Getting It Right the First Time: Belief Elicitation with Novice Participants

Li Hao and Daniel Houser

April 2010

Discussion Paper

Interdisciplinary Center for Economic Science
4400 University Drive, MSN 1B2, Fairfax, VA 22030
Tel: +1-703-993-4850 Fax: +1-703-993-4851
ICES Website: www.ices-gmu.org
ICES RePEc Archive Online at: http://edirc.repec.org/data/icgmuus.html

Getting It Right the First Time: Belief Elicitation with Novice Participants

Li Hao and Daniel Houser*

April 9, 2010

Abstract

The auction design literature makes clear that theoretically equivalent mechanisms can perform very differently in practice. Though of equal importance, much less is known about the empirical performance of theoretically equivalent mechanisms for belief elicitation. This is especially unfortunate given the increasing interest in eliciting beliefs from (often novice) respondents in large-scale surveys. Using laboratory experiments with novice participants endowed with heterogeneous beliefs, we compare the empirical merit of two belief elicitation mechanisms proposed by Karni (2009), which we denote as "declarative" and "clock." These mechanisms are of interest because incentive compatibility does not require strong assumptions such as risk neutrality or expected utility maximization. Our key findings are that under the clock mechanism, (i) subjects are more likely to report their beliefs truthfully; and (ii) the distribution of elicited beliefs more accurately characterizes the true belief distribution. Our findings have substantial practical value to anyone wishing to elicit beliefs from novice respondents, a goal of increasing importance to large-scale survey design.

Keywords: probabilistic belief elicitation, declarative mechanism, clock mechanism, proper scoring rules, laboratory experiment

JEL code: C91

---

* Interdisciplinary Center for Economic Science and Department of Economics, George Mason University, 4400 Unviersity Dr, MSN 1B2, Fairfax, VA 22030. Hao: lhao@gmu.edu. Houser: dhouser@gmu.edu. We thank James Andreoni, Jason Aimone, Yan Chen, Soo Hong Chew, Eric Danan, Glenn Harrison, Ron Harstad, Xiaofei Pan, Rong Rong, Jacob Sagi, Nat Wilcoxon, and seminar participants at ESA international 2009 (Washington DC), SEA 2009 (San Antonio, TX), Advanced Workshop in Experimental Economics 2009 (Sydney, Australia) for helpful discussions.

I. Introduction

The insight that theoretically equivalent mechanisms can perform very differently in practice provides the foundation for a significant literature on auction design. This insight is of equal importance in the context of belief elicitation, yet much less is known about the empirical merit of alternative elicitation approaches. This paper takes a small step toward filling this gap. Using laboratory experiments with novice participants[1] endowed with heterogeneous beliefs, we compare two belief elicitation mechanisms proposed by Karni (2009). The mechanisms, which we denote as "declarative" and "clock," are of interest because incentive compatibility does not require strong assumptions such as risk neutrality or expected utility maximization. Our key findings are that under the clock mechanism, (i) subjects are more likely to report their true beliefs; and (ii) the distribution of elicited beliefs more accurately characterizes the distribution of true beliefs.

Proper scoring rules[2], first used by the meteorological statistician Brier (1950), and later popularized by Savage (1971), are the most widely known probabilistic belief elicitation procedure. However, proper scoring rules are only incentive compatible under risk neutrality. An important advance was made by Allen (1987), who developed a general scoring rule that relaxes the risk-neutrality requirement by using binary lottery payoffs to induce risk-neutrality[3]. This approach has not gained popularity, and economists continue to use proper scoring rules in their research (e.g., Nyarko and Schotter, 2002, Palfrey and Wang, 2009).Recent effort to attenuate elicitation bias caused by risk preferences include Offerman et al's (2009) approach that first estimates risk attitudes and then adjusts elicitations accordingly, and Andersen et al's(2010) joint estimation of subjective beliefs and utility functions. In addition, Schlag and van der Weele (2009) and Hossain and Okui (2010) independently proposed extensions of Allen's (1987) generalized scoring rules, both involving using binary lotteries to induce risk neutrality.

Karni (2009) proposed two elicitation procedures that also employ the binary lottery payoff technique and are incentive-compatible irrespective of risk attitudes[4]. We denote these mechanisms as "declarative" and "clock" due to the fact that they can be written, respectively, as sealed-bid second-

[1] By "novice" we mean "zero-experienced," which is often the case in large-scale elicitations.

[2] A scoring rule is "proper" if one must report true belief to maximize expected score. Since the observed outcome is used to calibrate, proper scoring rules provide incentives for accuracy. The quadratic, spherical, and logarithmic scoring rules are examples of proper scoring rules.

[3] Allen's (1987) trick was treating the cash payoffs in standard quadratic scoring rules as probabilities of winning the prize in a binary lottery, since expected utility is a linear function of the probabilities.

[4] Schlag and van der Weele (2009) argue that although both approaches are incentive compatible and robust to risk preferences, Allen's (1987) mechanism provides stronger incentives than Karni's (2009).

price and English clock auctions with random bonus payments (assuming risk-neutrality). Although Karni (2009) suggests the mechanisms are equivalent, the auction literature points to important differences in the empirical properties of second-price and clock mechanisms (see, e.g., Kagel et al, 1987; Kagel and Levin, 1993 & 2009; Harstad 2000). In light of this, it seems of substantial practical importance to investigate whether and how elicitations stemming from a clock mechanism might differ from those generated by a declarative mechanism.

The mechanisms work as follows. An individual holds a probabilistic belief about the likelihood that an event E will occur. Her default payment is contingent on E: she receives $10 if E occurs and zero otherwise. Meanwhile, the mechanism selects a random number *r* from the uniform distribution on [0, 1].

In the ***declarative mechanism***, the individual simply "declares" her belief. If her declared belief is greater than r, she keeps her original contingent payment; otherwise, her contingent payment is swapped with a lottery that pays $10 with probability of *r* and zero otherwise. Karni (2009) showed that the dominant strategy is to declare one's true belief.

In the ***clock mechanism***, the individual "competes" with a dummy bidder in an ascending clock auction. The clock starts at 0, rises continuously, and stops when one of the bidders drops out or when the clock reaches 1, whichever comes first. The dummy bidder stays in the auction as long as the clock is below *r* and exits exactly at *r*. If the dummy bidder is the first to exit, the individual keeps her original contingent payment; otherwise, her contingent payment is swapped with the lottery that pays $10 with probability of *r* and zero otherwise. Karni (2009) showed that the dominant strategy is to exit exactly at the true belief.

It is immediately evident that these mechanisms are not isomorphic: The clock mechanism fails to elicit whenever the dummy bidder is the first to exit (that subject's response has been "filtered" by the mechanism). It is easy to see that this filtering is detrimental to the clock mechanism if the filtered elicitation reflects the true belief. However, under certain reasonable conditions that we develop below, the filtering actually improves the performance of the mechanism. Our experiment's results suggest that those conditions might hold at least for novice participants, in that the clock mechanism performs significantly better for that population.

