Cognitive heterogeneity and complex belief elicitation

Ingrid Burfurd and Tom Wilkening*

December 31, 2019

Introduction

The Stochastic Becker-DeGroot-Marschak (SBDM) mechanism is a theoretically elegant way of eliciting incentive-compatible beliefs under a variety of risk preferences. However, the mechanism is complex and there is concern that some participants may misunderstand the incentive properties of the mechanism. We use a two-part design in which we identify participants with high and low cognitive ability and elicit their beliefs in both easy and hard decision problems. Relative to Introspection, there is less variation in belief accuracy between easy and hard problems in the SBDM mechanism. However, there is a greater difference in the accuracy of reports between "high ability" participants and "low ability" participants. These results suggest that while the SBDM mechanism encourages individuals to think more carefully about beliefs, it is more sensitive to cognitive ability. Our results show that mechanism complexity is an important consideration in developing elicitation mechanisms, and identifies cognition as an important consideration when interpreting elicited beliefs.

*Ingrid Burfurd: Department of Economics, Finance and Marketing, RMIT. E-mail: ingrid.burfurd@gmail.com. Tom Wilkening: Department of Economics, The University of Melbourne. E-mail: tom.wilkening@unimelb.edu.au. We thank Amy Corman for assistance in running the experiments. We also thank Ralph-Christopher Bayer, Nikos Nikiforakis, and James Tremewan for helpful comments. We gratefully acknowledge the financial support of the Australian Research Council through the Discovery Early Career Research Award DE140101014 as well as the Faculty of Business and Economics at the University of Melbourne.

1 Introduction

Most economic theories describe the decision-making process as a confluence of preferences, beliefs, and cognitive ability. Disentangling these primitives is a challenge because they are all unobservable in most empirical data. An important advantage of experiments is that auxiliary revelation mechanisms can be used to elicit participants' beliefs. Accurate belief data can supplement choice data to facilitate stronger identification of the preferences and cognitive processes that guide choice.

It is well-known that heterogeneous preferences can make eliciting accurate beliefs difficult.[1] This is because heterogeneous preferences may also impact behavior in the revelation mechanism used to elicit beliefs. For example, individuals may misreport in unincentized introspection mechanisms if they find it arduous to think carefully about their beliefs or if revealing their true belief causes them discomfort. Offering explicit incentives can mitigate these issues, but incentive-compatible mechanisms must use lotteries and lotteries interact with risk preferences. This has led to the use of the sophisticated Stochastic Becker-Degroot-Marschak mechanism (SBDM), which is predicted to induce truthful revelation for a wide variety of preferences.[2]

Despite the impressive theoretical properties of the SBDM, the existing literature that compares belief elicitation methods has found little evidence that the SBDM mechanism outperforms introspection in terms of belief accuracy (Hollard et al., 2016; Trautmann and Kuilen, 2015).[3] Further, participants using incentive-compatible belief elicitation mechanisms often misreport their beliefs even when the probability of an event occurring is objectively known (Hao and Houser 2012; Burfurd and Wilkening 2018). These results suggest that there may be a second potential difficulty for belief elicitation: an interaction between belief elicitation mechanisms and cognitive ability.

Heterogeneous responses to belief elicitation based on cognition has potentially important implications for interpreting belief data and for deciding on which belief elicitation method to use. If decision-making and reporting behavior vary systematically with latent cognitive ability, cognitively demanding mechanisms might yield reliable reporting data

[1]For a discussion of belief elicitation techniques and preferences, see Schlag et al. (2013), Schotter and Trevino (2014), and Trautmann and Kuilen (2015).

[2]The mechanism that we refer to as the SBDM has a variety of names in the literature. Ducharme and Donnell (1973) is the first empirical paper we are aware of that uses the procedure and refers to the mechanism as "bets mode" for eliciting beliefs. Schlag et al. (2013) refer to the mechanism as "reservation probabilities" while Trautmann and Kuilen (2015) use the term "probability matching". Many other papers refer to the mechanism as the "Karni mechanism" due to the theoretical contributions of Karni (2009). We prefer SBDM due to the strong similarities between the mechanism and the mechanism proposed by Becker et al. (1964) for eliciting valuations. The use of probabilities to control for risk aversion is discussed as early as Smith (1961) and Savage (1971). Varieties of the mechanism have been studied by Grether (1981), Allen (1987), and Holt (2006).

[3]Hollard et al. (2016) finds that both the SBDM and introspection outperforms the quadratic scoring rule using a series of subjective tasks. Trautmann and Kuilen (2015) find that while accuracy between the SBDM mechanism and Introspection mechanism do not differ, their is some evidence that incentive-compatible belief mechanisms are better predictors of a participant's own actions.

from participants who make better decisions, and less reliable data from participants who make sub-optimal ones. This could have serious implications for analysis, since reporting errors would be correlated systematically with unobservable cognitive ability. It also suggests that researchers may face a tradeoff between catering for heterogenous preferences and heterogeneous cognitive abilities.

To explore the interaction between belief elicitation mechanisms and cognition we use a two-part design. We first identify high and low-cognitive ability participants and then examine how participants of both types respond to belief elicitation. To evaluate participants' cognitive ability we use a variant of an urn task introduced in Charness and Levin (2005) and Charness et al. (2007), which we refer to as the bucket task. In each period, participants are individually assigned one of two buckets (A or B) with equal probability. Each bucket is divided into two sides, and each side contains 20 black and/or white balls. Participants draw and replace a ball from the *left* side of their bucket at the start of each period and are paid $4 if they observe a black ball and $0 if they observe a white ball. The color of the ball is informative, and presents an opportunity for participants to update their belief that they have been given Bucket A. Participants then choose whether they would like to draw an additional ball from the left or the right side of their bucket. They are paid $4 if their second ball is black and $0 if their second ball is white.

"High-ability" participants are predicted to reason probabilistically. The game is structured so that it is optimal for these participants to *switch* to the right side of their bucket if their first ball successfully earned $4, and to *stay* with the left side of the bucket if their first ball was unsuccessful. "Low-ability" participants are predicted to make decisions consistent with a simple reinforcement-learning heuristic and an affective response to success, rather than probabilistic reasoning. Low-ability participants are predicted to *stay* with the left side after an initial success, and *switch* to the right side after an unsuccessful left-side draw. The bucket game thus allows us to identify types of individuals based on their proclivity to switch or stay when they are successful.

After 20 iterations of the bucket game, we begin part two of our experiment: players continue to play the bucket game, but belief elicitation is introduced. In our main treatments, half of the participants are exposed to an introspection mechanism; after observing their first ball, participants are asked to report the probability that they have been given Bucket A. The other half are exposed to the SBDM mechanism. This belief elicitation method is incentive-compatible under minimal assumptions about risk preferences but is fairly complex and likely unfamiliar to participants.

To allow for variation in the cognitive cost of forming correct beliefs, we vary the composition of balls in the urn across sessions and the number of balls drawn within the session. A feature of our design is that all participants participate in periods in which two balls are drawn and both a black and white ball are observed. In these periods, the combined signal is uninformative and thus it takes no effort to form a belief. Our design therefore allows us to observe belief accuracy in (i) problems in which signals are informative and beliefs are costly to compute and (ii) problems in which signals are uninformative and beliefs are easy to compute.

Optimal decision-making in the bucket task requires participants to use cognitive skills that are closely related to the assumptions that underpin the truth-telling properties of the SBDM mechanism. Ex-ante, we predicted that, relative to introspection, there would be a greater difference in accuracy between high and low-ability participants. Our results are consistent with these predictions: the average error of a high-ability participant is 37.8 percent smaller than a low-ability participant in the SBDM. However, it is only 2.9 percent smaller in the Introspection mechanism.

The main rationale for using incentive-compatible belief elicitation is to induce participants to think carefully about their beliefs and to provide high-quality information even when calculating beliefs is costly. Thus, we would predict that accuracy in the SBDM will be less sensitive to the cognitive difficulty of the decision task than the Introspection mechanism. Consistent with this second prediction, we find that in the SBDM mechanism, there is only a small difference in accuracy between decision problems in which beliefs are costly to compute and those in which beliefs are easy to compute. By contrast, in the Introspection mechanism, the average error in complex tasks is 87 percent larger than the average error in the easy tasks.

Taken together, our results suggest that while the SBDM mechanism encourages participants to think carefully about their beliefs in difficult elicitation problems, some individuals struggle with understanding the mechanism. This may lead to heterogeneity in belief accuracy across a population based on cognitive ability. Our results help to clarify why earlier studies have found mixed evidence regarding the relative efficiency of the SBDM mechanism and the Introspection mechanism. It also highlight a potential confound in designs that rely on disaggregated beliefs since reporting errors may be correlated systematically with unobservable cognitive ability.

The rest of the paper is arranged as follows. In Section 2 we describe the Stochastic Becker-DeGroot-Marschak mechanism and discuss the existing literature on belief elicitation and cognitive processes. In Section 3 we discuss the experiment, hypotheses, and analysis plan. Results are presented in Section 4.

2 The Stochastic Becker-DeGroot-Marschak mechanism

Consider a participant in an experiment who has a subjective belief about the distribution of a discrete random variable X, with range $\mathcal{X}$. Her true beliefs P_X describes the probability that $X = x$ for each $x \in \mathcal{X}$, and the researcher wants to know belief p that event $P(X = x)$ will occur.

If participants have an aversion to lying, and if there is no cognitive costs from identifying or reporting p, unincentivized introspection will be truth-telling. However, if the researcher is concerned that these conditions are not satisfied, she can use explicit incentives to induce truthful reporting. "Scoring rules" describe a payment schedule based on a participant's reported belief $r \in [0,1]$ and the realisation of the random variable X. For a single realisation of X, a scoring rule S is a mapping $S : [0,1] \times \mathcal{X} \to \mathbb{R}$. This means that $S(r,x)$ is paid when r is reported and outcome x is realized.

For a participant who has utility function u, in which u is a utility function in the class of von Neumann-Morgenstern Expected Utility functions, a rational participant faced with scoring rule S reports $r \in [0,1]$ to maximize $\mathbb{E}u(S(r,X))$ where, by the expected utility assumption,

$$\mathbb{E}u(S(r,X)) = \sum_{x \in \mathcal{X}} u(S(r,x))P(X = x).$$

Using the terminology introduced by Winkler and Murphy (1968), a "proper" scoring rule renders it optimal for risk-neutral agents to report their beliefs truthfully. That is, given a utility function $u(S(r,X)) = S(r,X)$, the scoring rule is "truth-telling" (or "incentive-compatible") in the sense that, for all $P_X \in \mathcal{P}_X$,

$$p \in \arg\max_{r \in [0,1]} \mathbb{E}u(S(r,X)).$$

As the definition suggests, truth-telling may not occur in cases in which $u(S(r,X)) \neq S(r,X)$. This may be problematic when participants have heterogeneous risk preferences that are unobservable to the researcher.[4]

As noted as far back as Smith (1961) and Savage (1971), moving from a deterministic scoring rule to a stochastic one makes it possible to induce truth-telling for all von Neumann-Morgenstern Expected Utility miximizers. Here, we discuss a stochastic scoring rule that has garnered significant interest in the literature: the Stochastic Becker-DeGroot-Marschak mechanism (SBDM).

In the SBDM mechanism, the experimenter presents a participant with a choice under risk, described by lottery $H_A L$, which pays H if Event A occurs and L if not. The participant forms a subjective belief that event A will occur. We denote this subjective probability p. The participant is then asked to issue report $r \in [0,1]$ about her belief p before making a decision based on her beliefs. A second lottery is created in parallel. A number z is realized from the distribution of random variable Z, which has distribution P_Z on support $[0,1]$. The participant does not know z, but does know that if z falls above her report r she will receive lottery $H_z L$, which makes a high payoff H with probability z.

[4]See Schlag et al. (2013) for a review of scoring rules and techniques that might be used to control for risk aversion. In addition to the stochastic elicitation techniques discussed below, researchers have also tried to separate risk preferences from beliefs econometrically. See, in particular Offerman et al. (2009) and Andersen et al. (2014).

