-specific IV estimate (vertical coordinate) and simulated expectation of the IV estimate assuming Bayesian behavior ($\delta = \beta = 1$).

Result 1 *Learning estimates of the IV approach will predictably respond to the scenario selection*

(a) *At the individual level, using more positively biased scenarios for the same individuals will generate estimates that are greater than estimates using negatively biased scenarios*

(b) *At the aggregate level, using data from individuals with more positively biased scenarios will yield estimates that are greater than individuals with lower biased scenarios*

Given the failures of the IV approach, we set out to find a method of addressing rounding in experimental data using alternative approaches. One such is the interval-valued model described in Section 3. For our approach, we combine this model with a hierarchical structure, allowing parameters δ, β, σ_ϵ, and σ_η to be participant-specific random parameters drawn from a multivariate normal distribution:

$$\begin{pmatrix} \delta_i \\ \beta_i \\ \log \sigma_{\epsilon,i} \\ \log \sigma_{\eta,i} \end{pmatrix} \sim N(\mu, \Sigma), \;\; \Sigma = \begin{pmatrix} \tau_1 & 0 & 0 & 0 \\ 0 & \tau_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \tau_n \end{pmatrix} \Omega \begin{pmatrix} \tau_1 & 0 & 0 & 0 \\ 0 & \tau_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \tau_n \end{pmatrix} \tag{7}$$

where, at the population level, μ is a vector of means, τ is a vector of standard deviations, and Ω is a correlation matrix.[5] The results of this approach are presented in Table 3. The table shows that the posterior population-level means of δ and β are both close to 1. In fact, the 95% credible region includes 1 for the case or β, meaning that on average the behavior in our experiment is consistent with Bayesian belief updating. We summarize this observation with Result 2:

Result 2 *Learning estimates of the Hierarchical model are close to Bayesian updating*

Table 3: Hierarchical model: posterior moments of population-level parameters

parameter	δ	β	$\log \sigma_\epsilon$	$\log \sigma_\eta$
μ	1.049 (0.019)	0.919 (0.064)	-0.185 (0.068)	-0.255 (0.084)
τ	0.180 (0.019)	0.748 (0.053)	0.752 (0.056)	0.954 (0.070)
Ω				
δ	1	0.641 (0.080)	0.390 (0.103)	0.260 (0.098)
β	0.641 (0.080)	1	0.538 (0.074)	0.617 (0.060)
$\log \sigma_\epsilon$	0.390 (0.103)	0.538 (0.074)	1	0.643 (0.091)
$\log \sigma_\eta$	0.260 (0.098)	0.617 (0.060)	0.643 (0.091)	1

Notes: Posterior means with posterior standard deviations in parentheses

5 Conclusion

In this paper, we investigate the impact of rounding bias on the estimates of belief updating. Specifically, we use simulations to show that rounding of reported probabilities can lead to unreliable estimates of the updating process when using common econometric techniques. In addition, we propose a model that accounts for rounding bias through the interval-valued Bayesian approach. We then use an experimental approach to complement simulations and provide additional evidence on the role of rounding. Because our experiment provides a rich dataset for assessing both econometric models, we are able to first evaluate each presented scenario using simulations and then relate the individual learning parameters from choices to the expected bias obtained from the simulations. Our results provide clear evidence that scenarios used for the experiment lead to biased results due to rounding when using the IV regression approach common in the literature.

[5] Details of the priors selected for this estimation and the data-augmentation process for latent beliefs can be found in Appendix A.

Our study provides several promising directions for future research. First, whereas we focus on the impact of rounding on belief updating, future work could investigate the role of rounding in other settings with continuous action space, such as in price competition and auctions. Second, while our results show that rounding may interact with base rate neglect and conservatism, investigating the interaction between rounding and other behavioral biases, such as overconfidence or ambiguity aversion, could provide a better understanding of decision biases. Third, extending our approach to field data, for example, such as estimates of forecasted inflation in a survey of consumer sentiment, would offer valuable insights into how rounding impacts real-world decision processes from financial markets to consumer behavior.

References

Allen, E. J., P. M. Dechow, D. G. Pope, and G. Wu (2017): "Reference-dependent preferences: Evidence from marathon runners," *Management Science*, 63(6), 1657–1672.

Barron, K. (2021): "Belief updating: does the 'good-news, bad-news' asymmetry extend to purely financial domains?," *Experimental Economics*, 24(1), 31–58.

Bayes, T. (1763): "LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S," *Philosophical transactions of the Royal Society of London*, (53), 370–418.

Beber, B., and A. Scacco (2012): "What the numbers say: A digit-based test for election fraud," *Political analysis*, 20(2), 211–234.

Benjamin, D. J. (2019): "Errors in probabilistic reasoning and judgment biases," *Handbook of Behavioral Economics: Applications and Foundations 1*, 2, 69–186.

Buser, T., L. Gerhards, and J. Van Der Weele (2018): "Responsiveness to feedback as a personal trait," *Journal of Risk and Uncertainty*, 56, 165–192.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell (2017): "Stan: A probabilistic programming language," *Journal of statistical software*, 76, 1–32.

Coutts, A. (2019): "Good news and bad news are still news: Experimental evidence on belief updating," *Experimental Economics*, 22(2), 369–395.

Danz, D., L. Vesterlund, and A. J. Wilson (2022): "Belief elicitation and behavioral incentive compatibility," *American Economic Review*, 112(9), 2851–2883.

Dempster, A. P., and D. B. Rubin (1983): "Rounding error in regression: The appropriateness of Sheppard's corrections," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 45(1), 51–59.

El-Gamal, M. A., and D. M. Grether (1995): "Are people Bayesian? Uncovering behavioral strategies," *Journal of the American statistical Association*, p. 1137–1145.

Erdem, T., and M. P. Keane (1996): "Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets," *Marketing science*, 15(1), 1–20.

Giustinelli, P., C. F. Manski, and F. Molinari (2022): "Tail and center rounding of probabilistic expectations in the health and retirement study," *Journal of econometrics*, 231(1), 265–281.

Grether, D. M. (1980): "Bayes rule as a descriptive model: The representativeness heuristic," *The Quarterly Journal of Economics*, 95(3), 537.

Grossman, S. J., and J. E. Stiglitz (1980): "On the impossibility of informationally efficient markets," *The American economic review*, 70(3), 393–408.

Hill, S. J. (2017): "Learning together slowly: Bayesian learning about political facts," *The Journal of Politics*, 79(4), 1403–1418.

Isaac, M. S., Y. Wang, and R. M. Schindler (2021): "The round-number advantage in consumer debt payoff," *Journal of Consumer Psychology*, 31(2), 240–262.

Kahneman, D., and A. Tversky (1972): "Subjective probability: A judgment of representativeness," *Cognitive psychology*, 3(3), 430–454.

Kahneman, D., and A. Tversky (1973): "On the psychology of prediction.," *Psychological review*, 80(4), 237.

Kendall, M. (1938): "The conditions under which Sheppard's corrections are valid," *Journal of the Royal Statistical Society*, 101(3), 592–605.

Manski, C. F., and F. Molinari (2010): "Rounding probabilistic expectations in surveys," *Journal of Business & Economic Statistics*, 28(2), 219–231.

Meng, C. C. (2023): "The price paid: Heuristic thinking and biased reference points in the housing market," *Journal of Urban Economics*, 134, 103514.

Möbius, M. M., M. Niederle, P. Niehaus, and T. Rosenblat (2022): "Managing Self-Confidence: Theory and Experimental Evidence.," *Manag. Sci.*, 68(11), 7793–7817.

Mortensen, D. (1986): "Job Search and Labor Market Analysis," *Handbook of labor economics/Elsevier*.