One might ask whether the advantage of the clock mechanism is due only to the "filtering" effect. We demonstrate that the answer is "no" by applying a clock-equivalent filter to the data obtained using the declarative mechanism. The resulting estimate of the population distribution of beliefs is far less accurate than the belief distribution implied by clock elicitations.

The purpose of this paper is threefold. First, we focus on novice respondents with heterogeneous beliefs, offering practical value to implementations in large-scale elicitation environments. In recent years, large-scale belief elicitations from novice respondents has become a particularly active area (e.g., Manski, 2004, Bellemare et al., 2008), in part due to the advantage of external validity. Our results take this one step further by indicating that eliciting beliefs using a clock format generates more accurate reports than the alternative approach[5].

Second, our results lend support to the auction literature by providing evidence that the (ascending) clock maintains its advantage of inducing truth-revealing dominant strategies. It is well known that the equivalence of sealed-bid second-price auctions and English clock auctions quickly breaks down in practice, as English clock auctions induce bids much closer to the true values (Kagel et al, 1987; Kagel and Levin, 1993 & 2009; Harstad 2000). We extend this literature by showing that the clock mechanism in belief elicitation maintains its "virtue" and induces beliefs truthfully.

Finally, it is important to emphasize that our subjects were endowed with objective beliefs[6]. By doing so, we examine the performance of the mechanisms absent confounds due to variations in subjects' abilities to predict uncertain events such as presidential election results (e.g., Andersen et al, 2010) and stock market prices (e.g., Offerman, et al, 2009). Winkler and Murphy (1968) made an important distinction between an individual's ability to make coherent probabilistic assessments that reflect her true beliefs and her knowledge about the event under consideration, or "normative goodness" and "substantive goodness" respectively[7]. According to this distinction, all of our participants possess "substantive goodness", and our experiment compares, using inexperienced respondents, the normative goodness induced by two theoretically equivalent elicitation mechanisms.

The paper proceeds as follows. Section 2 reviews Karni's (2009) theory, Section 3 formulates our hypothesis, Sections 4 and 5 report experimental design and results, and Section 6 concludes.

---

7 Sometimes incentives cannot be easily provided, for example, there are clear difficulties in providing incentives to elicit one's "true" belief regarding the chance of contracting cancer within the next 10 years.

[6] The optimal strategies are not obvious in Karni's mechanisms. In the post-experiment survey, many subjects answered there was no best strategy.

[7] For example, with respect to the probability of precipitation, a meteorologist possesses "substantive goodness." The reason is that she is an expert on the subject of interest, but she might not necessarily possess the "normative goodness" due to unfamiliarity with the concept of probabilities and consequently reports probabilities that are unintentionally different from her true beliefs. The contrary might hold for a statistician, who possesses the "normative goodness" due to understanding of probabilities but not necessarily the expertise in forecasting the weather.

2. Review of Karni's (2009) Theory

We briefly review Karni's (2009) mechanisms. In Savage's (1954) framework, the individual holds a belief that an event E will occur with a probability of $\pi(E)$. If E occurs, the individual gets the prize $\$x$; otherwise, she gets the non-prize $\$y$ ($\$x > \$y$). We call this mapping between the occurrence (and non-occurrence) of the event and monetary payoffs a "bet," denoted by $\beta := x_E y$.

Consider a lottery that pays $\$x$ with probability r or $\$y$ with probability $1 - r$; denote this lottery by L. The number r is randomly selected from the uniform distribution on [0, 1]. The individual knows only the distribution. She does not know the value of r when she makes the decision.

2.1. *Declarative Mechanism*

The individual submits a decision, $\mu \in [0,1]$, which is compared with the random number r . If $\mu \geq r$, she gets the bet $\beta := x_E y$; if $\mu < r$, she gets the lottery L.

Dominant Strategy. Karni (2009) demonstrates that the unique dominant strategy in the mechanism is to report truthfully: $\mu = \pi(E)$. This guarantees that the individual obtains either the bet β or the lottery L, whichever has the higher probability of winning the prize $x. The individual has no incentive to report a number greater than the truth, due to the fact that as soon as the random number r lies between the truth and her report, $\pi(E) < r < \mu$, she gets the bet β, and forgoes the lottery L that has a higher winning probability. The same logic applies when her report is lower than the truth.

2.2. *English Clock Mechanism*

In the English clock auction mechanism, the individual competes with a dummy bidder[8] who always exits the auction at number r. The clock starts at 0 and increases continuously as long as both the individual and a truth-revealing dummy bidder are "in the auction." The clock stops when at least one bidder drops out, or when the clock reaches 1, whichever occurs first. If the individual is the first to drop out, she gets the lottery L; if the dummy bidder is the first to drop out, the individual gets the bet β.

Dominant Strategy. Following Karni's (2009) argument, the dominant strategy is to stay in the auction as long as the clock is below $\pi(E)$, and drop out exactly at $\pi(E)$.

[8] In our experiments, the dummy bidder is not explicitly defined, but is represented by the random stopping of the individual's clock. Behavior differences may or may not exist between these two presentations; it is our intention to keep the environment as simple as possible.

2.3. Assumptions

Karni's (2009) mechanisms are built on two assumptions. First, the individual must have no other stakes in the events of the interest. That is, the wealth of the individual, excluding the elicitation-related payoffs, is independent of the occurrence (or nonoccurrence) of the event. Importantly, this no-stakes condition[9] (Kadane and Winkler, 1988) is also required for proper scoring rules. The second assumption is that the individual is "probabilistically sophisticated" (Machina and Schmeidler, 1992). In essence, probabilistic sophistication means that the individual ranks bets with subjective probabilities over outcomes in a similar fashion as she would rank lotteries with an objective probability distribution. A noteworthy advantage of Karni's (2009) mechanism is that a person displaying probabilistic sophistication need not be an expected utility (EU) maximizer, while EU is required for the validity of approaches based on scoring rules.

3. Testable Hypothesis

Since the clock mechanism fails to elicit probabilistic beliefs whenever the dummy bidder is the first to exit, the clock mechanism "filters" decisions. For the purpose of accuracy, this filtering can be detrimental or beneficial depending on whether the filtered decision is optimal. Considering that novice participants do not always form optimal decisions, a natural question emerges: *Does the clock mechanism filter more non-optimal decisions than optimal decisions?*

If the answer is yes, then provided that participants employ the same strategies in the two environments, beliefs elicited from the clock mechanism are evidently more likely to reflect the true beliefs than are those from the declarative mechanism.

This section derives the condition under which the clock mechanism does indeed filter more non-optimal than optimal decisions. It is built upon the following assumptions.