If z falls below r she receives lottery $H_A L$. The lotteries therefore offer identical payoffs with different probabilities. It is in the participant's best interest to report $r = p$, because a report of $r \neq p$ might mean the participant receives the less desirable lottery.

By construction, the SBDM uses the same two payoffs for the subjective and objective lotteries and thus the particular cardinal values assigned to the high and low payoffs are not predicted to influence reports. As a result, the SBDM induces truth-telling under minimal assumptions about preferences; namely, that $u(S(r, X))$ are consistent with stochastic dominance and probabilistic sophistication (Karni, 2009). As per Machina and Schmeidler (1992), probabilistic sophistication means that a participant will rank lotteries according to the implied probability distribution over outcomes. "Stochastic dominance" is the condition that a participant has preference relation $\succeq$ over lotteries such that $H_q L \succeq H_{q'} L$ for all $H > L$ if and only if $q \geq q'$.

2.1 Cognitive heterogeneity and belief elicitation

Although there is little research that empirically studies the interaction between participants' cognitive abilities and their reporting behavior in belief elicitation mechanisms, a few papers suggest that cognition may influence behavior in the SBDM. Hao and Houser (2012) evaluate two implementations of the SBDM mechanism: the standard implementation in which a participant directly reports her beliefs, and an ascending clock mechanism.[5] While both mechanisms are incentive-compatible, the ascending clock mechanism is also obviously strategy-proof and more easily understood by cognitively limited agents (Li, 2017). Thus, differences in the quality of reports across these two implementations would suggest that cognition may influence reporting. Hao and Houser identify "naive" subjects, who report $r \neq p$, and "sophisticated" subjects who report $r = p$. The clock mechanism reduces the sample of naive observations and improves the accuracy of reported beliefs.[6]

Freeman and Mayraz (2019) study how individuals choose between safe and risky lotteries in environments in which (i) they are shown exactly one lottery, (ii) they are given a choice list and one decision is randomly selected for payment, and (iii) they are given a choice list but informed about the decision that will be paid prior to making a choice. The paper finds more risk taking in the individual choice problem relative to the other two formats and conjectures that the choice list provides scaffolding that helps decision makers identify their true preferences. If cognition is an issue in the SBDM mechanism, then we should also find that accuracy in the SBDM is improved with choice

[5] The clock implementation of Hao and Houser (2012) has each participant compete against a dummy bidder that exits the auction at (unknown) probability z. The clock starts at 0 and rises continuously as long as both the participant and the bidder is in the auction. The clock stops when one of the two bidders drop out. If it is the participant, the participant receives lottery $H_z L$. If the dummy bidder exits first, the bidder receives the original lottery $H_A L$.

[6] Although the ascending clock auction leads to better reports in Hao and Houser (2012), it censors data when the dummy bidder wins. We thus use direct reporting methods in this paper.

lists. Holt and Smith (2016) compares behavior between a direct elicitation method and a choice list using an "induced value" urn task in which participants receive one or more signals from an urn. The probability that the balls are drawn from a particular urn can be calculated explicitly via Bayes rule. The paper does not find a significant difference in accuracy between the choice list approach and a direct elicitation implementation based on Holt and Smith (2009). However, it does find that the choice list reduces boundary reports. Burfurd and Wilkening (2018) also does not find differences in accuracy between a direct elicitation format based on Hao and Houser (2012) and a choice list format in urn problems with a single draw. However, Burfurd and Wilkening (2018) does find that there is significant heterogeneity in belief accuracy across individuals even when the probability of an event is objectively known.

In a concurrent project, Schlag and Tremewan (2019) studies a "frequency" based belief elicitation mechanism that can be used when multiple realisations of an outcome are available. The paper compares this mechanism to an SBDM mechanism based on the instructions of Dal Bó et al. (2017). The authors find that the frequency method outperforms the SBDM and that difference in performance is driven by a large number of participants who choose a focal report of 50% in the SBDM mechanism. These focal reports are correlated with poor performance in a cognitive reflection task. We do not find the same large spike of focal reports at 50% in our data, though we use a different analogy-based instruction format and include a control quiz.[7]

3 The experiment

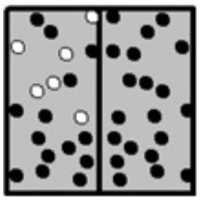
We utilize a two-part design in which we first identify high and low-cognitive ability participants using a computerized "bucket game" and then study how participants respond to different belief elicitation techniques. We describe the bucket game before introducing the treatments.

3.1 The Bucket Game

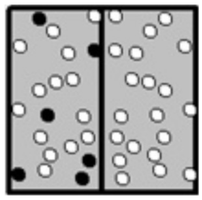
The bucket game is a variant of an urn task introduced in Charness and Levin (2005) and Charness et al. (2007) and is implemented as follows. In each period, a participant is allocated one of two buckets (A or B) with equal probability. Each bucket is divided into a left and a right side and each side holds 20 balls. Subjects are not initially told which bucket they have been given, but are provided an illustration that shows the composition of balls in the two buckets. An example illustration is given in Figure 1. As can be seen, the left hand side of each bucket is composed of a mixture of black and white balls and

[7]Burfurd and Wilkening (2018) find that a control quiz significantly increases accuracy in the SBDM mechanism when using the analogy-based instruction format of Hao and Houser (2012).

there are more black balls in the left hand side of Bucket A than Bucket B. The right hand side of Bucket A is filled with only black balls and the right hand side of Bucket B is filled with only white balls. The buckets used in all treatments share these features.



(a) Bucket A



(b) Bucket B

Figure 1: Illustrations of Bucket A and Bucket B, as presented to participants

In each period, the participant observes the color of a ball that is drawn (with replacement) from the left-hand side of her bucket. If the participant observes a black ball, she receive a stage-one payment of $4. If the ball is white, she receive $0. Next, the participant must decide whether to draw a second ball from the same (left) side of her bucket, or to switch to the other (right) side. The participant receives a payment of $4 if she observes a black ball in this second stage and receives $0 if she observes a white ball.

There are more black balls on the left hand side of Bucket A than Bucket B. Thus, the first draw from the bucket is informative about the bucket that has been allocated to the participant and high-ability participants are predicted to use this information in their choice. If a high-ability participant observes a black ball from the left-hand side of her bucket, the probability she has been given Bucket A will exceed 0.5 and she should choose to switch to the right side of the bucket. If a high-ability participant receives a white ball, the probability that she has been given Bucket A will be less than 0.5 and she should choose to continue to draw from the left side.

However, the game is designed so that the expected value maximizing choice is at odds with an intuitive reinforcement learning heuristic in which a decision maker repeats actions that are successful and changes actions when unsuccessful. When observing a black ball on the first draw, the participant is "successful" and receives $4. Thus, reinforcement learning would predict that the participant will continue to choose left. After observing a white ball, the participant receives $0 and reinforcement learning would predict that the participant switches to the right. Thus, we would predict that low-ability participants who use a reinforcement learning heuristic will always choose the side with the lower expected value.

3.2 Experimental Design and Treatments

We recruited 259 participants using ORSEE (Greiner, 2015) from the experimental economics subject pool at the University of Melbourne. Upon arriving at the lab, participants were greeted by an experimenter and randomly assigned to a computer terminal using a set of bingo balls. Each terminal was assigned one of six potential treatments, summarized in Table 1. These treatments differed in the number of black balls in the left hand side of Bucket A, and in the belief elicitation task that the participant was exposed to in later periods.

A session consisted of three blocks and each block consisted of 20 periods. In the first block, participants in all treatments received computerized instructions describing the bucket game and were required to successfully answer all questions of a computerized quiz before starting the experiment. Participants then played 20 periods of the bucket game. They were informed about whether they successfully drew a black ball from their chosen container in each period.

Treatment	**Belief Elicitation Method (Blocks Two and Three)**	**Number of Black Balls in Left Side of Bucket A**	**n**
SBDM - 14	SBDM	14 of 20	40
SBDM - 12	SBDM	12 of 20	41
Introspection - 14	Introspection	14 of 20	40
Introspection - 12	Introspection	12 of 20	38
No Beliefs - 14	No Elicitation	14 of 20	41
No Beliefs - 12	No Elicitation	12 of 20	39

Table 1: Summary of Treatments

In the second block, we elicited beliefs with the SBDM mechanism in one-third of treatments and with an Introspection mechanism in one-third of treatments. The remaining treatments were not exposed to any belief elicitation mechanism. As with the first block, all participants received computerized instructions at the start of the second block and were required to take a quiz before continuing. The instructions and quiz were identical to the first block in treatments without a belief elicitation mechanism. The instructions in the other treatments explained the belief elicitation task and added additional control questions to ensure participant comprehension. In describing the SBDM mechanism, we use an adaptation of the direct elicitation method developed in Hao and Houser (2012). This set of instructions was shown in Burfurd and Wilkening (2018) to yield high quality data and to be quick to implement relative to alternatives.

After reading the instructions for Block Two, participants played twenty more periods of the bucket game. We elicited beliefs after the participant had observed the draw from the bucket but before they made the decision whether to choose left or right. All beliefs were expressed as the "chance-in-100" the participant has been given Bucket A.

The third block of the experiment was identical to the second block, except that a participant drew two balls with replacement from the urn instead of one. Subjects were paid for each black ball they received from these draws. Instructions for Block Three were short and discussed only the additional draw that the participant received.

To avoid wealth effects and potential hedging strategies, participants are paid in cash for three randomly chosen periods announced at the end of the experiment—one chosen from each of the three blocks. In Block One, the participant's profit for the selected period was the value of her first ball plus the value of her second ball. In Block Two, we used this same payment rule for participants in the Introspection and No Belief treatments. For participants in the SBDM treatments, we 'tossed a coin' to determine whether profit for the second ball was determined by her left/right choice—in which case a second ball was drawn from her nominated side of the bucket—or her beliefs. If her profit was determined by her beliefs, then we used the outcome of the SBDM mechanism to determine payment. Subjects could therefore earn \$0, \$4 or \$8 in Block One, \$0, \$4 or \$8 in Block Two, and \$0, \$4, \$8, or \$12 in Block Three. Participants were allowed to proceed at their own pace through the experiment and most participants completed the experiment in under 45 minutes. Including a show-up fee of \$10, the average payment of a participant was \$24.40 AUD.[8]

3.2.1 Posteriors

To generate variation in the difficulty of the belief updating task, we used two different sets of buckets across the treatments and varied the number of balls drawn within a treatment. In our "high information" treatments, Bucket A contained 14 black balls and Bucket B contained 6 black balls. In Blocks One and Two of this treatment, receiving a single black signal results in a posterior of $\rho' = 0.7$ while receiving two black signals in Block Three results in a posterior of $\rho' = 0.84$. In the other half of the treatments, Bucket A contained 12 black balls and Bucket B contained 8 black balls. In these treatments, receiving a single black signal results in a posterior of $\rho' = 0.6$ and receiving two black signals results in a posterior of $\rho' = .69$.

A feature of our design is that all participants were exposed to periods in which they drew one black ball and one white ball in block three. In these periods, the prior was uninformative and required no Bayesian updating to report the true belief. We conjecture that reporting the correct beliefs was not cognitively challenging in these periods, and we use these periods to test whether accuracy in the Introspection mechanism is influenced by task difficulty.

All bucket games were designed so that posteriors were an equal distance from the prior whether the participant observes a white or a black ball (i.e., the posteriors are 0.7 and 0.3 after receiving a black ball or a white ball in the high information treatments). This symmetry allows us to cleanly aggregate participants' reported beliefs: for example, in Block Two of the high information treatment, a participant who reported $r = 0.5$ has a measured inaccuracy of 0.2 regardless of whether they observed a white or a black ball.

[8] The experiments were run in November and December of 2015. At the time of the experiments, \$1 AUD $\approx$ \$0.72 USD,

The expected payoffs for choosing left and right for each potential posterior are given in Table 2.

