Experimental Guidance for Eliciting Beliefs with the Stochastic Becker-DeGroot-Marschak Mechanism

Ingrid Burfurd and Tom Wilkening*

November 2017

Abstract

We compare different implementations of the Stochastic Becker-DeGroot-Marschak (SBDM) belief elicitation mechanism, which is theoretically elegant but challenging to implement. In a first experiment we compare three common formats of the mechanism in terms of speed and data quality. We find that all formats yield reports with similar levels of accuracy and precision, but that the instructions and reporting format adapted from Hao and Houser (2012) is significantly faster to implement. We use this format in a second experiment in which we vary the delivery method and quiz procedure. Dropping the pre-experiment quiz significantly compromises the accuracy of subjects' reports and leads to a dramatic spike in boundary reports. However, switching between electronic and paper-based instructions and quizzes does not affect the accuracy or precision of subjects' reports.

JEL Codes: C91 D81 D83
Keywords: Beliefs, Elicitation, Prediction Accuracy, Methodology

*Ingrid Burfurd: Department of Economics, The University of Melbourne. 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, Laboratory Manager at the University of Melbourne's Experimental Economics Lab. 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 theories of decision-making assume that choices are based on an individual's preferences and probabilistic beliefs. Economists who want to test the descriptive validity of these theories are hindered by the fact that preferences and beliefs are typically unobservable. An advantage of economic experiments over other sources of empirical data is that secondary measures such as probabilistic beliefs can be elicited. These secondary measures supplement choice data and allow for stronger identification of the forces governing the decision-making process.

A challenge faced by practitioners is that there is a potential tradeoff between practical considerations—such as speed—and data quality considerations, such as accuracy and separability between subjects' beliefs and preferences. To help practitioners assess the relative merits of different experimental techniques, we explore the practicality-quality trade-off with regard to the Stochastic Becker-DeGroot-Marschak (SBDM) belief elicitation mechanism. The SBDM mechanism has been chosen for two reasons. First, the SBDM mechanism is incentive-compatible for all subjects whose preferences respect probabilistic sophistication and dominance (Karni, 2009).[1] These properties are desirable because heterogeneous risk preferences have been well documented in the laboratory (see, for example Holt and Laury (2002)) and there is evidence that some subjects are not well described by the expected-utility model of decision-making (Harrison and Rutström, 2009). Second, the SBDM mechanism is quite complex.[2] This complexity has prompted practitioners to experiment with quite different instructions, reporting interfaces, and training methods. Given the current absence of standard procedures, we believe it is important to identify which format offers the best balance of practicality and data quality.

We report results from two experiments. In the first experiment we compare three isomorphic presentations of the SBDM mechanism, which are adapted from Holt and Smith (2009), Hao and Houser (2012) and Trautmann and van de Kuilen (2015). The first format presents careful and detailed descriptive instructions, the second introduces a simple analogy to explain a complex probabilistic concept, and the third uses a list-based format for reporting beliefs. Each has desirable features. In order to get at the practicality-quality trade-off we compare the formats in terms of accuracy, precision, and the time it takes for subjects to work through instructions, a quiz, and each iteration of the belief elicitation task. We find that all formats yield reports with similar levels of

[1]The term "probabilistic sophistication" is used as per Machina and Schmeidler (1992)— that is, that the subject ranks lotteries based purely on the implied probability distribution over outcomes. The practical implication is that a subject will rank bets with subjective probabilities over outcomes in the same manner as he would rank lotteries with an objective probability distribution. Epstein (1999) defines ambiguity neutrality as a decision maker who is probability sophisticated. Thus, SBDM is not in general incentive compatible when decision makers are ambiguity averse. "Dominance" is the condition that a subject has preference relation $\succeq$ over lotteries such that $H_pL \succeq H_{p'}L$ for all $H > L$ if and only if $p \geq p'$.

[2]Ducharme and Donnell (1973) present the first experimental test of the mechanism and observe that while it is "basically simple," the SBDM mechanism task "seems complicated at first exposure."

accuracy and precision, but that the instructions and reporting format adapted from Hao and Houser (2012) is significantly faster to implement.

In a second experiment we restrict attention to the Hao and Houser (2012) format and run three treatments that focus on the practicalities of implementation. One treatment drops the pre-experiment quiz, one delivers the instructions and quiz on paper, and the third delivers the instructions and quiz electronically. We find that dropping the pre-experimental quiz significantly compromises the accuracy of subjects' reports and leads to a dramatic spike in boundary reports. We also find that switching between electronic and paper-based instructions and quizzes does not affect the accuracy or precision of subjects' reports.

This paper contributes to the small but growing literature on belief-elicitation methodologies. Existing work has compared the quality of reports under different belief elicitation mechanisms, including Huck and Weizsäcker (2002), Palfrey and Wang (2009), Massoni et al. (2014), Trautmann and van de Kuilen (2015), and Hollard et al. (2016).[3] There has, however, been little work on the practicalities of implementation. The notable exception is Holt and Smith (2016), which is closest to our paper. Holt and Smith use a Bayesian updating task to compare direct elicitation and a list-based format for implementing the SBDM mechanism. Our paper partially replicates their list of formats but also tests analogy-based instructions which are promising in both speed and accuracy. We also provide guidance on the importance of quizzes and instruction format when implementing the SBDM mechanism.

2 The Stochastic Becker-DeGroot-Marshak Mechanism

The Stochastic Becker-DeGroot-Marschak mechanism is based closely on the Becker-DeGroot-Marschak (BDM) mechanism (Becker et al., 1964), which was originally conceived as a method for eliciting certainty equivalents for lotteries. In its original context the BDM mechanism works as follows. Let H_pL denote the lottery that pays H with probability p and L otherwise. In the first stage of the mechanism, the subject is asked to report a price r, which he is prepared to pay to acquire the lottery H_pL. In stage two, a number z is realised from the distribution of random variable Z, which has distribution P_Z with support $[0, H]$. The subject receives the outcome of lottery H_pL if $z \leq r$ and payment z otherwise.

For all expected-utility maximizing agents it is a dominant strategy to report one's certainty equivalent (CE). The intuition for this result is straightforward: a subject who reports $r > CE$ runs the risk that $CE < z < r$. He will be paid according to the outcome of the lottery which he values at CE, but would prefer to receive payoff z. If the subject

[3]For excellent reviews of belief elicitation in the lab, see also Schlag et al. (2013) and Schotter and Trevino (2014).

under-states their CE, with $r < CE$, this is also costly: if $r < z < CE$ the subject will receive z but would prefer to receive the lottery.

The Stochastic Becker-DeGroot-Marschak mechanism adapts this approach to elicit the probability p of a particular stochastic event A. As per the deterministic case, the subject is endowed with a lottery that pays H if event A occurs and L otherwise. Given a true belief p, this lottery corresponds to a lottery H_pL. The subject reports his belief r about p. A number z is realised from the distribution of random variable Z, which has distribution P_Z on support $[0,1]$. If $z \leq r$, the subject retains his original lottery; if $z > r$, the agent exchanges his original lottery for a new lottery H_zL. The lottery payoffs are identical, with the two lotteries distinguished only by their probabilities of winning. Not only is this mechanism robust to heterogeneous risk preferences, but also to preferences that do not conform with expected-utility maximisation. For subjects who do not have a stake in the event of interest (i.e. they have no incentive to hedge) and whose preferences are consistent with probabilistic sophistication and dominance, it is in their interest to report $r = p$, as they otherwise risk receiving their less-preferred lottery (Karni, 2009).