Oprea, R., and S. Yuksel (2022): "Social exchange of motivated beliefs," *Journal of the European Economic Association*, 20(2), 667–699.

Phillips, L. D., and W. Edwards (1966): "Conservatism in a simple probability inference task.," *Journal of experimental psychology*, 72(3), 346.

Pope, D., and U. Simonsohn (2011): "Round numbers as goals: Evidence from baseball, SAT takers, and the lab," *Psychological science*, 22(1), 71–79.

Pope, D. G., J. C. Pope, and J. R. Sydnor (2015): "Focal points and bargaining in housing markets," *Games and Economic Behavior*, 93, 89–107.

Rabin, M. (2002): "Inference by believers in the law of small numbers," *The Quarterly Journal of Economics*, 117(3), 775–816.

Rogoff, K. (1985): "The optimal degree of commitment to an intermediate monetary target," *The quarterly journal of economics*, 100(4), 1169–1189.

Rosch, E. (1975): "Cognitive reference points," *Cognitive psychology*, 7(4), 532–547.

Ruud, P. A., D. Schunk, and J. K. Winter (2014): "Uncertainty causes rounding: an experimental study," *Experimental Economics*, 17, 391–413.

Samuelson, P. (1963): "Risk and uncertainty: A fallacy of large numbers," *scientia*, 57(98).

Sheppard, W. F. (1897): "On the calculation of the most probable values of frequency-constants, for data arranged according to equidistant division of a scale," *Proceedings of the London Mathematical Society*, 1(1), 353–380.

Stan Development Team (2024): "RStan: the R interface to Stan," R package version 2.32.6.

Tricker, A. (1984): "Effects of rounding on the moments of a probability distribution," *Journal of the Royal Statistical Society Series D: The Statistician*, 33(4), 381–390.

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

Wiltermuth, S., T. Gubler, and L. Pierce (2022): "Anchoring on Historical Round Number Reference Points," in *Academy of Management Proceedings*, vol. 2022, p. 18018. Academy of Management Briarcliff Manor, NY 10510.

Appendix to "Rounding the (Non)Bayesian Curve" by James Bland and Yaroslav Rosokha

A Bayesian estimator

A.1 Data augmentation

Let $y_{i,t}$ be the latent, un-rounded logit belief with classical measurement error held by participant i in decision t. We can write this latent belief updating process as:

$$\begin{aligned} y_{i,t} &= \delta y_{i,t-1} + \beta\lambda_{i,t} + \epsilon_{i,t} + \eta_{i,t} - \delta\eta_{i,t-1} \\ \epsilon_{i,t} &\sim iidN(0, \sigma_\epsilon^2) \\ \eta_{i,t} &\sim iidN(0, \sigma_\eta^2) \end{aligned}$$

Assuming that $\eta_{i,0} = 0$ and $y_{i,0}$ is the (logit) objective prior, we can write:

$$y_i = \begin{pmatrix} y_{i,1} \\ y_{i,2} \\ y_{i,3} \\ y_{i,4} \end{pmatrix} \sim N(m_i, s) \tag{8}$$

$$\text{where: } m_{i,t} = \delta m_{i,t-1} + \beta\lambda_{i,t}, \quad \mu_{i,0} = y_{i,0} \tag{9}$$

$$s_{j,k} = \sigma_\eta^2 I(j = k) + \sigma_\epsilon^2 \sum_{p=0}^{j-1}\sum_{q=0}^{k-1} \delta^{p+q} I(j - p = k - q) \tag{10}$$

Where $I(\cdot)$ is the indicator function.

Given bounds of $y_{i,t} \in (\underline{y}_{i,t}, \overline{y}_{i,t})$, we can augment the data by y as follows:

$$y_i \sim \text{Truncated Normal}\left(m_i, s, \underline{y}_i < y_i < \overline{y}_i\right)$$

A.2 Hierarchical specification

The hierarchical models assume parameters δ_i, β_i, $\sigma_{\epsilon,i}$, and $\sigma_{\eta,i}$ are participant-specific (hence the i subscript), and drawn from a (transformed) multivariate Normal distribution:

$$\begin{pmatrix} \delta_i \\ \beta_i \\ \log\sigma_{\epsilon,i} \\ \log\sigma_{\eta,i} \end{pmatrix} \sim N(\mu, \Sigma),$$

$$\text{where: } \Sigma = \text{diag\_matrix}(\tau)\Omega\text{diag\_matrix}(\tau)$$

That is, μ and τ are vectors of means and standard deviations, respectively, and Ω is a correlation matrix.

We estimate the hierarchical model with the following hyper-priors on the population-level parameters:

$$
\begin{aligned}
\mu_\delta &\sim N(1, 0.25^2) \\
\mu_\beta &\sim N(1, 0.25^2) \\
\mu_{\sigma_\epsilon} &\sim N(0, 0.25^2) \\
\mu_{\sigma_\eta} &\sim N(0, 0.25^2) \\
\tau_\delta &\sim \text{Cauchy}^+(0, 1) \\
\tau_\beta &\sim \text{Cauchy}^+(0, 1) \\
\tau_{\sigma_\epsilon} &\sim \text{Cauchy}^+(0, 1) \\
\tau_{\sigma_\eta} &\sim \text{Cauchy}^+(0, 1) \\
\Omega &\sim \text{LKJ}(4)
\end{aligned}
$$