Posterior	Observed Color	Probability of Receiving $4 if choosing:		Expected payoff from choosing:		Expected gain from choosing optimally
		Left	Right	Left	Right	
$\rho' = 0.50$	Black-White	0.50	0.50	2.00	2.00	0.00
	White-Black	0.50	0.50	2.00	2.00	0.00
$\rho' \in \{0.40, 0.60\}$	Black	0.52	0.6	2.08	2.40	0.32
	White	0.48	0.4	1.92	1.60	0.32
$\rho' \in \{0.31, 0.69\}$	Black × 2	0.54	0.69	2.15	2.77	0.62
	White × 2	0.46	0.31	1.85	1.23	0.62
$\rho' \in \{0.30, 0.70\}$	Black	0.58	0.7	2.32	2.80	0.48
	White	0.42	0.3	1.68	1.20	0.48
$\rho' \in \{0.16, 0.84\}$	Black × 2	0.64	0.84	2.55	3.38	0.83
	White × 2	0.36	0.16	1.45	0.62	0.83

Notes: all payoffs are in $AUD

Table 2: Expected payoffs from choosing left or right after observing signals in the bucket game

3.3 Statistics and Hypotheses

Our per-period measure of **accuracy** is the absolute error of a participant's reports, relative to the objective Bayesian posterior. Thus, a lower number represents a higher accuracy. When comparing treatments, we use permutation tests that assigns all observations of an individual to either the SBDM or Introspection mechanism randomly in each permutation. All permutation tests are performed 10,000 times and the null hypothesis is that there are no differences between the test groups. Thus, the hypotheses described below are the alternative H1 hypotheses. Although many of these alternative hypotheses would indicate a one-sided test, we report the two-tailed p-values across the paper to avoid confusion and state explicitly whether the test statistic is significant in the proper one-sided or two-sided test.

Sample sizes were chosen to ensure that the permutation tests could reject the null at least 95 percent of the time for an effect size of 5 percentage points when comparing two groups of at least 30 participants. In the power calculations, we used a standard deviation that was twice the standard deviation that was ultimately observed in the full sample permutation tests. Thus, our main tests are well powered.

3.3.1 Heterogeneous response to belief elicitation

We classify participants as "high ability" and "low ability" based on their decisions in the last ten periods of Block One of the experiment. A participant is classified as "high ability" if they made 7 or more correct left/right decisions in periods 11-20. Our type cutoff is set to achieve as close to a median split across high and low-ability types as possible.[9] Based on this classification there are 84 high-ability participants and 75 low-ability participants in our treatments with a belief elicitation mechanism. The proportion of high-ability types is balanced across treatments, with 42 high-ability participants in the Introspection treatments (54 percent of subjects) and 42 high-ability participants in the SBDM treatments (52 percent).

As shown by Karni (2009), the SBDM mechanism is incentive-compatible when individuals' preferences over risk satisfy probabilistic sophistication and stochastic dominance. Thus, for high-ability participants, we would predict high accuracy in the SBDM regardless of the belief updating task.

By contrast, a participant who chooses the incorrect decision in the bucket task is actively choosing an urn with a lower expected value over one with a higher expected value. Such actions violate stochastic dominance. Thus, low-ability participants may have difficulty understanding and interacting with the SBDM mechanism. Using behavior in the Introspection treatments to control for inherent differences in accuracies between the two groups, we predict:

Hypothesis 1 *The SBDM mechanism is more sensitive to cognitive ability than the Introspection mechanism.*

If Hypothesis 1 is true, we should see a larger difference in accuracy between high and low-ability participants in the SBDM mechanism than in the Introspection mechanism. With a null hypothesis that there is no difference between elicitation techniques' sensitivity to participants' cognitive abilities, our test of the hypothesis is a difference-in-difference regression

$$Accuracy = \alpha_0 + \beta_1 HighAbility + \beta_2 SBDM + \beta_3 SBDM * HighAbility, \qquad (1)$$

in which *Accuracy* is the absolute difference between the objective posterior and a reported belief, *High-Ability* is a dummy variable based on the median split of behavior in the baseline game and *SBDM* is the SBDM treatment dummy. We predict that $\beta_3 < 0$, which would indicate that there is greater difference in belief accuracy between high and low-ability participants in the SBDM mechanism than in the Introspection mechanism.

While the Introspection mechanism may be easier for low-ability participants to understand, a concern is that participants may not have an incentive to think carefully about their belief when updating is cognitively costly. This would imply that the quality of data in the introspection mechanism may be strongly dependent on the difficulty of forming correct beliefs.

In our design, participants are exposed to decision problems in which signals are informative and in which Bayesian updating is challenging. Participants are also exposed

[9] Results are qualitatively robust to using a different threshold for the group split or if classification is based on all twenty Block-One decisions.

to simple problems in which signals are uninformative and no Bayesian updating is needed. Using behavior in the SBDM treatments to control for inherent differences in accuracy between these two types of problems, we would predict:

Hypothesis 2 *The Introspection mechanism is more sensitive to task difficulty than the SBDM mechanism.*

To test for Hypothesis 2, we compute the difference in accuracy between decision problems with informative signals and uninformative signals in each of the two mechanisms. We predict that the difference is greater in the Introspection mechanism than in the SBDM mechanism. Thus, our test statistic is given by:

$$Accuracy = \alpha_0 + \beta_1 Info + \beta_2 Introspection + \beta_3 Introspection * Info, \quad (2)$$

in which *Accuracy* is the absolute difference between the objective posterior and a reported belief and *Info* is a dummy variable that is one for any posterior that is different from the 0.5 prior and zero otherwise. We predict that $\beta_3 > 0$ as this would indicate that there is greater variation in report quality under introspection when participants encounter easy versus difficult decision problems.

Combining Hypotheses 1 and 2, we predict that the relative performance of the SBDM is likely to be greatest for high-ability types in problems with informative signals and lowest for low-ability types in problems with uninformative signals. A priori, we cannot order the other two combinations of types and decision problems since the relative importance of mechanism complexity and task difficulty are unknown.

3.3.2 Observer Effects

In the appendix, we use additional treatments with no belief elicitation to test for observer effects in the SBDM mechanism and Introspection mechanisms. An observer effect is a situation in which the introduction of belief elicitation helps participants perform optimally in the underlying task.[10]

To test for observer effects, we compare the proportion of correct left/right choices in our control treatments without belief elicitation to the proportion of correct left/right choices in the SBDM mechanism and Introspection mechanism. As discussed in Rutström

[10]There is mixed evidence that belief elicitation affects game play in other settings. Rutström and Wilcox (2009) compares behavior when participants do not have beliefs elicited, when participants participate in unpaid introspection, and when participants' beliefs are elicited using the QSR. They test the idea that the cognitively "intrusive" QSR might drive a sharper wedge between the "affective process" of belief formation and the "deliberative judgement" reporting process. In the case of the QSR they ultimately reject the hypothesis that belief elicitation does not affect game play, although they note than these effects are concentrated in earlier periods. Nyarko and Schotter (2002) also find that belief elicitation has unintended consequences, and that participants exposed to belief elicitation are more likely to use mixed strategies than pure strategies. However, Guerra and Zizzo (2004) and others have found no evidence that elicitation affects decision-making.

and Wilcox (2009), observer effects are predicted to occur when the belief elicitation mechanism shifts a decision maker away from "affective process" of belief formation and towards the "deliberative judgement" reporting process. Since the instructions for the SBDM required an explicit discussion of probability and incentives, we predicted that this mechanism would generate an observer effect and that this effect could improve belief accuracy in the belief elicitation task. As shown in the appendix, we find no observer effects in any of our treatments, and the proportion of correct left/right choices in treatments using the SBDM mechanism and the Introspection mechanism are very similar both before and after beliefs are introduced.[11]

4 Results

Result 1 *Consistent with Hypothesis 1, the SBDM mechanism is more sensitive to cognitive ability than the Introspection mechanism.*

Table 3 reports the average accuracy of beliefs under the SBDM mechanism and the Introspection mechanism for (i) high-ability participants, (ii) low-ability participants, and (iii) both high- and low-ability participants combined. We report average accuracy for each informative posterior pair starting with the most informative posteriors and ending with the least informative signal. Thus, for instance, the $\rho' \in \{0.16, 0.84\}$ column corresponds to data from block three of the high-information treatments when a participant has drawn either two black balls or two white balls. We then show average accuracy for all informative signals combined and for the uninformative case in which one black ball and one white ball is drawn. Finally, average accuracy over all decision problems is shown in the last column.

As seen in the last column, accuracy for high ability participants in the SBDM mechanism is 10.45 while accuracy for low-ability participants is 16.30. Thus, high-ability participants are 37.8 percent more accurate than low types in the SBDM mechanism, a 5.85 point difference.

Accuracy for high-ability participants in the Introspection mechanism is 14.58 while accuracy for low-ability participants is 15.01. Thus, high-ability participants are only 2.9 percent more accurate than low types in the Introspection mechanism, a 0.44 point difference. The difference in accuracy between high and low-ability types is significantly greater in the SBDM than in introspection in a permutation test using the specification provided in Equation 1 (p-value = .066).[12]

[11] In periods 11-20, participants make correct left/right choices 66.05 percent of the time in the SBDM treatments, 65.25 percent of the time in Introspection treatments, and 64.88 percent of the time in the No Belief treatments. In the ten periods after beliefs were introduced, participants make correct left/right choices 70.74 percent of the time in the SBDM treatments, 72.56 percent of the time in Introspection treatments, and 69.50 percent of the time in the No Belief treatments.

[12] As noted in the design section, we report the two-sided p-value throughout the paper and consider

Belief Elicitation Method	Cognitive Type	Informative Signals				All Informative Signals	Uninformative Signals	All Signals
		$\rho' \in \{0.16, 0.84\}$	$\rho' \in \{0.30, 0.70\}$	$\rho' \in \{0.31, 0.69\}$	$\rho' \in \{0.40, 0.60\}$	$\rho' \neq 0.5$	$\rho' = 0.5$	
SBDM	High	8.66	11.46	14.49	9.10	10.57	10.06	10.45
Introspection	High	15.34	17.02	15.26	15.37	15.93	9.82	14.58
- Permutation Test:		(p-value = 0.000)	(p-value = 0.128)	(p-value = 0.836)	(p-value = 0.111)	(p-value = 0.015)	(p-value = 0.938)	(p-value = 0.050)
SBDM	Low	16.30	18.22	19.30	14.44	16.85	14.42	16.30
Introspection	Low	16.91	18.68	19.39	15.06	17.26	7.77	15.01
- Permutation Test:		(p-value = 0.834)	(p-value = 0.888)	(p-value = 0.977)	(p-value = 0.887)	(p-value = 0.853)	(p-value = 0.040)	(p-value = 0.560)
SBDM	Full Sample	12.16	14.84	16.80	11.58	13.60	12.15	13.27
Introspection	Full Sample	15.97	17.73	17.34	15.22	16.54	8.84	14.78
- Permutation Test:		(p-value = 0.051)	(p-value = 0.242)	(p-value = 0.840)	(p-value = 0.213)	(p-value = 0.061)	(p-value = 0.133)	(p-value = 0.315)

Table 3: Average accuracy of beliefs under the SBDM mechanism and the Introspection mechanism for (i) high-ability participants, (ii) low-ability participants, and (iii) both high- and low-ability participants combined. The reported p-values are based on permutation tests using 10,000 iterations in which the subset of participants is held fixed and participants are randomly allocated to the SBDM or Introspection mechanism in each iteration of a regression on the treatment effect. The null hypothesis is that the treatment coefficient is equal to 0 (i.e. that there is no difference in accuracy between the SBDM and Introspection). The two-sided test statistic is reported.

Result 2 *Consistent with Hypothesis 2, the difference in accuracy with uninformative versus informative signals is greater in the Introspection mechanism than in the SBDM mechanism.*

Referring back to Table 3 and looking at the rows corresponding to the full sample, average accuracy in the SBDM is 12.15 in problems in which the signal is uninformative and 13.60 in problems in which the signal is informative. Accuracy under Introspection is 8.84 when the signal is uninformative and 16.54 when the signal is informative. Thus, under the SBDM, the difference in accuracy is 1.45 while it is 7.70 under Introspection. The difference in difference is significant in a permutation test using the specification provided in equation 2 (p-value = .004).