2.1 The SBDM in Practice

The SBDM mechanism is a complex procedure. Its incentive compatibility requires subjects to have a thorough understanding of the mechanism, or at least to trust a researcher who tells them that it is in their best interests to report beliefs accurately. Experimental economists have broadly taken one of three approaches when implementing the SBDM, varying in the ways they explain the SBDM and the way subjects report their beliefs.

Early implementations of the SBDM mechanism such as Ducharme and Donnell (1973) and Grether (1992) explained the SBDM mechanism rigorously and precisely, often alongside descriptions of probabilistic concepts and incentive-compatibility. They then ask subjects to report r directly—that is, to issue a numeric report about their belief. We refer to this as a "descriptive" approach to capture the faithful depiction of the underlying SBDM mechanism.

Our benchmark for the descriptive format is Holt and Smith (2009) (HS). Subjects are told that they must report their r-in-100 belief that a particular event ("Event A") has occurred. This event is worth \$x. HS explain that belief r is equivalent to a belief that a lottery has an r-in-100 chance of winning \$x. Subjects are then introduced to a stochastic "payoff lottery", in which the subject can win \$x. Subjects are told that the probability of winning the payoff lottery is t-in-100, with t drawn from a uniform distribution between 0 and 100. If the subject's reported belief r is above cutoff t, the subject will be paid \$x if Event A has occurred. If r is less than or equal to t the subject's payoff will be determined by the payoff lottery. Both lotteries potentially pay \$x, and—according to their reported belief r—the subject will play whichever game gives him a higher probability of winning.

Möbius et al. (2007), Hollard et al. (2016), and Möbius et al. (2011) also use direct reporting, but use analogies to explain the stochastic payoff mechanism. Our "analogy-based" format is adapted from the instructions presented in Hao and Houser (2012) (HH), which use a 'chips-in-a-bag' analogy to explain the stochastic payoff mechanism. Subjects are asked to report a belief r about the probability of an event occurring (with the event associated with payoff \$x). They are told that a number between 0 and 100 will be randomly selected, with each number equally likely to be chosen. If this number "*?*" is larger than r, the subject's payoff will be determined by the draw of a chip from a bag. This bag contains 100 chips: *?* are black and the remainder are white. A black chip is worth \$x. Subjects are told that after they report belief r they will be paid either according to the realisation of the event or the draw of a chip from the bag—whichever has a higher payoff according to their reported belief. Hao and Houser's subjects see a physical bag filled with chips; our chips-in-a-bag are computerised.

Trautmann and van de Kuilen (2015) and Holt and Smith (2016) move away from direct reporting and explore an alternative list-style reporting format for the SBDM mechanism. The format is similar to the lists that are common in risk and time-preference elicitation tasks: a subject is presented with a list of choice tasks in which he indicates his preference over two lotteries. In Trautmann and van de Kuilen (2015) (TK) the subject indicates whether he prefers to be paid according to "Asset A"—which makes a payment if a particular event is realised—or Option B, which offers an objective probability of winning with the outcome determined by the role of a die. Following (TK), our variant of the "list" format requires subjects to choose whether they would prefer to be paid according to the outcome of Event A, or alternatively according to the outcome of the Dice Lottery. Similar to Holt and Smith (2016), we use a two-step titration procedure. In step one subjects nominate the support for their switch-point, with supports expressed as ranges of 10 percentage points (e.g. "51 - 60 percent"). On a second screen subjects indicate precisely when they switch from preferring one lottery to the other. The experiment does not allow subjects to nominate more than one switch-point.[4]

3 Experiment 1

Experiment 1 was conducted at the University of Melbourne's Experimental Economics Laboratory in July 2015 and consisted of 125 subjects. Each subject was paid a \$15 show-up fee, and won \$15 or \$0 in the experiment.[5] The experiment used deliberately

[4]By restricting subjects to a single switch point we might prevent subjects from reporting their true preferences and/or imposing consistency when subjects are actually confused. However, as we did not allow for multiple reports in the other two mechanisms the cleanest comparison is to preserve a single switch point.

[5]Subjects' total completion time for Experiment 1 varied between 16 minutes and 58 minutes, and subjects received an average payoff of \$25.95.

high stakes to ensure that rewards were salient. Payment was based on one period chosen from the fifteen periods at random.

We use an "induced probability" approach in our design. Subjects are given a Bayesian updating task and asked to report their beliefs about a posterior which has an objective probability that is known to the researcher. The task is modelled on Holt and Smith (2009). Subjects are told that there are two buckets: Buckets A and B. Bucket A contains two dark balls and one light-colored ball, while Bucket B contains two light balls and one dark ball. Subjects are informed that each bucket is equally likely to be selected, and that a ball will then be drawn from this unknown bucket. Each ball is equally likely to be chosen. Subjects are shown the color of the ball and asked to nominate their belief that the ball has been drawn from Bucket A. We make minor adjustments to the instructions to accommodate our computerised format and the belief-formation task is called the "Bucket Game" for easy and consistent reference throughout the instructions.

Subjects all received identical instructions regarding the Bucket Game and the pay-one-period payment protocol. Subjects then read one of the three SBDM mechanism instructions. The HH and TK instructions are adapted to the context of the HS "Bucket Game", and all instructions use the same language. In particular, this means that probabilities are expressed as the "chance in 100" of an event occurring. Following HS, all instructions tell subjects to "think carefully" about their beliefs because it will affect the selection of payoff method.

After reading their instructions all subjects completed a computerised pre-participation quiz. Quiz formats for the HS and HH treatment were identical while the TK format was slightly modified to give subjects practice in making reports via the two-step procedure.

Neither the instructions nor quiz use verbal interactions. This is to minimise experimenter effects and so that the instructions can be easily used across experiments and laboratories. We also use "portable" instructions—that is, instructions that avoid reference to the experiment itself.

Each subject completed 15 repetitions of the belief-elicitation task. At the end of each period subjects learned whether they earned \$15 or \$0. Participants in the HS and HH Treatments learned z, were reminded of their report r, and were told whether they were paid according to the Bucket Game or Dice Lottery/Lottery Bag Game. Subjects in the TK Treatment were told which of their choices was randomly selected, and were reminded about their preferred payoff option. All subjects were told the outcome of the stochastic payoff lottery, or alternatively whether their ball was drawn from Bucket A or B.

Experiment 1 was conducted across eight sessions and two days, with four sessions held on each day. In each session roughly a third of subjects participated in each treatment. Subjects drew a numbered ball from a jar and were seated at the corresponding computer station, with a third of the laboratory's computers devoted to each treatment. A summary of treatments is shown in Table 1.