Assumption 1. Individuals employ exactly the same decision strategies in the two mechanisms, although in the clock mechanism when the dummy bidder is the first to exit, their decisions are filtered and thus not recorded. That is, we assume different mechanisms do not induce different strategies.

[9] As Kadane and Winkler (1988) put it, the elicited probabilities intertwine with utilities "not just through the explicit or implicit payoffs related to the elicitation process, but also through other stakes the individual may have in the events of interest." No-stakes condition is a strong assumption, but it is unclear to what extent this assumption matters empirically.

Assumption 2. We assume only two types of decisions: i) optimal decisions that reveal beliefs $\pi(E)$ truthfully; and ii) naïve decisions that are randomly picked[10] from a cumulative distribution function $F_n(\cdot)$. Let the proportions of the optimal and naïve decisions be α and $1-\alpha$ respectively ($0<\alpha<1$).

Assumption 3. The loss function is defined such that the clock mechanism, as compared to the declarative mechanism, gains a unit of benefit when it filters a naïve decision, $\Delta L = 1$, and suffers a unit of loss when it filters an optimal decision, $\Delta L = -1$.

$$\Delta L_i = \begin{cases} 1, & if\ the\ filtered\ decision\ i\ is\ naïve \\ -1, & if\ the\ filtered\ decision\ i\ is\ optimal \end{cases}$$

Assumption 4. Suppose the number *r*, the probability of winning the prize in the lottery, is randomly drawn from a cumulative distribution function $F_r(\cdot)$[11].

Theorem 3.1: Under the above assumptions,

Beliefs elicited from the clock mechanism are more likely to reflect true beliefs than those from the declarative mechanism as long as the net benefit (benefits minus losses) from filtered decisions in the clock mechanism is no less than zero.

$$\Delta L = (1-\alpha)\int_0^1 F_r(u)dF_n(u) - \alpha F_r(\pi) \geq 0 \tag{3.1}$$

In particular, we specified both $F_r(\cdot)$ and $F_n(\cdot)$ as uniform distribution on [0, 1] in our experiments. This means the number *r* and naïve decisions are randomly distributed on Uniform [0, 1], and thus equation (3.1) simplifies to

$$\Delta L = \frac{1}{2}(1-\alpha) - \pi\alpha \geq 0 \tag{3.2}$$

In (3.2), the first term $\frac{1}{2}(1-\alpha)$ represents the benefits from filtering naïve decisions: there are totally $(1-\alpha)$ naïve decisions in the population, and exactly half of them get filtered because both naïve decisions and the number *r* (that executes the filtering function) are uniformly distributed on [0,

[10] Random decision is one of the simplest models. The post-experiment survey lends support to this specification, as a significant number of subjects claimed that they picked their answers randomly, such as their lucky numbers.
[11] Although Karni specified the uniform distribution on [0, 1] for $F_r(\cdot)$, we note that the mechanisms hold for any continuous distribution $F_r(\cdot)$.

1]. The second term $\pi\alpha$ represents total losses from filtering optimal decisions: among a total of α optimal decisions that wish to drop out at exactly π, the uniformly distributed number *r* filters exactly $\pi\alpha$ optimal decisions. Intuitively, the larger the (positive) difference between the value of ΔL and zero, the greater the advantage the clock mechanism holds over the declarative mechanism.

In particular, holding the belief π fixed, the lower the proportion of optimal decisions (α), the greater the value of ΔL. The reason is that the more naïve decisions in the population, the more benefit the clock mechanism has by filtering them and thus improve elicitation accuracy. In contrast, when every decision is optimal, the clock mechanism has no advantage.

The proportion of optimal decisions in a population is likely to increase with experience. Thus, it is natural to conjecture that the proportion of optimal decision-makers is low among novice respondents, and so we obtain the following hypothesis.

Hypothesis: *With novice respondents, beliefs elicited from the clock mechanism are more likely to be accurate (subjects report the endowed beliefs truthfully) than those from the Declarative mechanism.*

On the other hand, holding the proportion of optimal decisions fixed, the lower the belief π, the greater the value of ΔL. The reason is that naïve decisions are always filtered by half, but the percentage of optimal decisions being filtered is proportional to the belief π. Therefore, setting beliefs closer to the zero (or the lower bound) creates a favorable environment for obtaining evidence to support our hypothesis; this guides us in choosing which probabilistic beliefs to give our subjects.

4. Experiment Design and Procedures

A key feature of our design is that we endow our subjects with objective probabilistic beliefs. We made this choice for two reasons. First, as noted above, our goal is to assess the ability of the mechanisms to induce truth-telling strategies; therefore, eliminating differences in the ability of participants to form accurate beliefs helps in this regard. Second, our research is connected to second-price and English clock auctions, where participants make decisions using known (induced) values. In our case, half of subjects were endowed with belief of 0.2 and the other half with belief of 0.3. It is important to point out that the dominant strategies are not obvious in either mechanism[12]. Instructions are attached in the appendix.

[12] However, the truth-revealing dominant strategy is obvious in proper scoring rules, and thus impractical for us to include in the comparison.

4.1. Declarative Mechanism

Endowed Belief = 0.2. The subject chooses between bag A and bag B. She knows that bag A has 10 chips in total: 2 white chips and 8 black chips. She also knows that bag B also has a total of 10 chips of the two colors, but the number of white chips (denoted by R) is equally likely to be {1, 2, 3, 4, 5, 6, 7, 8, 9}.

The subject declares a number between 1 and 9 (inclusive; integer only). The experimenter compares her stated number with R, i.e., the number of white chips in Bag B. If the stated number is greater than R, the subject gets to draw a chip from bag A; otherwise, she draws a chip from bag B. In either case, the subject is paid $10 if she draws a white chip, and is paid $1 if she draws a black chip.

Endowed Belief = 0.3. This proceeds as the above procedure, except now there are 3 white chips (and 7 black chips) in bag A.

Dominant Strategy*.* Take bag A as the default choice; the declarative mechanism is effectively asking the subject, "*What is minimum number of white chips in bag B so that you are willing to switch to bag B?"* The dominant strategy is to declare either the number of the white chips in bag A, or one more than the number of white chips in bag A. The presence of two equally advantageous actions stems from our using a discrete state space, but vanishes when probabilities are drawn from a continuum.

4.2. Clock Mechanism

Endowed Belief = 0.2. Bag A and bag B are exactly the same as in the declarative mechanism with endowed belief of 0.2. Instead of declaring a number, the subject participates in a clock auction. The clock starts at number 1, and rises by 1 every 5 seconds. The clock stops when the subject drops out, or when the clock reaches number R, whichever comes first[13]. If the clock stops due to reaching number R, the subject gets to draw a chip from bag A; if the clock stops due to the subject's dropout, she draws a chip from bag B. In either case, the subject is paid $10 if she draws a white chip and $1 if she draws a black chip.