Having found support that cognitive ability interacts with the SBDM at the aggregate level, and that accuracy in the Introspection treatment varies with task difficulty, we now take a deeper look at the data to understand what is driving the differences in mechanism performance. We begin by comparing high-ability participants' responses to both mechanisms when signals are informative.

Result 3 *Reports in the SBDM mechanism are more accurate than reports in the Introspection mechanism for high-ability participants when signals are informative. The difference is due in part to the larger number of focal reports observed in the Introspection mechanism.*

As seen by comparing the first two rows of Table 3, the SBDM is more accurate for high-ability participants when we combine the data from all the informative priors (p-value = 0.015) and when we combine all data from the treatments with belief elicitation p-value = 0.050). Thus reports in the SBDM mechanism are more accurate than reports in the Introspection mechanism for high-ability participants when signals are informative.

Figure 2 shows the distribution of reports for high-ability participants for each of the eight informative signals under the SBDM mechanism and Introspection. As can be seen in each panel, the Introspection mechanism has more focal reports of 0, 50, and 100 than the SBDM mechanism. Aggregating over the eight informative priors, focal reports occur in 41.1 percent of cases in the Introspection mechanism and in only 19.6 percent of cases in the SBDM mechanism. This difference is significant when we compare the average proportion of focal reports made in the two mechanisms in a permutation test using data from periods with uninformative signals (p-value = 0.017). Excluding the focal reports, the average error of high-ability participants in the Introspection mechanism is 7.81 in periods with an informative prior while the average error in the SBDM mechanism is 8.72 in the same periods. Thus, the larger number of focal reports in the Introspection mechanism appears to be the main driver of differences between the two mechanisms for high-ability participants.

a p-value < 0.1 to be significant when evaluating tests (like this one) that are one sided.

Result 4 *There is no significant difference in report accuracy between the SBDM mechanism and Introspection for high-ability participants when signals are uninformative. Participants in the Introspection mechanism make significantly more correct and incorrect focal reports.*

In periods with uninformative signals, accuracy for high-ability participants is 10.06 in the SBDM mechanism and 9.82 in the Introspection mechanism and there is no significant difference between the two mechanisms (p-value = 0.938). In the Introspection mechanism 71.24 percent of participants report the correct belief of 50 while only 51.83 percent of participants report the correct belief in the SBDM. This difference in correct focal reports is significant (p-value = 0.044). However, other focal reports are also common under introspection. In particular, 11.29 percent of reports are extreme reports of 0 or 100 in the Introspection mechanism while only 2.62 percent of reports are extreme reports of 0 of 100 in the SBDM mechanism.

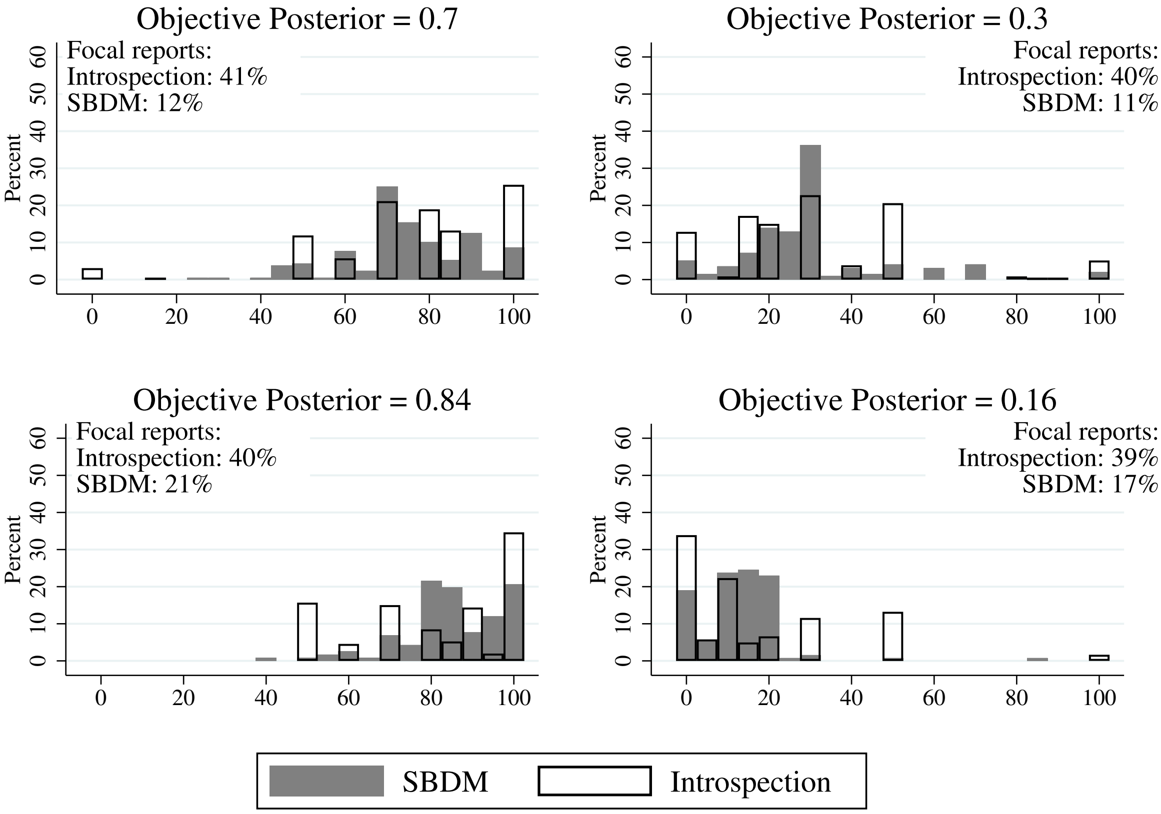
Result 5 *There is no significant difference in the accuracy of beliefs for low-ability participants in decision problems with informative signals. However, Introspection outperforms the SBDM for low-ability participants when signals are uninformative.*

As seen by comparing rows 3 and 4 of Table 3, low-ability participants are slightly more accurate in the SBDM mechanism than the Introspection mechanism in each of the four cases with informative signals. However, none of these differences are significant.

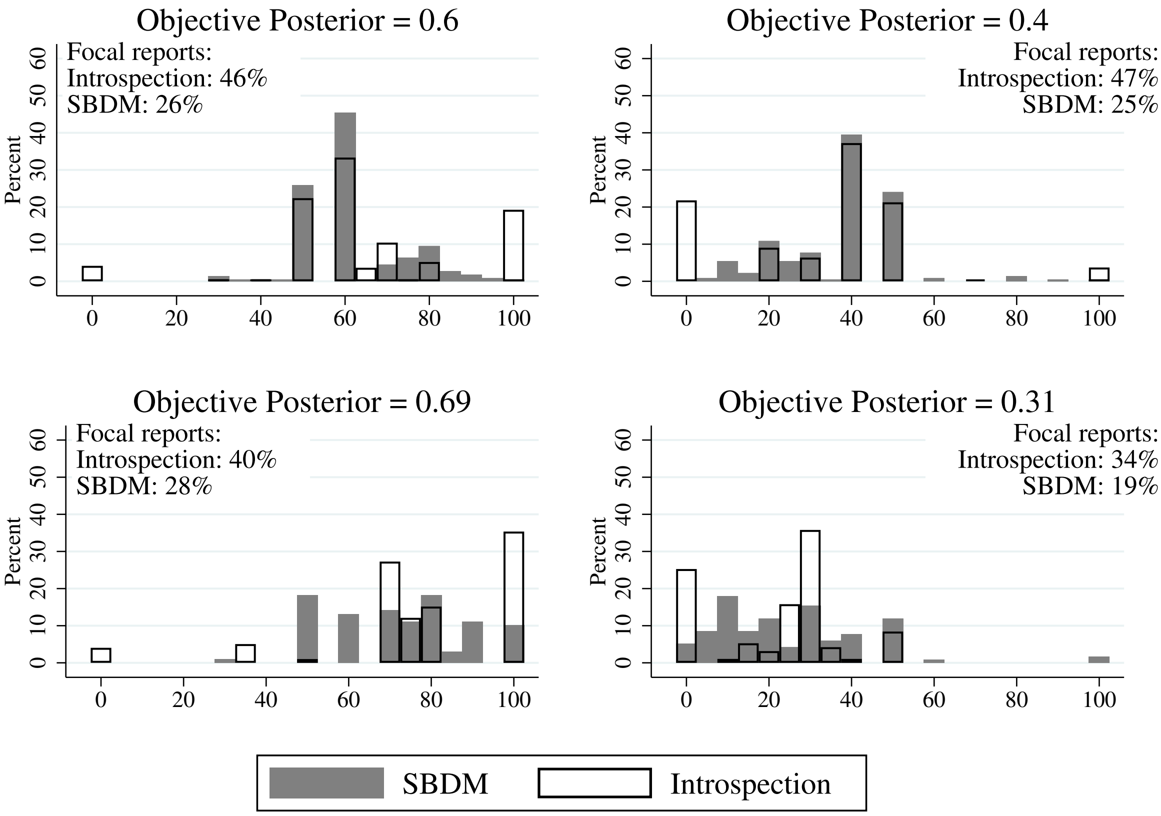
When the signals are not informative, the average error in the Introspection mechanism is 7.77 while the average error in the SBDM mechanism is 14.42. This difference is significant in a permutation test (p-value = 0.040). Looking at the distribution of reports, a correct report of 50 is made in 79.47 percent of cases in the Introspection mechanism and in only 38.52 percent of cases in the SBDM mechanism. This difference is significant when we compare the average proportion of correct reports made in the two mechanisms in a permutation test using data from periods with uninformative signals (p-value < 0.001). The strong reduction in correct focal reports of 50 suggests that some individuals do not understand the truth-telling properties of the SBDM mechanism and misreport as a result.

5 Conclusion

The Stochastic Becker-DeGroot-Marschak (SBDM) mechanism is a theoretically elegant way of eliciting incentive-compatible beliefs under a variety of risk preferences. However, the mechanism is complex and there is concern that some participants may misunderstand the incentive properties of the mechanism. We use a two-part design in which we identify participants with high and low cognitive ability and elicit their beliefs in both



(a) High information treatments with 14 black balls in the left side of Bucket A



(b) Low information treatments with 12 black balls in the left side of Bucket A

Figure 2: Distribution of reported beliefs by high-type participants

easy and hard decision problems. Relative to Introspection, there is less variation in belief accuracy between easy and hard problems in the SBDM mechanism. However, there is a greater difference in the accuracy of reports between "high ability" participants and "low ability" participants. These results suggest that while the SBDM mechanism encourages individuals to think more carefully about beliefs, it is more sensitive to cognitive ability. Our results show that mechanism complexity is an important consideration in developing elicitation mechanisms, and identifies cognition as an important consideration when interpreting elicited beliefs.

By identifying different channels by which errors occur in the two mechanism, we can better understand the mixed results from earlier studies that compare the two mechanisms. In particular, our finding that accuracy in the Introspection mechanism varies with belief-task difficulty implies that any horse race between the two mechanisms is likely to be strongly task dependent, and that cognitive costs may be an important consideration in whether to offer explicit incentives for beliefs.

Our finding that accuracy in the SBDM varies with participants' cognitive abilities suggests that researchers should be cautious when using individual beliefs to identify types. For example, in the literatures on overconfidence, it is common to use the difference between an agent's true ability and their belief about this ability as a proxy for overconfidence (Cesarini et al., 2006; Urbig et al., 2009). If errors are correlated with cognitive ability, then individuals who are assigned to the overconfident group may also include a large set of low-ability types who struggle to optimize in other situations.