Treatment	Periods	Computerised	n	Instructions		Quiz
				Word Count	Z-tree Screens	Reporting Format
HS	15	Yes	42	936	6	Direct
HH	15	Yes	41	397	2	Direct
TK	15	Yes	42	391	4	List-style

Table 1: Summary of Experiment 1

3.1 Outcome Measures and Statistical Tests

We consider three outcome measures when assessing the trade-offs that exist across formats: accuracy, precision, and brevity. Our measure of **accuracy** is the mean of the absolute error of a subject's reports, relative to the objective Bayesian posterior. Between-treatment variations in accuracy provides an indication of the incentive-compatibility characteristics of each treatment.

As a measure of **precision** we use the standard deviation of absolute errors for each individual.[6] The experiment centers around an objective Bayesian updating task, and there is no reason to suspect that individuals should vary systematically on their understanding of this task across formats. Between-format variation in precision may thus be a sign of different degrees of learning about the incentive properties of the mechanism and would suggest differences in initial understanding.

Finally, our measure of practicality is **brevity**, and we use the total time it takes a subject to go through the entire experiment. This includes the time taken to read the instructions, complete the quiz, and answer all 15 decision problems. It does not include time taken to complete the post-experiment questionnaire.

Throughout the analysis we perform the Kruskal–Wallis test over all three formats with each individual treated as a single observation. This test is the natural extension of the Mann-Whitney-Wilcoxon test when there are more than two treatments. The null hypothesis is that a random observation from subjects in each treatment is equally likely to be larger or smaller than an observation drawn from a different treatment. As a post-hoc test, we also use Dunn's test for stochastic dominance to compare pair-wise treatments and we adjust errors using the the Benjamini-Hochberg procedure to adjust for multiple hypotheses. All results in the paper have also been assessed using randomisation tests identical to those in Holt and Smith (2016). Any differences between the two approaches are noted in the main text.

[6] Typically precision is defined as the inverse of the variance. However, since some subjects have zero variance, this measure is unbounded.

4 Experiment 1: Results

Result 1 *The accuracy and precision of reports achieved with adaptations of the Holt and Smith (2009) format, the Hao and Houser (2012) format, and the Trautmann and van de Kuilen (2015) format are not significantly different from one another. The Hao and Houser format is significantly faster to run than the other two formats.*

Support for Result 1 is provided in Table 2, which provides summary statistics for our three outcome measures and the p-values from all treatment-level statistical tests. As can be seen in the first row, average accuracy in the HS, HH and TK is similar, with no apparent difference between the three formats. The Kruskal–Wallis test cannot reject the null hypothesis and there is no significant difference found in any of the pairwise tests. As can be seen in the second row, the precision of reports is similar across the three formats and there is no statistical evidence that the three formats differ at the aggregate level.

As can be seen in the third row of the table, the HH format takes subjects 850 seconds on average to complete, while the HS format takes 1089 seconds and the TK format takes 1212 seconds. The difference in time is significant according to the Kruskal–Wallis test. Looking at the pairwise tests, response time in the HH format is significantly different from both the HS and TK formats. There is no significant difference in time between the HS format and the TK format.

	Treatment Means			KW-test	Pairwise Dunn Tests		
	HS	*HH*	*TK*		*HS*v*HH*	*HS*v*TK*	*HH*v*TK*
Mean Abs. Error	12.4	13.1	14.7	0.546	0.326	0.410	0.393
	(1.53)	(1.43)	(1.61)				
Within-Sub. SD Errors	9.8	9.9	9.5	0.821	0.445	0.482	0.824
	(1.11)	(1.17)	(1.44)				
Total Time (seconds)	1089	850	1213	$<$**0.001**	**0.005**	0.069	$<$**0.001**
	(63.1)	(47.2)	(67.7)				

Table 2: Summary Statistics for the HS, HH and TK Treatments. The Kruskall-Wallis test is performed at the measure level and the Dunn pairwise tests adjusted for multiple hypotheses using the Benjamin-Hochbern adjustment. Standard errors are reported in parentheses.

Statistical tests at the aggregate level may mask the distributional features of subjects' reports that are likely to be of concern to practitioners. For instance, in many settings direct reports lead to groupings at round numbers—such as 10 or 20—and larger clusters at 0, 50, and 100. These groupings are likely to be obscured when averaged over multiple periods. We therefore examine the distribution of subjects' reports and the corresponding absolute errors.

Figure 1 shows the distribution of subjects' reported beliefs. In all three treatments there are pronounced spikes that are consistent with accurate Bayesian updating (posteriors of 67 percent in the wake of observing a dark ball, and 33 percent in the wake of a light ball). In the HS and HH Treatments there are also clusters of observations at each of the 10-point intervals nearest the true posterior and a small number of reports at 50. Boundary reports occur 5.7 percent of the time in the HS Treatment and 4.8 percent of the time in the HH Treatment. Clustering at 10-point intervals is less pronounced in the TK Treatment and boundary reports occur in only 1.7 percent of cases.[7] However, after the observation of a dark ball, 17 percent of TK reports are 33—the posterior that should occur after observing a light ball. This suggests that some subjects might be losing track of the signal they have observed.[8]

