Belief Elicitation:
Limiting Truth Telling with Information on Incentives

David Danz, Lise Vesterlund, Alistair J. Wilson*

August 2020

Abstract

We study truth telling within the current state-of-the-art mechanism for belief elicitation and examine how information on incentives affects reports on a known objective prior. We find that transparent information on incentives gives rise to error rates in excess of 40 percent, and that only 15 percent of participants consistently report the truth. False reports are conservative and appear to result from a biased perception of the BSR incentives. While attempts to debias are somewhat successful, the highest degree of truth telling occurs when information on quantitative incentives is withheld. Perversely the mechanism's incentives are shown to decrease truthful reporting.

Keywords: Incentive compatibility, belief elicitation, binarized scoring rule, experiments.

* All authors: Department of Economics, 230 Bouquet Street, Pittsburgh, PA. Danz: University of Pittsburgh and WZB, danz@pitt.edu. Vesterlund: University of Pittsburgh and NBER, vesterlund@pitt.edu. Wilson: alistair@pitt.edu. We are grateful to Tim Cason, Yoram Halevy, Paul J. Healy, Steffen Huck, Alex Imas, Dorothea Kübler, Ryan Oprea, Isabel Trevino, and seminar audiences at UCSB and WZB Berlin. We thank Felipe Araujo, Prottoy Akbar, Mallory Avery, Conor Brown, Ying Kai Huang, Matthew Raffensberger, Yuriy Podvysotskiy, and Tianyi Wang for help with the design and implementation of the experiment.

1. INTRODUCTION

Information on individual beliefs is central to our ability to draw inference on economic decisions (Manski, 2004). Absent data on what people think and expect we are often unable to discriminate between alternative models of choice, gauge the limits of rationality, or test new equilibrium concepts. While individual beliefs are of clear importance, eliciting them is not straightforward. To observe beliefs through revealed choice, incentive-compatible elicitation mechanisms take on a complex task: for each distinct belief there must be a corresponding choice in the mechanism that uniquely maximizes the agent's outcome; allowing the analyst to interpret the revealed choice as the agent's true belief.

Initial efforts to design incentive-compatible mechanisms centered on doing so for a narrow class of decision makers: risk-neutral expected-utility (EU) maximizers. However, conservative reporting and systematic deviations from truth telling initiated a search for mechanisms that were incentive compatible for a broader class of decision makers. Of concern was pull-to-center effects which suggested that risk aversion caused participants to report beliefs that were more conservative than those truly held.[1] The binarized quadratic scoring rule (Hossain and Okui, 2013, henceforth BSR) was uncovered as a particularly promising alternative, offering incentives that make truth telling incentive compatible for any decision maker who wishes to maximize the chance of winning a prize—nesting arbitrary EU preferences. Building on the insights of Roth and Malouf (1979) the flexibility is achieved by linking reported beliefs to a pair of state-contingent lotteries. For every specific belief, the mechanism provides a unique stochastically dominant lottery pair, whereby decision makers aiming to maximize the chance of winning are given strict incentivizes to reveal their true belief.

In addition to being theoretically incentive compatible for a wider set of preferences, initial empirical evidence showed that the BSR outperforms narrower mechanisms like the standard (non-binarized) quadratic scoring rule (Hossain and Okui, 2013; Harrison and Phillips, 2014), which in turn have been shown to outperform improper scoring rules (Nelson and Bessler, 1989; Palfrey and Wang, 2009; Schotter and Trevino, 2014) and unincentivized elicitations (Gächter and Renner, 2010; Wang, 2011; Trautmann and van de Kuilen, 2014). Better theoretical properties and evidence of superior performance quickly rendered the BSR the state-of-the-art in belief elicitation.[2] However, despite the

[1] For recent reviews see Schotter and Trevino (2014) and Schlag, Tremewan and van der Weele (2015).

[2] Recent applications of the BSR include studies on gender and coordination (Babcock et al., 2017), investment and portfolio choice (Hillenbrand and Schmelzer, 2017; Drerup et al., 2017), coordination (Masiliūnas, 2017), matching markets (Chen and He, 2017; Dargnies et al., 2019), biased information processing (Hossain and Okui, 2019; Erkal et al., 2019), cheap talk (Meloso et al., 2018), risk taking (Ahrens and Bosch-Rosa, 2019), information source choice (Charness, Oprea, and Yuksel, forthcoming), memory and uncertainty (Enke, Schwerter, and Zimmermann, 2020; Enke and Graeber, 2019), discrimination (Dianat, Echenique, and Yariv, 2018), correlated and motivated beliefs (Oprea and Yuksel, 2020; Cason, Sharma, and Vadovič, 2020).

many positives, problems persist. Disturbingly, the conservative reporting patterns that identified problems with risk-neutral quadratic scoring seem to also arise in BSR elicitations. For example, Babcock et al. (2017) use the BSR to elicit beliefs over four possible outcomes and while the qualitative comparative statics mirror those for behavior, the reported beliefs appear conservative.

This paper explores whether BSR incentives—in particular, subjects' knowledge of the quantitative incentives—encourage truth telling. Challenging for this pursuit is that we generally do not know participants' true beliefs. To identify the response to information on incentives we therefore focus on beliefs elicited over an induced prior probability. While we also elicit posterior probabilities, to evaluate whether incentives play a greater role when reports are more computationally demanding, we focus on the elicited priors where we know exactly what a well-incentivized participant should report, separate from their ability to perform Bayesian updating. Of interest is when and why participants deviate from reporting the induced prior; whether potential distortions are systematic and of a form that suggests that they arise from the incentives rather than from simple confusion; and finally, whether there are ways of modifying the BSR implementation to better encourage truth telling.

We address these questions by exploring the performance of the BSR in five different treatments using a between-subject design. An initial Information treatment provides transparent quantitative information on the incentives. Participants are informed that their chance of winning is maximized by truthful reporting, and they are provided a description of how the BSR mechanism is implemented, simple numerical information on the offered lotteries for all provisional responses, and end-of-period feedback given state realizations. Observed behavior in the Information treatment reveals a substantial and robust failure to report the objective prior. Further, we show greater false reporting for non-centered priors than for a centered prior of ½, and the misreported beliefs for non-centered priors tend to be conservative: a ‘pull-to-center’ effect.

Exploring what gives rise to the high false-report rate (and systematic patterns in false reports) we ask whether it results from confusion over the reporting task, or from the BSR incentives used in the elicitation. To address this question, we deploy two modifications to the Information treatment, one supplementing the quantitative incentive information, the other eliminating it. The first, a reduction-of-compound-lottery (RCL) treatment provides a further quantitative aid to participants, a calculator that helps them reduce the compound BSR lottery. While the RCL treatment reduces false reports and eliminates the pull-to-center effect, false reports continue to be observed, with greater misreporting for non-centered than centered priors.[3] In contrast, our No-Information treatment, which removes all quantitative information on the BSR incentives, secures a lower overall rate of false reports, and eliminates both the pull-to-center effect and differential false-reporting by the

[3] The RCL treatment points to divergence of the reported belief from the known prior, partially resulting from subjects’ misconception of the incentives (à la Cason and Plot, 2014).

location of the prior. The No-Information treatment demonstrates that participants do understand the task at hand, and that absent precise incentive information they report the induced priors at high rates. Indicative of false reports resulting from incentives is that excessive false reporting and pull-to-center effects arise only when participants are informed of the BSR's quantitative incentives.

To further test the effect, we conduct a Feedback treatment where quantitative incentive information is revealed gradually, through feedback on the earned lotteries at the end of each period. Confirming incentive information as driving false-reports we find that end-of-period feedback gradually erodes truth telling. While false reports start out at the same low-rate as our No-Information treatment, deviations from truth telling increase as the participants get more and more feedback on the quantitative incentives, eventually reaching the level of false reports in the Information treatment.

Finally, taking a pragmatic approach we explore a design that provides more-limited information on the incentives. Our Description treatment reveals that a high degree of truth telling is preserved when we present participants with a verbal, non-quantitative description of how the BSR mechanism assigns monetary rewards. While this treatment is somewhat successful, it still does not surpass the truth-telling rates observed in the No-Information treatment.

The paper is structured as follows. Section 2 presents our design and results from the Information treatment where participants get detailed information on the quantitative incentives. Section 3 contrasts the Information results with those of two treatments, one that removes all information about the incentives, and another that adds a calculator to help reduce the compound lottery inherent to the BSR. Section 4 focuses on gradual release of information on the quantitative incentives through end-of-period feedback, while Section 5 examines whether we can add limited information on the mechanism's incentives while maintaining the No-Information truth-telling rates. Section 6 shows that our main results extend to the elicited posterior beliefs, and Section 7 concludes with a discussion of what likely drives the deviation from truth telling under the BSR and the implications of our findings for belief elicitation more generally.

2. Baseline Design and Results

The core elicitation environment is held constant across our five experimental treatments. A total of 299 undergraduate students were recruited to participate in an individual decision-making task at the Pittsburgh Experimental Economics Laboratory (PEEL).[4] Each treatment consisted of three separate sessions (with a recruitment aim of 20 participants per session). The procedures of the experiment, the number of periods, the elicited belief scenarios, as well as the offered incentives are all held constant across

[4] In addition, subjects had to be 18+ years old to be eligible for participation. Invitations to all sessions were gender balanced.

treatments. In terms of exposition, we describe the common features of the experimental environment as we introduce the Information treatment.

2.1. Information Treatment

The Information treatment is designed as a baseline: an implementation of the BSR that presents participants with clear information on the quantitative incentives associated with any provisionally considered report on the belief. An Information treatment session (like all the treatments we examine) consists of ten periods, where each period has three sequential decisions. Participants are paid for one of the three decisions in two of the ten periods.[5]

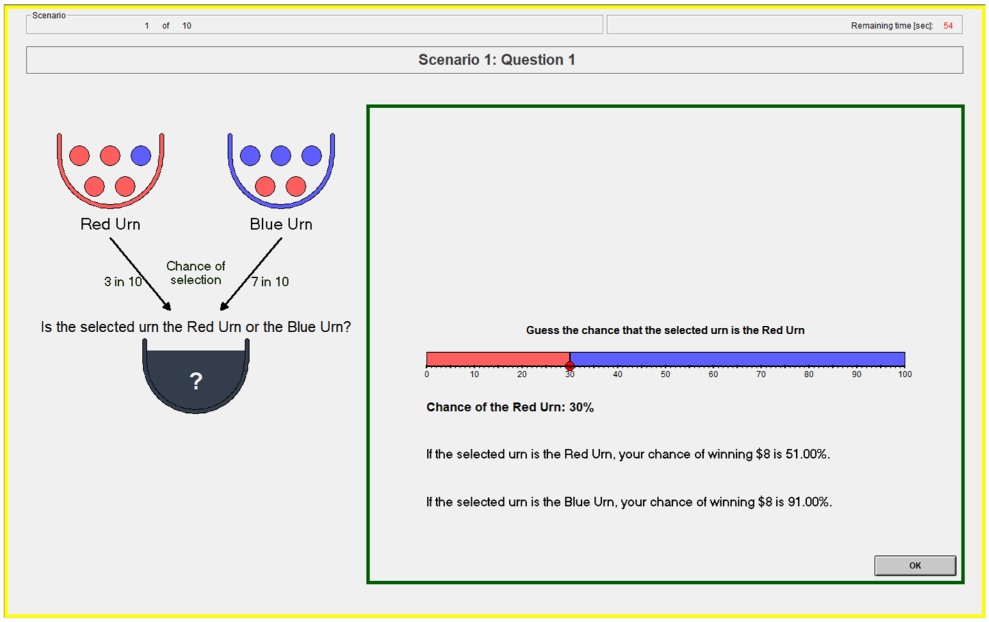
Within the session participants make choices in the interfaces shown in Figure 1. Panel A shows the decision screen and panel B the end-of-period feedback. At the beginning of each period participants are shown two urns, one Red and one Blue. Each urn contains five colored balls (either red or blue) where the Red Urn contains more red balls than the Blue Urn. One of the two urns is selected at random, where the main task for participants is to guess the likelihood that the selected urn is Red. Participants are informed of the composition of both urns and of the prior probability π_0 that the Red Urn is selected, presented as an X-in-10 chance (see panel A). Given this information, participants are asked to submit three sequential guesses on the chance that the selected urn is Red. Guess 1 is made without any additional signals, and Guesses 2 and 3 are made, respectively, after observing the colors of a first and then a second independent draw from the selected urn.[6]

The decision screen in Figure 1A shows a provisionally marked 30 percent guess on the Red Urn, secured by placing a cursor on a slider ranging between 0 and 100 percent (with one percentage-point increments). Each provisionally marked belief leads to an offered state-contingent lottery pair displayed on the screen: one if the selected urn is the Red Urn, another if the selected urn is the Blue Urn. Both lotteries are over a prize of \$8 if won, and \$0 otherwise. In particular, the BSR incentive given a stated probability of q on the Red Urn offers a $1\text{-}(1\text{-}q)^2$ chance of winning the prize if the Red Urn was selected, and $1\text{-}q^2$ if the Blue Urn was selected. Thus, the chance of winning is maximized by reporting the likelihood Red is selected, which for Guess 1 corresponds to the induced prior and for

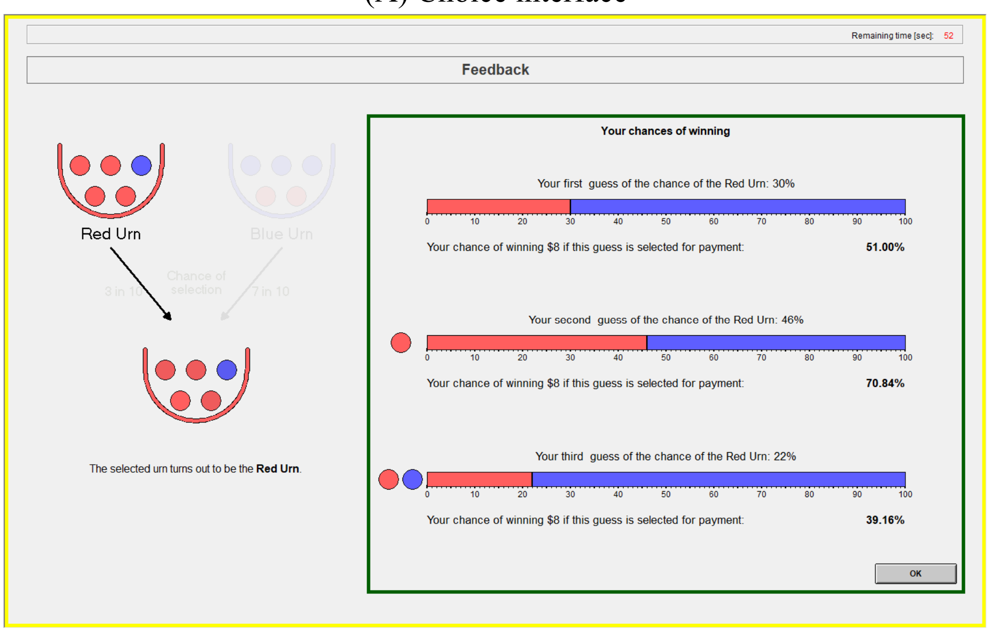
[5] The experimental interface is programmed in z-Tree (Fischbacher, 2007). Participants received printed instructions that were read out loud and summarized in a short-scripted presentation at the start of each session (see Appendix B for instructions for the Information Treatment, with exact language-deltas for all other treatments; cf. reporting best practices articulated in De Quidt et al. 2019; and Appendix C for the slides and script used in the summary presentation). Across all treatments the average duration of a session was 71 minutes with average earnings of \$20.08, including an \$8 show-up fee.

[6] Our task resembles the standard Bayesian updating task with the addition of the prior elicitation, see Benjamin (2019) for a recent review of the literature on belief updating.

Guesses 2 and 3 the Bayesian posteriors, updated in response to the draws from the selected urn.[7]



(A) Choice interface



(B) End-of-period feedback

FIGURE 1. INTERFACE SCREENSHOTS

[7] The evaluated scenarios and random realizations are held constant across treatments. Within each session all 20 participants see the same sequence of 10 scenarios, though in different random orders. While the signal realizations vary across participants and sessions of a treatment, the sequencing of scenarios considered (state, signal realizations, and sequence) is held matched across treatments.

Later in the paper we will introduce (sequentially) four additional treatments that vary the information on the quantitative incentives provided to participants. However, we begin by first outlining our results in the baseline Information treatment, which provides clear incentive information through four channels: (i) The instructions explicitly provide the qualitative information that truthful reporting is a dominant strategy (a common feature to the presentation in all of our treatments) stating that *"[t]he payment rule is designed so that you can secure the largest chance of winning the prize by reporting your most-accurate guess."* This statement was also emphasized in a slide presentation summarizing the instructions where it is the last thing participants see before making their first decision (see final slide of Appendix C). (ii) The written instructions provide a concise verbal description of how the state-contingent lotteries determine the prize realization.[8] (iii) Within the interface, as participants move their provisional belief, the screen is updated to provide clear quantitative information on the state-contingent probabilities of winning. This can be seen in Figure 1A in the two lines below the input slider. With $q = 0.3$ selected, the interface displays the associated chances of winning the prize for each realization of the selected urn; in this case 51 percent if Red, and 91 percent if Blue. (iv) Finally, as shown in Figure 1B, participants receive feedback information on the selected urn at the end of each period, as well as the realized quantitative chance of winning the \$8 prize given the state realization (the selected urn) and their submitted Guesses 1–3.

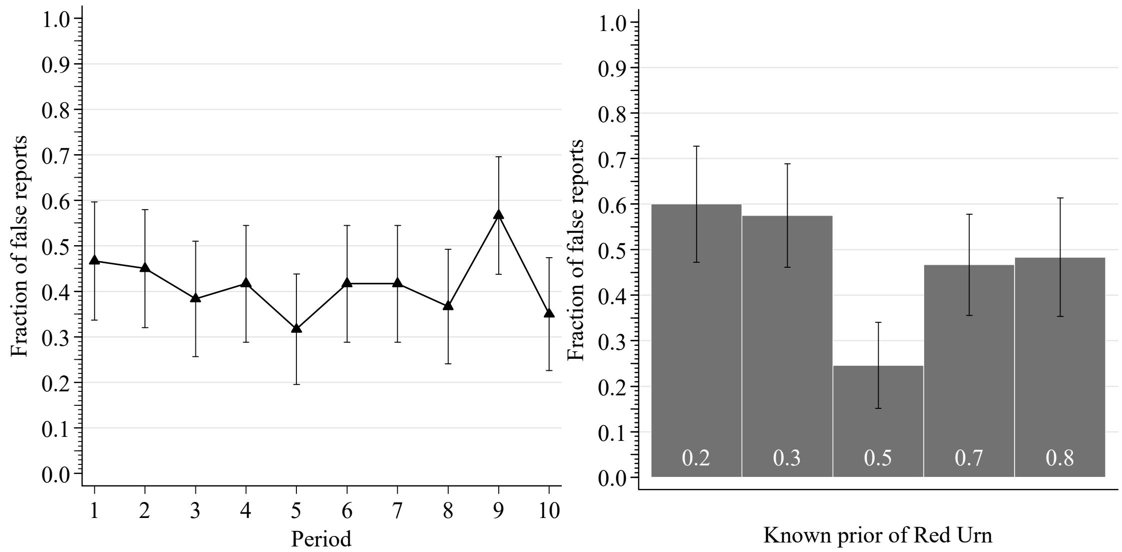
After the ten periods (30 elicitations total) we elicit risk attitudes (encoded as switch points on pricelists) and ask participants to respond to a Cognitive Reflection Test (Frederick, 2005).[9] One participant per session is randomly selected to be paid for these end-of-experiment elicitations. Finally, participants complete a post-experiment questionnaire on demographics, and provide a self-assessment of their comprehension of the incentives and the extent to which they reported their most-accurate guess.

2.2. Information Treatment Results

In examining whether BSR incentives secure truthful reporting we focus our analyses on Guess 1, the elicitation of the induced prior. To evaluate if information on incentives plays a greater role when eliciting the computationally demanding posterior beliefs, Section 6 presents a complementary analysis for Guesses 2 and 3, showing that the same qualitative results hold. However, the advantage of focusing on the prior elicitations is that it removes confounds with regards to the ability to Bayesian update. The induced prior is

[8] The explanation of how the chance of winning was determined in the state-contingent lotteries relied on a comparison of the reported guess to that of two (uniform) random numbers, thereby avoiding the presentation of formulas, or the understanding of a squared error (see Wilson and Vespa, 2018).

[9] In each row of the first table, participants chose between a sure payoff of \$4 and a lottery $p \cdot \$8 \oplus (1-p) \cdot \$0$ with p increasing in each row from 0 to 1 in steps of 0.1 (see Bruner, 2009). In the second table the lottery was the same in all rows with $\frac{1}{2} \cdot \$8 \oplus \frac{1}{2} \cdot \$0$ and the sure payoff increased over the rows from \$0 to \$8 in steps of \$0.80 (see Abdellaoui et al., 2011). Tables 3 and 4 were the same as 1 and 2, respectively, except that all prize payoffs were scaled by a factor of 1.25.



(A) By Period (B) By Prior

FIGURE 2. FALSE-REPORT RATE IN INFORMATION TREATMENT.

unambiguous and should be reported back by every participant who understands the offered incentives and seeks to maximize their chance of winning.[10]

Figure 2 illustrates the rate of false reports (any elicited belief q that differs from the induced prior π_0) by period across the session, and by the objective prior. Panel A shows a substantial rate of false reporting over the ten periods, averaging 41.5 percent, which is maintained without a time trend across the session (p=0.842).[11] False reports are widespread, with 85 percent of participants failing to report the induced prior in one or more of the ten periods. These deviations from truth telling are surprising considering the incentive compatibility for a general family of underlying preferences, the qualitative statement that truth telling will maximize participants' chances of winning, and the prior evidence on the comparative superiority of the BSR. Panel B illustrates how the rate of false reports varies with the location of the induced prior. For non-centered priors ($\pi_0 \in \{0.2, 0.3, 0.7, 0.8\}$) we find that false reports are the norm (52.8 percent), while they are significantly less likely to occur for the exact-centered prior (π_0 =0.5, with a 24.6 percent false report rate, p<0.001 from a participant-clustered t-test of difference).

Adding to the high frequency of false reports is the magnitude of the deviations. Conditional on a false report the average deviation from the prior is 0.167, and our results

[10] While Offerman et al (2009) use the elicitation of induced priors for *ex post* corrections, we use it to assess the elicitation procedure itself, see also Hao and Houser (2012) and Holt and Smith (2016). The simplistic elicitation eliminates belief formation and may help participants focus on the incentives provided (see for example, Avoyan and Schotter, 2020 on shared attention).

[11] Tests of time trends are based on probit regressions of false reports on period with participant-clustered standard errors.

are the same when we exclude small mistakes from consideration.[12] Further, false reports tend to be conservative with a pull-to-center effect. Among the false reports for non-centered priors we find that 53.7 percent lie between the objective prior and the exact center (a stated report of q=½), while only 32.6 percent fall between the objective prior and the nearest extreme (with the remaining 13.7 percent of misreports being somewhere between the exact center and the distant extreme). Considering the similarly-sized widths for defining each false-report type, the greater movement toward the center than the nearest extreme (p=0.058) leads to the conclusion that there is a significant pull-to-center effect. In contrast to belief-elicitation mechanisms such as the QSR, where the 'pull-to-center' effect was interpreted as resulting from risk aversion, the effect here is unexpected as the BSR is incentive compatible for arbitrary risk preferences. Indeed, we find no evidence that risk aversion is the culprit: Individual risk attitudes do not predict the propensity to deviate from the true prior, nor the conservative reporting tendency.[13]

Why are so many participants misreporting the prior? Is it the task at hand that participants fail to understand or that they simply object to reporting the given prior? Alternatively, are the offered incentives causing participants to distort their reports?

Table 1 provides two views into the incentives. In the first panel we indicate the state-contingent lottery pairs offered under the BSR for reports from 1 to 0.5 in 10 percentage-point increments; in the second panel we outline the expected earnings from a series of reporting strategies.

Table 1A indicates how a failure to reduce the offered compound lotteries might cause distorted reporting. Consider a participant with an induced prior of 0.8. Truthful reporting leads to a 96 percent chance of winning the $8 prize if the selected urn is Red, and a 36 percent chance if Blue. By (mis)reporting a more-conservative belief of 0.7, the chance of winning decreases by 5 percentage points conditional on Red and increases by 15 percentage points conditional on Blue. While the large increase in the chance of winning on the less likely event ('Urn is Blue' in Table 1) will not provide enough of an offset for an EU-decision maker who perfectly reduces the lottery through the true likelihood of each state (here 0.8/0.2), this need not be the case for decision makers who struggle with compound lottery reduction, who misperceive the incentives, or who have non-EU preferences.[14]

[12] The results are the same when shifting our binary definition of a false report to allow for small errors that deviate by no more than 5 percentage points from the prior (Appendix Table 4). For example, this less strict definition of false reports give rise to a false-reports rate of 40.6 percent for non-centered priors and 17.1 percent for centered priors (different with p<0.001). See Cason and Plot (2014) for a similar approach when assessing the BDM mechanism.

[13] Individual false-report rates and the extent to which these move toward the center are not significantly correlated with an individual being risk averse or loving (identified by whether willingness to pay for a 50 percent chance of winning $8 is below or above the certainty equivalent of $4). Tobits on the number of false reports made by each participant show no significant effect of being risk averse or risk loving (p=0.681 and p=0.883).

[14] Deviations can be interpreted as resulting from preferences, from calculation mistakes/ misinterpretations of the probabilities, from failures in critical thinking, from a misunderstanding of the task,

TABLE 1: BSR INCENTIVES

(A). OFFERED LOTTERIES

Submitted Belief on Red	Chance of receiving \$8 if: Urn is Red	Urn is Blue
1.0	100%	0%
0.9	99%	19%
0.8	96%	36%
0.7	91%	51%
0.6	84%	64%
0.5	75%	75%

(B). EXPECTED EARNINGS

Reporting rule	Earnings
Truthful ($q = \pi$)	\$6.27
Middle ($q = 1/2$)	\$6.00
Random ($q \sim U[0,1]$)	\$5.33
Minimizing	\$2.88
Information (data)	\$6.11

Moreover, by construction, the asymmetry in the magnitude of the gain/loss in the chance of winning on the less/more likely event intensifies with the certainty in the belief. For example, consider a participant reporting 0.89 instead of a true belief of 0.99 on Red. The deviation increases the chance of winning conditional on Blue from 2 to 21 percent, but it decreases the chance of winning conditional on Red from 99.99 to 98.79 percent. Participants who are lured by this asymmetry in the change in probabilities are expected to have greater rates of false reports on non-centered than centered priors, and to have their reports on non-centered priors pull to the center. Both reporting patterns observed in the Information treatment.

While misunderstanding the offered lotteries is a channel for mistakes, an exacerbating factor is that the penalties to such misperceptions are small. The cost of reporting a prior other than that given is a 1 percentage point decrease in the chance of winning the \$8 prize for a 10 percentage-point deviation in the report (and a 3 percentage-point reduction for a 20 percentage-point deviation). For an induced prior of 0.8 a truthful report generates a 91 percent overall chance of winning the \$8, so the chance of winning drops to 90 percent with a false report of 0.7, and to 88 percent with a false report of 0.6. Table 1B summarizes the consequences of the flat incentives by reporting the expected payment if a random prior elicitation were selected for payment. A strategy of truthfully

an unwillingness to undertake the costly computation or a combination. Our after-experiment survey asked participants how they made their decisions, and the responses provide anecdotal evidence (all responses verbatim) that they purposefully distort reports to secure a higher chance of winning on the less likely event, and were aware of the incentives for doing so: *"I generally erred on the side of caution when picking the urns. For example, if x=5, I would select 50% for the red urn. If sat x=8 then I would pick the red a little more opportunistically."*; *"I kept my initial answers at 50% because you get a 75% chance of getting the \$8 anyways. Then I adjusted as I saw the different outcomes."*; *"at first, I guessed based on probability probability the urn was picked based on the dice roll and then considered the balls that were drawn from the bag; however, I quickly realized that since I am pretty risk adverse, sticking to a 50-50 chance would result in being paid the \$8 75% of the time regardless of which urn was selected. I mostly stuck to that model as I proceeded through the experiment. When i felt daring, I would move my guesses a little bit around the 50-50 mark (but never very far)."*; *"I believe that leaving each chance at 75% was my best chance of making the most money in the experiment."*

reporting across all of the induced priors yields an expected payment of \$6.27. The table then provides the expected payoff from three alternative strategies: a \$6 expected payment from reporting the centered prior of $q = 1/2$ regardless of the induced prior; \$5.33 when choosing the report entirely at random (in particular, uniformly); and \$2.88 when choosing the furthest possible belief from the induced prior. Finally within Table 1B, as additional evidence of the limited costs to deviating, we note that despite the frequent failure to report the induced prior, participants in the Information treatment secured \$6.11 on average from each prior elicitation, missing out on only sixteen cents of the maximum truth-telling earnings amount of \$6.27.

In examining the drivers of the large degree of false-reporting in our Information treatment we consider three potential channels: (i) false reports inherent to the task of reporting the prior, for example, resulting from confusion about what is being asked; (ii) false reports driven by a failure to reduce the compound lottery inherent to the BSR incentive; (iii) false reports driven by other features of the BSR incentives, for example, through a non-EU preference for a greater chance of winning on the less likely event combined with the flat incentives for truth telling. While the first channel captures confusion inherent to our belief elicitation, the latter two are related directly to the BSR incentive structure.

2.3. RCL & No-Information Design

We use two additional treatments to provide insights on the source of false reporting. Both manipulate the participants' information on the incentives. In the first, we provide *additional* information specifically tailored to limit misunderstanding of the compound lottery, providing a calculator that reduces it to a simple lottery (the RCL treatment). In the second, we remove all information on the quantitative incentives (the No-Information treatment).

Our Reduction of Compound Lotteries (RCL) calculator provides participants with a tool to compute the total chance of winning the \$8 prize for any report (see Figure 3 for the interface). The RCL calculator asks participants to enter their true belief and determines the overall chance of winning for each potential report. The calculator therefore helps participants verify that truth telling maximizes their chance of winning by reducing the offered lottery pair for any true and stated belief.[15] Beyond the addition of the RCL

[15] Like the more-involved explanation of the BDM in Healy (2017, 2018) the hope was that this would help participants understand that the mechanism was incentive compatible. While more substantial explanations and training may have enhanced participants' comprehension of the mechanism's incentive compatibility, because belief elicitations are typically secondary measures in experiments, we opted for an aid that would not substantially increase the length of the instructions.

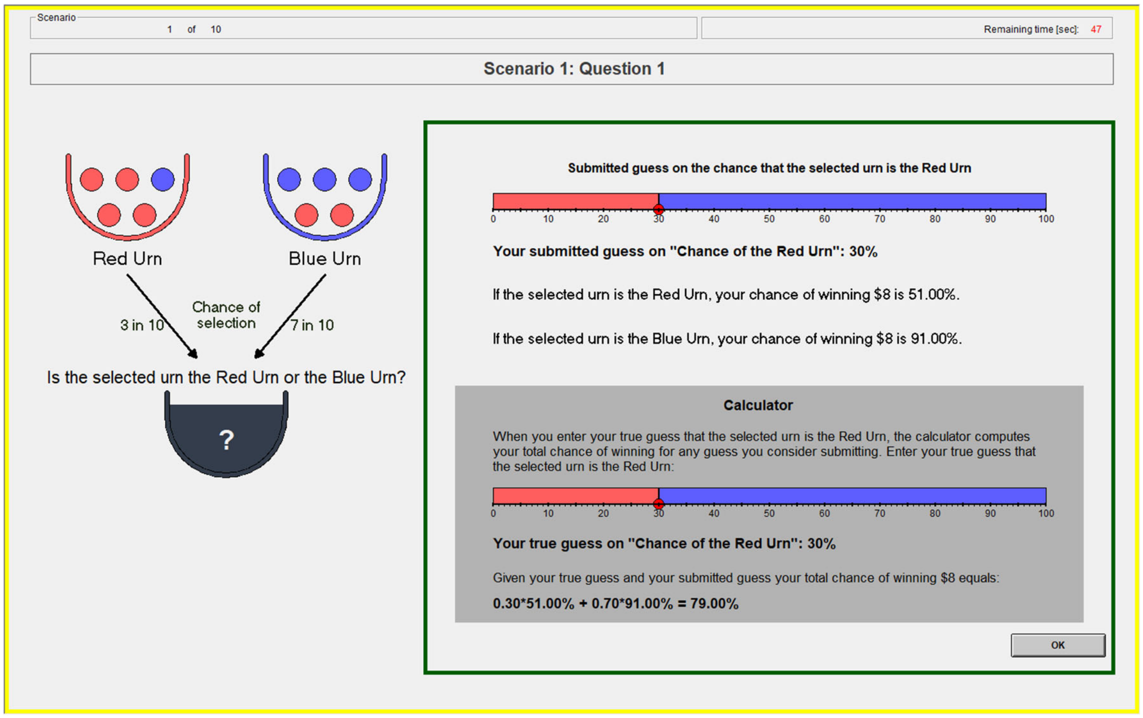


FIGURE 3. RCL TREATMENT SCREENSHOT

calculator (and supplementary instructions on how to use it) the treatment is otherwise identical to the Information treatment.

In contrast, while our No-Information design holds constant all qualitative details from the Information treatment it removes all quantitative information on the incentives. Participants are still told that the procedure was designed so that truthful reporting will maximize their chances of winning and that $8 is at stake as a prize, but they are uninformed on the chances of winning the prize.[16] In addition to removing the description of the mechanism in the instructions, the No-Information interface also removes the numerical information on the state-contingent lotteries at each provisionally selected belief (the figures in Figure 1A below the input bar) and the end-of-period feedback on the earned chance of winning for each Guess (the three ex post probabilities in Figure 1B).[17]

In terms of identification, the RCL treatment offers a channel to assess the extent to which an inability to reduce compound lotteries is driving the Information results. Given the lack of *any* incentive information in No Information, the level of false reports in this treatment serves to identify factors other than the incentives (for example, confusion with

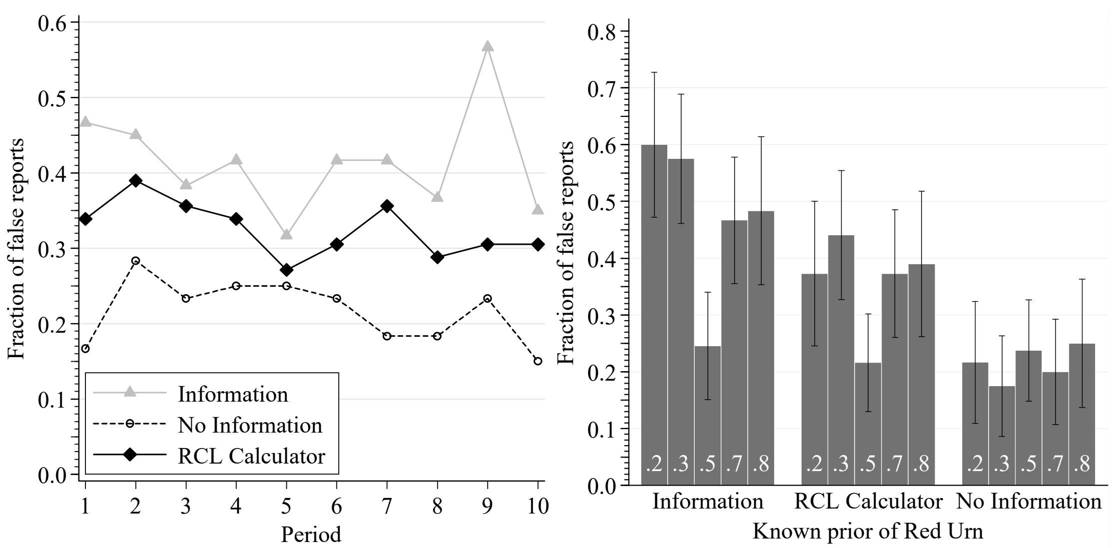
[16] Participants are informed that *"[t]he precise payment rule details are available by request at the end of the experiment."* Of the 60 subjects in the treatment, only one requested this information.

[17] The end-of-period feedback screen in No-Information instead provides feedback only on the realized selected urn.

the task). However, by removing all quantitative incentive information, any difference in false reports with the Information treatment is identified as coming from some feature of the BSR incentives (whether it be failure to reduce the inherent compound lottery, or some other feature). The relative differences between the three treatments can therefore help to decompose the effects on false reporting.

2.4. RCL & No-Information Results

Three sessions were run for each treatment, with 60 participants in No Information and 59 in RCL (one RCL session under-recruited and ran with 19 participants). Paralleling our data presentation for the Information treatment, Figure 4 reports the false-report rate by session period (panel A) and by the objective prior (panel B).[18] We note that while the RCL treatment reduces the frequency of false reports, the reduction is even greater in the No-Information treatment.[19] We also see that the pattern of greater false reporting for non-centered than centered priors is reduced but not eliminated in the RCL treatment, while it disappears entirely in No-Information.[20] Thus, although an improved ability to reduce



(A) By Period (B) By Prior

FIGURE 4. FALSE-REPORT RATE IN NO-INFORMATION AND RCL TREATMENTS

[18] While there are significant differences in the average absolute error across treatments, the difference is driven by the rate of false reports, hence our focus on this measure. Conditional on a false report there are no significant differences in the magnitude of the deviation from truth telling, with treatment-average deviations for false reports ranging between 0.150 and 0.183.

[19] For comparison we note that Holt and Smith (2016) use a similar two-urn-guessing paradigm to elicit beliefs under the QSR, the BDM, and the lottery choice method. Evaluating only one prior of 0.5 they find false report rates on the prior elicitation ranging between 20% to 33%.

[20] Further the patterns in the results continue to hold when eliminating small mistakes within 5 percent of the induced prior (see Appendix Table 4).

compound lotteries decreases false reports, the best results are obtained by eliminating quantitative information on the BSR incentives.

Table 2 shows the false-report rates with participant-clustered standard errors, where the first three rows confirm that the treatment effects illustrated in Figure 4 are significant.[21] The first data column in Table 2 indicates treatment-level false-report rates across all prior elicitations (pooling centered and non-centered priors).[22] The two *By-Prior* columns separate the false-report assessment into two subcategories: those occurring when the induced prior is centered ($\pi_0 = 0.5$), and those when it is non-centered ($\pi_0 \neq 0.5$). Finally, the last three columns decompose the false reports for non-centered priors into three regions, thereby assessing the extent to which agents' reported beliefs are skewed toward the center. We examine the proportion of non-centered priors for which a false report: (i) moves toward the center (false reports of $q \in (\pi_0, \frac{1}{2}]$ when $\pi_0 < \frac{1}{2}$, and of $q \in [\frac{1}{2}, \pi_0)$ when $\pi_0 > \frac{1}{2}$, respectively); (ii) moves to the nearest extreme (false reports of $q \in [0, \pi_0)$ and $q \in (\pi_0, 1\,]$, respectively); and (iii) moves between the exact center and the distant extreme (false reports of $q \in (\frac{1}{2}, 1]$ and $q \in [0, \frac{1}{2})$, respectively). Partitioning the non-centered false reports in this manner secures that deviations toward the center and the near extreme have the same widths, allowing us to fairly assess the extent to which participants deviate toward the center (as opposed to the near extreme). The results for the baseline Information treatment (the first data row) mirror Figure 2: more than 40 percent of the submitted beliefs misreport the prior; where the false-report rate is significantly greater for non-centered than centered priors (p<0.001); and that false reports on non-centered priors are more likely to be pulled toward the center than the nearest extreme (p=0.058).

[21] Unless otherwise stated, all treatment comparison p-values are obtained from two-sided t-tests derived from the difference in treatment-level false reports rates (with standard errors clustered over the 299 participants). As all conditioning variables are indicators that fully partition the data, the results are equivalent to comparisons of regression coefficients obtained from a (in this case, well-specified) LPM model. Probit estimates indicate almost identical quantitative marginal effects and qualitative inference, so we focus here on the easier to interpret measures.

[22] Table 2 also reports the average treatment levels from two additional treatments (details below).

TABLE 1. FALSE REPORTS AND TYPE BY TREATMENT

Treatment	False Reports			False-Report Type ($\pi_0 \neq 0.5$)		
	All Priors	By Prior				
		$\pi_0 = 0.5$	$\pi_0 \neq 0.5$	Center	Near Extreme	Distant Extreme
Information	0.415 (0.042)	0.246 (0.047)	0.528 (0.048)	0.283 (0.042)	0.172 (0.032)	0.072 (0.014)
RCL	0.325 (0.041)	0.216 (0.043)	**0.398** (0.048)	**0.169** (0.033)	0.164 (0.031)	0.065 (0.014)
No Information	**0.217** (0.039)	0.238 (0.040)	**0.203** (0.041)	**0.058** (0.018)	0.108 (0.027)	0.036 (0.014)
Feedback $_{(t=1,2)}$	**0.217** (0.045)	0.236 (0.060)	**0.200** (0.055)	**0.031** (0.021)	0.154 (0.047)	**0.015** (0.015)
Feedback $_{(t=9,10)}$	0.341 (0.053)	0.255 (0.064)	0.405 (0.071)	**0.087** (0.039)	0.275 (0.062)	0.043 (0.024)
Description	**0.245** (0.040)	0.196 (0.039)	**0.278** (0.046)	**0.108** (0.027)	0.131 (0.029)	0.038 (0.012)
N	2,630	2,630		1,568		

Note: Standard errors in parentheses clustered by participant (299 clusters) recovered from three separate joint estimates on the false report proportion in the prior elicitations: (i) *All priors*, dependent variable an indicator for $q \neq \pi_0$, with treatment level estimation; (ii) *By Prior* column pair, same dependent variable as All priors, but with separate treatment estimates for centered/non-centered prior location; and (iii) *False-Report type* column triple, treatment-level estimation for the division of non-centered false reports into three mutually exclusive regions: center (between the π_0 and ½), near extreme (between the closer extreme 0/1 and π_0), and distant extreme (the further of 0/1 and ½). Bold face coefficients are significantly different from the relevant Information coefficient with $p < 0.1$ (two-sided test).

Relative to Information we note that the RCL calculator does reduce the rate of false reports. Across all elicited priors the RCL treatment leads to a 9.0 percentage point reduction in false reports. While this reduction is not significantly different from zero (p=0.131), it is when we focus solely on the non-centered priors (a 12.9 percentage point reduction, p=0.057).[23] Despite the reduction under RCL, we still find that false reports are more likely for non-centered than centered priors (an 18.2 percentage point difference, p<0.001). However, the false reports on the non-centered priors no longer pull towards the center: the RCL treatment's 39.8 percent false-report rate for non-centered elicitations is more evenly distributed between those that move toward the center, and those that move to the nearest extreme (p=0.903). Specifically, while deviations made toward the nearest extreme in RCL continue to occur at a similar rate to Information (16.4 vs 17.2 percent, p=0.851), we find a significant reduction in the false reports moving toward the center (16.9 vs. 28.3 percent, p=0.034).

[23] While the working hypothesis motivating the RCL treatment was one-sided—that reducing the lottery will help participants understand the incentive compatibility—we report two-sided tests for consistency.

While helping participants reduce compound lotteries decreases false reporting and removes the pull-to-center effect, even greater truth telling is obtained through the elimination of quantitative incentive information in No Information. The proportion of false reports when participants have no quantitative information on the BSR incentives is significantly lower than both the Information and RCL treatments ($p<0.001$ and $p=0.056$, respectively).[24] Further false-report rates are no greater on non-centered than centered priors ($p=0.317$) and there is no evidence of false reports on non-centered priors pulling to the center ($p=0.175$ for a two-sided test, where the difference has the opposite sign).[25]

With No Information substantially reducing the rate of false reports and eliminating the pattern of greater false reporting on non-centered priors, we infer that both effects are causally linked to knowledge of the quantitative BSR incentives.[26] The No-Information treatment therefore demonstrates that false reporting in the Information treatment does not simply arise from the task. Participants report the objective prior at high rates, independent of its location, provided they are *uninformed* of the quantitative incentives from doing so.[27]

Comparing the three treatments we get a sense of what drives participants to falsely report non-centered priors in the Information treatment. While 38 percent of these false reports can be attributed to the task itself (for example, confusion), the remaining 62 percent are directly linked to the BSR incentives, whether it be through an inability to reduce compound lotteries or another feature of the incentives offered—accounting for

[24] The rate of false reports across the No-Information session decreases by 4 percentage points between the last and first two periods of the treatment, though the effect is insignificant ($p=0.448$).

[25] Intriguingly, the reduction in false reports is only seen for reports that pull-to-center (28.3 in Information vs. 5.8 percent in No-Information, $p<0.001$) and not in those toward the nearest extreme (17.2 vs. 10.8 percent, $p=0.160$).

[26] The difference in the rate of false reports across centered/non-centered priors in RCL is highly significant ($p<0.001$), where the frequency of pull-to-center reports is significantly larger than in No-Information ($p=0.004$). Curiously, RCL is the only treatment where false reports and pull-to-center correlates with CRT scores and attitudes toward risk. A Tobit on the number of false reports shows that risk-averse individuals are more likely to falsely report ($p=0.023$) but not to deviate toward the center ($p=0.209$); risk seeking individuals are no more likely to falsely report ($p=0.620$) but are more likely to pull-to-center ($p=0.098$). High CRT scores are predictive of fewer mistakes (0.003) and less pull-to-center ($p=0.078$). Risk attitudes and CRT scores are not significantly predictive of false reports in any other treatment (nor systematic in the assessed directions).

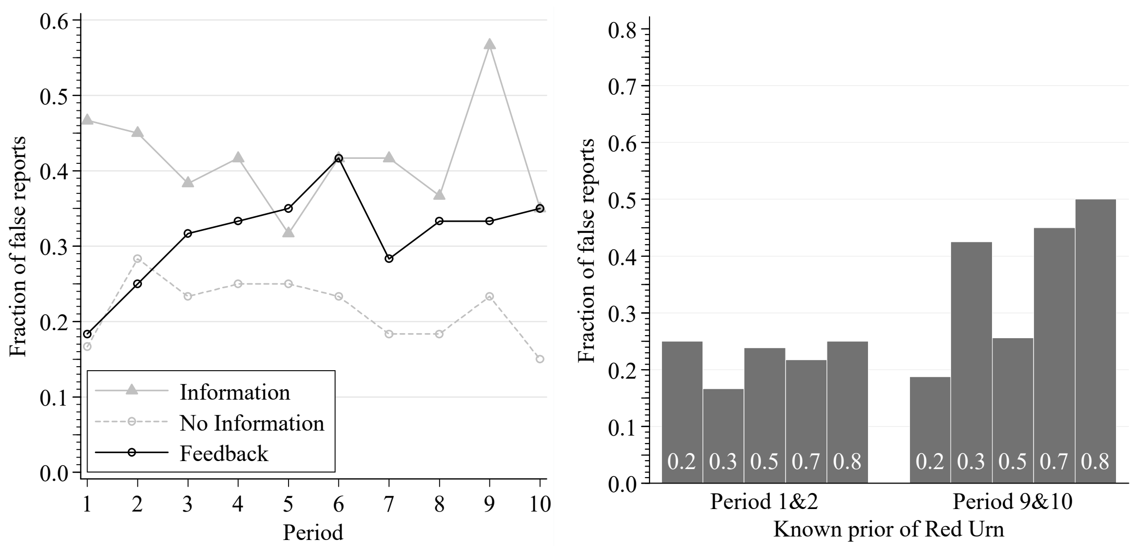
[27] The data in our post-experimental questionnaire further bolsters the case that it is the incentives that drives the false reports. Participants are asked to rate their agreement with "I always reported my most-accurate guess on the Red urn being the selected urn." Responses were collected on a 5-point Likert scale. Looking at the fraction of answers in the Strongly Agree/Agree categories, we find, 70 percent of respondents claiming they always reported their most-accurate guess in the Information treatment, and 85 percent in No-Information ($p=0.049$, χ^2-test of independence; see Figure 8 in the Appendix for further details). Self-assessment of truthful reports is (insignificantly) higher in RCL than Information (81 vs. 71 percent, $p=0.149$). Further, while there are no differences in comprehension of mechanism between RCL and Information, participants in No-Information are less likely to report that they understood how their pay would be calculated (72 percent) and how the submitted belief affected their pay (70 percent) than participants in the Information and RCL treatments (80 and 86 percent on pay, and 83 and 86 percent on beliefs, respectively).

approximately 25 and 37 percent, respectively, of the Information treatment's false reporting for non-centered priors.

3. FEEDBACK TREATMENT

The evidence thus far suggests that knowledge of the quantitative BSR incentives causes an increase in the frequency of false reports when the induced prior is non-centered. To further explore and identify the effect as coming from information on the quantitative incentives we conducted a Feedback treatment with 60 further participants where incentive information is gradually revealed through the end-of-period feedback screen in Figure 1B. That is, in the Feedback treatment we replicate the No-Information instructions and main decision screen, but after each period's elicitations we provide participants with the Information treatment's end-of-period feedback. The provided feedback informs the participants on the earned probability of winning (fixing the realized state) as a simple lottery. The quantitative incentive information provided in the Feedback treatment is therefore limited to the reported beliefs and realized state and is acquired slowly as the session progresses.

Figure 5A indicates the false-report rate by period across the Feedback sessions. While false reports start out at the same rate as No-Information, over time the fraction of false reports increases, eventually reaching a level that is indistinguishable from that of the Information treatment.



(A) By Period (B) By Prior

FIGURE 5. FALSE-REPORT RATE IN FEEDBACK TREATMENT

Referring to Table 2 for comparisons and inference, we find a false-report rate of 21.7 percent for the first two periods (the Feedback$_{(t=1,2)}$ row) which grows significantly

(p=0.003) to 34.2 percent for the final two periods (the Feedback$_{(t=9,10)}$ row). Thus, feedback on the quantitative incentives increases the frequency of false reports over the session, where the starting and ending points provide a strong match to the No-Information and Information treatments, respectively. In the first column of Table 2 we find that the false-report rate in the first two periods of Feedback is statistically inseparable from the overall No-Information rate (p=1.000) but significantly different from Information (p=0.001). Conversely, the final false-report rate in Feedback's last two periods is significantly different from the No-Information treatment (p=0.060) but inseparable from Information (p=0.282).[28]

Though only provided with three *ex post* measures per period on the quantitative incentive (one for each submitted belief) the fraction of false reports in Feedback reaches the Information-treatment level within four periods. Despite a loss of power with our focus on the first and last two periods, Figure 5B suggests that participants also begin to respond differentially to non-centered and centered priors. Comparing false reports in the last two Feedback periods we find a rate of 25.4 percent for the exact-center priors, and of 40.6 percent for the non-centered priors (p=0.089 from a two-sided test for differences).[29] While our comparison of the No-Information and Information treatments reveals a between-subject increase in false reporting when given quantitative-incentive information, the Feedback treatment illustrates a similar effect within subject. As participants slowly learn about the BSR incentives through the period feedback, the rate of false reports increases significantly, where the effect is most pronounced for the non-centered priors.[30]

4. DESCRIPTION TREATMENT

Considering the above, a dilemma emerges for belief elicitation. On one hand we want participants to be fully apprised of the offered incentives, as knowing them is a necessary condition for incentivizing truthful reporting. On the other, the point of the mechanism is to measure beliefs with minimal noise, and quantitative-incentive information is substantially distorting the reported beliefs. Can we inform participants partially on their incentives without distorting reports? One option is to simply rely on the (truthful) qualitative statement that the mechanism is incentive compatible. The data from our No-Information treatment certainly suggests that this is the better option in terms of the accuracy of collected belief data. However, advocating for what amounts to a black box

[28] Notably, there is no significant time trend at the 10 percent level in any treatment except for the Feedback. Responses to the exit survey indicate that participants learn the incentives over time (for example, "*I kept my initial answers at 50% because you get a 75% chance of getting the $8 anyways. Then I adjusted as I saw the different outcomes*").

[29] Using the last five periods of data instead of the last two, the difference between non-centered and centered priors (with participant-clustered errors) is significantly different with p=0.005.

[30] Similar to the RCL treatment, we do not detect differential pull-to-center effects at any point in the Feedback treatment.

from the point of participants is jarring to the general philosophy of incentivized decision making.

An intermediate option is to add a description of the mechanism's implementation rule without providing the more-precise quantitative details provided in our less-successful treatments. The point of the description would be to make the quantitative incentives ostensibly calculable by subjects, and to provide a skeleton structure to participants on the rule for how their earnings will be calculated. This approach is frequently used in other mechanisms; for example, consider the non-technical description of how a second-price bidding rule works (and equivalently, how many strategy methods like the BDM function), or how a complicated matching algorithm like top-trading cycles would be described to parents providing school-choice rankings.[31]

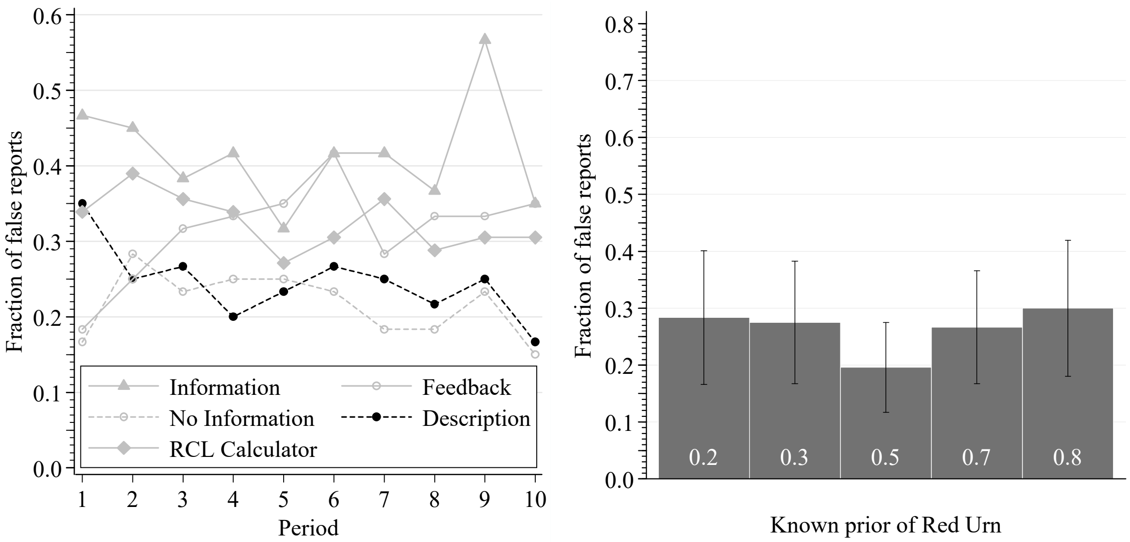
Our Description treatment pursues this approach with a further 60 unique participants. Participants in the treatment are provided with the same statement on the dominance of truth telling, but this is augmented by the short non-quantitative description of how the mechanism determines prize realizations. Mathematically inclined participants are thus informed on the mechanism's quantitative incentives, while the less mathematically inclined are provided with a concrete procedure for how reported beliefs are mapped into final earnings.

Figure 6 summarizes the false-report rate by period and prior for the Description treatment. Panel A reveals a moderate rate of false reports (24.5 percent) which does not differ significantly from the false-report rate in No Information (p=0.610, cf. All Prior column in Table 2) and is significantly lower than the Information treatment (p=0.004). Looking to how the false-report rate varies with the prior, panel B shows that the rate of false reports in Description does not vary substantially with the prior (p=0.211). Similarly, we find no evidence that false reports are more likely to be toward the center than the nearest extreme (p=0.564).

As such, the complete ambiguity over the chosen incentive compatible mechanism in No Information can be relaxed a little without damaging reports. However, given the distinctly different reporting behavior under the same incentives in Information—where participants are provided information on the precise quantitative incentives it is unlikely that participants comprehend the offered incentives in Description.[32]

[31] Our results (with a fixed mechanism) dovetail with Holt and Smith (2016) who find evidence across mechanisms for the superiority of a BDM-based crossover elicitation. Similar to our Description treatment, their crossover mechanism does not spell out the marginal effects on the probability of winning, focusing on the qualitative compatibility. In comparison, their QSR elicitation uses a table to make clear the marginal effects on the monetary prize.

[32] Looking at the response to survey questions and coding these as agreeing to the statement on understanding how payoffs were calculated, how a stated belief affected pay, and whether they truthfully reported we find that the Description and the No-Information treatments are statistically indistinguishable from one another in participants' self-reported understanding of the mechanism (72 vs. 77 percent, p=0.532, χ^2-test) but that there are differences in understanding how beliefs affected pay (70 vs. 83 percent, p=0.084), and indications of differences in self-reported inclination to report truthfully (85 vs. 75 percent, p=0.171).



(A) By Period (B) By Prior

FIGURE 6. FALSE-REPORT RATE IN DESCRIPTION TREATMENT

5. POSTERIOR REPORTS

Our analysis has focused on elicitations of the induced prior—as this provides the cleanest test on truthful reporting without concern for participants' ability to Bayesian update. However, it may be that quantitative-incentive information is necessary when eliciting beliefs that require effort on behalf on the participant. Complementing our analysis of the priors we therefore examine the frequency and pattern of distorted reports in Guesses 2 and 3, where participants receive signals on the state and are asked to report a posterior belief.

After observing the period's scenario information—the composition of the two urns and the prior probability—participants first report their belief on the objective prior; they are then shown two independent draws from the selected urn and asked to report an updated posterior belief after each realization. While the objective Bayesian posterior is easily determined by the analyst from the provided details, such inference requires probabilistic sophistication and non-trivial calculation on the part of participants.[33] As such, the elicited posterior beliefs are expected to deviate from the Bayesian posteriors. Indeed, the number of cases in which participants exactly report non-boundary Bayesian posteriors is just 6.8 percent across all treatments; focusing on 'truthful' reports would therefore capture only a

Reporting an understanding of how beliefs affected pay may result from understanding that truthful reporting maximized the chance of winning the prize.

[33] See Benjamin (2019) for a survey of the literature on behavior in Bayesian updating tasks.

tiny fraction of participant decisions, and would exclude many that were attempting to reveal their true (non-Bayesian) posterior belief.[34] To assess false reporting on elicited posterior beliefs, we instead characterize reported beliefs by whether they are *distant* from the objective Bayesian posterior and assess the pattern of distant reports across treatments. We find that although the treatment differences are smaller for distant-posterior reports, the qualitative pattern of results mirror those for the false reports on the priors: in terms of the total rates, the sensitivity to location, and the pull-to-center effect.

Classifying distant reports as those that differ from the Bayesian posterior by more than 15 percentage points (the approximate average size of the deviation for false reports in our prior elicitations) we find that distant reports account for 33 percent of observations in the Information treatment.[35] The rate of distant reports is reduced to 25-27 percent for the other treatments (except for the last rounds of the Feedback treatment). Further, the difference in distant-report rates between Information and our other treatments mirrors our results from the elicited priors, in that the difference in distant-reports is driven by elicitations where the true Bayesian posterior is non-central ($\pi \leq 0.35$ or $\pi \geq 0.65$). Non-central Bayesian posteriors in the Information treatment indicate a 35 percent distant-report rate, which decreases to 28 percent in RCL (different from Information with p=0.113), 27 percent in No Information (p=0.048), and 25 percent in Description (p=0.016). By contrast there are no treatment differences for posteriors in the central region.

Further, in an attempt to parallel our Table 2 examination of pull-to-center deviations, we examine the rates of three distinct behaviors when the true posterior is intermediate (non-central according to our prior definition, but also non-extreme so we additionally enforce $0.05 \leq \pi \leq 0.95$): (i) posterior beliefs at the exact center ($q = 1/2$); (ii) posterior beliefs as the nearest extreme belief of $q = 0$ or $q = 1$ to the Bayesian posterior; and (iii) posteriors beliefs in the wrong half of the elicitation interval (so any $q > 1/2$ if $\pi < 1/2$ and vice versa). While we find no differences in the rates of near extreme or wrong-half posterior reports between Information, No Information, RCL and Description, there are significant differences in the frequency of exact-center reports.[36]

[34] For boundary posteriors, the realized signal perfectly reveals the state through a simple inference without the need for calculation. For boundary cases where the Bayesian posterior is either 0 or 1, 84.8 percent of the elicited posteriors report the true boundary belief.

[35] The Appendix, Figure 9 shows that the results are not sensitive to what we define as 'distant,' while Table 5 provides statistics for inference (paralleling Table 2).

[36] However, we did find that participants were significantly less likely to make wrong-half posterior reports in the first two rounds of the Feedback treatment. Concerned that this might be driven by attention declining over the treatment, we examined wrong-half posteriors in the early periods of other treatments. However, this pattern does not show up in any other treatment.

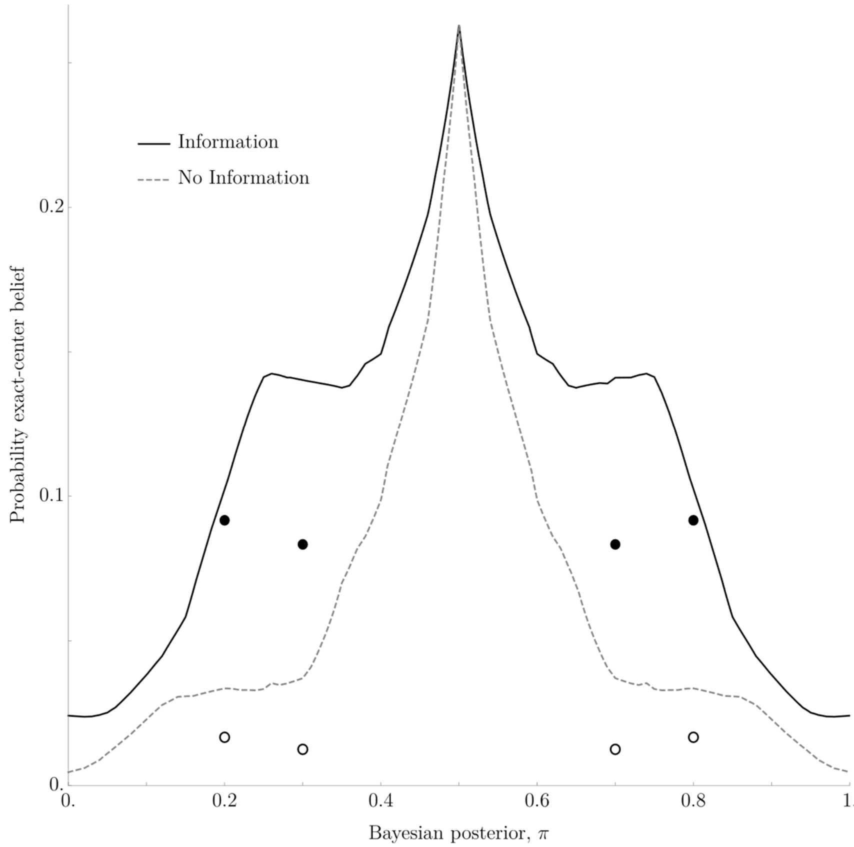


FIGURE 7. FREQUENCY OF EXACT-CENTER REPORTS BY OBJECTIVE PROBABILITY

Note: Results from a kernel-smoothed estimation of rate of exact-center reports in Guesses 2 and 3, conditional on the Bayesian posterior, enforcing symmetry around 0.5. Circles show exact-center report rates in the prior elicitations of 0.2/0.8 and 0.3/0.7 (solid for Information, hollow for No Information). Exact-center reports for prior elicitations at $\pi_0 = 1/2$ not shown due to scale (0.754 for Information, 0.762 for No-Information).

In the Information treatment 8.5 percent of intermediate-posterior reports are exact center, compared to less than half of this rate in our other treatments: 3.7 percent in RCL (p=0.010), 2.7 percent in No Information (p=0.002), and 3.4 percent in Description (p=0.008). Figure 7 illustrates the differential rates of exact-center reports in our Information and No-Information treatments through a smoothed plot of the exact-center response rate as a function of the true Bayesian posterior (where the additional plotted points indicate the exact-center reports by prior). Across both prior and posterior beliefs, false reports are less likely to be at the exact center in No Information than in the Information treatment.

While clearer identification in the posteriors requires us to focus on more specific measures of false reporting—given substantial unobserved heterogeneity over each participant's true updating rule—the general patterns mirror our findings when eliciting the induced priors.

6. Discussion and Conclusion

In a treatment providing participants with precise quantitative information on the BSR incentivizes we document large rates of false reports over an objective prior. These false reports are significantly more likely for non-centered than centered priors, and the false reports for non-centered priors are more likely to be toward the center than the extremes.

While a calculator that helps reduce compound lotteries decreases false reports, the rate remains high, and the lowest degree of false reporting (and the absence of other systematic patterns in the distortions) is found in a treatment where participants have no information at all on the quantitative incentives. A summary of the results is seen by the share of participants who report the induced prior for every one of the ten scenarios (Appendix Figure 10). While the rate of participants who consistently report the induced prior is a mere 15 percent in the Information treatment, the rate increases to 25 percent when provided with a calculator to reduce compound lotteries. However relative to the Information treatment, the share of participants who consistently report the given prior is more than three times higher (50 percent) in the No-Information treatment where all quantitative-incentive information is withheld.

The presentation of quantitative information on the incentives leads to distorted reports for both the prior and posterior elicitations, pushing reported beliefs to the center. While the pattern of false reports is evident when participants hold clear information on the incentives, we use additional treatments to assess *why* the BSR incentives gives rise to these types of deviations.[37]

The higher rate of false reports on non-centered than centered priors, along with the pull-to-center tendency for non-centered priors, suggests that participants are drawn toward more-conservative reporting by the BSR incentives, as deviations toward the center secure a substantial increase in winning on the less likely event, relative to the smaller decrease in winning for the more likely event. While failure to reduce compound lotteries contributes to the Information treatment findings, false reports remain high when participants get a compound-lottery calculation aid, and participants continue to have higher rates of false reports on non-centered than centered elicitations. The sensitivity to the location of the prior is only eliminated when quantitative information on the incentives is fully removed.

While the asymmetric changes in the chance of winning the prize may encourage false reports, such deviations are further encouraged by the very limited cost associated with the

[37] Importantly, information on the quantitative incentives does not cause participants to report that they are more confused about their offered incentives. Participants were asked if they understood how their experimental payments were calculated, and if they understood how their stated beliefs were used in that calculation (Likert scale). In both questions the Information treatment has the lowest fraction of participants disagreeing (5 percent and 1.7 percent, in comparison to the average across the other treatments of 8.9 percent and 7.1 percent). A third survey question asked whether they always reported their most-accurate guess on the elicitation. Here the results mirror the belief data, with the most disagreements found in Information (10 percent), and the least in No-Information (6.7 percent).

deviation. As with many other mechanisms that attempt to fully separate a continuous type space, the BSR suffers from a flat-maximum problem. That is, the marginal changes in payoffs that generate strict incentive compatibility are weak in economic terms, where relatively large deviations from truthful reporting produce negligible effects on final earnings.

By manipulating the availability of information on incentives we find (both between- and within-subject) that truth telling is distorted by the provision of information on the quantitative BSR incentives. Our results suggest that the BSR incentives along with the failure to reduce the associated compound lottery accounts for close to two-thirds of the false reports on non-centered priors. Combined with the evidence of growing false reports in our Feedback treatment (where subject receive feedback on the probability of winning a simple lottery given the state realization) the evidence confirms the causal effects of information on the quantitative incentives. Disturbingly most of the deviations from truth telling result from being well-informed on the quantitative incentives offered.

Although scoring rules like the BSR allow us to truthfully state to participants that their chances of winning the prize are maximized by truthfully reporting, it is disturbing that despite this prominent (literally bold-faced) statement before they begin, provision of quantitative information on the incentives causes participants to distort their reports. With the BSR being the current state of the art what does this imply for belief elicitation more generally? The underlying theoretical requirements for fully separating incentive compatible mechanisms makes the decision to collect beliefs at any resolution seem costless. However, our experiments show that the mechanism has incentives that increases false reporting.

As an alternative to pursuing ever-more sophisticated belief elicitations, it may be time to ask whether it is reasonable to assume that participants in our studies hold exact probabilistic beliefs, and that we are able to provide monetary incentives to elicit them at arbitrary precision. Instead of taking our results as a call for the development of mechanisms that are incentive compatible for an ever-more-general class of decision maker, we might instead ask whether the necessary economic inferences could be drawn with less-precise measurements, where the incentives for truthful reporting can be simpler and starker.[38]

For example, in discrete settings it may be sufficient to elicit the event the participants deem most likely and incentivize the elicitation by offering compensation only in the event that the report is correct.[39] In continuous settings, the same can be achieved by paying

[38] Ex post corrections of beliefs a la Offerman et al (2009) similarly relies on individuals holding exact beliefs. If precise beliefs are a requirement, then dynamic mechanisms that elicit the same belief through multiple (adaptive) coarse elicitations may be used to make incentives appear stark (see Schmidt and Zankiewicz, 2016).

[39] See, for example, Bhatt and Camerer, 2005, Hurley and Shogren, 2005, Niederle and Vesterlund, 2007, Vanberg, 2008, Blanco et al., 2010, Dargnies, 2012, Di Tella et al. 2015, LeCoq et al., 2015, Toussaert, 2018, Bordalo et al. 2019, Cantoni et al., 2019, Wilcox and Feltovich, 2000, and Huffman et al., 2019

participants if the true outcome falls within some bounds around their guess ().[40] Alternatively, it may be sufficient to determine whether a belief lies within a certain *fixed* interval. This allows for deviations between the potential intervals to come at a high cost and may still provide the information necessary for inference.[41] For example, suppose that in understanding individual behavior we wish to elicit the belief that an opponent will select action A or B, and that the individual's predicted behavior theoretically depends on the belief on action A exceeding 30 percent. Rather than eliciting the precise belief that action A is chosen, it may secure more reliable and truthful reporting to instead focus the elicitation on whether or not the belief on A exceeds the theoretical cutoff. If elicited beliefs are collected primarily as controls or for auxiliary tests of a behavioral mechanic, it may also be preferable for inference to collect coarser measures with starker incentives. While there are many paths to improve belief elicitation, we certainly should be hesitant to adopt mechanisms with incentives that when clearly outlined to the participant, distort their truthfulness.

[40] For example, Charness and Dufwenberg, 2006, Abeler et al, 2019, Danz et al., 2018. As Abeler et al (2019) points out, "[t]his mechanism is very simple and easier to explain and understand than proper scoring rules. It elicits in an incentive-compatible way the mode (or more precisely, the mid-point of the [x]-percentage point interval with the highest likelihood) of a subject's distribution of estimates."

[41] Asking participants to select one of a several fixed ranges of probabilities acknowledges both the ambiguity associated with providing an exact probabilistic belief (Manski, 2004) and makes it possible to provide starker incentives for doing so.

REFERENCES

Abdellaoui, Mohammed, Driouchi, Ahmed, and L'Haridon, Olivier, "Risk aversion elicitation: reconciling tractability and bias minimization," *Theory and Decision* 71, 1 (2011), pp. 63–80.

Abeler, Johannes, Nosenzo, Danielle, Raymond, Collin, "Preferences for Truth-Telling," *Econometrica*, 87, 4, (2019), pp. 1–118.

Ahrens, Steffen and Bosch-Rosa, Ciril, "Heads we both win, tails only you lose: the effect of limited liability on risk-taking in financial decision making" Discussion Paper, No. 162, Ludwig-Maximilians-Universität München und Humboldt-Universität zu Berlin, Collaborative Research Center Transregio 190 - Rationality and Competition, München und Berlin (2019).

Avoyan, Ala and Schotter, Andrew, "Attention in Games: An Experimental Study" (2020), *European Economic Review*, 124, 103410 (2020), pp. 1–28.

Babcock, Linda, Recalde, Maria P, Vesterlund, Lise, and Weingart, Laurie, "Gender differences in accepting and receiving requests for tasks with low promotability," *American Economic Review* 107, 3 (2017), pp. 714–47.

Benjamin, Daniel, "Errors in Probabilistic Reasoning and Judgment Biases," in Bernheim Douglas, DellaVigna, Stefano, Laibson, David, eds., *Handbook of Behavioral Economics: Applications and Foundations 1*, Volume 2, North-Holland, (2019), pp. 69–186.

Bhatt, M, Camerer, CF. "Self-referential thinking and equilibrium as states of mind in games: fMRI evidence," *Games and Economic Behavior* 52, 2, (2005), pp.424–459.

Blanco, M, Engelmann, D, Koch, AK, Normann H-T, "Belief elicitation in experiments: is there a hedging problem?" *Experimental Economics*, 13, 4, (2010), pp.412–438.

Bordalo, Pedro, Coffman, Katherine, Gennaioli, Nicola, and Shleifer Andrei, "Beliefs about Gender," *American Economic Review* 109, 3 (2019), pp.739–773.

Bruner, David M, "Changing the probability versus changing the reward," *Experimental Economics* 12, 4 (2009), pp. 367–385.

Cantoni, D., Yang, D.Y., Yuchtman, N, Zhang, YJ., "Protests as Strategic Games: Experimental Evidence from Hong Kong's Antiauthoritarian Movement," *Quarterly Journal of Economics*, 134, 2, (2019), pp. 1021–1077.

Cason, Timothy N. and Plott, Charles R. "Misconceptions and Game Form Recognition: Challenges to Theories of Revealed Preference and Framing," *Journal of Political Economy*, 122 (2014), pp. 1235–1270.

Cason, Timothy N., Sharma, Tridib, and Vadovič, Radovan. "Correlated beliefs: Predicting outcomes in 2× 2 games." *Games and Economic Behavior*, 122 (2020), pp. 256–276.

Charness, Gary, Dufwenberg, Martin "Promises and Partnership," *Econometrica*. 74, 6, (2006), pp. 1579–1601.

Charness, Gary, Oprea, Ryan, and Yuksel, Sevgi. "How do people choose between biased information sources? Evidence from a laboratory experiment." *forthcoming at Journal of the European Economic Association,* (2020).

Chen, Yan and He, Yinghua, "Information Acquisition and Provision in School Choice: An Experimental Study." (2017).

Danz, David, Madarász, Kristóf, Wang, Stephanie, "The Biases of Others: Projection Equilibrium in an Agency Setting," *CEPR Discussion Paper*, DP12867, (2018).

Dargnies, Marie-Pierre, Hakimov, Rustamdjan, and Kübler, Dorothea, "Self-Confidence and Unraveling in Matching Markets," *Management Science* 65, 12 (2019), pp. 5603–18.

Dargnies, Marie-Pierre, "Men Too Sometimes Shy Away from Competition: The Case of Team Competition," Management Science 58, 11, (2012), pp. 1982–2000.

DeQuidt, Jonathan, Vesterlund, Lise, and Wilson, Alistair, "Experimenter Demand Effects," *Handbook of Research Methods and Applications in Experimental Economics* (Edward Elgar Pub., edited by Arthur Schram and Aljaz Ule).

Dianat, Ahrash, Echenique, Federico, and Yariv, Leeat, "Statistical Discrimination and Affirmative Action in the Lab," *CEPR Discussion Paper No. DP12915,* (2018).

Di Tella, Rafael, Perez-Truglia Ricardo, Babino Andres, Sigman Mariano, "Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism," American Economic Review 105, 11 (2015), pp. 3416–3442.

Drerup, Tilman, Enke, Benjamin, and Von Gaudecker, Hans-Martin, "The precision of subjective data and the explanatory power of economic models," *Journal of Econometrics* 200, 2 (2017), pp. 378–389.

Enke, Benjamin, and Graeber, Thomas, "Cognitive Uncertainty". *NBER Working Paper No. 26518*, (2019).

Enke, Benjamin, Schwerter, Frederik, and Zimmermann, Florian, "Associative memory and Belief Formation," *NBER Working Paper No. 26664*, (2020).

Erkal, Nisvan, Gangadharan, Lata, and Koh, Boon Han, "Attribution biases in leadership: Is it effort or luck?" (2019).

Fischbacher, Urs, "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics* 10, 2 (2007), pp. 171–178.

Frederick, Shane, "Cognitive reflection and decision making," *Journal of Economic Perspectives* 19, 4 (2005), pp. 25–42.

Gächter, Simon and Renner, Elke, "The effects of (incentivized) belief elicitation in public goods experiments," *Experimental Economics* 13, 3 (2010), pp. 364–377.

Hao, Li and Houser, Daniel, "Belief Elicitation in the Presence of Naïve Respondents: An Experimental Study," *Journal of Risk and Uncertainty* 44, 2 (2012), 161–80.

Harrison, Glenn W and Phillips, Richard D, "Subjective beliefs and statistical forecasts of financial risks: The chief risk officer project," in *Contemporary Challenges in Risk Management* (Springer, 2014), pp. 163–202.

Healy, Paul J, "Epistemic Game Theory Experiments: Utility Elicitation and Irrational Play" mimeo, (2017).

Healy, Paul J, "Explaining the BDM—Or any Random Binary Choice Elicitation Mechanism—To Subjects", mimeo, (2018).

Hillenbrand, Adrian and Schmelzer, André, "Beyond information: Disclosure, distracted attention, and investor behavior," *Journal of Behavioral and Experimental Finance* 16 (2017), pp. 14–21.

Holt, Charles A. and Smith, Angela M., "Belief Elicitation with a Synchronized Lottery Choice Menu That Is Invariant to Risk Attitudes," *American Economic Journal: Microeconomics* 8, 1 (2016), pp. 110–39.

Hossain, Tanjim and Okui, Ryo, "The binarized scoring rule," *Review of Economic Studies* 80, 3 (2013), pp. 984–1001.

Hossain, Tanjim and Okui, Ryo, "Belief formation under signal correlation" (2019).

Huffman, David, Raymond, Collin, Shvets, Julia, "Persistent Overconfidence and Biased Memory: Evidence from Managers," mimeo, (2019).

Hurley, TM, Shogren, JF. "An Experimental Comparison of Induced and Elicited Beliefs," *Journal of Risk and Uncertainty*, 30, 2, (2005), pp. 169–188.

LeCoq, C, Tremewan, J, Wagner, AK., "On the effects of group identity in strategic environments," *European Economic Review*, 76, (2015), pp. 239–252.

Manski, Charles F. "Measuring expectations." *Econometrica* 72.5 (2004): 1329-1376.

Masiliūnas, Aidas, "Overcoming coordination failure in a critical mass game: strategic motives and action disclosure," *Journal of Economic Behavior & Organization* 139 (2017), pp. 214–251.

Meloso, Debrah, Nunnari, Salvatore, and Ottaviani, Marco, "Looking into crystal balls: a laboratory experiment on reputational cheap talk" (2018).

Nelson, Robert G and Bessler, David A, "Subjective probabilities and scoring rules: Experimental evidence," *American Journal of Agricultural Economics* 71, 2 (1989), pp. 363–69.

Niederle, Muriel and Vesterlund, Lise, "Do women shy away from competition? Do men compete too much?" *Quarterly Journal of Economics* 122, 3 (2007), pp.1067–1101.Offerman, Theo, Sonnemans, Joep, Van De Kuilen, Gijs, and Wakker, Peter P, "A Truth Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes," *The Review of Economic Studies,* 76, 4 (2009), pp.1461–89.

Oprea, Ryan, and Yuksel, Sevgi. "Social Exchange of Motivated Beliefs," mimeo, (2020).

Palfrey, Thomas R and Wang, Stephanie W, "On eliciting beliefs in strategic games," *Journal of Economic Behavior & Organization* 71, 2 (2009), pp. 98–109.

Roth, Alvin E and Malouf, Michael W, “Game-theoretic models and the role of information in bargaining,” *Psychological Review* 86, 6 (1979), pp. 574.

Schlag, Karl H and van der Weele, Joël J, “A Method to Elicit Beliefs as Most Likely Intervals,” *Judgment and Decision Making* 10, 5 (2015), 456–68.

Schmidt, Tobias and Zankiewicz, Christian, “Binary Choice Belief Elicitation: An Adaptively Optimal Design” (2016).

Schotter, Andrew and Trevino, Isabel, “Belief Elicitation in the Laboratory,” Annual Review of Economics 6,1 (2014) pp. 103–28,

Toussaert S. “Eliciting Temptation and Self-Control Through Menu Choices: A Lab Experiment,” *Econometrica*, 86, 3, (2018), pp. 859–889.

Trautmann, Stefan T and van de Kuilen, Gijs, “Belief elicitation: A horse race among truth serums,” *Economic Journal* 125, 589 (2014), pp. 2116–35.

Vanberg, C. “Why Do People Keep Their Promises? An Experimental Test of Two Explanations,” *Econometrica*, 76, 6, (2008), pp. 1467–1480.

Wang, Stephanie W, “Incentive effects: The case of belief elicitation from individuals in groups,” *Economics Letters* 111, 1 (2011), pp. 30–33.

Wilcox NT, Feltovich N. “Thinking Like a Game Theorist: Comment,” *Mimeo, University of Houston*, (2000).

Wilson, Alistair J. and Vespa, Emanuel, “Paired-uniform scoring: Implementing a binarized scoring rule with non-mathematical language” (2018).

APPENDIX A: ADDITIONAL TABLES AND FIGURES

TABLE 4. ϵ-FALSE REPORTS BY TREATMENT

Treatment	False Reports		
	All Priors	By Prior	
		$\pi_0 = 0.5$	$\pi_0 \neq 0.5$
Information	0.312 (0.036)	0.171 (0.035)	0.406 (0.046)
RCL	**0.227** (0.034)	0.127 (0.034)	**0.294** (0.041)
No Information	**0.162** (0.034)	0.183 (0.040)	**0.147** (0.036)
Feedback $_{(t=1,2)}$	**0.133** (0.032)	0.182 (0.056)	**0.092** (0.046)
Feedback $_{(t=9,10)}$	**0.2** (0.046)	0.117 (0.052)	**0.260** (0.061)
Description	**0.182** (0.034)	0.138 (0.036)	**0.211** (0.039)
N	2,630	2,630	

Note: Standard errors in parentheses clustered by participant (299 clusters) recovered from three separate joint estimates on the false report proportion in the prior elicitations: (i) *All priors*, dependent variable an indicator for $|q - \pi_0| > 5$ with treatment level estimation; and (ii) *By Prior* column pair, same dependent variable as *All priors*, but with separate treatment estimates for centered/non-centered prior location.

TABLE 5. POSTERIOR INFERENCE (GUESS 2 AND 3): FALSE REPORTS AND TYPE BY TREATMENT

Treatment	Distant Reports				Distant Report Movement $\pi \in [0.15,0.35] \cup [0.65,0.85]$		
	All	By Posterior Location					
		Center	Extreme	Interm.	Center	Near Extreme	Distant Extreme
Information	0.370	0.275	0.285	0.471	0.255	0.075	0.141
	(0.026)	(0.039)	(0.040)	(0.033)	(0.027)	(0.015)	(0.022)
RCL	**0.304**	0.212	0.258	**0.377**	**0.187**	0.092	0.098
	(0.025)	(0.035)	(0.036)	(0.033)	(0.025)	(0.021)	(0.020)
No Information	**0.313**	0.317	**0.210**	**0.386**	**0.142**	0.109	0.134
	(0.026)	(0.033)	(0.037)	(0.034)	(0.020)	(0.029)	(0.021)
Feedback $_{(t=1,2)}$	**0.279**	0.300	**0.159**	**0.361**	0.194	0.111	**0.056**
	(0.040)	(0.076)	(0.052)	(0.052)	(0.042)	(0.031)	(0.024)
Feedback $_{(t=9,10)}$	0.313	0.396	0.216	**0.339**	**0.127**	0.110	0.102
	(0.039)	(0.066)	(0.057)	(0.047)	(0.031)	(0.031)	(0.032)
Description	**0.287**	0.254	0.235	**0.338**	**0.166**	0.054	0.118
	(0.026)	(0.035)	(0.033)	(0.034)	(0.019)	(0.016)	(0.019)
N	5,260		5,260			2,458	

Note: Standard errors in parentheses clustered by participant (299 clusters) recovered from three separate joint estimates on the distant-report rate: (i) *Distant Reports*, *All*, proportion of distant reports ($|q - \pi| \geq 0.15$) over treatment; (ii) *Distant Reports, By Posterior Location*, proportion of distant reports over treatment and prior location (centered, $\pi \in (0.35,0.65)$; extreme, $\pi \in [0, 0.15) \cup (0.85, 1]$; or intermediate, all other); and (iii) *Distant Report Movement* proportion of distant reports by location of the movement, conditioning on an intermediate posterior. Types in (iii) for $\pi < 0.5$ (with symmetric definition for $\pi > 0.5$) defined as: movements to the center ($q \in (\pi, 0.5]$); movements to the near extreme ($q \in [0, \pi)$); and to the distant extreme ($q \in (0.5, 1]$). Bold face coefficients are different from the relevant Information coefficient with $p < 0.1$ (two-sided test).

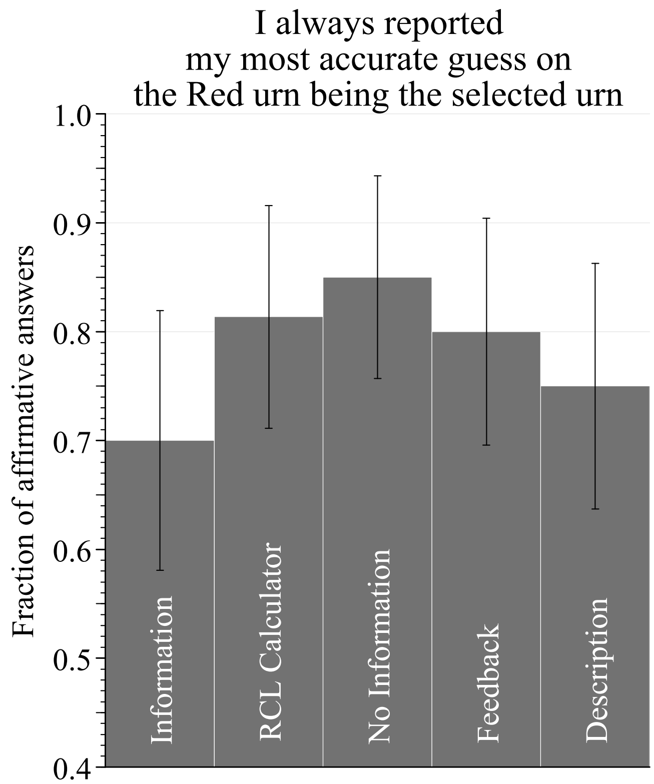


FIGURE 8. RESPONSES TO POST-EXPERIMENTAL QUESTIONNAIRE

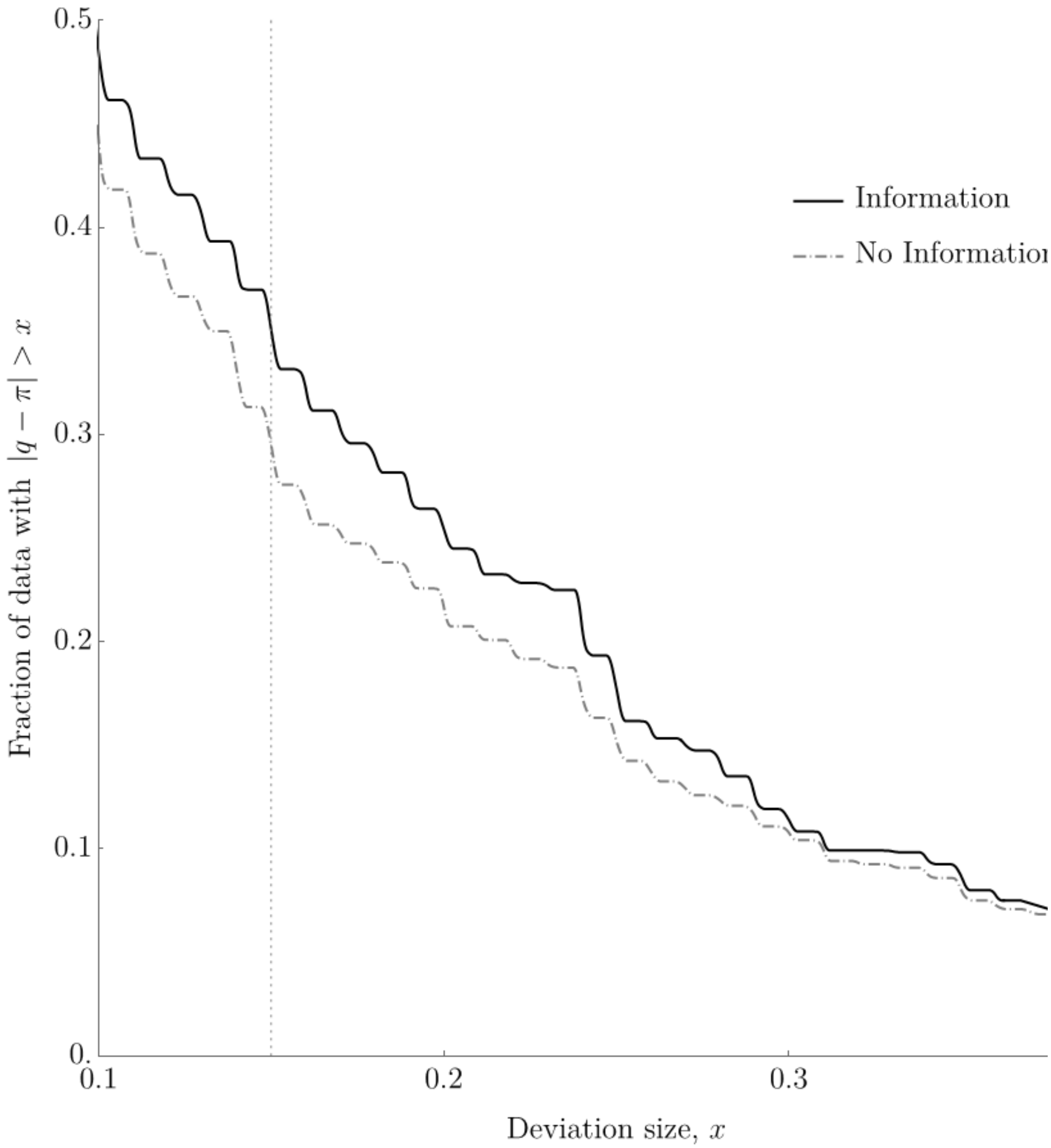


FIGURE 9. PROPORTION OF POSTERIOR REPORTS BY DISTANCE FROM BAYESIAN POSTERIOR

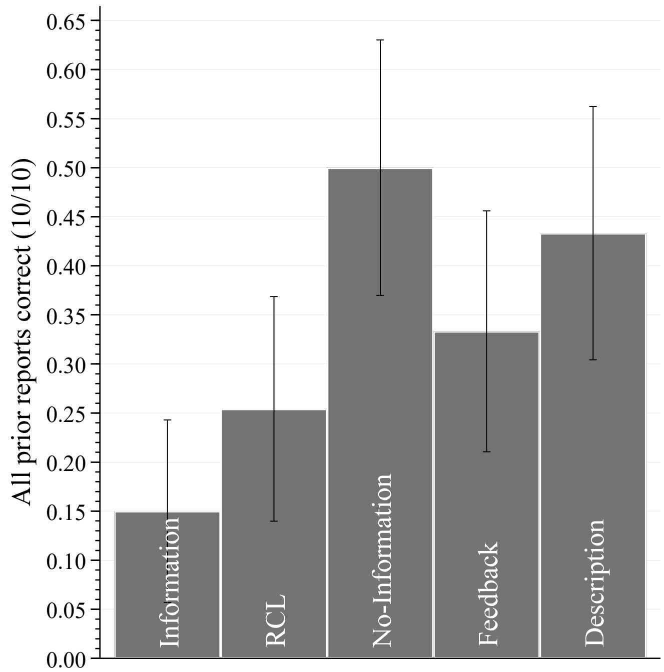


FIGURE 10. SHARE OF PARTICIPANTS WHO CONSISTENTLY REPORT THE PRIOR

APPENDIX B: EXPERIMENT INSTRUCTIONS[42]

Instructions

Thank you for participating in our study. This is an experiment on decision making. The other people in this room are also participating in the experiment, and you may not talk to them. If you have a question, please raise your hand and an experimenter will come and answer you in private.

You will receive $8 for participating in this experiment, but the decisions you make can further increase these earnings. Any money you make will be paid privately and in cash at the end of the experiment.

Explanation of your task

The experiment will consist of **ten scenarios**. In each scenario the computer will fill two urns with five 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. Within each scenario you will make a total of three guesses.

Each scenario proceeds as follows:

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

Computer Selects an Urn: The computer selects the Red or the Blue urn by rolling a fair 10-sided die and comparing it to a number **X** between 1 and 10. The selected urn is determined as follows:

- If the die roll is less than or equal to **X** then the Red urn is selected.
- If the die roll is greater than **X** then the Blue urn is selected.

Once the computer selects an urn it is fixed and stays the same for the entire scenario. The die-roll selection rule **X** means that the chance the computer selects the Red urn is **X**-in-10. For example, suppose **X**=6, then there is a six-in-ten chance (60 percent) that the computer selects the Red urn, and a four-in-ten chance (40 percent) that the computer selects the Blue urn.

The number **X** will vary across the 10 scenarios. After the computer has filled the two urns and rolled the 10-sided die to determine which urn is selected, you will be asked to make your guesses. At the beginning of each scenario you will learn how many red and blue balls

[42] Treatment differences are indicated in `[square brackets]`. In addition, the RCL treatment uses the term "*Your Submitted Guess*" instead of "*Your Guess*" throughout, and the term "submit" instead of "provide" whenever the instructions referred to the participant reporting.

there are in each urn, and the rule the computer used to select an urn (the number **X**). However, you will not learn which of the two urns has been selected until after you have made your guesses.

You are asked to provide your best guess that the computer has selected the Red urn for the scenario. The three questions are ordered as follows:

Guess 1 Knowing only the rule **X** that the computer used to select an urn, you provide your first guess that the selected urn is the Red one.

Guess 2 The computer fairly draws one of the five balls from the selected urn. After seeing the color of this ball you provide your second guess that the selected urn is the Red one.

Guess 3 After replacing the first-drawn ball back into the selected urn and mixing it, the computer fairly draws a second ball from the five. After seeing the color of the second ball you provide your third guess that the selected urn is the Red one.

Note that the draws from the selected urn in questions 2 and 3 are independent from one another: After the first draw is made, it is as if 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 five balls has the same chance of being drawn in each question.

Feedback After you have answered the scenario's three questions you learn which urn the computer selected and drew balls from. Your three 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.

Your Guess

For each question you have to guess the chance that the selected urn is the Red one. Your guess is a percentage probability from 0 to 100—with 0 indicating a 0-out-of-100 chance that the selected urn is the Red urn, and 100 indicating a 100-out-of-100 chance. The number you provide is called *Your Guess*.

You choose *Your Guess* by clicking the response bar on your screen. The width of the red part of the bar indicates your guess that the Red urn was selected.

- Larger values of *Your Guess* represent a greater chance that the Red urn was selected and a smaller chance that the Blue urn was selected
- Smaller values of *Your Guess* represent a smaller chance that the Red urn was selected and a greater chance that the Blue urn was selected

The width of the blue part of the bar is 100−*Your Guess*, and represents your guess that the Blue urn was selected.

[Information, RCL, Description treatments:

Payment Rule

We now explain how *Your Guess* is used to determine whether you win the $8 prize.

- The computer chooses two numbers between 1 and 100, where each number is equally likely, as if rolling two 100-sided dice. These numbers are called *Computer Number A* and *Computer Number B*.
- The computer determines whether you win the $8 prize according to which urn was selected:

The selected urn is the Red urn: You will win the $8 prize if *Your Guess* is **greater than or equal** to **either** of the two *Computer Numbers*.

The selected urn is the Blue urn: You will win the $8 prize if *Your Guess* is **less than either** of the two *Computer Numbers*.

]
[Information, RCL treatments:

To help you understand the payment rule, as you move *Your Guess* the computer will inform you of:

- The probability of winning the $8 if the Red urn was selected
- The probability of winning the $8 if the Blue urn was selected

]
[RCL treatment:

As mentioned above, we designed the payment rule to make sure that your greatest total chance of winning is secured by letting *Your Submitted Guess* equal to your most-accurate guess that the urn is Red (what we will call *Your True Guess* on Red). We provide a calculator to help you determine your total chance of winning the prize given any *True* and *Submitted Guesses*.

The calculator will appear in a gray box on the bottom of your screen. When you have entered *Your True Guess* that the urn is Red the calculator will use *Your Submitted Guess* to compute your total chance of winning. The formula used to calculate your total chance of winning is given by:

$$\begin{gathered}\text{(True Guess on Red)} \times \text{(Prob. of Winning if Red given Submitted Guess)} \\ + \\ \text{(True Guess on Blue)} \times \text{(Prob. of Winning if Blue given Submitted Guess).}\end{gathered}$$

]

Final Payment

The payment rule is designed so that you can secure the largest chance of winning the prize by reporting your most-accurate guess. [No-Information treatment: The precise payment rule details are available by request at the end of the experiment.]

At the end of the experiment, the computer will randomly choose two of the ten scenarios for payment. From each of these two scenarios, one of the three guesses will be randomly chosen for payment. Every guess has the same chance of being selected for payment. [Information, RCL, Feedback treatments: At the end of each scenario you find out which urn was actually selected, and learn your chance of winning the $8 if the guess is selected for payment.]

For the selected questions we will use *Your Guess* and whether the selected urn was the Red urn to determine your chance of winning $8. After determining your chance of winning, the computer will conduct the lottery for the prize to see if you won the $8.

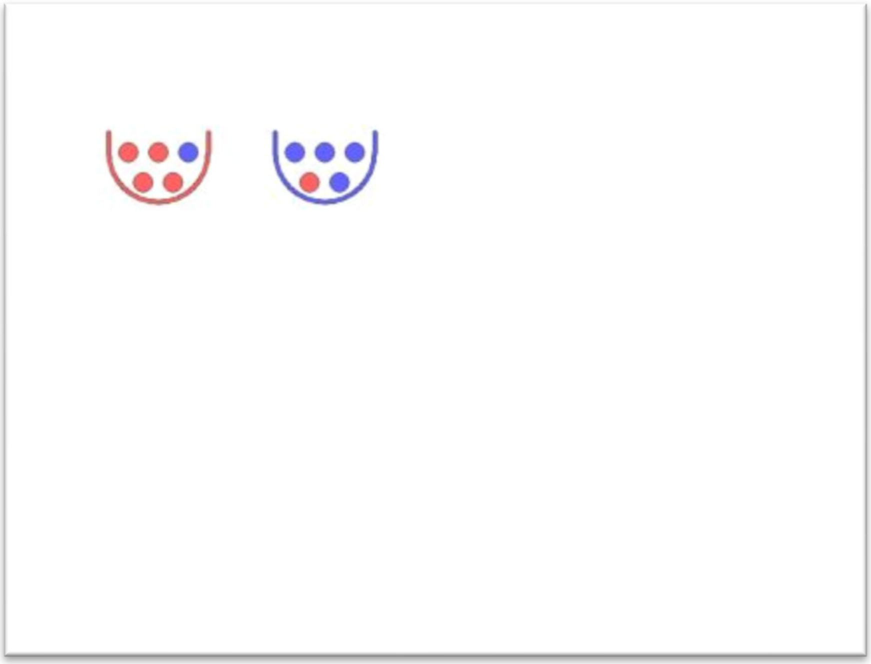
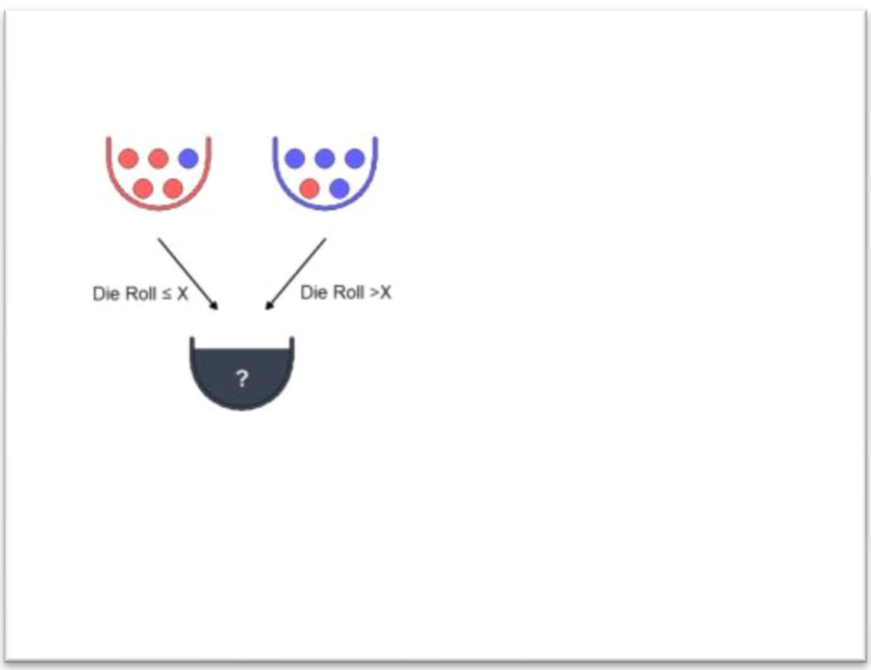
Your payment for this experiment will therefore be:

- $8 if you do not win the $8 on either guess.
- $16 if you win the $8 prize on one of the two selected guesses.
- $24 if you win the $8 prize on both selected guesses

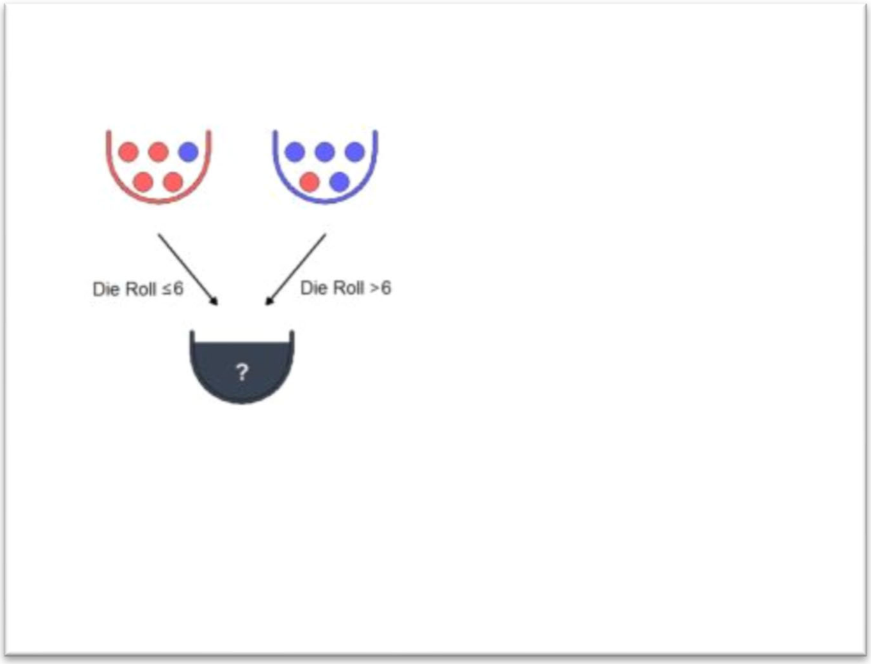
Summary

For a brief summary please take a look at the presentation at the front of the lab.

APPENDIX C: SLIDES AND SCRIPT USED IN SUMMARIZING INSTRUCTIONS[43]

Slides	**Script (read out loud by experimenter)**
	We now summarize the task in each scenario. To begin with the computer fills the two urns. Each urn is filled with five balls, which are either blue or red. The red urn is the urn with more red balls in it.
	Next the computer selects one of the two urns for the scenario. It does this using the rule X and a 10-sided die roll. If the die roll is equal to less than X the red urn is selected. If it's greater than X, the blue urn is selected. Because of this rule, the chance of selecting the red urn is X-in-10.

[43] Treatment differences are indicated in `[square brackets]`. In addition, the RCL treatment uses the term "*Your Submitted Guess*" instead of "*Your Guess*" throughout, and the term "submit" instead of "make" whenever the instructions referred to the participant reporting.

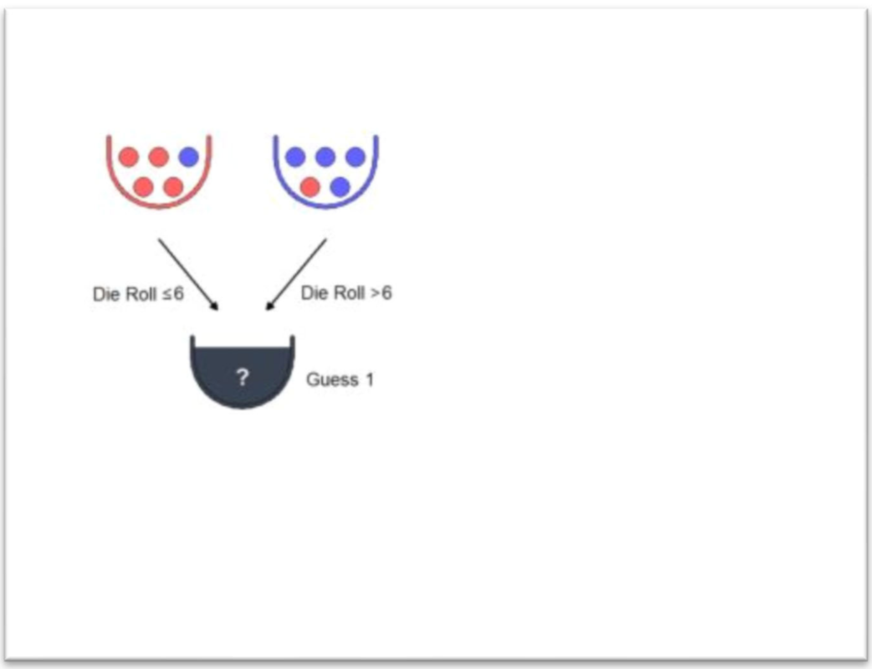


Suppose that X is equal to 6. So for die rolls of 1 to 6 the Red urn is selected.

And for die rolls from 7 to 10 the Blue urn is selected.

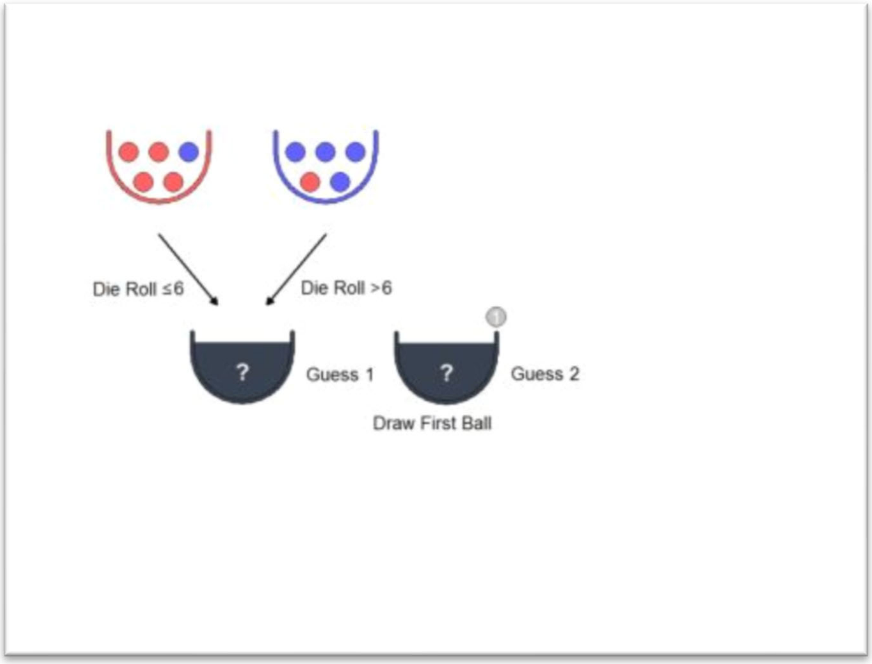
So the chance the red urn is selected is 6-in-10, or 60 percent.

The selected urn remains the same for the entire scenario.



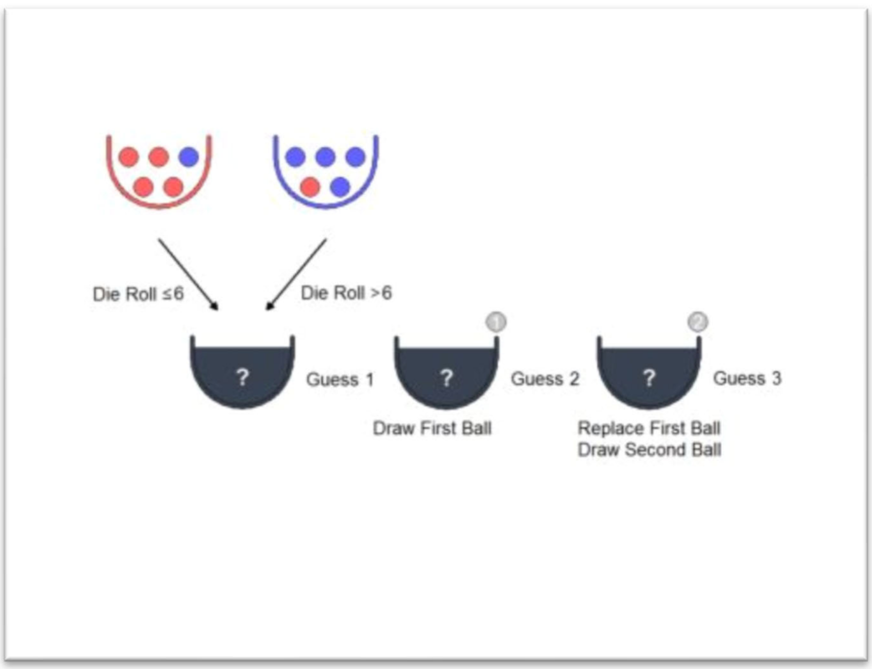
After the computer has selected one of the two urn you make your first guess.

You make your first guess only knowing the die roll rule (here 6) and how many red and blue balls are in each urn.



After you make your first guess, you then get to see a drawn ball from the selected urn. The drawn ball can be either red or blue, where the chance of this depends on which urn was selected for the scenario.

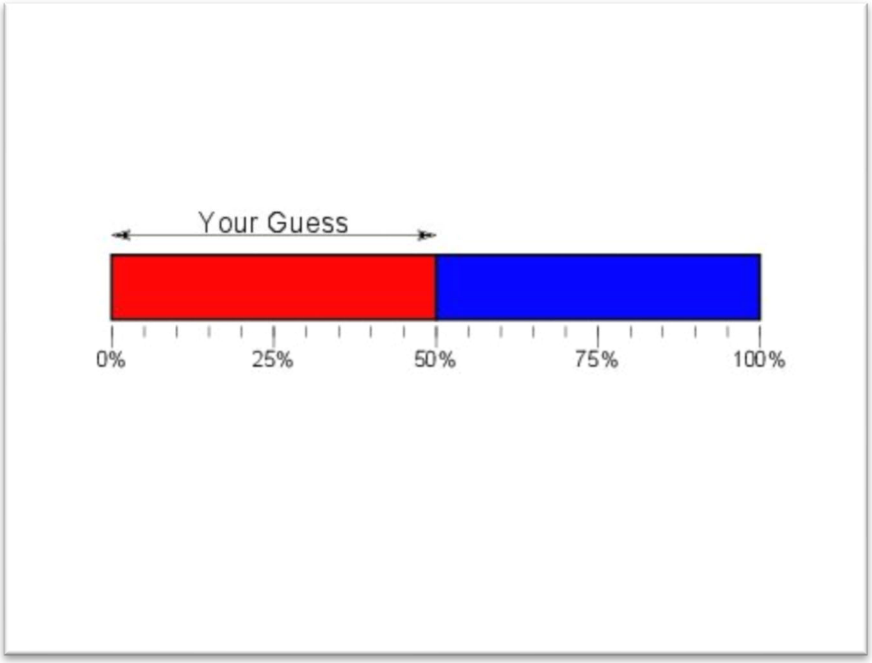
After seeing the color of the drawn ball, you make your second guess.



The first ball is put back into the selected urn, and the balls mixed.

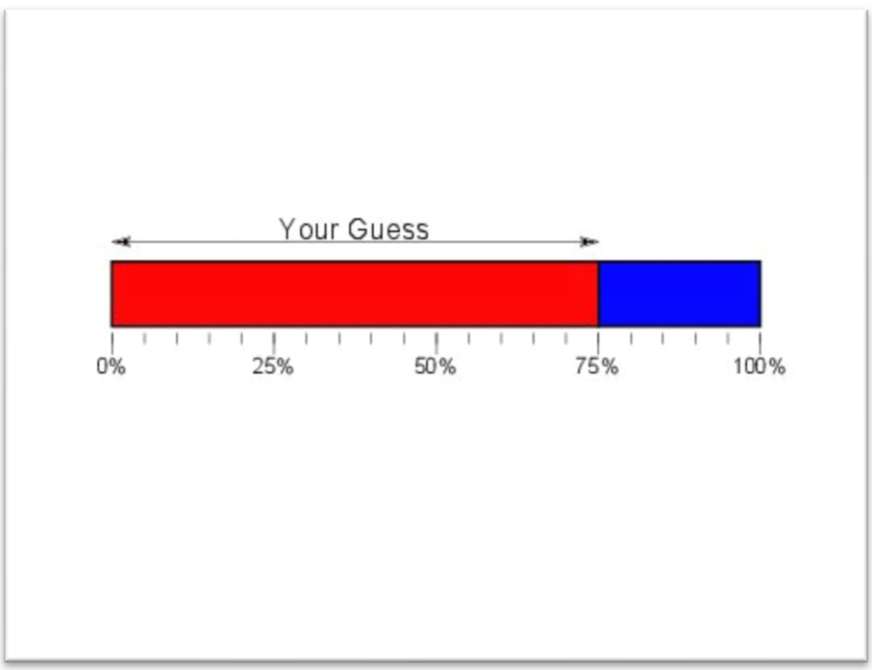
You then draw a second ball from the urn and see what color it is.

After seeing the color, you make your third and final guess.

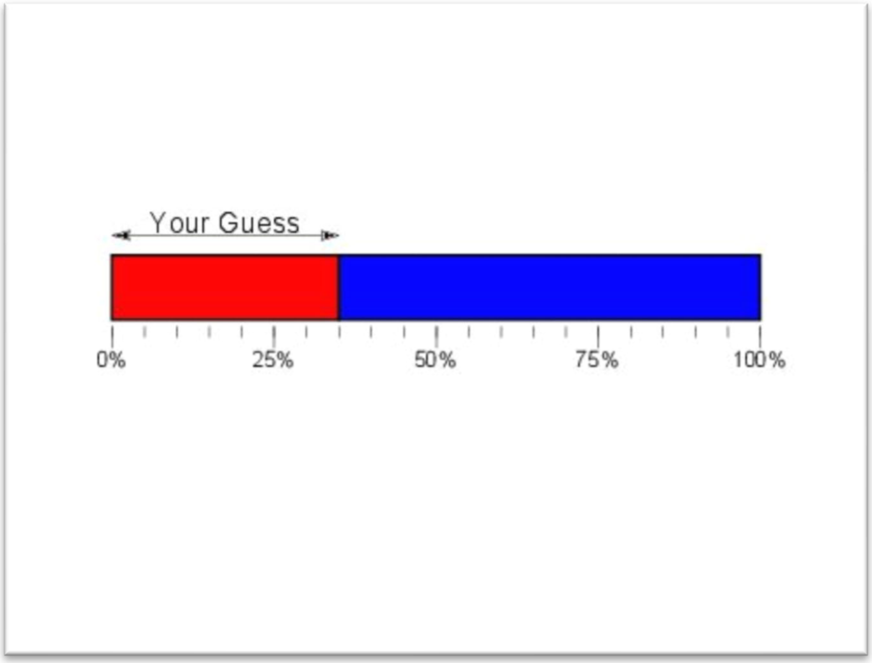


You enter Your Guesses by clicking the response bar on your screen.

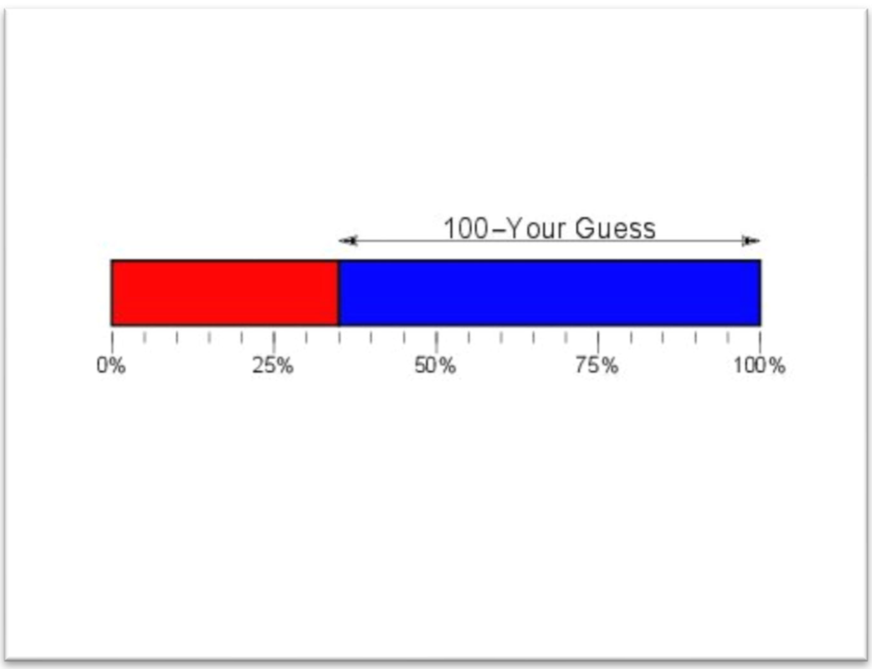
The width of the red part of the bar indicates your percentage chance that the red urn was selected.



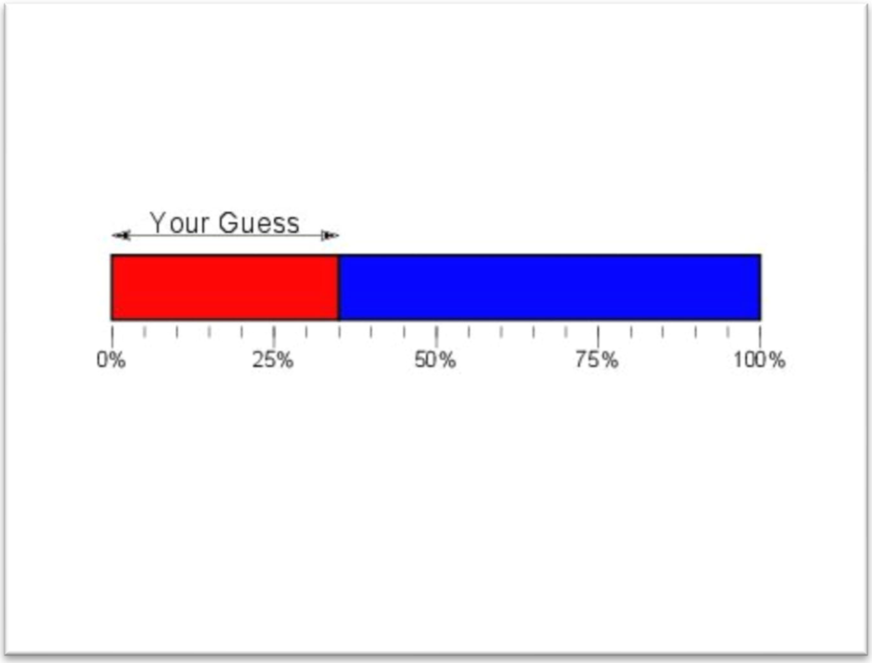
Your Guess is the width of the red part of the bar, and so wider selection represents a greater chance that the Red urn was selected.



A thinner red selection represents a smaller chance that the red urn was selected.

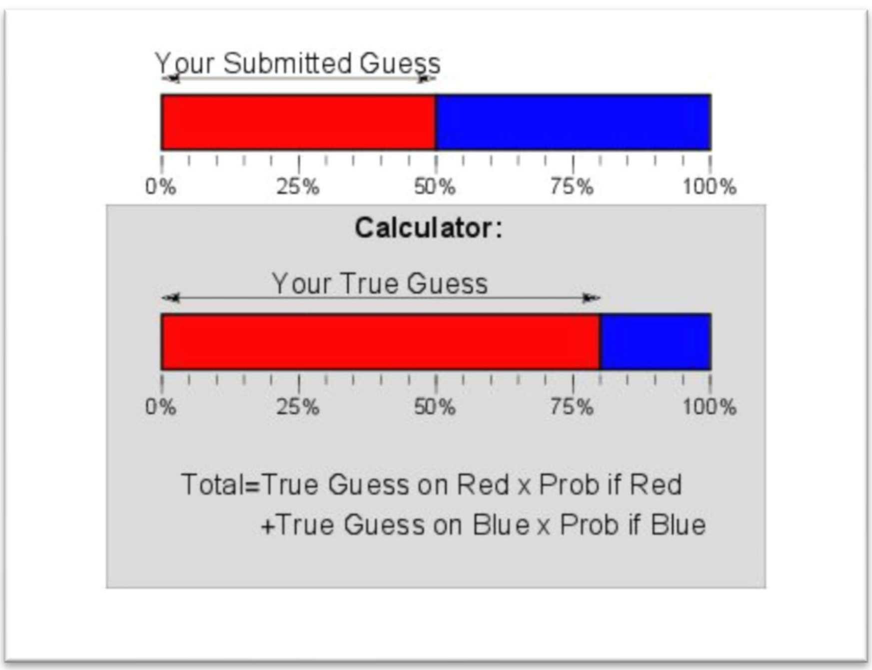


The width of the blue part represents 100-Your Guess and is the percentage chance that the blue urn was selected.



Remember, in every question we ask you for a guess that the Red urn is the selected urn.

[RCL treatment only]



In addition to the bar where you enter Your Submitted Guess, we also provide you with a calculator.

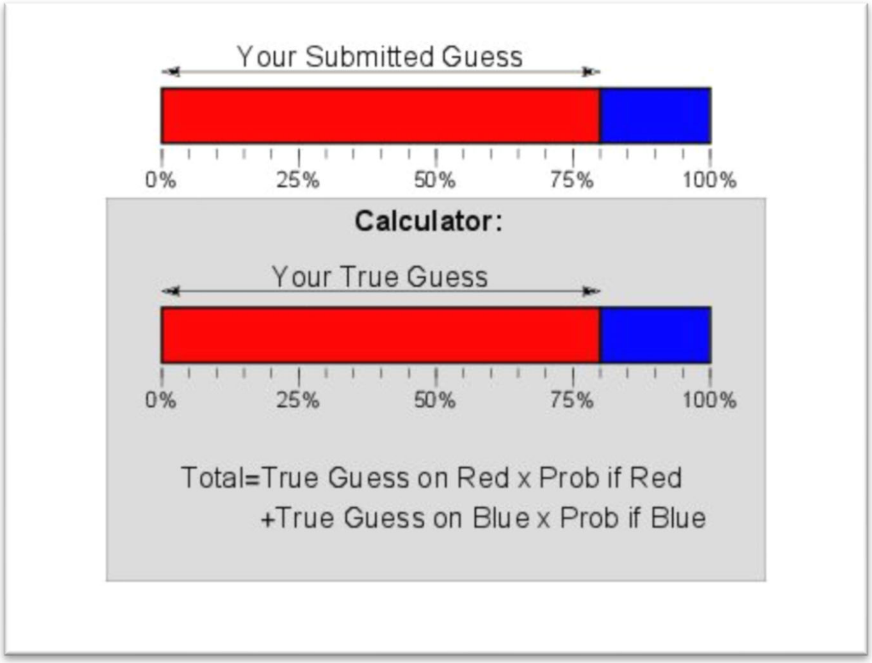
To use the calculator, you enter *Your True Best* guess.

For any selection of Your Submitted Guess and Your True best Guess the calculator will provide you with your total chance of winning.

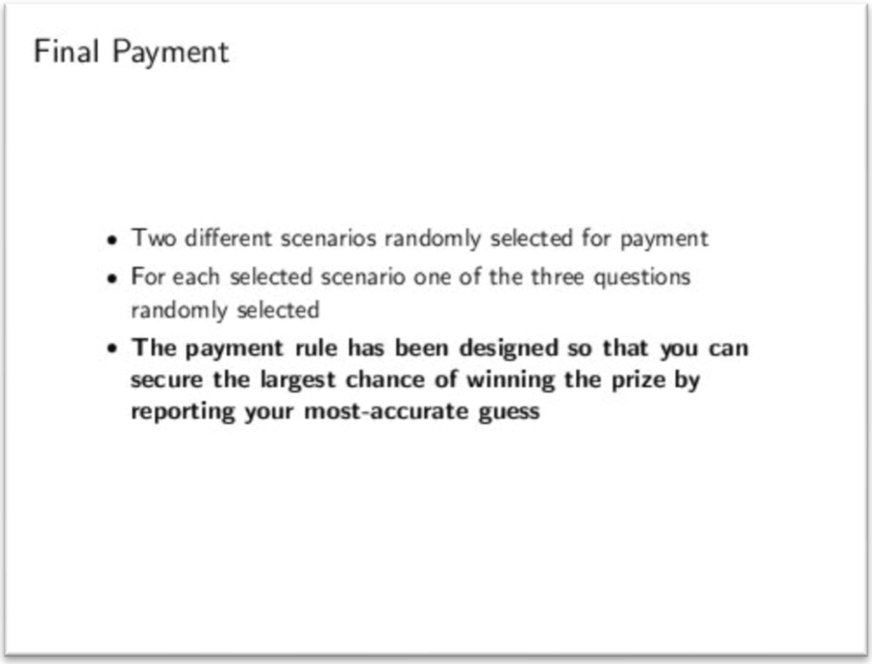
Your total chance of winning is calculated as

> Your True Best Guess on Red times the Likelihood that you Win if Red is Selected, given Your Submitted Guess
>
> + Your True Best Guess on Blue times the Likelihood that you Win if Blue is Selected, given Your Submitted Guess

[RCL treatment only]



The calculator allows you to verify that whatever your True Best Guess might be, the payment rule ensures that you will maximize your total chance of winning by setting your submitted guess equal to your True Guess.



Final Payment for the experiment will be $8 plus payment for two different scenarios.

For each selected scenario one of the three guesses is selected for payment.

The payment rule we use is designed so that you can secure the largest chance of winning the prize by reporting your most-accurate guess [RCL treatment: (Your True Guess)].

We will now start the experiment.