B Experimental Instructions and Interface

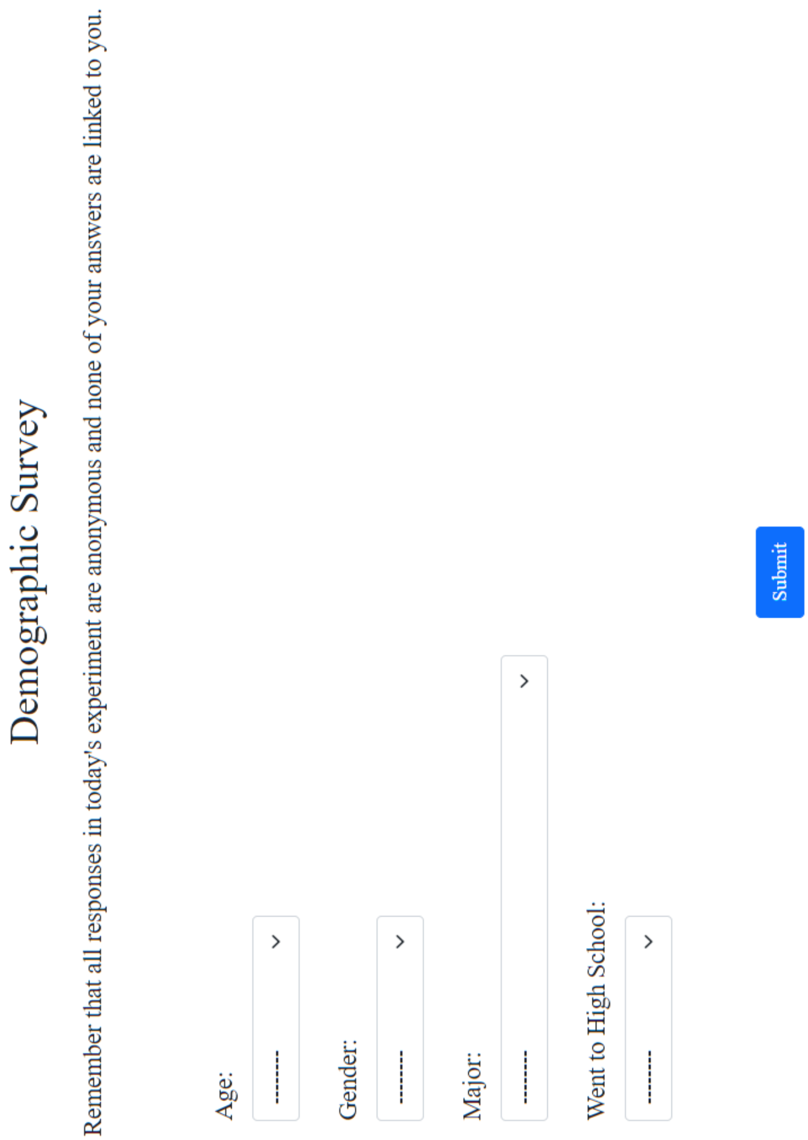


Figure B1: Demographic Survey

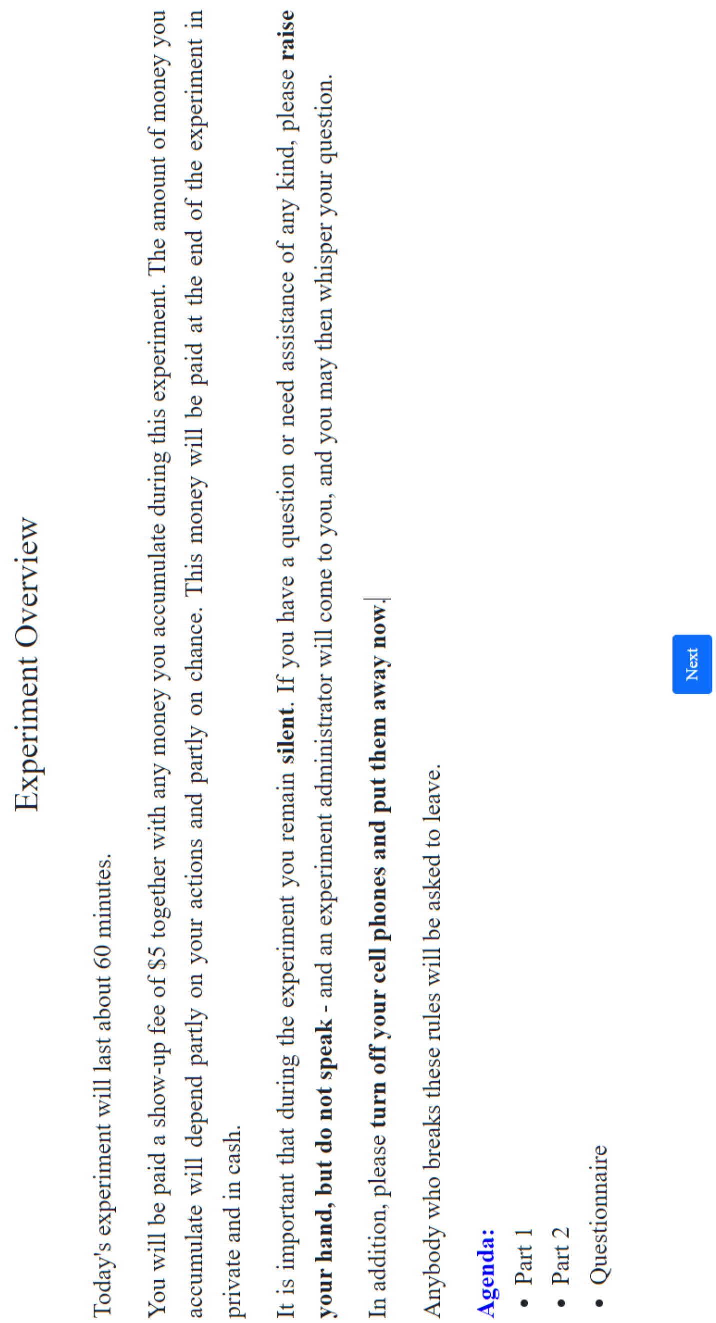


Figure B2: Experiment Overview

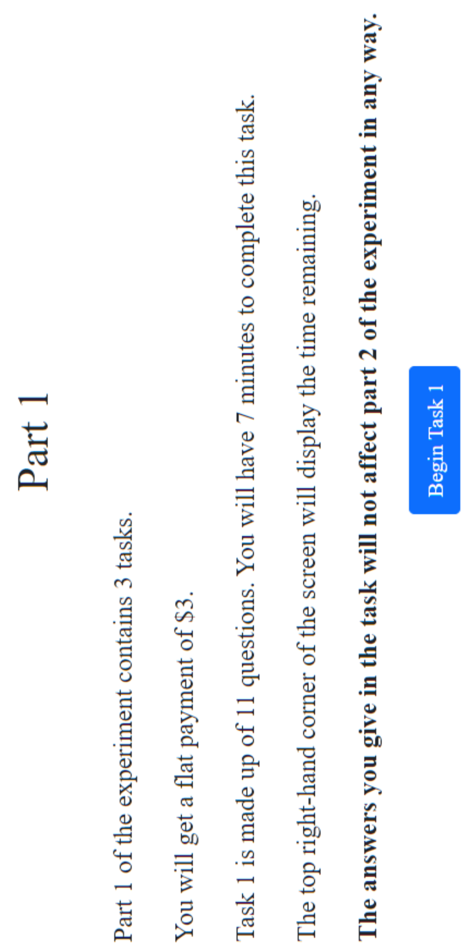


Figure B3: Part 1 Introduction

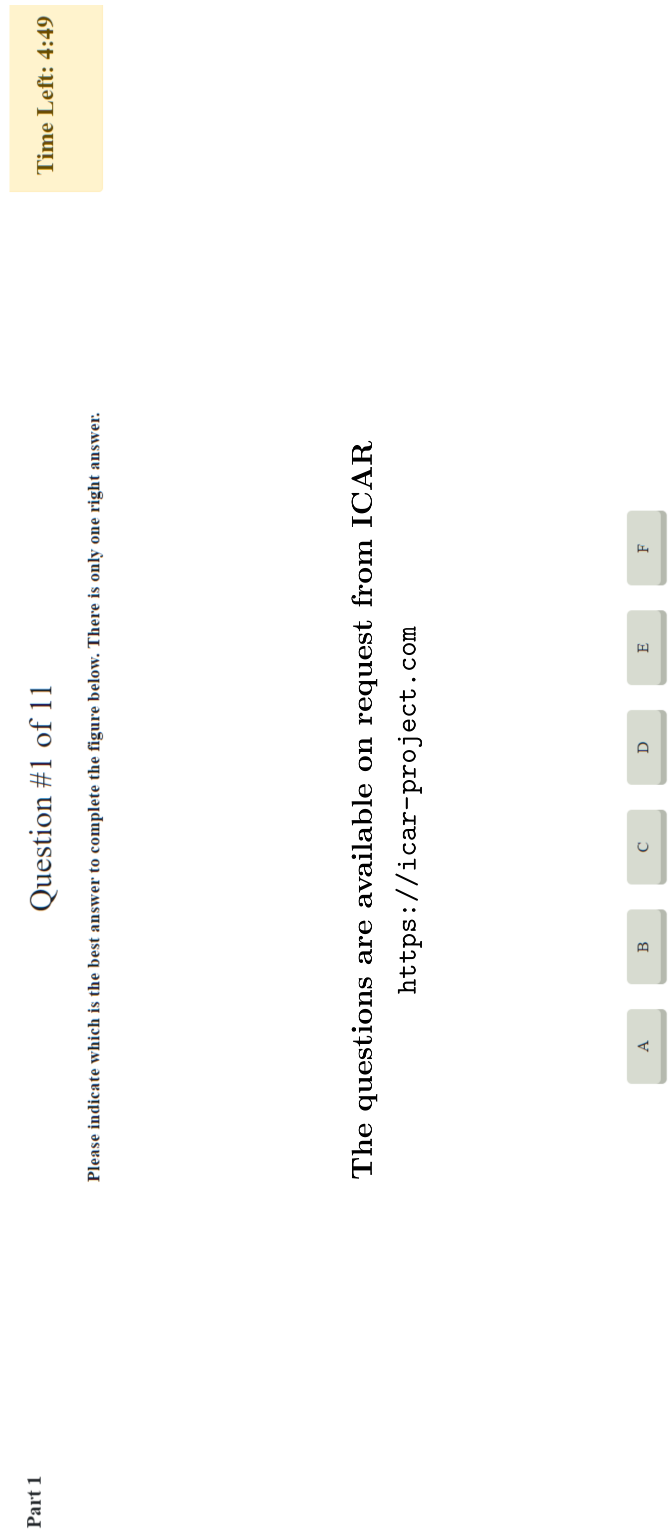


Figure B4: Part 1: Example Task


Part 1

Task 2

Here are a number of characteristics that may or may not apply to you.
Please respond to the following 22 items by selecting the extent to which you agree or disagree with that statement.

Item	I see Myself as Someone Who...	
#1	Is a reliable worker	Select
#2	Likes to reflect, play with ideas	Select
#3	Remains calm in tense situations	Select
#4	Tends to be quiet	Select
#5	Can be cold and aloof	Select
#6	Is curious about many different things	Select
#7	Is sometimes shy, inhibited	Select
#8	Is relaxed, handles stress well	Select
#9	Is helpful and unselfish with others	Select
#10	Is generally trusting	Select
#11	Does things efficiently	Select

Item	I see Myself as Someone Who...	
#12	Is easily distracted	Select
#13	Tends to be lazy	Select
#14	Does a thorough job	Select
#15	Can be somewhat careless	Select
#16	Likes to cooperate with others	Select
#17	Is inventive	Select
#18	Has an assertive personality	Select
#19	Has few artistic interests	Select
#20	Tends to be disorganized	Select
#21	Is sophisticated in art, music, or literature	Select
#22	Starts quarrels with others	Select

Figure B5: Big 5 Questionnaire Page 1

Part 1

Task 3

Here are a number of characteristics that may or may not apply to you.
Please respond to the following 22 items by selecting the extent to which you agree or disagree with that statement.

Item I see Myself as Someone Who...	
#1 Is full of energy	Select
#2 Is depressed, blue	Select
#3 Is sometimes rude to others	Select
#4 Makes plans and follows through with them	Select
#5 Prefers work that is routine	Select
#6 Worries a lot	Select
#7 Has a forgiving nature	Select
#8 Tends to find fault with others	Select
#9 Is ingenious, a deep thinker	Select
#10 Values artistic, aesthetic experiences	Select
#11 Can be tense	Select

Item I see Myself as Someone Who...	
#12 Gets nervous easily	Select
#13 Is reserved	Select
#14 Is emotionally stable, not easily upset	Select
#15 Is original, comes up with new ideas	Select
#16 Is considerate and kind to almost everyone	Select
#17 Is talkative	Select
#18 Has an active imagination	Select
#19 Generates a lot of enthusiasm	Select
#20 Perseveres until the task is finished	Select
#21 Is outgoing, sociable	Select
#22 Can be moody	Select

Figure B6: Big 5 Questionnaire Page 2