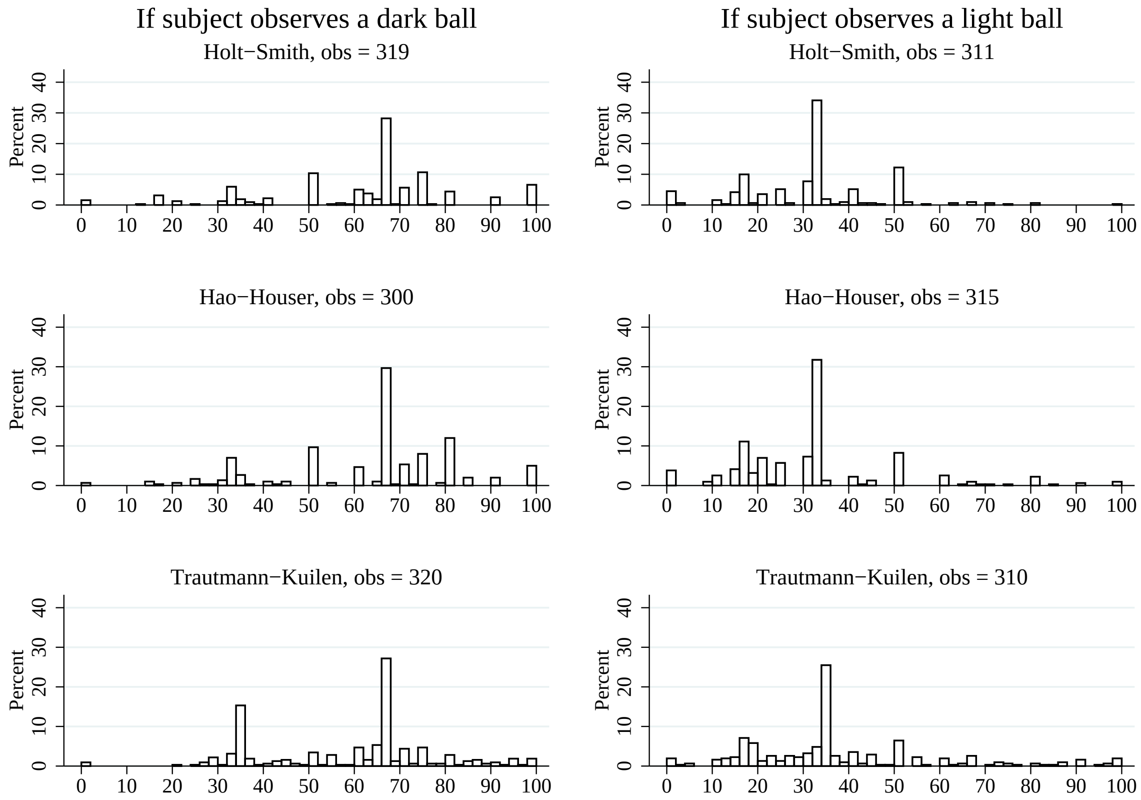

Figure 1: Reported Beliefs

Table 3 reports mean and median completion times for each major component of the experiment. Subjects in the HS and HH Treatments share the same quiz and period formats, and have similar mean and median completion times for these components of the experiment. Instruction times differ quite dramatically, however, with mean times of

[7]As in Holt and Smith (2016), the difference in boundary reports is significant in a randomisation test at the 0.01 level. We note, however, that the proportion of these reports is much smaller in our sample than in theirs. This is due in part to restricting our Bayesian task to a single draw.

[8]Note that every screen in the TK format reminds subjects of the color of the ball they have observed. Thus, the reverse reporting is unlikely to be due to recall and is more likely due to distraction or a lack of salience.

480 (HS) versus 305 seconds (HH), and median times of 333 versus 288 seconds (Kruskall-Wallis test: p < 0.001; Dunn test comparing HS and HH: $p = 0.000$). The mean subject therefore takes nearly three minutes longer to work through the HS instructions than the HH instructions.

	Mean Completion Time			Median Completion Time		
	HS	HH	TK	HS	HH	TK
Period (mean)	26	22	40	22	17	36
Instructions	480	306	419	433	288	373
Quiz	212	212	187	160	178	155
Total Time	**1089**	**850**	**1212**	**988**	**815**	**1112**

Table 3: Summary of Completion Times (seconds)

Figure 2 presents subjects' completion times across 15 periods. Subjects in the TK Treatment exhibit greater dispersion in period completion times than their peers, particularly in early periods. Recall that these subjects have to indicate their preferences over multiple lottery choices, which is reflected in significantly longer mean and median period completion times. Focusing on subject-level mean period completion times, TK subjects have a mean period completion time of 40 seconds, versus 26 and 22 in HS and HH; the medians of subject-level means are 36 in TK, 22 in HS, and 17 in HH (Kruskall-Wallis test: $p = 0.001$; Dunn test comparing TK and HS: $p = 0.001$; Dunn test comparing TK and HH: $p < 0.001$; Dunn test comparing HS and HH: $p = .061$).

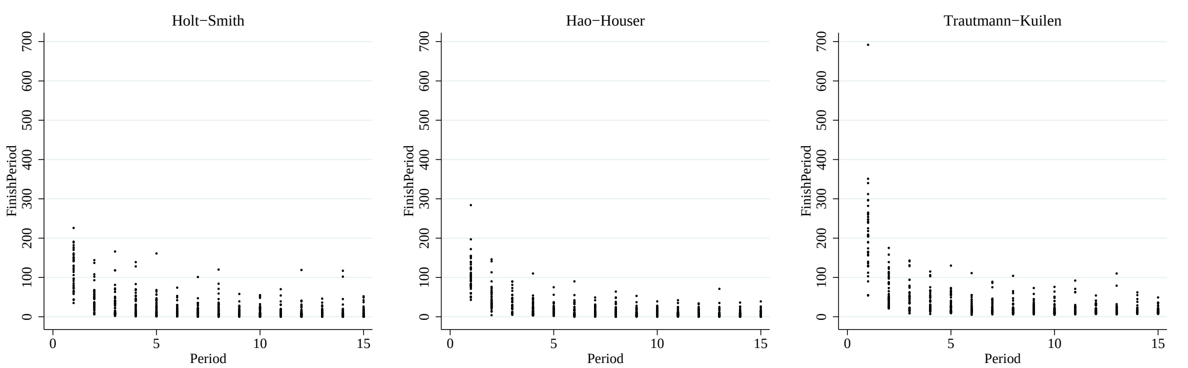

Figure 2: Completion Times by Period

As a result of HS's longer instructions and TK's two-stage reporting interface the total time taken to complete the experiment is significantly faster when subjects complete the HH Treatment. The HH format therefore stands out as the most immediately appealing due to its improved speed and the lack of evidence that precision and accuracy are improved in either of the longer formats. We use this treatment as the basis of our second experiment, which tests whether the format of quizzes influences performance and speed.

5 Experiment 2

Part 2 of our study varies the implementation of the instructions adapted from Hao and Houser (2012). The Hao-Houser Quiz Treatment (abbreviated to Q) is identical to the Hao and Houser Treatment from Experiment 1. The Hao-Houser No Quiz Treatment (abbreviated to NQ) drops the computerised quiz, and the Paper Treatment (P) administers the instructions and quiz in hard copy. The Q, NQ and P Treatments are compared using the same criteria Experiment 1: accuracy, precision and brevity.

Treatment	**Periods**	**Computerised**	**n**	**Quiz**
				Reporting Format
Quiz	15	Yes	28	Direct
No Quiz	15	Yes	30	No Quiz
Paper	15	No	30	Direct

Table 4: Summary of Experiment 2

Experiment 2 was conducted across three days. Three sessions were held on the first day, and one on each of the two subsequent days. Times were varied across the mornings and afternoons. The Quiz and No Quiz Treatments were both computerised and run jointly across three sessions with random assignment within each session.[9] Because of the need to distribute hard-copies, the Paper Treatment was conducted in separate sessions so that subjects were not concerned that some participants might be completing different experiments.[10] Subjects' total completion time for Experiment 2 varied between 14 minutes and 54 minutes, and subjects received an average payoff of $23.55.

5.1 Results

Result 2 *Reports in the computerized quiz treatment are significantly more accurate that reports in the no quiz treatment. Thus, using a quiz is important for ensuring accuracy in the computerised analogy-based Hao and Houser format. There are no significant differences in the accuracy of reports in the computerized and paper based quiz treatments, but the no quiz treatment is significantly faster than both electronic treatments.*

[9]The Q treatment is identical to the HH treatment and was repeated to allow for within session randomization. Completion times in the Q treatment were slightly faster than the HH treatment with a mean session time of 766 seconds and a median session time of 695 seconds. However, the difference in session times is not significant using a Mann-Whitney-Wilcoxon test (p-value = 0.13).

[10]Times for all treatments are measured precisely, with the exception of the Paper Treatment. When running the Paper Treatment, the laboratory assistant noted the times at which instructions were distributed, the time at which instructions were swapped for the quiz, and the time when the subject completed the quiz successfully. These times were noted in minutes rather than seconds, with all time-based analysis using the mid-point of the minute in question. There was 1 lab assistant and 15 subjects in each P treatment.