Endowed Belief = 0.3. This proceeds just like the above procedure, except now there are 3 white chips (and 7 black chips) in bag A.

Dominant Strategy*.* Similarly, consider the bag A as the default choice; the clock mechanism is effectively asking the subject, "*Taking the current displayed number as the number of white chips in bag B, do you want to switch to bag B now?"* The dominant strategy is to indicate the willingness-to-switch

[13] The clock displays the number R for 5 seconds, so the subject can indicate to drop out at number R, which means she obtains bag A. In this case, bag A and B have exactly the same amount of white chips.

by exiting as soon as the displayed number is the same as the number of the white chips in bag A, or one more than the number of white chips in bag A.

4.3. Treatment Design

Subjects participated in two (independent) rounds of decision making. The second round was a "surprise" as we announced it only after finishing the first round[14]. Table 1 summarizes our two-by-two treatment design.

Table 1: Treatment Design

Treatment	Number of Subjects (total=130)	First Round	Second Round
D+D	24	Declarative	Declarative
D+C	29	Declarative	Clock
C+D	37	Clock	Declarative
C+C	40	Clock	Clock

Each session consists of 4 to 8 participants, half of them were endowed with belief of 0.2 and the other half with belief of 0.3, for a heterogeneous belief environment. Subjects were given new instructions and quizzes if they participated in a new mechanism in the second round (i.e., in treatments "D+C" and "C+D").

4.4. Procedures

All sessions were conducted between April and October 2009 at the Interdisciplinary Center for Economic Science (ICES) laboratory of George Mason University in Fairfax, VA. Subjects were recruited with standard ICES procedures. A total of 130 undergraduates participated, with 4-8 people per session. Participants were paid a guaranteed $5 for showing up on time, in addition to their earnings in the experiment. The average earnings were $16, and sessions lasted between 30 and 60 minutes. The experiment was programmed using E-prime.

To facilitate the understanding of the mechanisms, subjects were first given abundant time to read the instructions on their own; afterwards, the experimenter read the instructions aloud to them. Each subject then had to finish a quiz, for which their answers were recorded. The experimenter then

[14] Subjects were not told there was a second round at the beginning of the experiment. The experimenter announced, "That was the end of the experiment. However, we still have some time left; let us do another experiment so you can make more money."

announced and explained the correct answers, emphasizing justifications behind the correct answers. We found that a majority of the subjects correctly answered all quiz questions. Subjects were required to ask questions privately, and to think carefully about the quiz answers. Note that we avoided using words such as “probability” or “distribution.” In fact, subjects were presented with only whole numbers, which should improve comprehension.

Also, we generated the random number R (number of white chips in bag B) using the following three steps. In step one, the experimenter showed subjects a deck of 9 cards, numbered from 1 to 9. The experimenter then put each card into one of nine opaque envelopes and shuffled the envelopes thoroughly. In step two, each subject was asked to pick an envelope and immediately return it to the experimenter without opening it. At this point, the experimenter wrote the subject’s ID on the envelope. Finally, in step three, the experimenter publicly opened each envelope from a distance (so that no subjects could read the numbers inside the envelopes), transcribed the random number R for each subject ID, and then sealed the envelope. The reason for these steps was to prove that the number R was determined prior to the subjects’ decisions, as well as to make clear that R was an integer between 1 and 9, each with equal probability. For the surprise second round, a new random number was generated for each subject from a new set of nine opaque envelopes, in exactly the same procedures described above.

Finally, we implemented the payment procedure individually to ensure that subjects knew they were making independent decisions. After the choice of bag was determined, the experimenter went to the subject with the appropriate bag and chips, which the subject examined. The experimenter put the 10 chips into the opaque cloth bag. The subject then drew *one* chip from the bag while keeping her head turned away. The subject earned $10 if she drew a white chip and $1 if she drew a black chip.

5. Results

We organize our results in three parts. The first subsection describes decisions from the first round, and reports the first result supporting our hypothesis that subjects are more likely to report their true beliefs in the clock mechanism. We follow up with our second result and show that the advantage of the clock mechanism is not purely driven by the filtering effect.

The second subsection presents decisions from the second round, and our third result shows that the clock mechanism does not perform worse than the declarative mechanism with experienced participants. Finally, in the third subsection, we classify subjects into four types based on decisions from the two rounds in order to investigate how behavior changes with experience.

5.1. Responses from Novice Participants

With heterogeneous beliefs of 0.2 and 0.3, Table 2 describes individual decisions in the first round[15].

Table 2: Novice Responses: Descriptive Statistics

	Declarative	Clock
Observations	53	77
Optimal decisions	25/47%	30/39%
Non-optimal decisions	28/53%	17/22%
Filtered decisions	--	30/39%
Mean of deviation from truth	.0604** (.0264)	.0362* (.0203)
Mean of absolute deviation from truth	.1208*** (.0221)	.0787*** (.0175)

Note: In parentheses are standard errors. All tests are two-sided one-sample *t*-test, *indicates 10% significance level, ** is 5%, and *** is 1%.

A total of 130 subjects participated in our experiments. Among the 53 and 77 observations in the declarative and clock mechanisms respectively, the proportions of optimal decisions are 47% versus 39%, and non-optimal decisions are 53% versus 22%.

Remark 1. *Slightly less than half of the novice responses use optimal strategies, even with our explicit effort for simplicity and transparence*. This suggests that the dominant strategies in this environment are indeed not trivial to subjects. Moreover, the percentage of optimal strategies in the declarative mechanism is consistent with that of the second-price auctions in Cooper and Fang (2008, p. 1583).

The clock mechanism has an extra category of "filtered decisions," which takes up 39% of total decisions. If we assume that two optimal decisions are equally likely to be chosen, then only 30% of optimal decisions are filtered[16]. This indicates that there are more than 39% of naïve decisions in the filtered category.

[15] The two equally optimal decisions are both set as deviation of 0. In particular, when endowed belief is 0.2, decisions {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9} are converted into deviations {-0.1, 0, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6}; when endowed belief is 0.3, they are {-0.2, -0.1, 0, 0, 0.1, 0.2, 0.3, 0.4, 0.5}.

[16] If endowed with belief of 0.2, there should be (0.2+0.3)/2=25% of optimal decisions filtered; similarly, if endowed with belief of 0.3, (0.3+0.4)/2=35% of optimal decisions are filtered. Since the population consists of the equal share of the two beliefs, 30% of the population's optimal decisions are filtered.