Part 2

Part 2 of the experiment will consist of twelve scenarios. In each scenario the computer will fill two urns with 10 balls, either red (●) or blue (●). We call the urn with more red balls the **Red urn**, and the one with more blue balls the **Blue urn**. One of these two urns is selected to be used in the scenario. Your task is to guess how likely it is that the selected urn is the Red urn.

Each scenario proceeds in steps:

Step 1. Computer Fills the Urns: The two urns are filled with 10 balls each, some blue, some red. You will always see the exact number of blue and red balls in the two urns.

Example on the right presents a Red urn that contains 7 red and 3 blue balls and a Blue urn that contains 4 red and 6 blue balls.

Step 2. Computer Selects an Urn: The computer selects the Red or the Blue urn according to the probability presented under each urn. The selected urn is determined as follows:

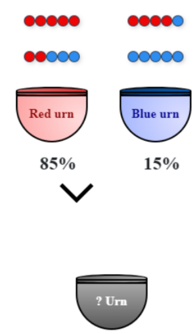
- The computer draws a random number between 1 and 100.
- If the random number is less than or equal to the percent listed under the Red urn, then the Red urn is selected.
- If the random number is greater than the percent listed under the Red urn, then the Blue urn is selected.

For the example on the right, if there is a 85 percent chance that the computer selects the Red urn, and a 15 percent chance that the computer selects the Blue urn.

The selected urn is labeled '? Urn' and is displayed in gray.

Step 3. Computer Draws Balls from the Selected Urn: All draws from the selected urn are independent from one another: After each draw is made, the ball is returned to the selected urn before the next draw is made. The contents of the selected urn are therefore always the same when a draw is made, and each of the balls in the urn has the same chance of being drawn.

To test the drawing procedure click on the 'Draw' button' nine times.



History of Draws:

Remember, once the computer selects an urn it is fixed and stays the same for the entire scenario. Your task will be to **provide your best guess that the computer has selected the Red urn for the scenario**. In particular, you will provide your guesses before and after draws are made.

After you have made decisions for all scenarios, you will learn which urn the computer selected and drew balls from. Your guesses will be used to determine your chances of winning an $8 prize. Your chance of winning the prize is set so that more-accurate guesses lead to a higher chance of winning. Next we will describe how you will make your guess and the compensation procedure.