Table 5 reports accuracy, precision, and brevity for each of the three quiz treatments. Average accuracy in the Quiz Treatment is 11.2 and it is 13.4 in the Paper Treatment, and this difference is not significant. Accuracy in the No Quiz Treatment is 20.4, which is significantly different from the Quiz Treatment at the 5 percent level (p = 0.04) and from the Paper Treatment at the 10 percent level (p = 0.06). The difference between the Paper and No Quiz Treatments is significantly different at the 5 percent level when using the alternative randomization test.[11]

Precision in the No Quiz Treatment is 15.2, while it is 10.4 and 8.6 in the Quiz and Paper Treatments. The three-way Kruskall-Wallis test is not significant, but we note that a pairwise randomization test finds that the difference between the Paper and No Quiz Treatments is significant at the 0.05 level. We interpret this difference to be due to differences in initial understanding and learning: in the No Quiz treatment, a large portion of individuals begin by making boundary reports and then revising their actions towards the objective probabilities. No such learning dynamic is observed in the other treatments.

	Treatment Means			KW-test	Pairwise Dunn Tests		
	Q	*NQ*	*P*		*Q*v*NQ*	*Q*v*P*	*NQ*v*P*
Mean Abs. Error	11.2	20.4	13.4	**0.042**	**0.017**	0.230	0.062
	(1.66)	(2.78)	(1.81)				
Within-Subject SD of Errors	10.4	15.2	8.6	0.158	0.238	0.187	0.082
	(1.54)	(2.35)	(1.52)				
Total Time (Seconds)	766.3	613.7	1079.7	**<0.001**	**0.035**	**0.004**	**<0.001**
	(62.5)	(37.7)	(86.4)				

Table 5: Summary Statistics for the Quiz (Q), No-Quiz (NQ) and Paper (P) Treatments. The Kruskall-Wallis test is performed at the measure level and the Dunn pairwise tests adjusted for multiple hypotheses using the Benjamin-Hochbern adjustment. Standard errors are reported in parentheses.

As can be seen in the third row, the quiz increases the overall time of the experiment from an average of 613.7 seconds to 766.3 seconds. Moving from an electronic quiz to a paper-based quiz increases the total time of the experiment to 1079.7 seconds. All differences are significant.

Figure 3 presents the aggregate distribution of subject's reports for the Quiz, No Quiz, and Paper Treatments. Accurate reports are much more common in the Quiz Treatment, while reports of 50 are more common in the Paper-based Treatment than the computerised Quiz Treatment. Reports in the No Quiz Treatment are frequently inconsistent with Bayesian updating: while boundary reports are uncommon in the Quiz and Paper treatments, they occur 118 times in the No Quiz Treatment and account for 26.22% of observations.

[11] All Randomisation Test results are included in the Appendix.

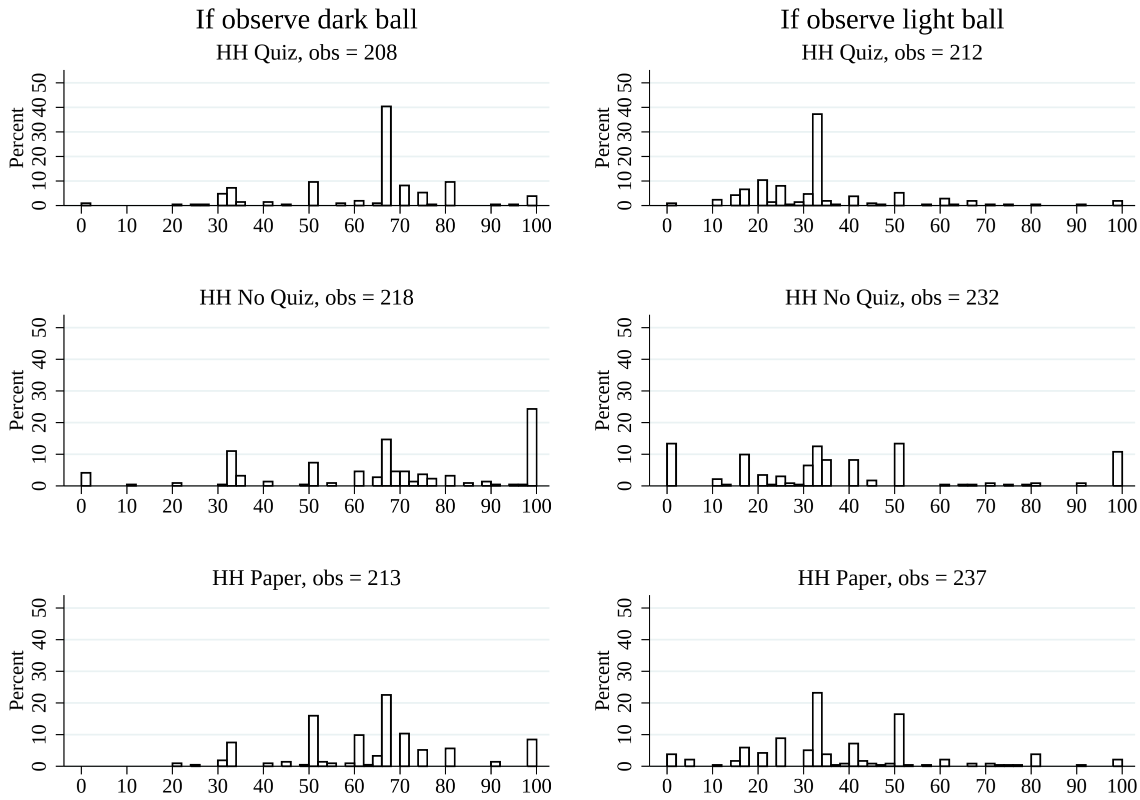

Figure 3: Reported Beliefs

As seen in Table 6, subjects' mean period completion times do not differ significantly across the three treatments. This is not unexpected given that all treatments use the same computerised reporting interface. Subjects take significantly longer to complete the instructions if they participate in the Paper Treatment: the mean completion time is nearly 8 minutes, in contrast with about 5 minutes for the computerised instructions (Quiz and No Quiz Treatments).

	Mean Completion Time			Median Completion Time		
	Q	NQ	P	Q	NQ	P
Period (mean)	18.6	20.5	20.4	15.2	17.5	19.6
Instructions	295	306	474	245	281	420
Quiz	191	0	300	157	0	240
Total Time	**766**	**613**	**1079**	**695**	**596**	**951**

Table 6: Summary of Completion Times (seconds)

6 Conclusion

While belief elicitation is increasingly popular there are no widely adopted or standard procedures. To help practitioners assess the relative merits of different experimental techniques, we explore the practicality-quality trade-off with regard to the SBDM belief elicitation mechanism. We study behavior in three formats of the SBDM: a "descriptive" instruction format with direct reporting, adapted from Holt and Smith (2009); an "analogy-based" instruction format with direct reporting, adapted from Hao and Houser (2012); and a "list-style" format adapted from Trautmann and van de Kuilen (2015). We find that accuracy and precision of reports in the three formats are remarkably similar but that the format adapted from Hao and Houser (2012) is quicker to run than the other two formats. We use this format in a second experiment in which we vary the delivery method and quiz procedure. Dropping the pre-experiment quiz significantly compromises the accuracy of subjects' reports and leads to a dramatic spike in boundary reports. Switching between electronic and paper-based instructions, however, does not affect the accuracy or precision of subjects' reports.