It is an open question as to how to improve the SBDM to reduce the impact of cognition. Holt and Smith (2016) and Burfurd and Wilkening (2018) suggest that choice lists can reduce focal reports but neither paper finds accuracy improvements from using multiple choice lists. By contrast, Hao and Houser (2012) suggests that combinatorial clocks might play an important role if researchers can overcome the censoring which results from using a single ascending clock. This result is consistent with the notion of obviously strategy-proof mechanisms (Li, 2017). One potential solution would be to conduct both an ascending and decreasing clock auctions against a dummy player with a common cutoff point $\hat{p}$ and pay the participant for the outcome of one of these clock auctions. In the ascending clock auction, the clock probability z goes from zero to one and the participant receives $H_A L$ if z reaches $\hat{p}$. If the participant drops out, she receives $H_{\hat{p}} L$. In the descending clock auction, the participant receives $H_{\hat{p}} L$ if z reaches $\hat{p}$ and she receives $H_A L$ if she drops out. In both mechanisms, it is a dominant strategy to drop out at ones true value. Further, at least one of the two mechanisms will have no censoring.

It is also an open question as to how cognition interacts with other proper scoring rules, particularly those that are robust to heterogeneity in risk preferences. The initial results in Schlag and Tremewan (2019) suggest that the frequency method may be easier to understand than the SBDM since it elicits beliefs in terms of natural frequencies. Thus this method has promise when multiple realisations of an outcome are available.

Bibliography

Allen, F. (1987). Notes-discovering personal probabilities when utility functions are unknown. *Management Science 33*(4), 542–544.

Andersen, S., J. Fountain, G. W. Harrison, and E. E. Rutström (2014). Estimating subjective probabilities. *Journal of Risk and Uncertainty 48*(3), 207–229.

Becker, G. M., M. H. DeGroot, and J. Marschak (1964). Measuring utility by a single-response sequential method. *Behavioral Science 9*(3), 226–232.

Burfurd, I. and T. Wilkening (2018). Experimental guidance for eliciting beliefs with the Stochastic Becker–DeGroot–Marschak mechanism. *Journal of the Economic Science Association 4*(1), 15–28.

Cesarini, D., O. Sandewall, and M. Johannesson (2006). Confidence interval estimation tasks and the economics of overconfidence. *Journal of Economic Behavior & Organization 61*(3), 453–470.

Charness, G., E. Karni, and D. Levin (2007). Individual and group decision making under risk: An experimental study of Bayesian updating and violations of first-order stochastic dominance. *Journal of Risk and Uncertainty 35*(2), 129–148.

Charness, G. and D. Levin (2005). When optimal choices feel wrong: A laboratory study of Bayesian updating, complexity, and affect. *The American Economic Review 95*(4), 1300–1309.

Dal Bó, E., P. Dal Bó, and E. Eyster (2017). The demand for bad policy when voters underappreciate equilibrium effects. *The Review of Economic Studies 85*(2), 964–998.

Ducharme, W. M. and M. L. Donnell (1973). Intrasubject comparison of four response modes for "subjective probability" assessment. *Organizational Behavior and Human Performance 10*(1), 108–117.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with orsee. *Journal of the Economic Science Association 1*(1), 114–125.

Grether, D. M. (1981). Financial incentive effects and individual decision-making. Technical report, California Institute of Technology, Workiing Papers 401.

Guerra, G. and D. J. Zizzo (2004). Trust responsiveness and beliefs. *Journal of Economic Behavior & Organization 55*(1), 25–30.

Hao, L. and D. Houser (2012). Belief elicitation in the presence of naïve respondents: An experimental study. *Journal of Risk and Uncertainty 44*(2), 161–180.

Hollard, G., S. Massoni, and J.-C. Vergnaud (2016). In search of good probability assessors: An experimental comparison of elicitation rules for confidence judgments. *Theory and Decision 80*(3), 363–387.

Holt, C. A. (2006). *Markets, games, and strategic behavior: recipes for interactive learning.* Pearson Addison Wesley.

Holt, C. A. and A. M. Smith (2009). An update on Bayesian updating. *Journal of Economic Behavior & Organization 69*(2), 125–134.

Holt, C. A. and A. M. Smith (2016). Belief elicitation with a synchronized lottery choice menu that is invariant to risk attitudes. *American Economic Journal: Microeconomics 8*(1), 110–39.

Karni, E. (2009). A mechanism for eliciting probabilities. *Econometrica 77*(2), 603–606.

Li, S. (2017). Obviously strategy-proof mechanisms. *American Economic Review 107*(11), 3257–3287.

Machina, M. J. and D. Schmeidler (1992). A more robust definition of subjective probability. *Econometrica 60*(4), 745–780.

Nyarko, Y. and A. Schotter (2002). An experimental study of belief learning using elicited beliefs. *Econometrica 70*(3), 971–1005.

Offerman, T., J. Sonnemans, G. van de Kuilen, and P. P. Wakker (2009). A truth serum for non-Bayesians: Correcting proper scoring rules for risk attitudes. *The Review of Economic Studies 76*(4), 1461–1489.

Rutström, E. E. and N. T. Wilcox (2009). Stated beliefs versus inferred beliefs: A methodological inquiry and experimental test. *Games and Economic Behavior 67*(2), 616–632.

Savage, L. J. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association 66*(336), 783–801.

Schlag, K. H. and J. Tremewan (2019). Simple belief elicitation. Technical report, mimeo.

Schlag, K. H., J. Tremewan, and J. J. Van der Weele (2013). A penny for your thoughts: a survey of methods for eliciting beliefs. *Experimental Economics 18*(3), 1–34.

Schotter, A. and I. Trevino (2014). Belief elicitation in the laboratory. *Annual Review of Economics 6*(1), 103–128.

Smith, C. A. (1961). Consistency in statistical inference and decision. *Journal of the Royal Statistical Society. Series B (Methodological) 23*(1), 1–37.

Trautmann, S. T. and G. Kuilen (2015). Belief elicitation: A horse race among truth serums. *The Economic Journal 125*(589), 2116–2135.

Urbig, D., J. Stauf, and U. Weitzel (2009). What is your level of overconfidence? A strictly incentive compatible measurement of absolute and relative overconfidence. Technical report, Utrecht School of Economics, Working Papers 09-20.

Winkler, R. L. and A. H. Murphy (1968). "Good" probability assessors. *Journal of Applied Meteorology 7*(5), 751–758.

Appendix A: Observer Effects

An observer effect is a case in which the introduction of belief elicitation changes how a participant understands the underlying game. As discussed in Rutström and Wilcox (2009), observer effects are predicted to occur when the belief elicitation mechanism shifts a decision maker away from "affective process" of belief formation and towards the "deliberative judgement" reporting process. Since the instructions for the SBDM required an explicit discussion of probability and incentives, we predicted that this mechanism would generate an observer effect and that this effect could improve belief accuracy in the belief elicitation task. We find:

Result 6 *There is no statistically significant observer effect in the data.*

Support for Result 6 is provided in Table 4, which shows the proportion of correct left/right choices in blocks one and two of the experiment with the data split into subsets of 10 periods. We focus the analysis on Periods 11-30 since these are the ten periods directly before and after the introduction of beliefs.

If an observer effect exists in the data, we would expect to see a larger improvement in the proportion of correct left/right choices at the start of Block 2 in the treatments in which beliefs have been introduced relative to treatments in which belief elicitation did not occur. There is no such pattern in the data: in the treatments with no belief elicitation, participants make mistake in 35.1 percent of cases in Periods 11-20 and in 30.5 percent of cases in Periods 21-30. This difference of 4.63 percentage points is not significantly different than the difference of 4.7 percentage points observed in the SBDM mechanism in a difference-in-difference permutation tests in which we restrict data to periods 11-30 (p-value = 0.898). It is also not different than the difference of 7.3 percentage points observed in the Introspection mechanism (p-value = 0.533).

Finally, we note that in Block Two, the mistake rates in the SBDM mechanism are not significantly different than the Introspection mechanism (p-value = 0.761). This suggests that differences in belief accuracy between these treatments is unlikely to be driven by differences in their understanding of the bucket game.

	Block One		**Block Two**	
	Periods 1-10	**Periods 11-20**	**Periods 21-30**	**Periods 31-40**
No Belief Elicitation	0.405	0.351	0.305	0.346
SBDM	0.335	0.340	0.293	0.275
Introspection	0.400	0.347	0.274	0.274

Table 4: Comparison of mistake rates

Appendix B: Additional Figures

In the main text, we provided the histogram of reported beliefs for the high-ability participants across all of the informative priors. In this appendix, we provide the histograms of reported beliefs for the other cases. Figure 3 shows the reported beliefs of high and low-ability participants in the case of an uninformed prior of 0.5 for both the SBDM mechanism and the Introspection mechanism using data from both the high-information treatments with 14 black balls in the left side of Bucket A and the low-information treatments with 12 black balls. Figure 4 shows the reported beliefs for the low-type participants across the eight potential informative posteriors. As in the main text, the posteriors are grouped into the high-information treatments and low-information treatments.

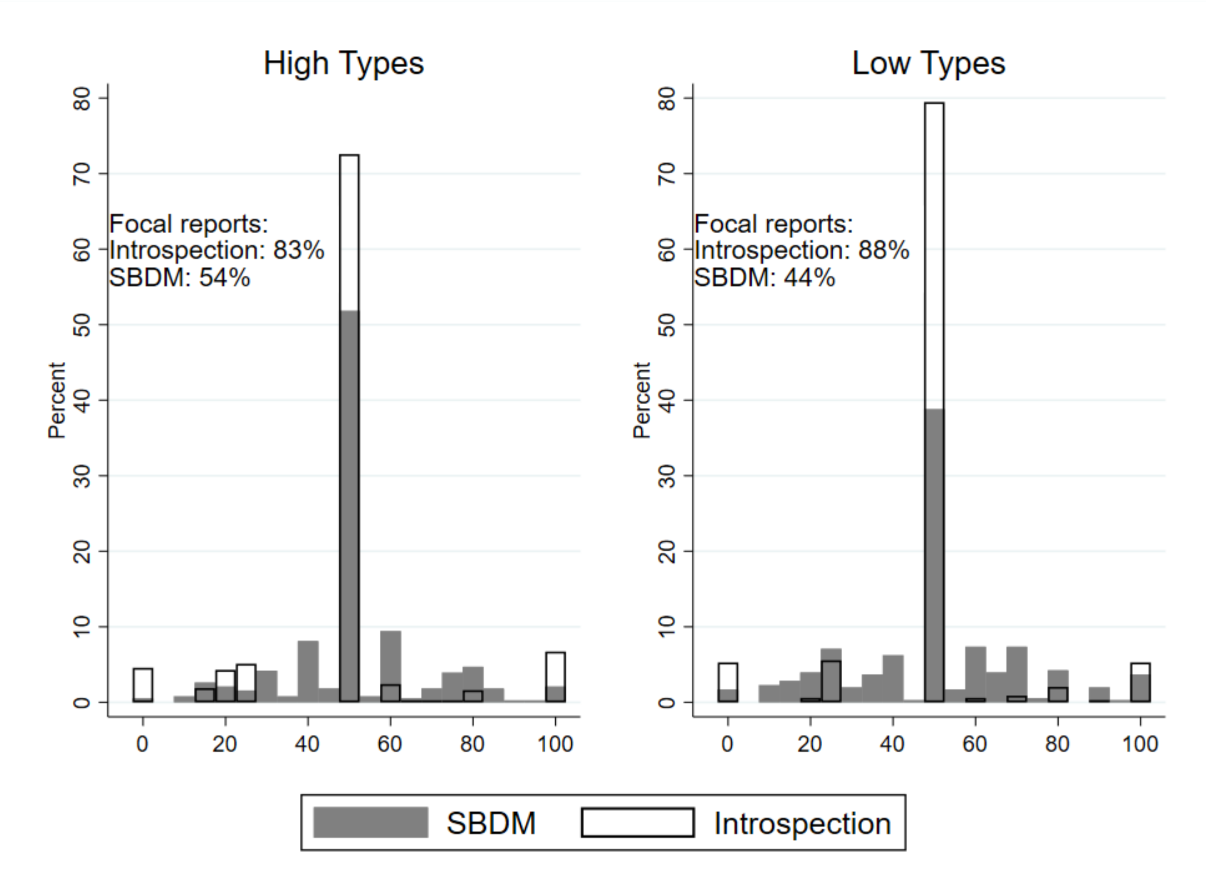
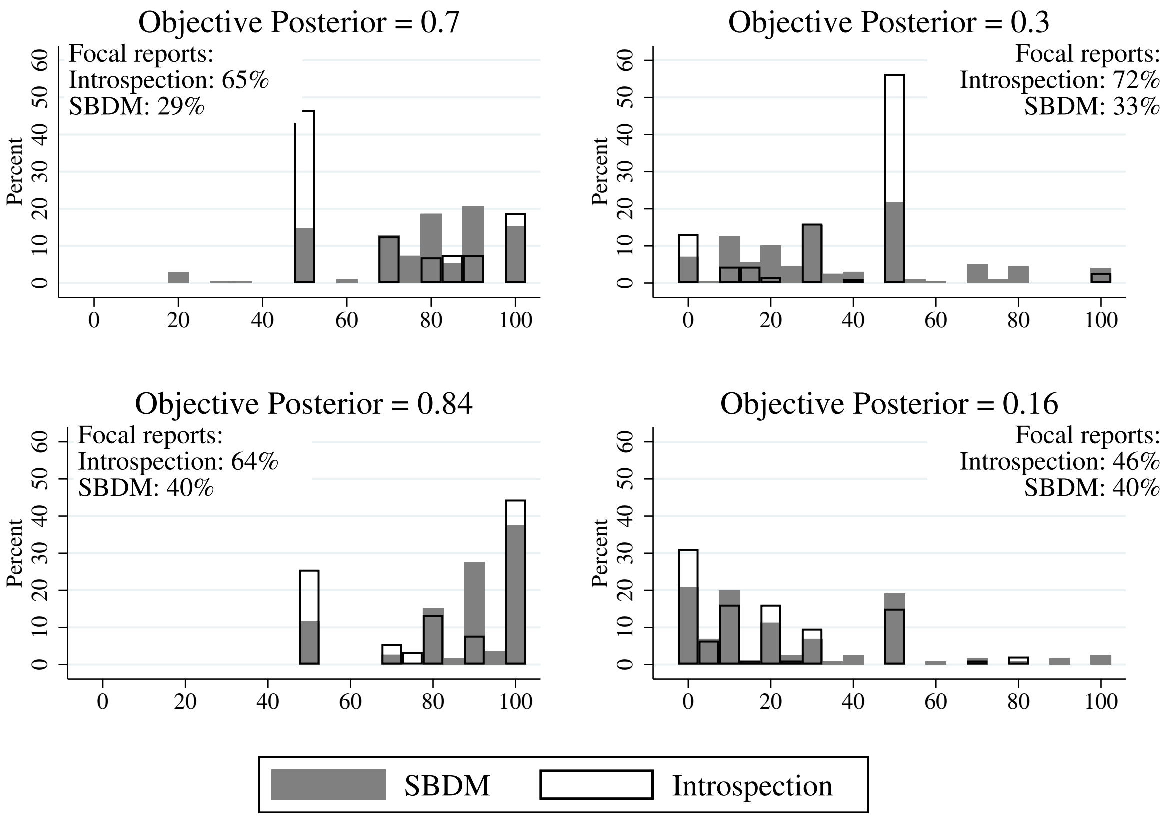
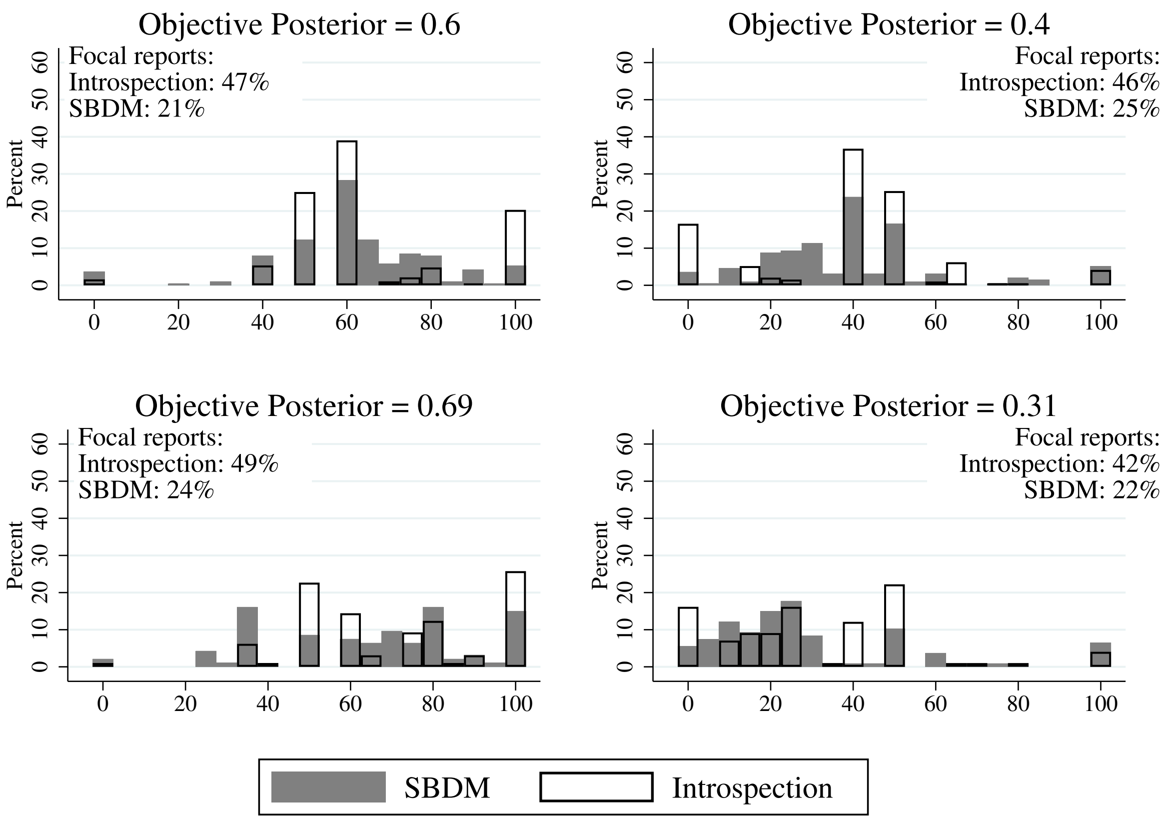


Figure 3: Distribution of reported beliefs when the Posterior is 0.50



(a) High information treatments with 14 black balls in the left side of Bucket A



(b) Low information treatments with 12 black balls in the left side of Bucket A

Figure 4: Distribution of reported beliefs by low-type participants

Appendix C: Instructions

A Instructions and Quizzes

The experiment included 3 blocks of 20 periods, which were referred to in the instructions as Experiments 1, 2 and 3. Statements in parentheses and italics provide additional details or discuss differences between the treatments and do not form part of the experiment instructions.

Experiment One

Thank you for choosing to participate today. We appreciate your time. This experiment is an opportunity to earn money. You will be paid in cash at the end of the experiment. You will be paid a $10 attendance fee. You will also receive payments based on the outcome of three experiments. You will not learn your total payoff until the end of the experiment.

There is a very short, anonymous questionnaire at the end of the experiment. You will be paid when the questionnaire is completed.

If you have any questions during the experiment, please sit quietly and raise your hand. An experiment assistant will be with you as soon as possible.

Payment for the first experiment: You will play the first experiment 20 times. Each repetition is called a "period." In each period you will get a payoff of $0, $4, or $8. At the end of the experiment, 1 of the 20 periods will be chosen randomly by the computer. Each period is equally likely to be chosen. Your cash payment for the first experiment will be your payoff in the randomly chosen period.

*(In bold text:)*Although you will play 20 periods in the first experiment, you are only paid in cash for the payoff you earn in a single period.

You are going to participate in a decision-making task, which is referred to as the "Choose-A-Side Game." There are two buckets: Bucket A and Bucket B. Each bucket contains 40 balls. Each bucket is divided in half, with 20 balls in each side.

There is a 50-in-100 chance (50% chance) that you have been given Bucket A. The left side of Bucket A contains 12 black balls and 8 white balls. The right side of Bucket A contains 20 black balls and 0 white balls.

(Stylized illustration of Bucket A: a rectangle divided vertically in two, with black or white dots to illustrate the ratio of black and white balls in each half of the bucket.)

There is a 50-in-100 chance (50% chance) that you have been given Bucket B. The left side of Bucket A contains 8 black balls and 12 white balls. The right side of Bucket A contains 0 black balls and 20 white balls. (The buckets and balls are all computerized.)

(*Stylized illustration of Bucket B: a rectangle divided vertically in two, with black or white dots to illustrate the ratio of black and white balls in each half of the bucket.*)

One of the buckets will be randomly chosen by the computer. Both buckets have an equal chance of being chosen. This means that both buckets have a 50-in-100 chance of being chosen (50%). (You might imagine that the computer tosses a coin to decide which bucket will be used.) You will not be told which bucket has been chosen by the computer. The computer will randomly select a ball from the left hand side of your bucket. Each ball has an equal chance of being chosen. You will be told the colour of the ball. After you see the ball, it is put back in the left hand side of your bucket. If the ball is black, you receive $4. If it is white, you receive $0 (nothing). This is your Stage-1 payoff.

You then have a second chance to draw a ball from your bucket. As before, black balls are worth $4. White balls are worth $0 (nothing). You must decide whether you would like the computer to draw the ball from the left hand side of your bucket, or the right hand side. The computer randomly selects a ball from the side you choose. If it is black, you receive $4. If it is white, you receive $0 (nothing). This is your Stage-2 payoff.

Your payoff for the period is your Stage-1 payoff plus your Stage-2 payoff. In total you might have a payoff of $0, $4, or $8 across both stages of the Choose-A-Side Game.

In each period there is a 50-in-100 (50%) chance of being given Bucket A or Bucket B. Your bucket is randomly determined by the computer and is not affected by the bucket you have been given in previous periods.

Summary: Choose-A-Side Game

There are 2 buckets, Bucket A and Bucket B. Each bucket has a 50-in-100 chance (50%) of being chosen. Each bucket is divided in half, with 20 balls in each half. The computer randomly selects a bucket for you. You do not know which bucket you have been given. You will see a randomly chosen ball from the left-hand side of your bucket. If it is black, your Stage-1 payoff is $4. If it is white, your payoff is $0 (nothing). You then choose whether you want a second ball drawn from the left or right side of your bucket. The computer draws a ball from your chosen side. If it is black, your Stage-2 payoff is $4. If it is white, your payoff is $0 (nothing). Your period payoff is your Stage 1 plus your Stage 2 payoff. In each period you might get a payoff of $0 (nothing), $4, or $8 in the Choose-A-Side Game. 1 of the 20 periods will be randomly chosen. Each period has an equal (1-in-20) chance of being chosen. Your will be paid your earnings from that period in cash at the end of the experiment.

When you have finished Experiment 1 you will be given instructions for a second experiment.

Quiz

At the start of a period the computer randomly selects a bucket for you.

1. What is the chance-in-100 that you get Bucket A? (50)
2. What is the chance-in-100 that you get Bucket B? (50)

The bucket has 20 balls in each side; 40 in total. The computer shows you a ball from the left-hand side of your bucket, tells you its colour, and tells you whether it is worth $4 or $0. This is your payoff for Stage 1. The computer puts the ball back in your bucket. The computer asks whether you want the next ball drawn from the left-hand side or the right-hand side of the bucket.

3. How many balls are there in the left-hand side? (20)

4. How many balls are there in the right-hand side? (20)

The computer draws a ball from the side you choose, tells you its colour, and tells you if you have won $4 or $0. This is your payoff for Stage 2.

5. What is the minimum payoff possible in a period (Stage 1 + Stage 2)? (0)

6. What is the maximum payoff possible in a period (Stage 1 + Stage 2)? (8)

You then finish the period.

7. How many periods are there in this experiment? (20)

8. Do you receive a cash payment for your payoff in every period? (No)

9. Every period has a 1-in-? chance of being paid? (20)

10. When the next period starts, what is the chance-in-100 that you get Bucket A? (50)

11. What is the chance-in-100 that you get Bucket B? (50)

(Experiment begins when all questions are answered correctly. At the end of Experiment One:)

Thank you! You have now played 20 periods and finished the first experiment. At the end of the third experiment you will find out which period was randomly chosen. You will be paid your payoff from the randomly chosen period. You will be pain in cash. You will now read instructions for the second experiment.

Experiment Two

You will play the second experiment 20 times. Each repetition is called a 'period.' In each period you get a payoff of $0, $4, or $8. At the end of the second experiment, 1 of the 20 periods will be chosen randomly by the computer. Each period is equally likely to be chosen. Your cash payment for the second experiment will be your payoff in the randomly chosen period. Your total payment today will include:

- Your show-up fee of $10
- A cash payment for a randomly chosen period from the first experiment

- A cash payment for a randomly chosen period from the second experiment
- A cash payment for a randomly chosen period from the third experiment

Although you will play 20 periods in this second experiment, you only receive cash for your payoff from a single period.

The set-up for Experiment 2 is the same as Experiment 1. There are two buckets: Bucket A and Bucket B. Each bucket contains 40 balls. Each bucket is divided in half, with 20 balls in each side.

There is a 50-in-100 chance (50% chance) that you have been given Bucket A. The left side of Bucket A contains 12 black balls and 8 white balls. The right side of Bucket A contains 20 black balls and 0 white balls.

(*Stylized illustration of Bucket A: a rectangle divided vertically in two, with black or white dots to illustrate the ratio of black and white balls in each half of the bucket.*)

There is a 50-in-100 chance (50% chance) that you have been given Bucket B. The left side of Bucket A contains 8 black balls and 12 white balls. The right side of Bucket A contains 0 black balls and 20 white balls. (The buckets and balls are all computerized.)

(*Stylized illustration of Bucket B: a rectangle divided vertically in two, with black or white dots to illustrate the ratio of black and white balls in each half of the bucket.*)

One of the buckets will be randomly chosen by the computer. Both buckets have an equal (50-in-100) chance of being chosen. You will not be told which bucket has been chosen by the computer. The computer will randomly select a ball from the left hand side of your bucket. Each ball has an equal chance of being chosen. You will be told the colour of the ball. After you see the ball, it is put back in the left hand side of your bucket. If the ball is black, you receive $4. If it is white, you receive $0 (nothing). This is your Stage-1 payoff.

(The three treatments involve different instructions from this point.)

SBDM mechanism treatment

After seeing the colour of the ball, you need to think about the chance that the ball was drawn from Bucket A. This is your "belief" that the ball was drawn from Bucket A. Your "belief" is a number between 0 and 100, to indicate the chance-in-100 that the ball has been drawn from Bucket A.

For example: If you are sure that Bucket A is being used, your belief is that there is a 100-in-100 chance that Bucket A is being used. If you are sure that Bucket A is not being used, your belief is that there is a 0-in-100 chance that Bucket A is being used. If you believe that it is equally likely that Bucket A is being used as Bucket B, then your belief is that there is a 50-in-100 chance that Bucket A is being used. (These are just examples. You can enter any chance-in-100 belief between 0 and 100.)

You then answer 2 questions. Question 1: What is your belief that the ball was drawn from Bucket A? Question 2: Do you want the computer to draw a second ball from the left or right-hand side of your bucket?

The computer then tosses a coin to determine which question is used to determine your Stage-2 payoff. Tails: Question 1. Heads: Question 2. If the computer throws a Heads, your Stage-2 payoff will be determined the same way as Experiment 1 (the Choose-A-Side Game). The computer will draw a ball from the side of the bucket you choose. As before, black balls are worth $4. White balls are worth $0 (nothing).

We will now explain how Stage-2 payoffs are determined if the computer throws "Tails."

In Question 1 you tell the computer your belief (the chance-in-100) that the first ball was drawn from Bucket A. If the computer throws "Tails", this is how we determine your Stage-2 payoff:

The computer creates a Lottery Bag. The computer randomly chooses a number between 0 and 100. Each number is equally likely to be chosen. Although the computer knows this number, you do not. We call this randomly chosen number "?". The computer fills a bag with 100 chips. "?" chips are black, and the rest are white. ?-in-100 chips are black. There are now two ways to get a payoff of $4: The 'Belief about Bucket A Game," and the Lottery Bag Game.

(Table/illustration here.)

THE BELIEF ABOUT BUCKET A GAME: Prize of $4 if the ball was from Bucket A. Prize of $0 if the ball was from Bucket B.	**THE LOTTERY BAG GAME:** Prize of $4 if you draw a black chip. Prize of $0 if you draw a white chip.
Chance-in-100 of winning $4: Belief (chance-in-100) that ball is from Bucket A	**Chance-in-100 of winning $4:** "?"-in-100

The computer knows the chance of winning $4 in the Lottery Bag Game. Based on your reported belief that the ball was drawn from Bucket A, the computer will select the game that gives you the highest chance of winning $4. (If the games give you an equal chance of winning, you will play the Lottery Bag Game.)

You should think carefully about your belief that the ball has been drawn from Bucket A, as the computer will use your reported belief to decide whether you are paid according to your "Belief about Bucket A" or the "Lottery Bag" Game. This experiment might feel very detailed and complicated, but it is set up this way so that it is in your best interest to report your beliefs honestly and carefully. If you make a report that is not your true belief, your payoff might be determined by the Lottery Bag Game when you would prefer to be paid based on your belief that the ball was drawn from Bucket A (or vice-versa).

The best thing you can do is report your belief honestly, so that you are given the game with the highest chance of a payoff of $4.

Summary: Experiment 2

You have a 50-in-100 chance of being given Bucket A or Bucket B in each period. You will be shown a ball from the left-hand side of your bucket. You will answer 2 questions. Question 1: What is your belief that the ball was drawn from Bucket A? Question 2: Do you want the computer to draw a second ball from the left or right side of your bucket? The computer then tosses a coin to determine which question is used to determine your Stage-2 payoff. Tails: Question 1. Heads: Question 2.

If the computer throws "Heads" your payoff is determined in the same way as Experiment 1. A second ball will be drawn from your bucket, from the side you choose. If the computer throws "Tails" your payoff will be determined by the "Belief about Bucket A" Game or a Lottery Bag Game. The best thing you can do is report your belief honestly, so that you are given the game with the highest chance of a payoff of $4.

Your period payoff is your Stage-1 payoff plus your Stage-2 payoff. In each period you might get a payoff of $0, $4, or $8. 1 of the 20 periods will be randomly chosen. Each period has an equal (1-in-20) chance of being chosen. Your will be paid your earnings from that period in cash at the end of the experiment.

Quiz

Imagine that you are shown a ball. Based on its colour, you report your belief that there is a 20-in-100 chance that the ball is from Bucket A. The computer flips a coin and it lands on "Tails." The computer creates a Lottery Bag game, and randomly includes 25 black chips. It has a 25-in-100 chance of winning $4. Based on your report, the computer chooses the game that gives you a higher chance of winning $4.

1. Which game will be used to determine your payoff for the period? (Lottery Bag Game)

2. What is your chance-in-100 of winning $4? (25)

3. What is your chance-in-100 of winning $0? (75)

Imagine you start a new period. You are shown a new ball. This time, you believe there is an 81-in-100 chance that the ball was taken from Bucket A... but you make an error! You type 18 by mistake. This is your reported belief.

The computer doesn't know your belief, only your reported belief. The computer thinks you believe there is an 18-in-100 chance of winning $4 in the Belief-About-Bucket-a Game.

The computer creates a Lottery Bag Game and randomly includes 36 black chips. It has a 36-in-100 chance of winning $4.

4. What do you believe is your chance-in-100 of winning $4 if you play the Belief-About-Bucket-A Game? (81)
5. What does the computer think you believe is the chance-in-100 of winning $4 if you play the Belief-About-Bucket-A Game? (18)
6. What is your chance-in-100 of winning $4 if you play the Lottery Bag Game? (36)

Based on your report, the computer chooses the game that it thinks will give you a higher chance of winning $4.

7. Which game will be used to determine your prize for the period? (Lottery Bag Game)

(Experiment begins when all questions are answered correctly.)

Unpaid Introspection Treatment

After seeing the colour of the ball, you need to think about the chance that the ball was drawn from Bucket A. This is your "belief" that the ball was drawn from Bucket A. Your "belief" is a number between 0 and 100 to indicate the chance in 100 that the ball has been drawn from Bucket A.You should think carefully about your belief that the ball has been drawn from Bucket A.

For example: If you are sure that Bucket A is being used, your belief is that there is a 100-in-100 chance that Bucket A is being used. If you are sure that Bucket A is not being used, your belief is that there is a 0-in-100 chance that Bucket A is being used. If you believe that it is equally likely that Bucket A is being used as Bucket B, then your belief is that there is a 50-in-100 chance that Bucket A is being used. (These are just examples. You can enter any chance-in-100 belief between 0 and 100.)

You then answer 2 questions.

Question 1: What is your belief that the ball was drawn from Bucket A

Question 2: Do you want the computer to draw a second ball from the left or right side of your bucket?

The computer randomly selects a ball from the side you choose. If it is black, you receive $4. If it is white, you receive $0 (nothing). This is your Stage-2 payoff. Your payoff for the period is your Stage-1 payoff plus your Stage-2 payoff. In total you might have a payoff of $0 (nothing), $4 or $8 across both stages of the Choose-A-Side Game. In each period there is a 50-in-100 (50%) chance of being given Bucket A or Bucket B.Your bucket is randomly determined by the computer and is not affected by the bucket you have been given in previous periods.

Summary: Experiment 2 You have a 50-in-100 (50%) chance of being given Bucket A or Bucket B. Each bucket is divided in half, with 20 balls in each half. You will be shown a ball from the left hand side of your bucket. If it is black, your Stage-1 payoff is \$4. If it is white, your payoff is \$0 (nothing). You will answer 2 questions:

Question 1: What is your belief that the ball was drawn from Bucket A?

Question 2: Do you want the computer to draw a second ball from the left or right side of your bucket?

You should think carefully about your belief that the ball has been drawn from Bucket A. If it is black, your Stage-2 payoff is \$4. If it is white, your payoff is \$0 (nothing). Your period payoff is your Stage 1 payoff plus your Stage 2 payoff. In each period you might get a payoff of \$0 (nothing), \$4 or \$8. 1 of the 20 periods will be randomly chosen. Each period has an equal (1-in-20) chance of being chosen. You will be paid your payoff from that period in cash at the end of the experiment. When you have finished Experiment 2 you will be given instructions for a third experiment.

Quiz

At the start of a period the computer randomly selects a bucket for you. The bucket has 20 balls in each side: 40 in total. The computer shows you a ball from the left-hand side of the bucket, tells you its colour, and tells you whether it is worth \$0 or \$4. This is your payoff for Stage 1. The computer puts the ball back in your bucket. The computer asks whether you want the next ball drawn from the left hand side or the right hand side of the bucket. The computer also asks your belief about the chance-in-100 that the ball was drawn from Bucket A.

Imagine that you're sure the ball is drawn from Bucket A. How would you report this as a chance-in-100 that the ball is drawn from Bucket A?

1. The chance-in-100 of the Ball being from Bucket A is: (100)

Imagine that you're sure the ball is not drawn from Bucket A. How would you report this as a chance-in-100 that the ball is drawn from Bucket A?

2. The chance-in-100 of the Ball being from Bucket A is: (0)

Imagine that you think there's an equal chance the ball is drawn from Bucket A. How would you report this as a chance-in-100 that the ball is drawn from Bucket A?

3. The chance-in-100 of the Ball being from Bucket A is: (50)

The computer draws a ball from the side you choose, tells you its colour, and tells you if you have won \$0 or \$4. This is your payoff for Stage 2.

4. What is the minimum payoff possible in a period (Stage 1 + 2)? (0)

5. What is the maximum payoff possible in a period (Stage 1 + 2)? (8)

You then finish the period.

6. How many periods are there in this experiment? (20)
7. Do you receive a cash payments for your payoff in every period? (No)
8. Every period has a 1-in-? chance of being paid? 1-in: (20)

(Experiment begins when all questions are answered correctly.)

No-Elicitation Treatment

You must then decide whether you would like the computer to draw a second ball from the left-hand side of your bucket, or the right-hand side.

The computer randomly selects a ball from the side you choose. If it is black, you receive $4. If it is white, you receive $0 (nothing). This is your Stage-2 payoff. Your payoff for the period is your Stage-1 payoff plus your Stage-2 payoff. In total you might have a payoff of $0 (nothing), $4, or $8 across both stages of the Choose-A-Side Game. In each period there is a 50-in-100 (50%) chance of being given Bucket A or Bucket B. Your bucket is randomly chosen by the computer and is not affected by the bucket you have been given in previous rounds.

Summary: Experiment 2 You have a 50-in-100 (50%) chance of being given Bucket A or Bucket B. Each bucket is divided in half, with 20 balls in each half. You will be shown a ball from the left-hand side of your bucket. If it is black, your Stage-1 payoff is $4. It if is white, your payoff is $0 (nothing). You must then decide whether you would like the computer to draw a second ball from the left-hand side of your bucket, or the right-hand side. A second ball will be drawn from your bucket, from the side you choose. If it is black, your Stage-2 payoff is $4. If it is white, your payoff is $0 (nothing). Your period payoff is your Stage 1 playoff plus your Stage 2 payoff. In each period you might get a payoff of $0 (nothing), $4, or $8 in the Choose-A-Side Game. 1 of the 20 periods will be randomly chosen. Each period has an equal (1-in-20) chance of being chosen. You will be paid your payoff from that period in cash at the end of the experiment. When you have finished Experiment 2 you will be given instructions for a third experiment.

Quiz

(The quiz for the No-Elicitation Treatment is the same as the quiz for Experiment 1—that is, the "Choose-A-Side" Experiment.)

(Experiment begins when all questions are answered correctly.)

Experiment Three

You are about to start Experiment 3. This is the final experiment. You will repeat the experiment 20 times. Each repetition is called a "period." In each period you get a payoff of \$0, \$4, \$8 or \$12. At the end of the third experiment, 1 of the 20 periods will be chosen randomly by the computer. Each period is equally likely to be chosen. Your cash payment for the third experiment will be your payoff in the randomly chosen period. Your total payment today will include:

- Your show-up-fee of \$10.
- A cash payment for a randomly chosen period from the first computerized experiment
- A cash payment for a randomly chosen period from the second computerized experiment
- A cash payment for a randomly chosen period from the third computerized experiment

Although you will play 20 periods in this third experiment, you are paid cash for your payoff in a single period. Experiment 3 is the same as Experiment 2, except you will see **two** balls drawn from your bucket

The computer will randomly select a ball from the left-hand side of your bucket. The computer will tell you the colour of the ball, and whether your payoff is \$0 or \$4. The computer will put the ball back in the left-hand side of your bucket. The computer will randomly select a **second** ball from the left-hand side of your bucket. Because the ball is randomly chosen, this might be the same ball (chosen a second time) or it might be a new ball. The computer will tell you the colour of the second ball, and whether your payoff is \$0 or \$4.

You have two chances to get a payoff of \$4 in Stage 1. This means you can secure a payoff of \$0, \$4 or \$8 in Stage 1.

(The three treatments involve different instructions from this point.)

SBDM Treatment

You then answer 2 questions:

- Question 1: What is your belief that the two balls were drawn from Bucket A?
- Question 2: Do you want the computer to draw a third ball from the left or right side of your bucket?

The computer then tosses a coin to determine which Question is used to determine your Stage-2 payment.

- Tails: Question 1

- Heads: Question 2.

Just like Experiment 2: If you throw a Heads, your Stage-2 payoff will be determined by the "Choose-A-Side Game." The computer will draw a ball from the side of the bucket you choose. You will get a payoff of $4 if the ball is black, and $0 (nothing) if it is white. If you throw a Tails, your Stage-2 payoff will be determined by the Lottery Bag Game, or the Belief-About-Bucket-A Game (whether the two balls were drawn from Bucket A). These games are played in exactly the same way as in Experiment 2. Based on your reported belief that the ball was drawn from Bucket A, the computer will select the game that gives you the highest chance of winning $4.

You should think carefully about your belief that the balls have been drawn from Bucket A, as the computer will use your reported belief to decide whether you are paid according to your "Belief-About-Bucket A" or the Lottery game. The experiment is set up so that it is in your best interest to report your belief honestly and carefully. If you make a report that is not your true belief your payoff might be determined by the Lottery Game when you would prefer to be paid based on your belief that the ball was drawn from Bucket A.

Summary: Experiment 3

You have a 50-in-100 (50%) chance of being given Bucket A or Bucket B. You will be shown **2 balls** from the left hand side of your bucket. You will answer 2 questions:

- Question 1: What is your belief that the 2 balls were drawn from Bucket A?
- Question 2: Do you want the computer to draw a third ball from the left or right side of your bucket?

The computer then tosses a coin to determine which question is used to determine your Stage-2 payment:

- Tails: Question 1
- Heads: Question 2

If the computer throws "Heads" your payoff is determined in the same way as Experiment 1. A second ball will be drawn from your bucket, based on the side you choose. If the computer throws "Tails" your payoff will be determined by the "Belief about Bucket A" game or a Lottery Game. The best thing you can do is report your belief honestly, so that you are given the game with the highest chance of a payoff of $4. Your period payoff is your Stage 1 payoff plus your Stage 2 payoff. In each period you might get a payoff of $0 (nothing), $4, $8 or $12 in the third experiment. 1 of the 20 periods will be randomly chosen. Each period has an equal (1-in-20) chance of being chosen. You will be paid your payoff from that period in cash at the end of the experiment.

Quiz

At the start of a period the computer randomly selects a bucket for you. Both buckets are equally likely to be chosen. The bucket has 20 balls in each side: 40 in total. The computer shows you a ball from the left side of your bucket, tells you its colour, and tells you whether your payoff is \$4 or \$0. The computer puts the ball back in the left-hand side of your bucket.

1. How many balls are there in the left hand side? (20)

The computer draws a second ball from the left-hand side of your bucket and tells you whether your second payoff is \$4 or \$0.

2. Is it possible that the computer drew the same ball twice? (Yes)

The computer puts the second ball back in your bucket.

3. How many balls are there in the left hand side?(20)

The computer asks whether you want the third ball drawn from the left hand side or the right hand side of the bucket. The computer also asks you to report your belief that the ball is from Bucket A.

4. What is the minimum payoff possible in a period (both balls from Stage 1 + ball from Stage 2)? (0)
5. What is the maximum payoff possible in a period (both balls from Stage 1 + ball from Stage 2)? (12)

(Experiment begins when all questions are answered correctly.)

Unpaid Introspection Treatment

After seeing the colour of the **two balls**, you need to think about the chance that the balls were drawn from Bucket A. This is your "belief" that the balls were drawn from Bucket A. Your "belief" is a number between 0 and 100 to indicate the chance-in-100 that the ball has been drawn from Bucket A. You should think carefully about your belief that the ball has been drawn from Bucket A. You then answer 2 questions:

- Question 1: What is your belief that the ball was drawn from Bucket A?
- Question 2: Do you want the computer to draw a second ball from the left or right side of your bucket?

The computer randomly selects a ball from the side you choose. If it is black, you receive \$4. If it is white, you receive \$0 (nothing). This is your Stage-2 payoff. Your payoff for the period is your Stage-1 payoff plus your Stage-2 payoff. In total you might have a payoff of $0, $4, $8 or $12 across both stages of the third experiment. In each period there is a 50-in-100 (50%) chance of being given Bucket A or Bucket B. Your bucket is randomly determined by the computer and is not affected by the bucket you have been given in previous periods

Summary: Experiment 3

You have a 50-in-100 (50%) chance of being given Bucket A or Bucket B. You will be shown **2 balls** from the left hand side of your bucket. You will answer 2 questions:

- Question 1: What is your belief that the 2 balls were drawn from Bucket A?
- Question 2: Do you want the computer to draw a third ball from the left or right side of your bucket?

You should think carefully about your belief that the ball has been drawn from Bucket A. A second ball will be drawn from your bucket, from the side you choose. If it is black, your Stage-2 payoff is $4. If it is white, your payoff is $0 (nothing). Your period payoff is your Stage 1 payoff plus your Stage 2 payoff. In each period you might get a payoff of $0 (nothing), $4, $8 or $8 in the third experiment. 1 of the 20 periods will be randomly chosen. Each period has an equal (1-in-20) chance of being chosen. You will be paid your payoff from that period in cash at the end of the experiment.

Quiz

At the start of a period the computer randomly selects a bucket for you. Both buckets are equally likely to be chosen. The bucket has 20 balls in each side: 40 in total. The computer shows you a ball from the left side of your bucket, tells you its colour, and tells you whether your payoff is $4 or $0. The computer puts the ball back in the left-hand side of your bucket.

1. How many balls are there in the left hand side? (2)

The computer draws a second ball from the left-hand side of your bucket and tells you whether your second payoff is $4 or $0.

2. Is it possible that the computer drew the same ball twice? (Yes)

The computer puts the second ball back in your bucket

3. How many balls are there in the left hand side? (20)

The computer asks whether you want the third ball drawn from the left hand side or the right hand side of the bucket. The computer also asks you to report your belief that the ball is from Bucket A.

4. What is the minimum payoff possible in a period (both balls from Stage 1 + ball from Stage 2)? (0)

5. What is the maximum payoff possible in a period (both balls from Stage 1 + ball from Stage 2)? (12)

(Experiment begins when all questions are answered correctly.)

No-elicitation Treatment

You then have a third chance to draw a ball from your bucket. As before, black balls are worth $4. White balls are worth $0 (nothing). You must decide whether you would like the computer to draw the ball from the left hand side of your bucket, or the right hand side. The computer randomly selects a ball from the side you choose.

If it is black, you receive $4. If it is white, you receive $0 (nothing). This is your Stage-2 payoff.

Your payoff for the period is your Stage-1 payoff plus your Stage-2 payoff. In total you might have a payoff of $0, $4, $8, or $12 across both stages of the third experiment.

In each period there is a 50-in-100 (50%) chance of being given Bucket A or Bucket B. Your bucket is randomly determined by the computer and is not affected by the bucket you have been given in previous periods.

Summary: Experiment 3

You have a 50-in-100 (50%) chance of being given Bucket A or Bucket B. You will be shown **2 balls** from the left hand side of your bucket You will answer a question: Question: Do you want the computer to draw a third ball from the left or right side of your bucket?

A third ball will be drawn from your bucket, from the side you choose. If it is black, your Stage-2 payoff is $4. If it is white, your payoff is $0 (nothing).Your period payoff is your Stage 1 payoff plus your Stage 2 payoff. In each period you might get a payoff of $0 (nothing), $4, $8, or $12 in the third experiment. 1 of the 20 periods will be randomly chosen. Each period has an equal (1-in-20) chance of being chosen. You will be paid your payoff from that period in cash at the end of the experiment

Quiz

At the start of a period the computer randomly selects a bucket for you. Both buckets are equally likely to be chosen. The bucket has 20 balls in each side: 40 in total. The computer shows you a ball from the left side of your bucket, tells you its colour, and tells you whether your payoff is $4 or $0. The computer puts the ball back in the left-hand side of your bucket.

1. How many balls are there in the left hand side? (20)

The computer draws a second ball from the left-hand side of your bucket and tells you whether your payoff is $4 or $0.

2. Is it possible that the computer drew the same ball twice? (Yes)

The computer puts the second ball back in your bucket.

3. How many balls are there in the left hand side? (20)

The computer asks whether you want the third ball drawn from the left hand side or the right hand side of the bucket.

4. What is the minimum payoff possible in a period (both balls from Stage 1 + ball from Stage 2)? (0)
5. What is the maximum payoff possible in a period (both balls from Stage 1 + ball from Stage 2)? (12)

(Experiment begins when all questions are answered correctly.)