Next

Figure B7: Part 2 Introduction

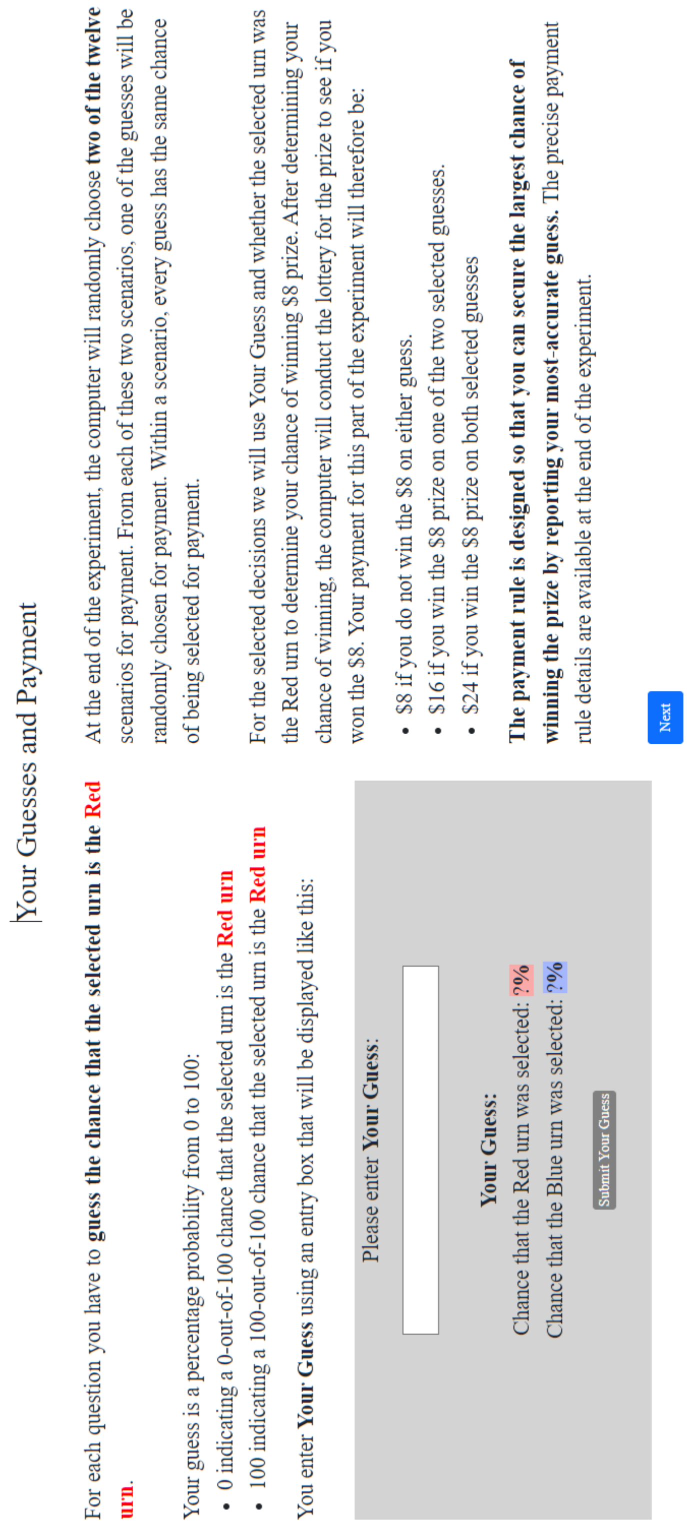


Figure B8: Part 2 Continued

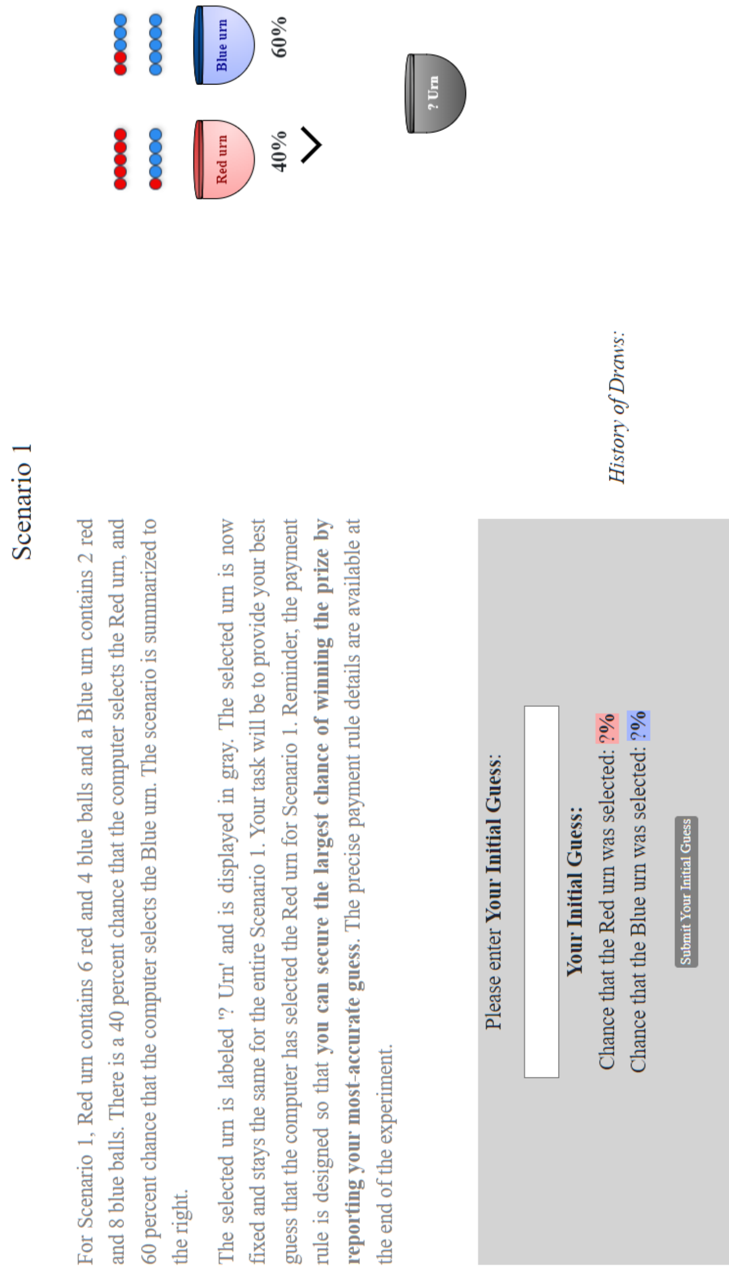


Figure B9: Example Scenario: Initial Guess

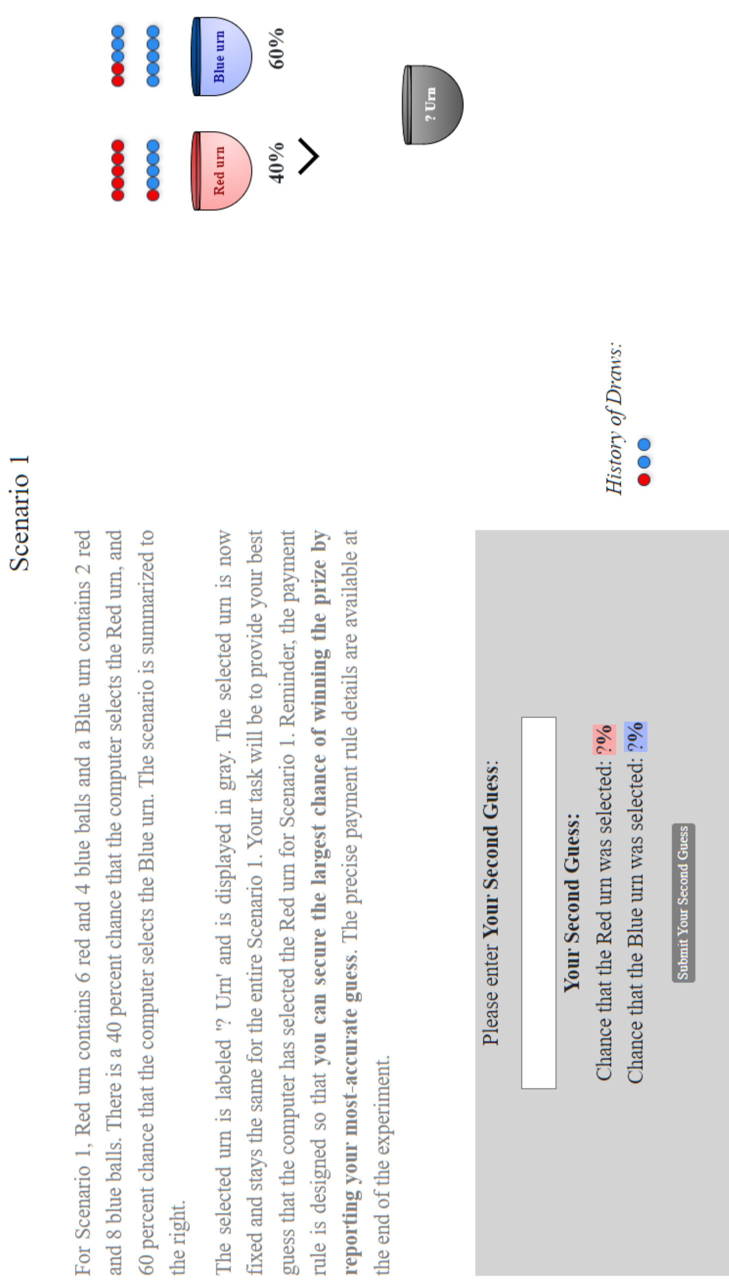