Brevity and efficient communication in experiment instructions tends to be undervalued, and should be taken more seriously given the limited attention span of subjects. Our HH format is the shortest, yet it helps subjects make quick decisions without compromising accuracy. We thus view it as a promising format for eliciting beliefs.

Recent work by Holt and Smith (2016) also compares direct-elicitation and list-based formats of the SBDM. As with our experiment, they do not find significant differences in accuracy between formats. However, there is evidence that there is a difference in accuracy for simple situations when the true probability is .5. In the online appendix, we report on an additional robustness experiment where we use a spinner task in which objective probabilities are easily calculated. We do not find differences in accuracy across formats using this alternative simple task.

Holt and Smith (2016) also find large differences in boundary reports across formats. When restricting the data in Holt and Smith (2016) to events that match ours, the overall boundary rate is only 2.8%, which is very similar to ours. The large difference in boundary reports in other segments of their data suggest that there may be an interaction between the complexity of the task and the importance of the elicitation format. In particular, it would be interesting to understand how subjects use the elicitation format to scaffold their probabilistic reasoning.

Bibliography

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

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.

Epstein, L. (1999). A definition of uncertainty aversion. *The Review of Economic Studies 66*(3), 579–609.

Grether, D. M. (1992). Testing bayes rule and the representativeness heuristic: Some experimental evidence. *Journal of Economic Behavior & Organization 17*(1), 31–57.

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.

Harrison, G. W. and E. E. Rutström (2009). Expected utility theory and prospect theory: One wedding and a decent funeral. *Experimental Economics 12*(2), 133–158.

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. and S. K. Laury (2002). Risk aversion and incentive effects. *American economic review 92*(5), 1644–1655.

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.

Huck, S. and G. Weizsäcker (2002). Do players correctly estimate what others do? evidence of conservatism in beliefs. *Journal of Economic Behavior & Organization 47*, 71–85.

Karni, E. (2009). A theory of medical decision making under uncertainty. *Journal of Risk and Uncertainty 39*(1), 1–16.

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

Massoni, S., T. Gajdos, and J.-C. Vergnaud (2014). Confidence measurement in the light of signal detection theory. *Frontiers in Psychology 1455*(5), 1–13.

Möbius, M. M., M. Niederle, P. Niehaus, and T. Rosenblat (2007). Gender differences in incorporating performance feedback. *draft, February*.

Möbius, M. M., M. Niederle, P. Niehaus, and T. S. Rosenblat (2011). Managing self-confidence: Theory and experimental evidence. Technical report, National Bureau of Economic Research.

Palfrey, T. and S. Wang (2009). On eliciting beliefs in strategic games. *Journal of Economic Behavior & Organization 71*, 98–109.

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. *Annu. Rev. Econ. 6*(1), 103–128.

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

Appendix A: Randomisation Test Results

Table 7 reports the results of pairwise randomization tests which compare outcomes from treatments in Experiments 1 and 2. All randomization tests are based on 500,000 simulations for comparability with the randomization test results reported in Holt and Smith (2016).

	Experiment 1			Experiment 2		
	HS-HH	HS-TK	HH-TK	Q-NQ	Q-P	NQ-P
Subject Abs. Error.	0.735	0.299	0.457	0.007	0.356	0.042
Within-subject S.D. Errors	0.949	0.846	0.806	0.105	0.395	0.022
Total Time (Seconds)	0.003	0.187	0.000	0.034	0.005	0.000

Table 7: Results of Pair-Wise Randomisation Tests comparing Treatments

Online Appendix B: Additional Control Treatments

Recent work by Holt and Smith (2016) compares direct-elicitation and list-based formats of the SBDM using a larger set of Bayesian tasks. Consistent with our results, they do not find significant differences in accuracy between formats when all data is used. However, there is some evidence that the list-based format performs better in a subset of simple decision problems where the true probability of the paid event is .5. This suggests that there may be an interaction between the difficulty of the Bayesian task and the accuracy of reports.

As an additional robustness check, we ran a follow-up experiment with a simplified task that does not require Bayesian updating. In the task, subjects are shown a wheel with six equidistant sections numbered 1 to 6. Subjects are told that we will spin the wheel in each period and that each of the sections is equally likely to be selected. In each period, a subject is assigned a set of either two or four numbers. They are paid \$15 if the spinner lands on one of their numbers and \$0 otherwise. As in our original experiment, the true probability of the assigned event is either .33 or .66 and the mechanism is designed to induce a report of this probability.

Our follow up "Spinner Task" experiments consisted of 102 subjects who participated in six experiments over three days. Subjects were randomized into formats within each session. In total we had 34 subjects in the HS format, 35 subjects in the HH format, and 33 subjects in the TK format. All other features of our protocol were identical to the original treatment.

Table 8 shows the accuracy rates from our Original Task, the Spinner Task, and the Task used in Holt and Smith (2016). As can be seen in the second row, accuracy in spinner task ranges between 10.5 and 15.0 between treatments. Similar to the Original Task, there is no significant difference in accuracy in the Kruskal–Wallis test and no difference in any of the pairwise comparisons using Dunn's test for stochastic dominance. Boundary reports were also similar to our Original Task and were observed in 1.6% of cases in the HS treatment, in 5.1% of cases in the HH treatment, and in 8.5% of cases in the TK treatment.

Online Appendix C: Instructions and Quizzes

The Bucket Game

Thank you for choosing to participate in today's experiment. 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 \$15 attendance fee plus earnings from a computerised experiment. If you have any questions during the experiment please sit quietly and raise your hand. An

	Treatment Means			KW-test	Pairwise Dunn Tests		
	HS	*HH*	*TK*		*HS*v*HH*	*HS*v*TK*	*HH*v*TK*
Original Task	12.4	13.1	14.7	0.546	0.326	0.410	0.393
(N = 125, T = 15)	(1.53)	(1.43)	(1.61)				
Spinner Task	10.5	12.5	15.0	0.365	0.299	0.265	0.376
(N = 102, T = 15)	(1.85)	(1.58)	(2.44)				
Holt and Smith (2016)	9.9	N/A	7.0	N/A	N/A	0.109	N/A
(N = 72, T = 12)	(0.67)		(0.53)				

Table 8: Mean Abs. Errors for the HS, HH and TK Treatments using the original task, the spinner task, and the task in Holt and Smith (2016). The Kruskall-Wallis test is performed at the measure level and the Dunn pairwise tests adjusted for multiple hypotheses using the Benjamin-Hochbern adjustment. Standard errors are reported in parentheses.

experiment assistant will be with you as soon as possible.