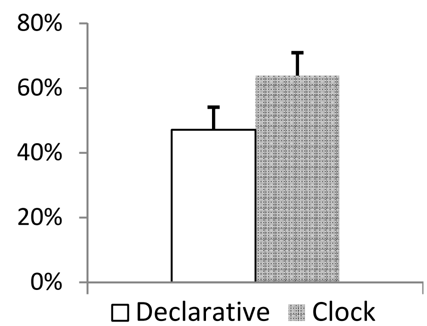
The mean of deviations and the mean of absolute deviations from optimal decisions are both significantly different from zero in both mechanisms[17]. They are smaller in the clock mechanism than the declarative mechanism, but insignificantly.

We take a closer examination of the two mechanisms by excluding filtered decisions. The logic behind this is that: (i) filtered decisions do not elicit information about the respondents; and (ii) they create the "inconsistent" problem, as the declarative mechanism has a higher proportion of optimal decisions and also a higher proportion of non-optimal decisions. We present the first result:

Result 1. *Elicited beliefs from novice respondents in the clock mechanism are more likely to be accurate than those from the declarative mechanism*.

Evidence: Among elicited beliefs, i.e., decisions that are not filtered, the proportions of optimal decisions are 64% and 47% in the clock and declarative mechanisms respectively (Figure 1). A two-sided Wilcoxon-Mann-Whitney test on binary data (1 if a decision is optimal, 0 otherwise) found a statistical different at p=.096.[18]

Figure 1: Proportion of Optimal Decisions in First Round: Excluding Filtered Decisions



Note: Error bars reflect the standard deviations of the samples.

However, *is the accuracy of the clock mechanism purely driven by its '"filtering" effect?* If respondents adopt exactly the same strategy under the two mechanisms, as assumed in Theorem 3.1, we can obtain the same beliefs elicited using the clock mechanism by applying an "artificial" clock-equivalent filter to the data obtained using the declarative mechanism. Imagining that the subject uses a pre-determined strategy to make her decisions, we should obtain the same data if her decision is

[17] Filtered decisions are excluded when calculating statistics.
[18] The result in a one-sided test is significant at the 5% level; after all, we have a clear ordered hypothesis illustrated in section 3.

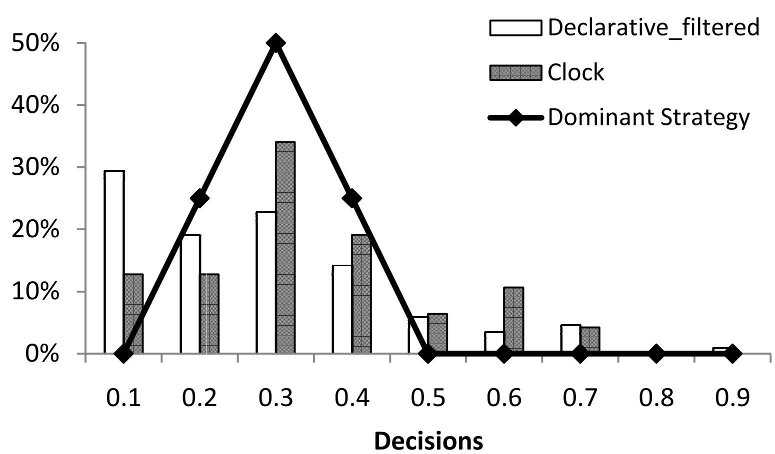
filtered during the experiment (in the clock mechanism) or after the experiment (by the artificial filter applied to the declarative mechanism).

In particular, for each decision in the declarative mechanism, we randomly select a number that is equally likely to be 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, or 0.9. The decision is filtered and discarded if it is greater than the random number. The sample of 53 independent decisions from the declarative mechanism is filtered 10,000 times, and the average frequency of each decision remaining is calculated. We obtain our second result.

Result 2. *The improved accuracy of the clock mechanism is not driven solely by the filtering effect; decisions of the declarative mechanism using artificial filtering are significantly different from those of the clock mechanism.*

Evidence: Shown in Figure 2, the decision distribution of the declarative mechanism using artificial filtering is significantly different from the decision distribution of the clock mechanism, as the Chi-squared goodness-of-fit test reports the p-value at p=.014.[19]

Figure 2. Distributions of First-Round Decisions



Note: Half are endowed with belief of 0.2, and the other half endowed with the belief of 0.3.

In addition, Figure 2 plots the decision distribution (in black line) when all participants adopt optimal strategies[20], which has a single mode at 0.3. In comparison, the elicited beliefs from the clock

[19] The Chi-square goodness-of-fit test between the decision distributions of the clock mechanism and the declarative mechanism without artificial filtering reports a p-value at 0.063.

mechanism (black bars) also have a single mode at 0.3, and thus characterize the dominant strategy distribution fairly well. However, the mode of the filtered elicitations from the declarative mechanism (white bars) is the non-optimal decision 0.1, and the distribution is far less informative about the population beliefs than that of the clock mechanism.

5.2. Responses from One-Time Experienced Participants

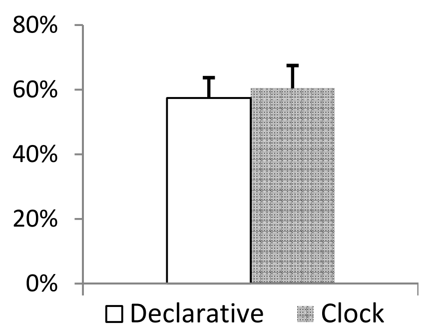
We are interested in whether experience increases the proportion of optimal decisions, so we examine the second-round decisions and present our third result,

Result 3. *Using one-time experienced participants, the likelihood that an elicited belief equals the endowed belief is the same using the declarative or the clock mechanism.*

Evidence: Figure 3 shows that the proportions of optimal decisions in the declarative and clock mechanisms (excluding filtered decisions) are 57% and 60% respectively, and that they are not statistically distinguishable (p=.84, two-sided Wilcoxon-Mann-Whitney). Combining results 1 and 3, we conclude that

Remark 2. *Overall, beliefs elicited using the clock mechanism are at least as likely to be accurate as those elicited using the declarative mechanism*[21].

Figure 3: Proportion of Optimal Decisions in Second Round: Excluding Filtered Decisions



Note: Error bars reflect the standard deviations of the samples.

We apply the artificial filtering on the second-round decisions from the declarative mechanism and obtain:

[20] Recall that half of the population holds belief of 0.2 and the other half holds belief of 0.3, and assume the two dominant strategies (i.e., belief and belief+0.1) are equally likely to be chosen; then there are 25% of decisions are 0.2, 50% are 0.3 and 25% are 0.4.