Figure B10: Example Scenario: Second Guess

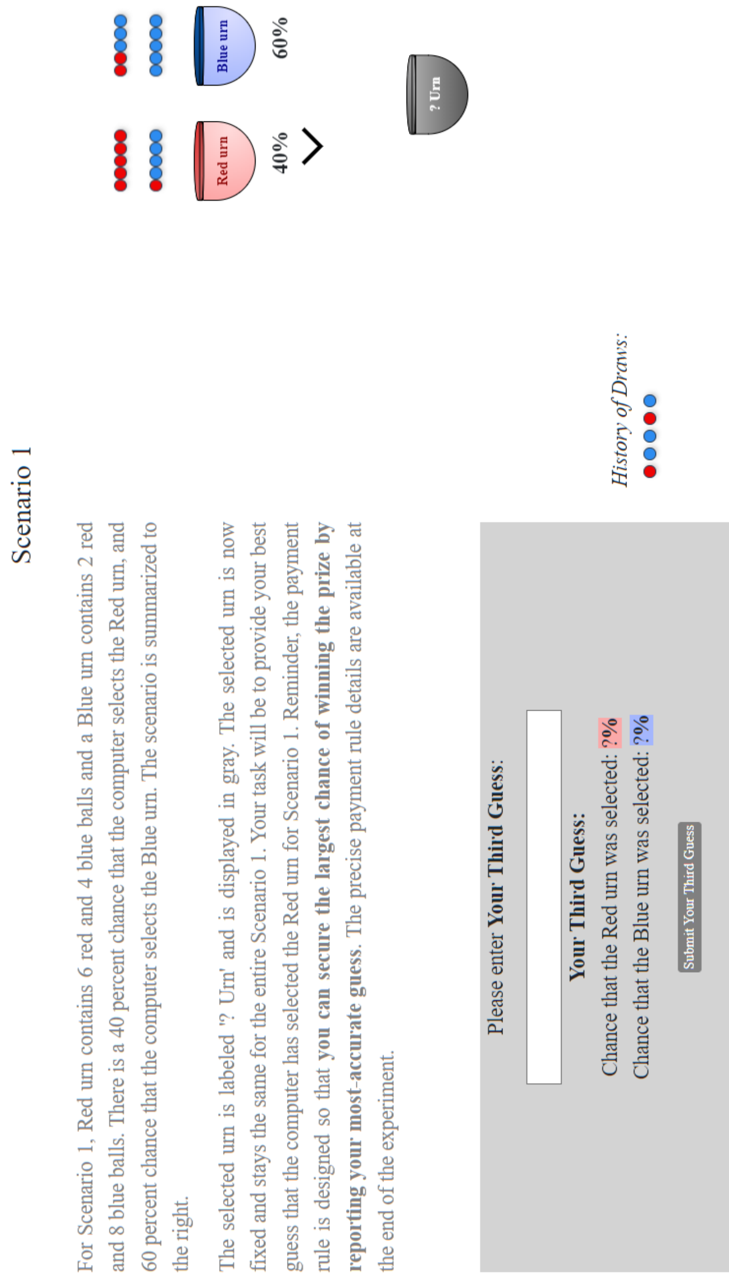


Figure B11: Example Scenario: Third Guess

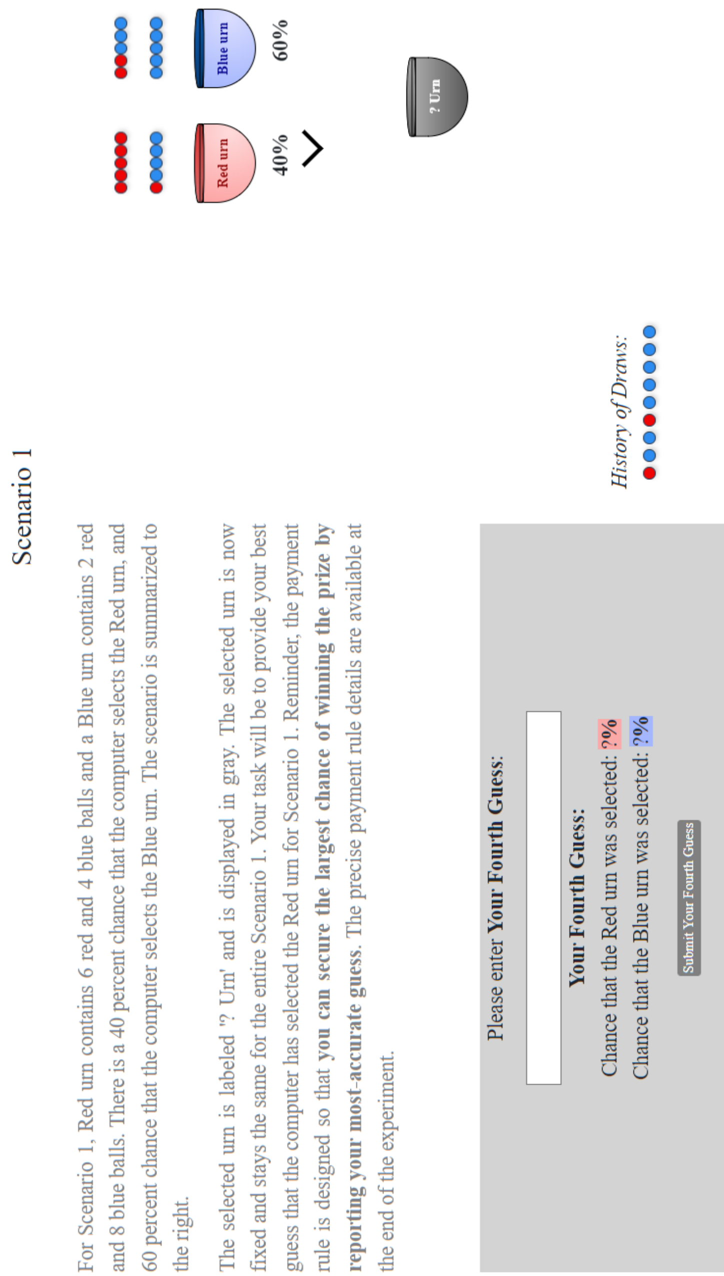
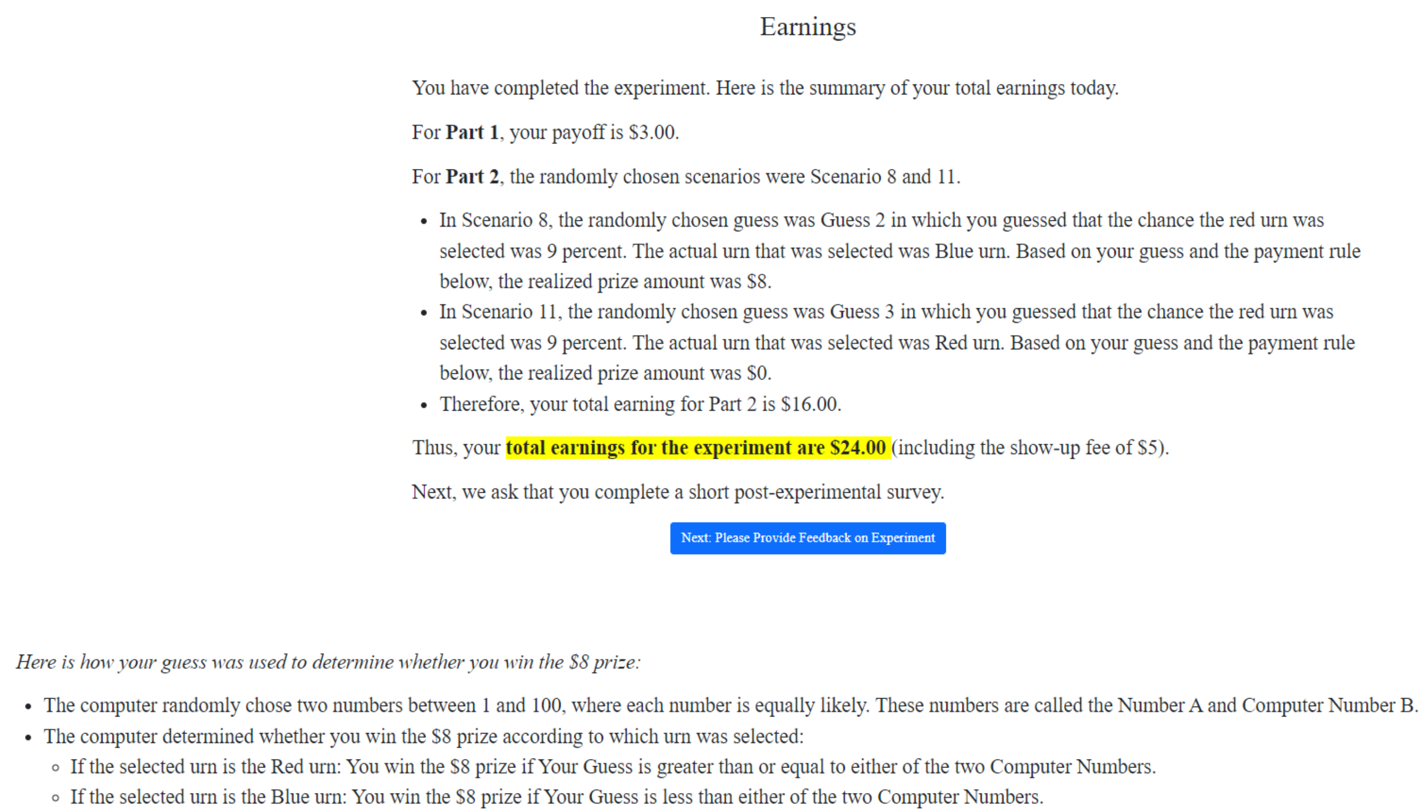