Payment for the computerised experiment: You will play a computerised game 15 times. Each repetition of the game is called a "period." In each period you will win a prize of \$15 or \$0. At the end of the experiment, 1 of the 15 periods will be chosen randomly by the computer. Each period is equally likely to be chosen. Your cash payment for the computerised experiment will be your prize from the randomly chosen period. Although you will play 15 periods, you are only paid in cash for the prize you earn in a single period.

The Bucket Game: You are going to participate in a game which is referred to as "The Bucket Game." There are two buckets: Bucket A and Bucket B. Bucket A contains 2 light balls and 1 dark ball. Bucket B contains 1 light ball and 2 dark balls. (These buckets and balls are all computerised.)

One of the buckets will be randomly chosen by the computer. Both buckets have an equal chance of being chosen. (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 chosen bucket. Each ball has an equal chance of being chosen. You will be told the color of the ball.

Holt and Smith Treatment Instructions

(Subject receives standard Bucket Game instructions.)

Bucket Game: reporting your beliefs

After seeing a random ball from one of the buckets, you will report your belief about the chance that Bucket A is being used. You will indicate a number between 0 and 100, which we will call P. This means that you think that the chance that Bucket A is being used is "P out of 100."

For example: If you could be sure that Bucket A is being used, you should choose P

= 100, which would indicate that you believe the chances are 100 out of 100 that Bucket A is being used. If you could be sure that Bucket A is not being used, you should choose P = 0 to indicate that the chances are 0 out of 100 that Bucket A is being used. Thus the magnitude of P corresponds to the chance that Bucket A is being used. For example, if you think that it is just as likely that Bucket A is being used as Bucket B, then you should choose P = 50, indicating that the chances are 50 out of 100 that Bucket A is being used.

If you indicate that the chances are P out of 100 that Bucket A is being used, then you should be indifferent between:

- Lottery A: being paid a prize ($15) if Bucket A is in fact being used, and $0 otherwise...; and
- A "P" lottery: being paid a prize ($15) in a lottery with P chance of getting $15, and $0 otherwise.

Notice that in each of these two options, the chances of earning the $15 prize are P out of 100, and this is the sense in which they are equivalent. Let me summarize. If you indicate that the chances of Bucket A being used are P out of 100, then you should be indifferent between:

1. getting $15 if Bucket A is being used; and
2. getting $15 with chance P out of 100.

The following procedure will be used to help you choose the value of P that makes you indifferent between the A lottery in (1) and the P lottery in (2) above.

After you record the value of P (between 0 and 100) that represents your beliefs about the chances of Bucket A being used, the computer will randomly select a number "N" between 0 and 100. Each number is equally likely to be chosen. Because it's like rolling a dice to randomly choose a number, this is called the "Dice Lottery." The "Dice Lottery" pays a cash prize ($15) with chance N out of 100, and $0 otherwise.

Recall that you will have told us a number P that represents the chance that the Lottery will pay 1 $15 prize, i.e. the chance that Bucket A is being used. In determining your payoffs for the period, we will use whichever is better for you, the Dice Lottery or the A Lottery. We will make the decision on which lottery is better for you by comparing the randomly determined N and the P that you tell us represents your beliefs about the chances that Bucket A is beling used.

Case of N less than P: If the "dice throw" results in N less than P, then the Dice Lottery offers a lower chance of the cash prize than the A lottery and Lottery P. We will reject the Dice Lottery and your earnings for the period will be determined by the A lottery: $15 if Bucket A is being used, $0 otherwise.

Case of N greater than P: If N is greater than or equal to P, then the Dice Lottery offers an equal or higher chance of the cash prize than the A Lottery and the P Lottery. We will accept the Dice Lottery and it will determine your earnings for the period: $15 with N in 100 chance.

Think of it this way: you can either take the A lottery, which is equivalent to a chance of P out of 100 of earning $15, or you can take the Dice Lottery. We will make the decision of whether to accept or reject this Dice Lottery by comparing N with the value of P that you nominated. If you tell us the value of P that best represents your beliefs about the chance (out of 100) that Bucket A is being used, then we can make the best decision for you about whether to accept or reject the Dice Lottery.

To summarize: You will be told a bucket has been randomly chosen. Then you will be told the color of the ball that has been randomly chosen from that bucket, and you write the number P between 0 and 100 that represents your beliefs about the chances out of 100 that Bucket A is being used.

There are two alternative ways that you can earn the $15 prize instead of the $0 prize. Your earnings will either be determined by the A Lottery ($15 if A is being used) or by the Dice Lottery ($15 if the computer randomly selects a number less than N).

You should think carefully about the value P that represents your beliefs about getting the $15 prize under the A Lottery, since we will use P to decide whether or not to replace the A Lottery with the Dice Lottery for determining your earnings.

You will play this game a total of 15 times. In each period you will earn a period prize of $15 or $0. At the end of the experiment 1 of the 15 periods will be randomly chosen to determine your cash payment from the experiment.

Holt and Smith Treatment Quiz

Imagine that you are shown a ball. Based on its color you report your belief that there is a 20-in-100 chance that the ball is from Bucket A. The computer randomly selects N = 25 for the Dice Lottery.

1. Which game will be used to determine your prize for the Period? (Subject chooses between the A Lottery and Dice Lottery?)

2. What is your chance in 100 of winning $15? (Subject enters integer.)

3. What is your chance in 100 of winning $0? (Subject enters integer.)

Imagine you start a new period. You are shown a new ball. This time you believe there is an 81-in-100 chance the ball was taken from Bucket A... but you make an error! You type “18” by mistake. The computer thinks you believe there is an 18-in-100 chance of winning $15 in the A Lottery.

The computer randomly selects N = 25 for the Dice Lottery.

4. What do *you* believe is your chance in 100 of winning $15 if you play the A Lottery? (Subject enters integer.)
5. What is your chance in 100 of winning $15 if you play the Dice Lottery? (Subject enters integer.)
6. Which game will be used to determine your prize for the period? (Subject chooses between the A Lottery and Dice Lottery?)

Thank you. You have completed the Quiz. The experiment is about to begin.

Hao and Houser Treatment: Instructions

(Subject receives standard Bucket Game instructions.)

After seeing the color 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. You will then report 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.

We will use your reported belief to help determine your prize in each period. This is how we determine your prize:

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 win a prize of $15: the Bucket Game and the Lottery Bag Game.

THE BUCKET GAME:
Prize of $15 if the ball was from Bucket A.
Prize of $15 if the ball was from Bucket B.

Chance-in-100 of winning $15:
Belief that ball is from Bucket A

THE LOTTERY BAG GAME:
Prize of $15 if you draw a black chip.
Prize of $15 if you draw a white chip.

Chance-in-100 of winning $15:
"?"-in-100

The computer knows the chance of winning $15 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 $15. (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 we will use your reported belief to decide whether you are paid according to the Bucket Game or the Lottery Bag Game.

Summary: You will be told the color of a ball. You will report your belief that the ball was drawn from Bucket A. There are two ways to win $15:

- The Bucket Game awards $15 if the ball was drawn from Bucket A, and $15 if it was drawn from Bucket B.

- The Lottery Bag Game award $15 if a black chip is drawn from the bag, and $0 if a white chip is drawn. There is an unknown, random chance of winning $15.