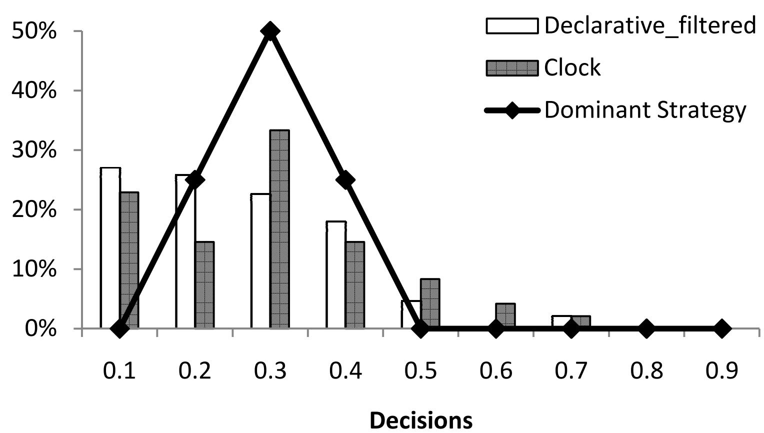
[21] In the declarative mechanism, deviations from dominant strategies in the second round are significantly smaller than they are in the first round (p=.04, two sided Wilcoxon-Mann-Whitney).

Result 4. *The second-round decisions of the declarative mechanism using artificial filtering are significantly different from those of the clock mechanism.*

Evidence: The Chi-squared goodness-of-fit test reports that the distribution of decisions from the declarative mechanism using artificial filtering are significantly different from the decision distribution of the clock mechanism [22] (p=.052).

Figure 4 has the exactly same setting as Figure 2. We notice that in the second round, elicited beliefs from the clock mechanism continue to reveal the mode of the underlying truth, whereas the declarative mechanism does not.

Figure 4. Distributions of Second-Round Decisions



Note: Half are endowed with belief of 0.2, and the other half endowed with belief of 0.3.

5.3. Cross-Learning and Types

Does first-time experience in different mechanisms affect decisions in the second round? Our two-by-two treatment design allows us to investigate the affect on decisions of experience with different mechanisms.

Result 5. *The likelihood that a second-round decision is optimal using either the declarative or clock mechanism does not depend on the first-round mechanism experience.*

Evidence: The likelihood that a second-round decision obtained using the declarative mechanism is optimal is independent of the first-round experience being declarative or the clock mechanism (p=.928, two-sided Wilcoxon-Mann-Whitney). Similarly, the likelihood that a second-round decision obtained

[22] The decision distributions of the declarative mechanism without artificial filtering versus the clock mechanism are not significantly different from each other: p=.392, Chi-squared goodness-of-fit test.

using the clock mechanism is also independent of the first-round experience being declarative or the clock mechanism (p=.180, two-sided Wilcoxon-Mann-Whitney).

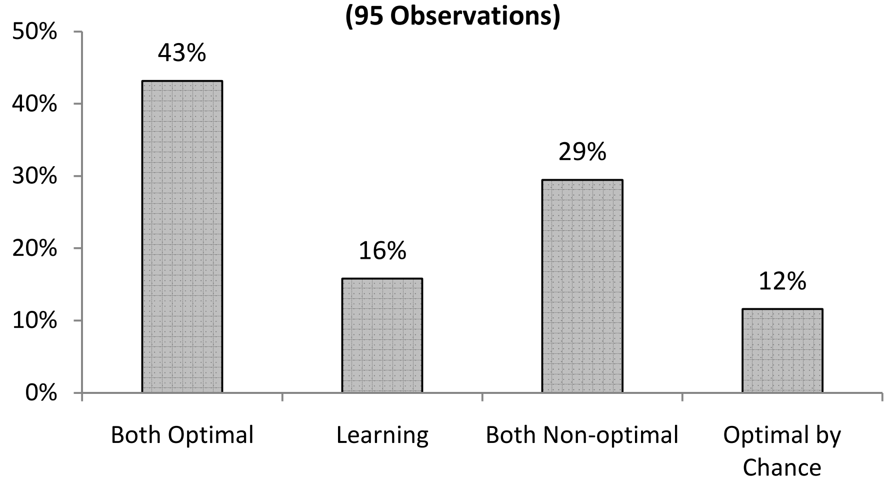
Combining results 3 and 5, we conjecture that the improvement of the accuracy of the declarative mechanism can be obtained from experiences of either mechanism. However, our result 5 might be limited by the fact that our subjects have only one round of experience. Other studies (e.g., Harstad, 2000) show that prior experiences with different mechanisms have a significant effect on subjects' strategies in the subsequent mechanisms.

Finally, we categorize subjects into four types based on their decisions in the two rounds. We drop observations from anyone whose decisions had been filtered at least once by the clock mechanism.

The first type "Both Optimal" adopts dominant strategies in both rounds; 43% of our subjects are this type. The second type, "Learning," (16% of all subjects) are those who learned the dominant strategies with experience, as they made non-optimal decisions in the first round, but switched to optimal decisions in the second round.

The last two types are people who at no point during the experiments understood the dominant strategies. This type makes up 41% of our subjects. "Both Non-optimal" consists of subjects who did not follow the dominant strategy in either round, and "Optimal by Chance" describes those who made optimal decisions in the first round but non-optimal decisions in the second round. We conjecture that the fourth type made optimal decision in the first round only by chance. Indeed, our post-experiment survey confirms that the majority of the fourth type (nine out of thirteen) did not believe there was a best answer.

Figure 5: Distribution of Decisions in Both Rounds (95 Observations)



6. Concluding Remarks

In a laboratory study using novice participants endowed with heterogeneous beliefs, we compared the declarative and clock belief elicitation mechanisms proposed by Karni (2009). These mechanisms are of interest because incentive compatibility does not require strong assumptions such as risk neutrality or expected utility maximization. We found that, in relation to the declarative mechanism, under the clock mechanism elicited beliefs are more likely to be accurate and the distribution of elicited beliefs more accurately characterizes the underlying (endowed) beliefs. These findings resonate with the auction literature by providing evidence that the ascending clock mechanism continues to induce truth-telling in a belief elicitation context. This result seems to have implications for the practical design of incentive-compatible belief elicitation mechanisms.

A limitation of our study stems from our choice of parameters. We induce beliefs that are nearer to zero than one, and we explained that doing this provides a favorable environment for the clock mechanism. While the clock may perform less well when actual beliefs are closer to the relevant upper bound, this is not necessarily a problem in practice. In particular, the investigator is free to choose the clock's range and increments arbitrarily, and can always include extra ticks at larger values. In doing so, the investigator can be more confident that actual beliefs lie well below the clock's upper limit.

Future research might investigate similar questions as addressed here in the context of the mechanisms suggested by Schlag and van der Weele (2009) and Hossain and Okui (2010). It would also be important to know whether the truth-inducing advantage of the clock mechanism persists in a simple belief elicitation environment where incentive-compatible mechanisms are difficult to implement. Of particular interest here are large-scale phone or internet surveys of respondents' beliefs regarding their risk of contracting diseases, losing their jobs, or having another baby. Investigating such questions will be an additional step towards a better understanding of alternative belief elicitation procedures.

References

Allen, Franklin (1987): “Discovering Personal Probabilities When Utility Functions Are Unknown,” *Management Science*, 33 (4), 542-544

Andersen, Steffen, John Fountain, Glenn. W. Harrison and E. Elisabet Rutström (2010): “Eliciting Subjective Probabilities,” (working paper 2010-06, Center for the Economic Analysis of Risk, Georgia State University)

Andreoni, James, Yeon-Koo Che and Jinwoo Kim (2007): “Asymmetric information about rivals’ types in standard auctions: An experiment,” *Games and Economic Behavior*, 59, 240-59

Andreoni, James and Tymofiy Mylovanov (2008): “Diverging Opinions”, (working paper: http://econ.ucsd.edu/~jandreon/WorkingPapers/disagreement.pdf)

Bohm, Peter, Johan Linden and Joakin Sonnegard (1997): “Eliciting Reservation Prices: Becker-DeGroot-Marschak Mechanisms vs. Markets,” *The Economic Journal*, 107(443), 1079-89

Becker, Gordon M., Morris H. Degroot, and Jacob Marschak (1964): “Measuring Utility by a Single-Response Sequential Method,” *Behavioral Science*, IX, 226-32

Berg, Joyce E., Lane A. Daley, John W. Dickhaut, and John R. O’Brien (1986): “Controlling Preferences for Lotteries on Units of Experimental Exchange,” *The Quarterly Journal of Economics*, 101(2), 282-306

Brier, G.W. (1950): “Verification of Forecasts Expressed in Terms of Probability”, *Monthly Weather Review*, 78, 1-3.

Chesley, G.R. (1975): “Elicitation of Subjective Probabilities: A Review”, *The Accounting Review*, April 1975, 325-337

Cooper, David J. and Hanming Fang (2008): “Understanding Overbidding in Second Price Auctions: An Experimental Study,” *The Economic Journal*, 118, 1572-95

DeMarzo, Peter M., Kremer, Ilan and Skrzypacz, Andrzej (2005): “Bidding with Securities: Auctions and Security Design,” *American Economic Review*, 95 (4), 936-59

Delavande, Adeline and Susann Rohwedder (2008):“Eliciting Subjective Probabilities in Internet Surveys,” *Public Opinion Quarterly*, 72(5), 866-91

Fuj, Edwin T., Michael Hennesy and James Mak (1985): “An Evaluation of the Validity and Reliability of Survey Response Data on Household Electricity Conservation”, *Evaluation Review*, 9(1), 93-104

Gilboa, Itzhak, Andrew W. Postlewaite, and David Schmeidler (2008): “Probability and Uncertainty in Economic Modeling, ” *Journal of Economic Perspective*, 22 (3), 173-88

Harstad, Ron (2000): “Dominant strategy adoption and bidders' experience with pricing rules,” *Experimental Economics*, 3(3), 261-280

Hossain, Tanjim and Okui Ryo (2010): “The Binarized Scoring Rule of Belief Elicitation,” working paper.

Kadane, Joseph B. and Robert L. Winkler (1988): “Separating Probability Elicitation from Utilities,” *Journal of the American Statistical Association*, 83(402), 357-63

Kadane, B. Joseph and Lara J. Wolfson (1998): “Experiences in Elicitation,” *Journal of the Royal Statistical Society, Series D (The Statistician)*, 47(1), 3-19

Kagel, John, Levin, Dan and Harstad, Ron (1987): "Information Impact and Allocation Rules in Auctions with Affiliated Private Values: A Laboratory Study," *Econometrica*, 55(6), 1275-1304

Kagel, John and Levin, Dan (1987): "Independent Private Value Auctions: Bidder Behaviour in First- Second- and Third-Price Auctions with Varying Numbers of Bidders," *The Economic Journal*, 103(419), 868-879

Kagel, John and Levin, Dan (2009): "Implementing Efficient Multi-object Auction Institutions: An Experimental Study of the Performance of Boundedly Rational Agents," *Games and Economic Behavior*, 66, 221-237

Karni, Edi (2009): "A Mechanism for Eliciting Probabilities," *Econometrica*, 77(2), 603-6

Karni, Edi (2008): "A Mechanism Design for Probability Elicitation," Unpublished Manuscript, Johns Hopkins University

Machina, Mark J. (2004): "Almost-objective uncertainty," *Economic Theory,* 24, 1-54

Machina, Mark J. and D. Schmeidler (1992): "A more robust definition of subjective probability," *Econometrica*, 60, 745-80

Machina, Mark J. (2009): "Risk, Ambiguity, and the Rank-Dependence Axioms," *American Economic Review*, 99(1), 385-92

Nyarko, Yaw and Andrew Schotter (2002): "An Experimental Study of Belief Learning Using Elicited Beliefs," *Econometrica*, 70 (3), 971-1005

Halevy, Yoram (2007): "Ellsberg Revisited: An Experimental Study," *Econometrica*, 75(2), 503-536

Offerman, Theo, Joep Sonnemans, Gijs Van de Kuilen and Peter P. Wakker (2009): "A Truth-Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes," *Review of Economic Studies*, 76(4), 1461-89

Palfrey, Thomas R. and Wang Stephanie W. (2009), "On Eliciting Beliefs in Strategic Games", *Journal of Economic Behavior & Organization*, 71, 98-109

Prasnikar, Vesna (1993): "Binary Lottery Payoffs: Do They Control Risk Aversion?" (Discussion Paper No. 1059, Center for Mathematical Studies in Economics and Management Science, Northwestern University: http://www.kellogg.northwestern.edu/research/math/papers/1059.pdf)

Roth, Alvin E. and Malouf W. K. (1979): "Game-Theoretic Models and the Role of Information in Barganing," *Psychological Review*, 86, 574-94

Roth, Alvin E. and J. Keith Murninghan (1982): "The role of Information in Bargaining: An Experimental Study", Econometrica, 50(5), 1123-42

Roth, Alvin E. and Francoise Schoumaker (1983): "Expectations and Reputations in Bargaining: An Experimental Study," *American Economic Review*, 73(3), 362-72

Rutström, E. Elisabet (1998): "Home-grown Values and Incentive Compatible Auction Design," International Journal of Game Theory, 27, 427-41

Rutström, E. Elisabet and Nathaniel T. Wilcox (2009): "Stated Beliefs versus Inferred Beliefs: A Methodological Inquiry and Experimental Test," *Games and Economic Behavior*, 67, 616-32

Savage Leonard J. (1971): “Elicitation of Personal Probabilities and Expectations,” *Journal of American Statistical Association*, 66(336), 783-801

Selten, R., A. Sadrieh, and K. Abbink (1999): “Money does not induce risk neutral behavior, but binary lotteries do even worse,” *Theory and Decision*, 46, 211-49

Smith, Vernon (1994): “Economics in the Laboratory”, *Journal of Economic Perspectives*, 8 (1), 113-31

Winkler, Robert L. and Allan H. Murphy (1968): “‘Good’ Probability Assessors,” *Journal of Applied Meterorology*, 7, 751-8

Yang, Zhijun, Coble, Keith H. and Hudson, M. Darren (2009): “The role of individual personality type in subjective risk elicitation outcomes”, *Journal of Risk Research*, 12(2), 209-22

Appendix:

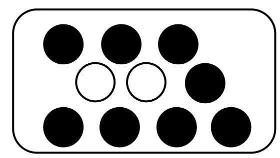
Instructions for Declarative mechanism with endowed belief of 0.2:

Welcome to this experiment! In addition to the $5 for showing up on time, you will be paid in cash based on your decisions in the experiment. Please note that no other participant's decisions in this experiment will affect your earnings, and vice versa. Please read these instructions carefully. Raise your hand if you have any questions, and the experimenter will come to assist you.

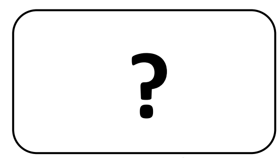
<u>Overview:</u>
The procedure is simple. You will first submit a number, and then you will draw a chip from one of two bags. If the chip you draw is white you will earn $10, if it is black you will earn $1.

<u>Details:</u>
Bag A has **2** white chips and **8** black chips for a total of **10**. **Bag B** also has **10** chips, some white, some black, but you do not know how many of each. *The number of white chips in Bag B* is on the card in the sealed envelope at your desk. This card was drawn in advance from a deck of 9 cards, labeled from 1 to 9. Please do not open the envelope until you are told to do so.

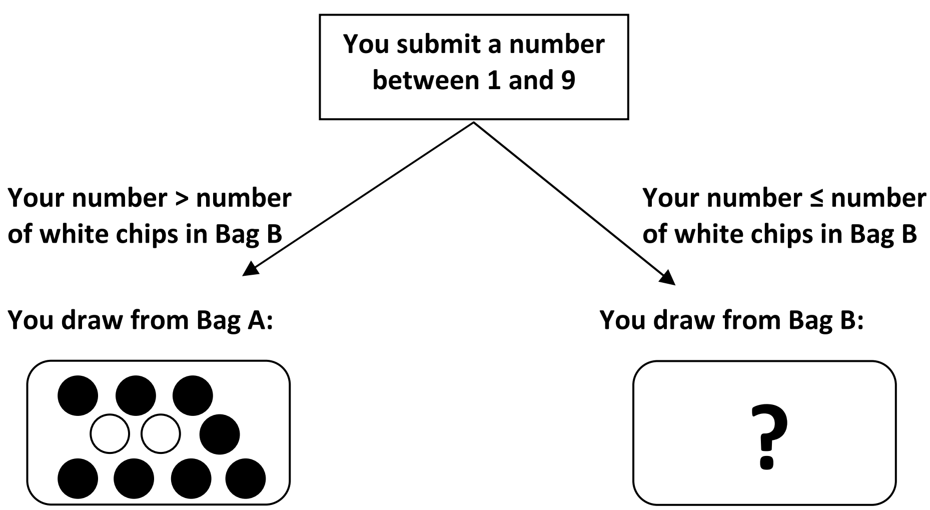
Bag A:



Bag B:



To determine the bag you'll draw from, you will first submit a number between 1 and 9. If the number you submit is less than or equal to *the number of white chips in Bag B*, you will draw from **Bag B**, otherwise you will draw from **Bag A**.



<u>Your payment:</u> If you draw a white chip you earn $10; a black chip earns you $1.

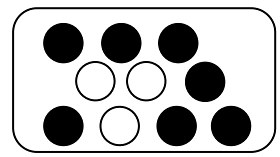
Instructions for Clock mechanism with endowed belief of 0.3:

Welcome to this experiment! In addition to the $5 for showing up on time, you will be paid in cash based on your decisions in the experiment. Please note that no other participant's decisions in this experiment will affect your earnings, and vice versa. Please read these instructions carefully. Raise your hand if you have any questions, and the experimenter will come to assist you.

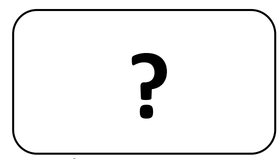
<u>Overview:</u>
The procedure is simple. You will first participate in an exercise, and then you will draw a chip from one of two bags. If you draw a white chip you will earn $10, if it is black you will earn $1.

<u>Details:</u>
Bag A has **3** white chips and **7** black chips for a total of **10**. **Bag B** also has **10** chips, some white, some black, but you do not know how many of each. *The number of white chips in Bag B* is on the card in the sealed envelope at your desk. This card was drawn in advance from a deck of 9 cards, labeled from 1 to 9. Please do not open the envelope until you are told to do so.

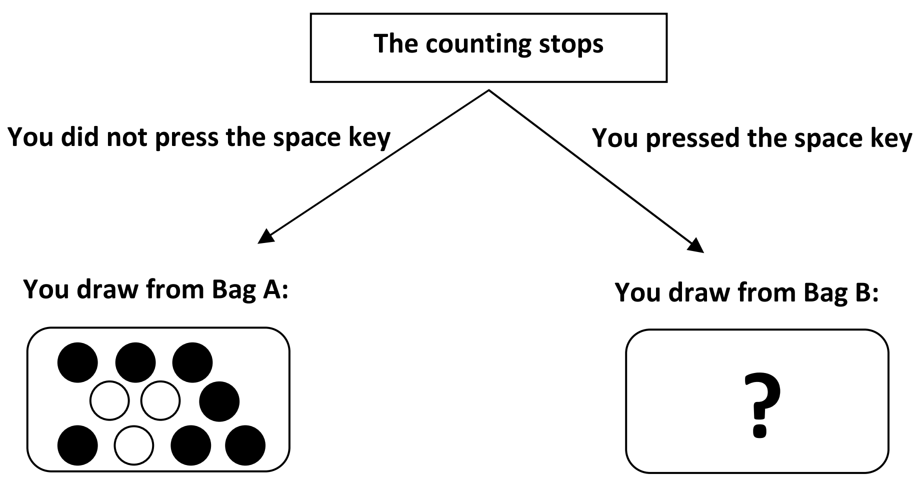
Bag A:



Bag B:



To determine the bag you'll draw from, you will first participate in an exercise. The computer screen in front of you will start counting from number **1**, and increase by **1** every 5 seconds until it reaches the number in the sealed envelope. You can stop the counting at any point by pressing the space key. If you press the space key before the counting stops, you draw from **Bag B**, otherwise you draw from **Bag A**.



<u>Your payment:</u> If you draw a white chip you earn $10; a black chip earns you $1.