Figure B12: Example Scenario: Fourth Guess



Earnings

You have completed the experiment. Here is the summary of your total earnings today.

For **Part 1**, your payoff is $3.00.

For **Part 2**, the randomly chosen scenarios were Scenario 8 and 11.

- In Scenario 8, the randomly chosen guess was Guess 2 in which you guessed that the chance the red urn was selected was 9 percent. The actual urn that was selected was Blue urn. Based on your guess and the payment rule below, the realized prize amount was $8.
- In Scenario 11, the randomly chosen guess was Guess 3 in which you guessed that the chance the red urn was selected was 9 percent. The actual urn that was selected was Red urn. Based on your guess and the payment rule below, the realized prize amount was $0.
- Therefore, your total earning for Part 2 is $16.00.

Thus, your **total earnings for the experiment are $24.00** (including the show-up fee of $5).

Next, we ask that you complete a short post-experimental survey.

Next: Please Provide Feedback on Experiment

Here is how your guess was used to determine whether you win the $8 prize:

- The computer randomly chose two numbers between 1 and 100, where each number is equally likely. These numbers are called the Number A and Computer Number B.
- The computer determined whether you win the $8 prize according to which urn was selected:
  - If the selected urn is the Red urn: You win the $8 prize if Your Guess is greater than or equal to either of the two Computer Numbers.
  - If the selected urn is the Blue urn: You win the $8 prize if Your Guess is less than either of the two Computer Numbers.