Based on your reported belief that the ball has been drawn from Bucket A you will play whichever game gives you a higher chance of winning $15. You should think carefully about your belief that the ball has been drawn from Bucket A.

You will play this game a total of 15 times. In each period you will earn a period prize of $15 or $0. At the end of the experiment 1 of the 15 periods will be randomly chosen to determine your cash payment from the experiment.

Hao and Houser Treatment: Quiz

This is identical to the Holt and Smith Quiz, but with references to the Lottery Bag Game rather than the Dice Lottery, and with references to the Bucket Game rather than the A Lottery.

Trautmann and van de Kuilen Treatment: Instructions

(Subject receives standard Bucket Game instructions.)

Before you learn which bucket was used you will choose how your prize is determined in this period. You need to think about whether the ball has been taken from Bucket A or Bucket B. We ask you to think about the chance-in-100 that the ball has been taken from Bucket A. This is your "belief" that the ball is from Bucket A." Your belief is important, because you will use it to make a decision in a Payment Game.

The Payment Game

You will see a list of choices. One is labelled "Bucket Game" and the other "Lottery Game." In each choice the Bucket Game yields:

- $15 prize if the ball was pulled from Bucket A

- $0 prize if the ball was pulled form Bucket B

In each choice the Lottery Game yields $15 with a particular probability, and $0 otherwise.

In the first choice, the Lottery Game gives you a 0-in-100 chance of winning $15. We imagine that most people would prefer the Bucket Game in Choice 1, because the Bucket Game has a chance of winning $15, whereas the Lottery Game has no chance of winning $15. In the last choice the Lottery Game gives you a 100-in-100 chance of winning $15. We imagine that most people would prefer the Lottery Game in the final choice, since the Lottery Game wins $15 for sure, while the Bucket Game only has a chance of winning $15. We therefore imagine that most people will switch from choosing the Bucket Game to the Lottery Game at some point in the list.

Bucket Game		**Lottery Game**		**Example Decision**
If ball drawn from Bucket A	If ball drawn from Bucket B	Chance in 100 of winning $15	Chance in 100 of winning $0	Prefer Bucket Game or Lottery Game?
$15	$0	0	100	Bucket Game
$15	$0	1	99	Bucket Game
$15	$0	2	98	Bucket Game
...etc	...etc	...etc	...etc	...
$15	$0	98	2	Lottery Game
$15	$0	99	1	Lottery Game
$15	$0	100	0	Lottery Game

There are 101 Lottery Game choices. We need to know whether you prefer the Bucket Game or Lottery Game for each choice. We try to make this easier by using a two-stage process. In Stage 1 we ask you to indicate *roughly* the point where you switch to preferring the Lottery Game. In Stage 2 we ask you to indicate *exactly* when you prefer the Lottery Game to the Bucket Game.

For example, imagine that you believe the Bucket Game has a 23-in-100 chance of winning a $15 prize. We imagine that you would prefer the Lottery Game if it has a 24-in-100 chance of paying $15.

In Stage 1 you would indicate that you want to switch from the Bucket Game to the Lottery Game when the Lottery Game has a probability of winning that lies between 20 and 29 in 100.

Bucket Game		**Lottery Game**		**Switch?**
If ball drawn from Bucket A	If ball drawn from Bucket B	Chance in 100 of winning $15	Chance in 100 of winning $0	Switch from Bucket Game to Lottery Game?
$15	$0	0-9	91-100	
$15	$0	10-19	81-90	
$15	$0	20-29	71-80	(Example) **Switch**

In Stage 2 you would refine your choice and indicate that you switch to preferring the Lottery Game when it has a 24-in-100 chance of winning $15.

Although we imagine that most people would switch from the Bucket Game to the Lottery Game at some point in the list, it is entirely up to you what to do in each of the

Bucket Game		Lottery Game		Example Decision
If ball drawn from Bucket A	If ball drawn from Bucket B	Chance in 100 of winning $15	Chance in 100 of winning $0	Prefer Bucket Game or Lottery Game?
$15	$0	20	80	Bucket Game
$15	$0	21	79	Bucket Game
$15	$0	22	78	Bucket Game
$15	$0	23	77	Bucket Game
$15	$0	24	76	Lottery Game
$15	$0	25	75	Lottery Game
...etc	...etc	...etc	...etc	...etc

choices. After you have made your choices the computer will randomly select one of the Lottery Games. The computer will check whether you preferred to play that particular Lottery Game or the Bucket Game. If you preferred the Bucket Game, you will get $15 if the ball was drawn from Bucket A, and $0 otherwise. If you preferred the Lottery Game, the computer will conduct the Lottery. You will get $15 if you win, and $0 otherwise. Remember that you will play 15 Periods, and that 1 Period will be randomly chosen to be pain in cash at the end of the experiment.

Summary: You will be told the color of a ball. It has been drawn from Bucket A or B. Your prize will be determined by the Bucket Game or the Lottery Game. There are 101 versions of the Lottery Game. Across 2 stages you will indicate when you switch from preferring the Bucket Game to the Lottery Game.

The computer randomly chooses one of the Lottery Games. If you preferred the Bucket Game, your prize will be $15 if the ball was from Bucket A, and $0 if it was from Bucket B. If you preferred the Lottery Game, your prize is $15 if you win with a particular probability, or $0 otherwise.

You will play this game a total of 15 times. Each time you will win $15 or $0. At the end of the experiment one period will be randomly chosen. Your winnings from that period will be paid in cash. This means that each one of your choices could eventually determine your payment.

Trautmann and van de Kuilen Treatment: Quiz

Imagine you believe that there is a 20-in-100 chance the ball was drawn from Bucket A.

1. Based on your belief, does a Lottery with a 21-in-100 chance of $15 give you a higher or lower chance of winning $15? (Subject chooses between "Higher" and "Lower.")

2. Based on your belief, does a Lottery with a 19-in-100 chance of $15 give you a higher or lower chance of winning $15? (Subject hooses between "Higher" and "Lower.")

3. If you wanted to switch from the Bucket Game to the Lottery Game to maximise your chance of winning $15, what would you choose in Stage 1? (Subject nominates switch point in table below.)

Bucket Game		Lottery Game		Switch?
If ball drawn from Bucket A	If ball drawn from Bucket B	Chance in 100 of winning $15	Chance in 100 of winning $0	Switch from Bucket Game to Lottery Game?
$15	$0	0-9	91-100	(Subject indicates)
$15	$0	10-19	81-90	(Subject indicates)
$15	$0	20-29	71-80	(Subject indicates)

4. And what would you choose in Stage 2? (Subject indicates Bucket or Lottery Game preferences in table below.)

Bucket Game		Lottery Game		Example Decision
If ball drawn from Bucket A	If ball drawn from Bucket B	Chance in 100 of winning $15	Chance in 100 of winning $0	Prefer Bucket Game or Lottery Game?
$15	$0	20	80	(Subject indicates)
$15	$0	21	79	(Subject indicates)
$15	$0	22	78	(Subject indicates)
$15	$0	23	77	(Subject indicates)
...etc	...etc	...etc	...etc	...etc

Thank you. You have completed the Quiz. The experiment is about to begin.