Figure B13: Compensation Screen

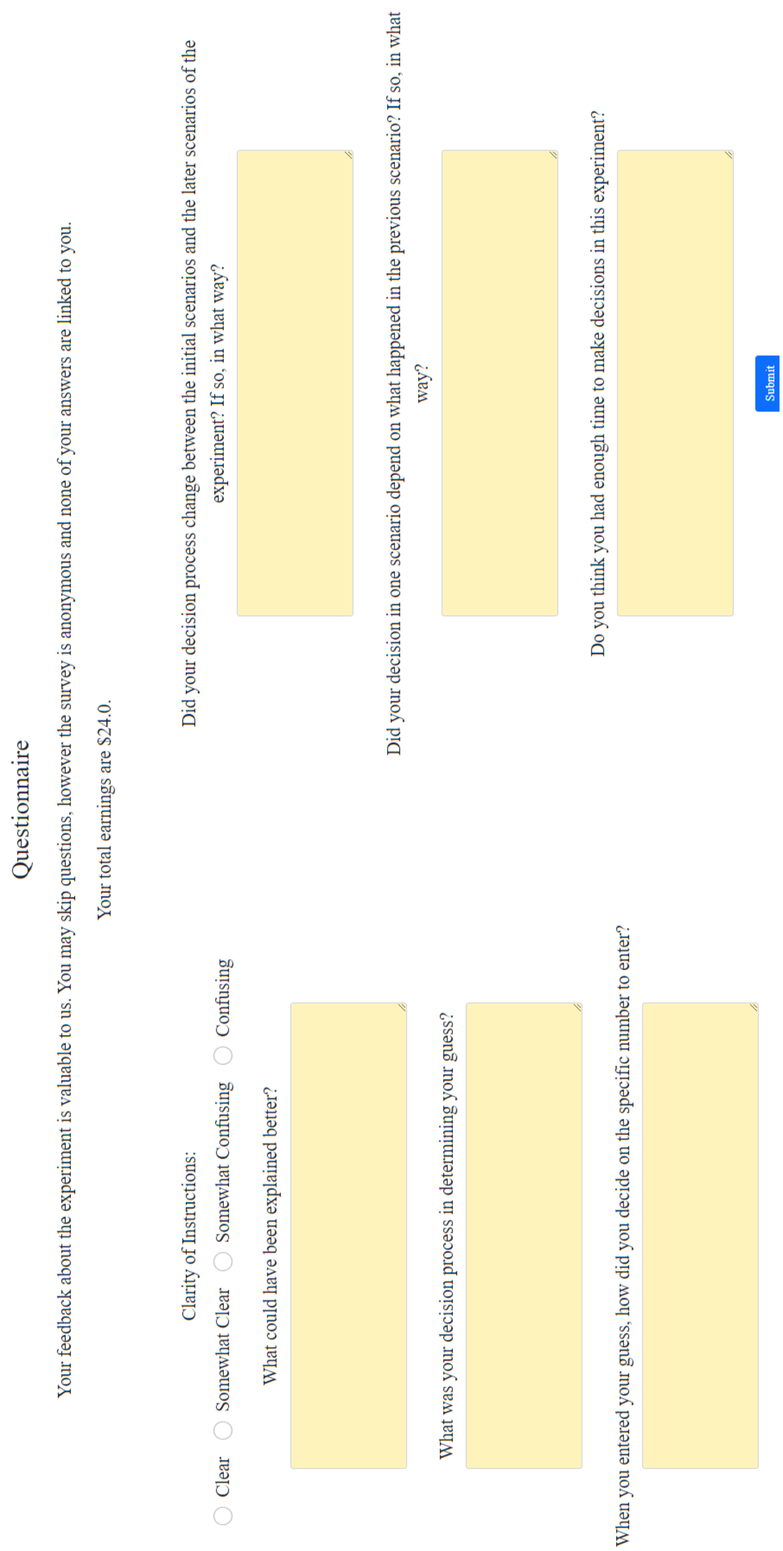


Figure B14: Feedback Questionnaire

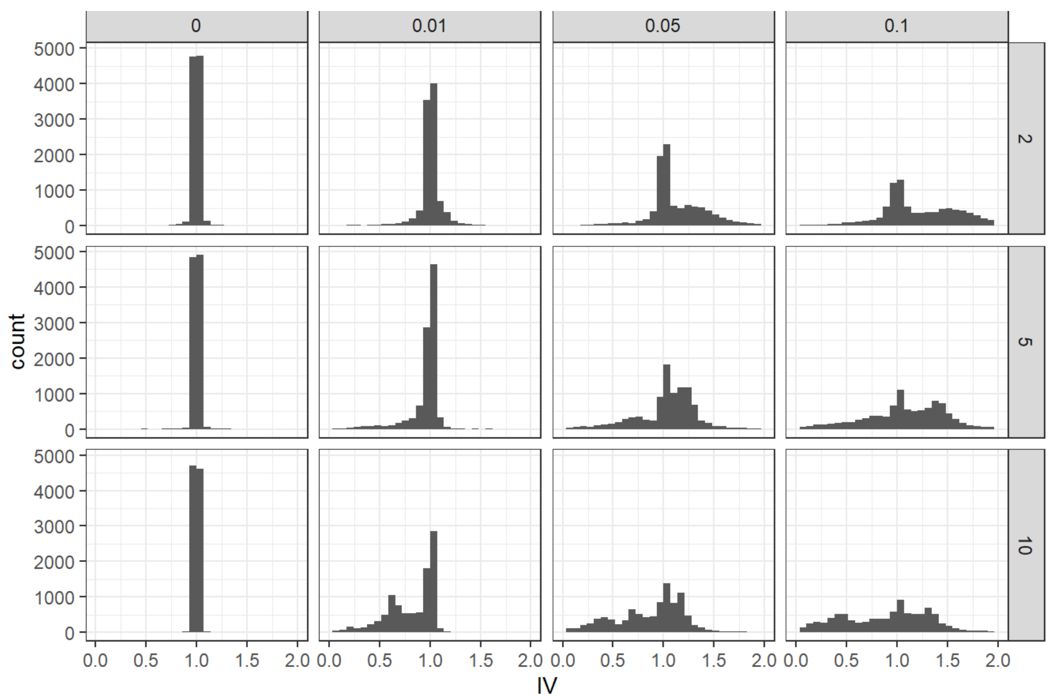
C Additional Results

Table C1: Summary of rounding behavior

decision	0	1-4	50	96-99	100	M10	M5	M1 (remainder)
1	0.011	0.003	0.092	0.003	0.023	0.377	0.303	0.187
2	0.062	0.010	0.089	0.010	0.073	0.384	0.239	0.132
3	0.130	0.022	0.068	0.020	0.138	0.306	0.216	0.099
4	0.192	0.034	0.041	0.034	0.205	0.232	0.183	0.078
Overall	0.099	0.017	0.073	0.017	0.110	0.325	0.235	0.124

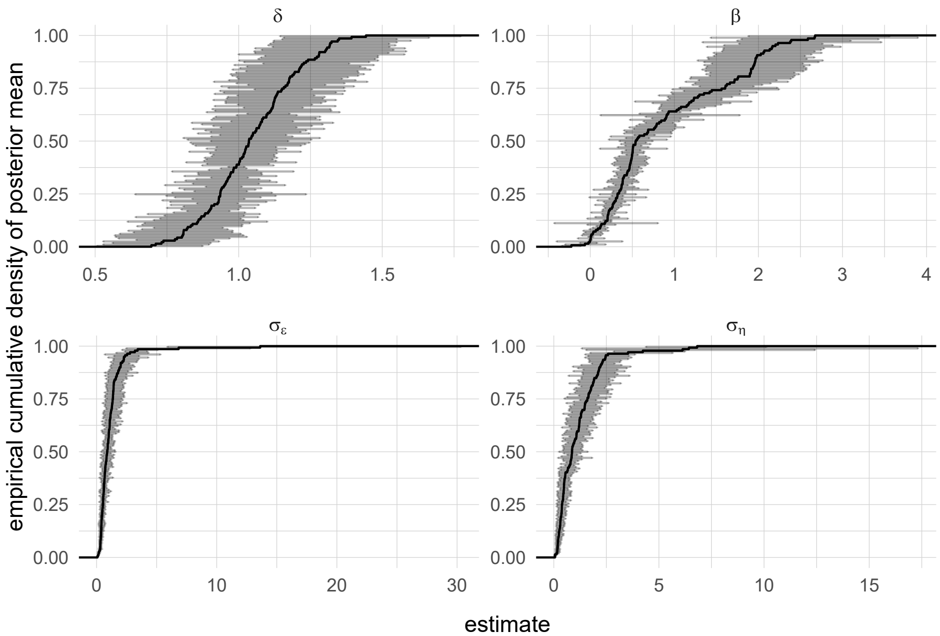
Notes:

Figure C15: Additional Simulations: Impact of Rounding on Estimates of β



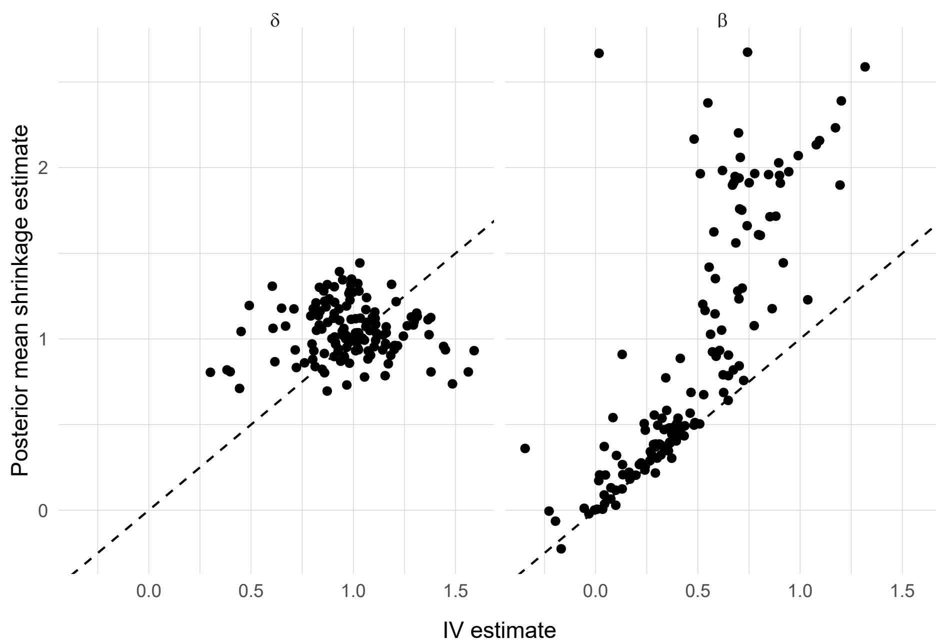
Notes: Histograms present estimates from 10,000 simulated experiments. For each experiment, we randomly pick a prior $\in [.001, .999]$, signal strength $\in [.5, 1]$, and let 100 Bayesian agents make decisions after each randomly drawn signal. We use the same data across columns of the figure but round responses to either .01, .05, and .10 and re-estimate.

Figure C16: Individual shrinkage estimates from the hierarchical model



Notes: Error bars show 95% Bayesian credible regions (2.5th-97.5th percentile)

Figure C17: Comparison of shrinkage estimates to individual IV estimates



Notes: