



---

INTERDISCIPLINARY CENTER  
FOR ECONOMIC SCIENCE  
GEORGE MASON UNIVERSITY

---

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

Li Hao and Daniel Houser

August 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](http://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\*

August 1st, 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. 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 their 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 endowed beliefs; and (ii) the distribution of elicited beliefs more accurately characterizes the distribution of endowed beliefs. Our findings have substantial practical value to anyone with interest in eliciting beliefs from novice respondents, a goal of increasing importance, for example, when designing large-scale surveys.

**Keywords:** 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 University Dr, MSN 1B2, Fairfax, VA 22030.

Li Hao: email: [lhao@gmu.edu](mailto:lhao@gmu.edu), office: (+1) 703 993 4858; fax: (+1) 703 993 4851.

Daniel Houser: email: [dhouser@gmu.edu](mailto:dhouser@gmu.edu), office: (+1) 703 993 4856; fax: (+1) 703 993 4851.

## I. Introduction

The insight that theoretically equivalent mechanisms can perform very differently in practice provides the foundation for a significant literature on auction design (see e.g., surveys by Kagel, 1995 and Kagel and Levin, 2008). 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 step toward filling this gap. Using laboratory experiments with novice participants<sup>1</sup> endowed with heterogeneous beliefs, we compare the two belief elicitation mechanisms proposed by Karni (2009). The mechanisms, which we denote as “declarative” and “clock,” are of interest because their 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 endowed beliefs; and (ii) the distribution of elicited beliefs more accurately characterizes the distribution of endowed beliefs.

Proper scoring rules<sup>2</sup>, first used by the meteorological statistician Brier (1950), and later popularized by Savage (1971), are widely used belief elicitation procedures (e.g., Nyarko and Schotter, 2002, Palfrey and Wang, 2009). Proper scoring rules, however, are not generally incentive compatible. For example, incentive compatibility fails whenever an individual is either risk-averse or risk-seeking. In light of this, Offerman et al. (2009) proposed that one first estimate risk attitudes and then adjust elicited beliefs accordingly. Alternatively, Andersen et al. (2010) suggested that one estimate jointly subjective beliefs and risk attitudes using maximum likelihood methods.

Allen (1987) was first to generalize the quadratic scoring rule by utilizing binary lottery payoffs to induce risk-neutrality<sup>3</sup>. McKelvey and Page (1990) implemented a similar generalized quadratic scoring rule in an early laboratory study of belief elicitation. More recently, Hossain and Okui (2010) showed that quadratic scoring rules with and

---

<sup>1</sup> “Novice participants” refer to people who have no experience with the mechanism, which is common in large-scale elicitations.

<sup>2</sup> A scoring rule is “proper” if one must report true beliefs to maximize expected score. Reported beliefs are compared against realized outcomes, so that proper scoring rules provide incentives for accuracy. The quadratic, spherical, and logarithmic scoring rules are examples of proper scoring rules.

<sup>3</sup> Allen’s (1987) idea is to treat the scores as probabilities of winning a prize in a binary lottery. This technique was first investigated by Roth and Malouf (1979).

without binary lottery payoffs perform equally well in an environment involving assets that provide uncertain returns. Also recently, Schlag and van der Weele (2009) extended Allen's (1987) mechanism in order to elicit parameters characterizing subjective probability distributions.

Our study builds on Karni (2009), which proposes two belief elicitation procedures that employ a binary lottery payoff technique and are thus incentive compatible irrespective of risk attitudes. We denote these mechanisms as “declarative” and “clock” because they can be written, respectively, as sealed-bid second-price or English clock auctions with random bonus payments (assuming risk-neutrality). Although Karni (2009) suggests the two mechanisms are equivalent<sup>4</sup>, 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; Rutström, 1998; Harstad, 2000). In light of this, it is of substantial practical importance to investigate whether and how elicitation from declarative or clock mechanisms might differ<sup>5</sup>.

Each of the two mechanisms provides incentives for an individual to report her true belief regarding the likelihood that an event will occur (for details see section 2). In the declarative mechanism, the individual reports her belief directly, while in the clock mechanism she participates in an ascending clock, and her stopping point is her belief. In the latter procedure the only other participant is a computerized player who exits stochastically. The auction ends when at least one bidder drops out or when the clock reaches its terminal value, whichever occurs first.

Note that the two mechanisms are not isomorphic in theory. The reason is that the clock mechanism fails to elicit an exact probability whenever the computerized bidder is the first to exit, because the respondent did not have an opportunity to respond. It is easy to see that this data censoring is detrimental when the censored response would have revealed the respondent's true belief, and beneficial when the censored response would

---

<sup>4</sup> On page 604, Karni (2009) introduces the clock mechanism by saying, “An equivalent probability-elicitation auction mechanism is as follows...”

<sup>5</sup> Previous studies including Grether (1992), Möbius et al (2007) and Holt and Smith (2009) have implemented the declarative mechanism in laboratory experiments; Köszegi and Rabin (2008) also discussed this mechanism.

not have revealed her true belief. Building on this intuition, we develop below the conditions under which censoring leaves the clock mechanism with an accuracy advantage over the declarative mechanism. Our experiment's results reveal, at least in our environment, that the clock mechanism elicits beliefs significantly more accurately than the declarative mechanism. Moreover, our analysis reveals that the advantage of the clock mechanism is not due solely to the censoring of decisions. Rather, we find respondents' use different strategies between the two mechanisms.

The contributions of this paper are threefold. First, by focusing on novice respondents with heterogeneous beliefs, we provide practical guidance for belief elicitation in contexts including large-scale survey environments. Indeed, in recent years large-scale belief elicitation (presumably from novice respondents) has become a particularly active area (e.g., Manski, 2004; Bellemare et al., 2008). Our investigation sheds light on how to implement incentive-compatible belief elicitation efficiently in environments that include inexperienced respondents<sup>6</sup>.

Second, our results contribute to the auction literature by demonstrating another environment where the (ascending) clock maintains its advantage in inducing truth-revealing dominant strategies. In particular, it is well known that the equivalence of sealed-bid second-price auctions and English clock auctions quickly breaks down in practice, in the sense that bids in English clock auctions are much closer to the truth in both induced (Kagel et al., 1987; Kagel and Levin, 1993 & 2009; Harstad, 2000) and home-grown value settings (Rutström, 1998).

The third contribution of this research is that it studies belief elicitation when participants are endowed with objective beliefs<sup>7</sup>. This allows us to shed light on the performance of the mechanisms absent noise due to variations in subjects' abilities to predict uncertain events. Garthwaite et al. (2005) point to the importance of doing this

---

<sup>6</sup> However, there exist practical difficulties when using incentive compatible mechanisms in the field to elicit one's beliefs regarding, for example, the chance of changing jobs within the next 5 years. As we discuss in the conclusion, future research could address these important issues.

<sup>7</sup> Some subjects may fail to adopt the beliefs with which they are endowed. However, this would occur in both mechanisms, and thus presents no difficulty for our comparative analysis (other than adding noise and leaving it more difficult to discover differences).

when they write, “it is important to distinguish between the quality of an expert’s knowledge and the accuracy with which that knowledge is translated into probabilistic form”, a distinction Winkler and Murphy (1968) refers to as “substantive goodness” and “normative goodness” respectively. Our subjects are perfectly informed and thus possess “substantive goodness”, so that our data imply the clock mechanism is better than the declarative at facilitating probabilistic formulations by novice respondents.

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, Section 6 discusses the results, and Section 7 concludes.

## 2. Review of Karni’s (2009) Theory

This section briefly reviews Karni’s (2009) mechanisms. In Savage’s (1954) framework, an individual holds the belief that an event  $E$  will occur with probability  $\pi(E)$ . If  $E$  occurs, the individual receives the prize  $x$ ; otherwise, she receives  $y$  ( $x > y$ ). We call a mapping between the occurrence (and non-occurrence) of an 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  $L(r)$ . The number  $r$  is randomly selected from a uniform distribution on  $[0, 1]$ . The individual knows the distribution, but she does not know  $r$  when she makes her 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 plays the bet  $\beta$ ; if  $\mu < r$ , she plays the lottery  $L(r)$ .

**Dominant Strategy.** Karni (2009) demonstrates that the unique dominant strategy in this mechanism is to report truthfully:  $\mu = \pi(E)$ . Doing so guarantees that the individual obtains either the bet  $\beta$  or the lottery  $L(r)$ , whichever has the higher probability of winning the prize  $x$ . The individual has no incentive to report a number greater than the truth, because as soon as the random number  $r$  falls between the truth and her

report,  $\pi(E) < r < \mu$ , she receives the bet  $\beta$ , and forgoes the lottery  $L(r)$  that has a higher winning probability. The same logic applies when her report is smaller than the truth.

## 2.2. English Clock Mechanism

In the English clock auction mechanism, the individual competes with a dummy bidder and knows that the dummy bidder exits the auction at (unknown) number  $r$ . The clock starts at 0 and rises continuously as long as both the individual and the 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 exit, she receives the lottery  $L(r)$ ; if the dummy bidder exits first, the individual receives the bet  $\beta$ .

**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 exit exactly at  $\pi(E)$ .

## 2.3. Assumptions

For Karni’s (2009) mechanisms, as well as proper scoring rules, a necessary condition for incentive compatibility is the no-stakes condition (Kadane and Winkler, 1988). The no-stakes condition requires the wealth of an individual, excluding elicitation-related payoffs, to be independent of the occurrence (or nonoccurrence) of the event<sup>8</sup>.

Incentive compatibility in Karni’s (2009) mechanisms requires also that an individual’s preferences exhibit dominance and probabilistic sophistication (Machina and Schmeidler, 1992)<sup>9</sup>. This condition is weaker than the requirement that preferences satisfy expected utility.

---

<sup>8</sup> As noted by Kadane and Winkler (1988), 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.” Previous work including Karni (1999) and Jaffray and Karni (1999) proposed elicitation procedures when the no-stakes condition is violated. However, we are not aware of any evidence informing the extent to which this violation matters empirically.

<sup>9</sup> 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.

### 3. Comparison of the Mechanisms

#### 3.1. Censoring Can Lead to Greater Accuracy in Elicitations

In this section we demonstrate that, under certain conditions, the clock’s censoring can lead to accuracy advantages over the declarative mechanism. We say that a mechanism is more accurate if its fraction of truthful elicitations is greater. Assuming that people use identical decision strategies with the two mechanisms<sup>10</sup>, the general conditions under which the clock displays greater accuracy are that (i) the population includes a sufficient fraction of naïve decision makers and (ii) the clock censors sufficiently few “optimal” decisions. We show below that these specifics can vary depending on the nature of the environment.

To develop this point, suppose there are two types of respondents: i) truth-revealers who report  $\pi(E)$  and ii) naïve agents whose responses are characterized by a distribution with c.d.f.  $F_n(\cdot)$ . Let the proportions of the optimal and naïve types be  $\alpha$  and  $1 - \alpha$  respectively ( $0 < \alpha < 1$ )<sup>11</sup>. Suppose also that the random number  $r$  (the probability of winning the lottery prize) has continuous and strictly increasing c.d.f.  $F_r(\cdot)$ .

By definition, the accuracy of the declarative mechanism equals  $\alpha$ , the fraction of optimal decisions in the population.

The accuracy of the clock mechanism is given by the following expression:

$$\frac{\alpha(1 - F_r(\pi))}{\alpha(1 - F_r(\pi)) + (1 - \alpha)(1 - \int_0^1 F_r(u)dF_n(u))}$$

---

<sup>10</sup> Our analysis suggests people use different strategies, but our development below shows that the clock can have an accuracy advantage even when this is not the case.

<sup>11</sup> If all decisions are optimal (or if all decisions are naïve), the clock mechanism has no advantage from censoring.

The numerator is the fraction of optimal decisions that are not censored by the clock. The denominator is the total fraction of non-censored decisions. Thus, the value of the expression is the fraction of truthful elicitations in the population<sup>12</sup>.

It immediately follows that the clock is more accurate than the declarative mechanism when the following condition holds:

$$\frac{\alpha(1 - F_r(\pi))}{\alpha(1 - F_r(\pi)) + (1 - \alpha)(1 - \int_0^1 F_r(u)dF_n(u))} - \alpha > 0 \quad (3.1)$$

Inequality (3.1) makes clear that whether the clock holds an accuracy advantage in relation to the declarative mechanism depends critically on the fraction of optimal decision makers  $\alpha$  as well as the value of the true belief  $\pi$  in relation to the distribution of the random number  $r$  (which influences the fraction of censored optimal decisions).

### 3.2. Experiment Hypothesis

Our interest is in testing the following hypothesis:

**Hypothesis:** *Beliefs elicited using the clock mechanism are more likely to be accurate than beliefs elicited using the declarative mechanism, especially with novice participants.*

To develop this, suppose now that  $F_r(\cdot)$  and  $F_n(\cdot)$  are both  $U(0,1)$ <sup>13</sup>. Under these assumptions it is easy to show that inequality (3.1) simplifies to

$$\frac{\alpha(1 - \pi)}{\alpha(1 - \pi) + (1 - \alpha)/2} - \alpha > 0 \quad (3.2)$$

---

<sup>12</sup> Note that this approach does not make any use of censored observations. An alternative is to take a censored belief as the interval between the clock's stopping point and the clock's upper limit. However, the transformation from interval estimates to point estimates is arbitrary (e.g., using the mean of the interval as the estimate of the belief), and results would generally be sensitive to this transformation. We do not pursue this approach here.

<sup>13</sup> Karni specified the uniform distribution on  $[0, 1]$  for  $F_r(\cdot)$ , but we note that properties of the mechanisms remain the same for any continuous and strictly increasing distribution  $F_r(\cdot)$ .

Given  $0 < \alpha < 1$ , (3.2) holds if and only if

$$0 < \pi < 0.5 \quad (3.3)$$

Thus, the clock has an advantage in this environment when true beliefs are less than 0.5, and not everybody makes optimal decisions. This result guides our experiment design, as we detail below.

#### 4. Experiment Design and Procedures

A key feature of our design is that we endow subjects with objective beliefs. We made this choice for two reasons. First, as noted earlier, our goal is to assess the mechanisms' ability to facilitate accurate translation from participants' knowledge to probabilities, so we eliminate differences in quality of their knowledge. In addition, doing this better connects our research to second-price and English clock auctions, where participants typically make decisions using known (induced) values. A transcript of subjects' instructions can be found in the appendix.

##### 4.1. Declarative Mechanism

**Endowed Belief = 0.2.** The subject is presented with two opaque bags (physical; made of cloth), 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 any integer from 1 to 9.

The participant submits a number between 1 and 9 (inclusive; integer only<sup>14</sup>) on a computer terminal<sup>15</sup>. This number is then compared with  $R$ , i.e., the number of white chips in Bag B. If the submitted number is greater than  $R$ , the subject draws a chip from

---

<sup>14</sup> Since both lotteries are presented using integers only, we chose to constrain the decisions also to integers for simplicity and transparency. Moreover, words such as “probabilities” or “distributions” were not used during our experiment.

<sup>15</sup> To maintain symmetry with the clock procedure, all decisions were submitted via computers.

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<sup>16</sup>.

**Endowed Belief = 0.3.** This proceeds exactly 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 the 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<sup>17</sup>. The presence of two equally advantageous actions stems from our using a discrete state space.

#### 4.2. Clock Mechanism

**Endowed Belief = 0.2.** Bag A and bag B are exactly the same as in the declarative mechanism with the endowed belief of 0.2. Instead of declaring a number, the subject participates in a computerized clock auction. Similar to an English clock auction (e.g., Kagel et al., 1987), it starts as the computer screen displays the number 1 for five seconds, and then the number 2 for five seconds, and so on. The subject exits the auction by pressing the space key. The clock stops when the subject exits, or after reaching and displaying the number  $R$  for 5 seconds, whichever comes first<sup>18</sup>. If the clock stops due to reaching number  $R$ , the subject draws a chip from bag A; if the clock stops due to the subject’s exit, she draws a chip from bag B<sup>19</sup>. 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 exactly as the above procedure, except now there are 3 white chips (and 7 black chips) in bag A.

---

<sup>16</sup> The subject physically drew a chip from the appropriate cloth bag. See section 4.4 for details.

<sup>17</sup> The two dominant strategies are equivalent because the individual ends up with the same bag except when  $R$  is the same as the number of white chips in bag A, as the former strategy leads to bag A while the latter leads to bag B. However, in this case the two bags have the same number of white chips, so the chance of winning \$10 is identical.

<sup>18</sup> The screen displays number  $R$  for 5 seconds, and the subject were aware that they were able to drop out at  $R$  and obtain bag A.

<sup>19</sup> The dummy bidder in our experiment is implemented by stopping the clock when it reaches number  $R$ .

**Dominant Strategy.** Considering bag A as the default choice, the clock mechanism is effectively asking the subject, “*The number displayed on the screen is the minimum 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 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 (see footnote 17).

#### 4.3. Treatment Design

Each subject participated in two independent elicitation tasks, which occurred in round one and round two, respectively. The second round was a “surprise” as we announced it only after completing the first round<sup>20</sup>. 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 consisted of 4 to 8 participants and a heterogeneous belief environment: half were endowed with beliefs equal to 0.2 and the other half with beliefs equal to 0.3. Subjects were given new instructions in the second round.

#### 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,

<sup>20</sup> Subjects were not told there was a second round at the beginning of the experiment. Upon finishing the first round, 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.”

VA. Subjects were invited via emails sent to a large undergraduate subject pool, and a total of 130 participated. Subjects were paid a guaranteed \$5 plus their earnings in the experiment. Average total earnings were \$16, and sessions lasted between 30 and 60 minutes. The experiments were partially computerized: we used physical bags and chips to illustrate the lotteries and perform the random draws, whereas the ticking clock was programmed using E-prime<sup>21</sup>.

To facilitate understanding of the mechanisms, subjects were first given abundant time to read the instructions. Following this, the experimenter read the instructions aloud to them. Each subject then took a quiz<sup>22</sup>, and their answers were recorded. The majority of the subjects correctly answered all questions. The experimenter then announced and explained the correct answers. Throughout of the experiment, we did not use words such as “probability” or “percent chance,” and subjects made all decisions in whole numbers<sup>23</sup>.

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 nine 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 subject could read the numbers inside), transcribed the random number  $R$  for each subject ID, and then sealed the envelope. The reason for these steps was to demonstrate 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, using exactly the same procedures as described above.

---

<sup>21</sup> E-Prime is a commercial software for computerized experiment design commonly used in psychology research: <http://www.pstnet.com/eprime.cfm>

<sup>22</sup> The quiz was designed to test whether subjects understood how various hypothetical scenarios and decisions are translated into payoffs. The quiz is available upon request.

<sup>23</sup> Hoffrage et al. (2000) showed that natural frequencies are much better than percent chances at facilitating statistical reasoning of people including experts and non-experts.

Finally, we implemented the payment procedure individually to ensure that subjects knew they were making independent decisions. After the bag of payoff was determined, the experimenter went to a 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 his or her head turned away. The subject earned \$10 if he or she drew a white chip and \$1 otherwise.

## 5. Results

We organize our results in two subsections. The first describes decisions from the first round, and our first result is that novice subjects are more likely to report their endowed beliefs in the clock mechanism. Our second result shows that the subjects use different strategies between the two mechanisms, and that the distribution of clock data more accurately characterizes the distribution of endowed beliefs.

The second subsection presents decisions from the second round, and our results show that the clock and declarative mechanisms are equally likely to elicit endowed beliefs; however, the distribution of clock data continues to reflect more closely the distribution of endowed beliefs.

### 5.1. Responses from Novice Participants

With heterogeneous beliefs of 0.2 and 0.3, Table 2 describes individual decisions in the first round<sup>24</sup>.

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%. *In the declarative mechanism, fewer than half of novice responses are optimal.* This suggests that the dominant strategies in this environment are

---

<sup>24</sup> The two equally dominant strategies 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}.

not trivial to subjects<sup>25</sup>. This is especially significant in light of our explicit effort for simplicity and transparency<sup>26</sup>.

Table 2: Novice Responses: Descriptive Statistics

	Declarative	Clock
Observations	53	77
Optimal decisions	25/47%	30/39%
Non-optimal decisions	28/53%	17/22%
Censored decisions	--	30/39%
Mean of deviation from truth	0.0604** (0.0264)	0.0362* (0.0203)
Mean of absolute deviation from truth	0.1208*** (0.0221)	0.0787*** (0.0175)

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

Overall, the clock mechanism censors 39% of decisions. In comparison, if the population consisted of only optimal decisions, the proportion would be between 25% and 35%. This suggests that not everybody in our experiment makes optimal decisions.

The mean deviations, as well as the mean absolute deviations from optimal strategies, are significantly different from zero in both mechanisms<sup>27</sup>. They are smaller in the clock mechanism than in the declarative mechanism, but the difference is insignificant.

We now take a closer look at the two mechanisms by excluding censored decisions. Note that we do not consider alternative approaches that make use of information in censored decisions (see footnote 12). Our first result is as follows:

**Result 1.** *With novice respondents, beliefs elicited using the clock mechanism are more likely to be accurate than beliefs elicited using the declarative mechanism.*

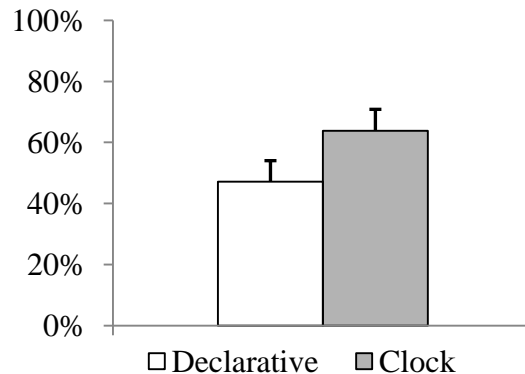
<sup>25</sup> Post-experiment surveys indicate many subjects did not recognize the existence of optimal strategies.

<sup>26</sup> Note the proportion of optimal decisions in our declarative mechanism is consistent with proportions of optimal decisions in the second-price auction in Cooper and Fang (2008, p. 1583).

<sup>27</sup> Censored decisions are excluded from these statistics (see footnote 12).

*Evidence.* Among elicited beliefs, 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 found a statistical difference at  $p=0.096$  (binary data: 1 if a decision is optimal, 0 otherwise)<sup>28</sup>.

**Figure 1: Proportion of Optimal Decisions in First Round**



Note: Error bar is one s.e. of the mean.

However, is the accuracy of the clock mechanism driven by data censoring? If respondents use identical pre-determined strategies in the two mechanisms, then beliefs elicited using the clock mechanism should be identical to beliefs elicited using the declarative mechanism after applying a filter that is equivalent to clock censoring. That is, we should obtain the same data if they are “naturally” censored during the experiment in the clock mechanism or “artificially” filtered after the experiment in 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 strictly greater than the random number. The sample of 53 independent decisions from the declarative mechanism is filtered 10,000 times, and the frequency of each decision remaining is calculated. Doing this yields our second result,

---

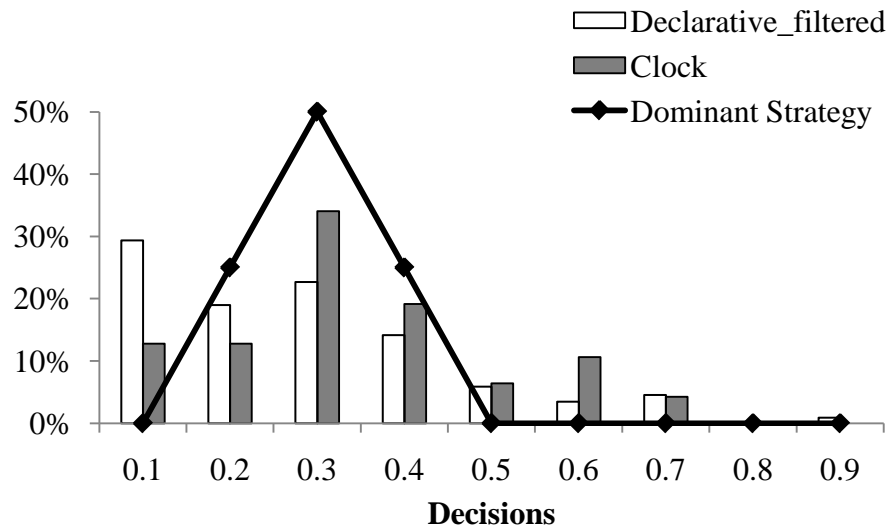
<sup>28</sup> The one-sided test is significant at 5% level; we have a clear ordered hypothesis as illustrated in section 3.

**Result 2.** *With novice respondents, filtered declarative data differ significantly from clock data.*

*Evidence.* As shown in Figure 2, the distribution of filtered declarative data is significantly different from the distribution of clock data ( $p=.014$ , Chi-squared test)<sup>29</sup>.

In addition, Figure 2 plots the distribution that would emerge under only dominant strategies<sup>30</sup>. It has a single mode at 0.3. In comparison, beliefs elicited using the clock mechanism also have a single mode at 0.3, and seem to characterize the dominant strategy distribution reasonably well. In contrast, the mode of filtered declarative data is 0.1, far less accurate regarding the underlying beliefs.

**Figure 2: Distributions of First-Round Decisions**



Note: Half holds beliefs equal to 0.2 and the other half holds beliefs equal to 0.3.

## 5.2. Responses from One-Time Experienced Participants

<sup>29</sup> The distribution of unfiltered declarative data is marginally significantly different from the distribution of the clock data ( $p=0.063$ , Chi-squared test).

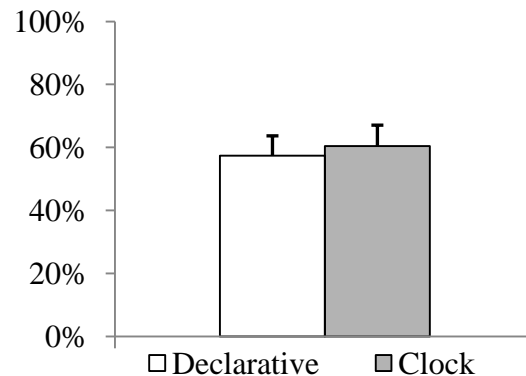
<sup>30</sup> We assume the two dominant strategies are equally likely to be chosen. Hence, the ratios of subjects who choose decisions 0.2, 0.3 and 0.4 are 1:2:1.

We next report an analysis of decisions made in the second round. Our third result is as follows.

**Result 3.** *With one-time experienced participants, the declarative and clock mechanisms are equally accurate.*

*Evidence.* Figure 3 shows that proportions of optimal decisions in the declarative and clock mechanisms are 57% and 60% respectively, and these values are not significantly different ( $p=.84$ , two-sided Wilcoxon-Mann-Whitney)<sup>31</sup>.

**Figure 3: Proportion of Optimal Decisions in Second Round**



Note: Error bar is one s.e. of the mean.

Similarly, we apply clock-equivalent filtering to second-round decisions from the declarative mechanism and compare them with the second-round clock data:

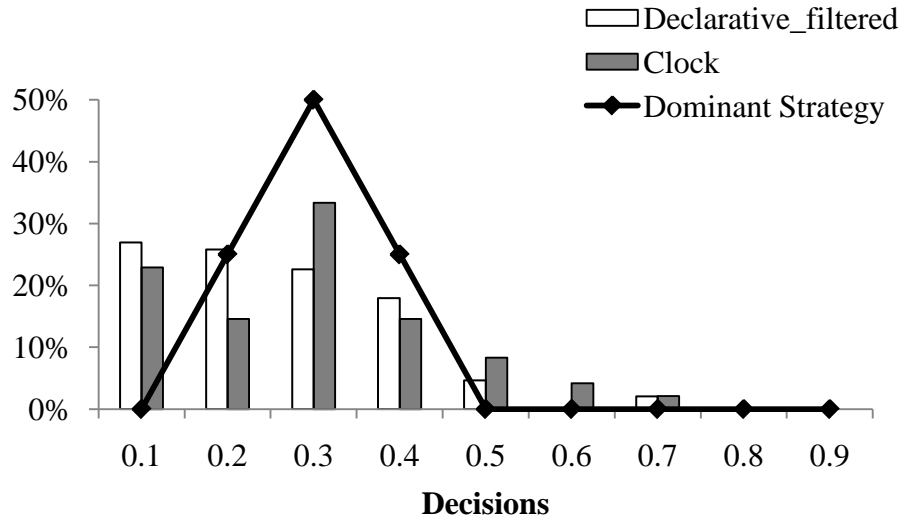
**Result 4.** *With one-time experienced participants, filtered declarative data differ significantly from clock data.*

---

<sup>31</sup> 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).

*Evidence.* As shown in Figure 4, the distribution of filtered declarative data is significantly different from the distribution of clock data ( $p=.052$ , Chi-squared test)<sup>32</sup>.

**Figure 4: Distributions of Second-Round Decisions**



Note: Half holds beliefs equal to 0.2 and the other half holds beliefs equal to 0.3.

Combining results 3 and 4, we observe that the proportions of truthful elicitation are not distinguishable in the second round, but the distribution of clock data continues to characterize underlying beliefs more accurately than does the distribution of filtered declarative data.

## 6. Discussion

The results of this study have implications for the practical design of incentive-compatible belief elicitation mechanisms. This can be important, for example, to the rapidly emerging literature reporting experiments from large-scale surveys (see, e.g., Andersen et al., 2008; Dohmen et al., 2009). In such studies it is often of interest to elicit participants' beliefs related, for instance, to the chance a responder might accept an offer in an ultimatum game (Bellemare et al., 2008). Incorporating incentive-compatible

<sup>32</sup> Although the distribution of unfiltered declarative data is not different from the distribution of the clock data ( $p=.392$ , Chi-squared test).

elicitation mechanisms into this literature would seem worthwhile to explore, and our results suggest that a first step might be to consider the clock, as it has substantial advantages when dealing with novice participants.

We found the accuracy of the declarative mechanism to improve significantly with participant experience, while the accuracy of the clock is high regardless of the level of participants' experience. This highlights that one can be confident in inferences from the clock mechanism using novice participants, while confidence with the declarative mechanism seems possible only with experienced participants.

Despite similarity in accuracy, we were surprised to find that differences in decision strategies continued to be apparent between the two mechanisms even with experienced participants. This finding resonates with the benefits of the clock reported in the auction literature (e.g., Kagel et al., 1987; Kagel and Levin, 1993; Harstad, 2000), raising the important fundamental questions of why and how a clock presentation improves decision making. Distilling the source of the clock's advantage might allow one to implement procedures to improve decision making in a wide variety of environments, even those that do not easily admit a clock representation of decision alternatives.

## **7. Conclusion**

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 their incentive compatibility does not require strong assumptions such as risk neutrality or expected utility maximization. We found that, in relation to the declarative mechanism, with the clock mechanism (i) elicited beliefs are more likely to be accurate and (ii) 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 induces truth-telling in a belief elicitation context. This result has implications for the practical design of incentive-compatible belief elicitation mechanisms.

A limitation of our study stems from our choice of parameters. We induced 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 (e.g., when true beliefs are above 0.5 in our experiment), 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, one can be more confident that actual beliefs are near the clock's starting point and thus minimize data loss.

Future research might investigate whether the truth-inducing advantage of a clock procedure persists in environments with subjective beliefs, or where incentive-compatible mechanisms are difficult to implement. This includes cases where outcomes are impractical or impossible to verify. Of particular interest here are large-scale surveys of respondents' beliefs regarding life style choices and consequences related to, for example, changes in health or job status. In addition, recent work by Charness et al. (2007, 2010) suggests that the frequency of mistakes declines when subjects can consult with others. A profitable next step would be to extend our analyses in the direction of group decision making. Investigating such questions is an additional step towards a better understanding of how belief elicitation mechanisms can be implemented efficiently.

### **Acknowledgements:**

We thank Glenn Harrison, Edi Karni, Eric Danan, Jacob Sagi, Ron Harstad, James Andreoni, Yan Chen, Soo Hong Chew, Nat Wilcox, colleagues at ICES, and seminar participants at ESA international 2009 (Washington DC), Advanced Workshop in Experimental Economics 2009 (Sydney, Australia) and SEA 2009 (San Antonio, TX) for helpful discussions.

## 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)
- Andersen, Steffen, Glenn W. Harrison, Morten I. Lau and E. Elisabet Rutström (2008): "Eliciting Risk and Time Preferences," *Econometrica*, 76 (3), 583-618
- Andreoni, James, Yeon-Koo Che and Jinwoo Kim (2007): "Asymmetric information about rivals' types in standard auctions: An experiment," *Games and Economic Behavior*, 59, 240-259
- 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), 1079-1089
- Becker, Gordon M., Morris H. Degroot, and Jacob Marschak (1964): "Measuring Utility by a Single-Response Sequential Method," *Behavioral Science*, IX, 226-232
- Bellemare, Charles, Sabine Kroger, and Arthur Van Soest (2008): "Measuring Inequality Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities," *Econometrica*, 76(4), 815-839
- 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.
- Charness, G., Karni, E., and Levin, D. (2007): "Individual and Group Decision Making Under Risk: An Experimental study of Bayesian Updating and Violations of First-order Stochastic Dominance." *Journal of Risk and Uncertainty*, 35, 129-148
- Charness, G., Karni, E., and Levin, D. (2010): "On the Conjunction Fallacy in Probability Judgment: New Experimental Evidence Regarding Linda," *Games and Economic Behavior*, 68, 551-556
- 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, 1572-1595

- DeMarzo, Peter M., Kremer, Ilan and Skrzypacz, Andrzej (2005): “Bidding with Securities: Auctions and Security Design,” *American Economic Review*, 95 (4), 936-959
- Delavande, Adeline and Susann Rohwedder (2008): “Eliciting Subjective Probabilities in Internet Surveys,” *Public Opinion Quarterly*, 72(5), 866-91
- Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp and Gert G. Wagner (2009): “Individual Risk Attitudes: Measurement, determinants and Behavioral Consequences,” *Journal of European Economic Association*, forthcoming
- 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
- Garthwaite, Paul H, Joseph B. Kadane and Anthony O’Hagan (2005): “Statistical Methods for Eliciting Probability distributions,” *Journal of the American Statistical Association*, 100 (47), 680-700
- Gilboa, Itzhak, Andrew W. Postlewaite, and David Schmeidler (2008): “Probability and Uncertainty in Economic Modeling,” *Journal of Economic Perspective*, 22 (3), 173-188
- Grether, David M.(1992): “Testing Bayes Rule and the Representative Heuristic: Experimental Evidence,” *Journal of Economic Behavior and Organization*, 17, 31-57
- Harstad, Ron (2000): “Dominant strategy adoption and bidders' experience with pricing rules,” *Experimental Economics*, 3(3), 261-280
- Hoffrage, U., S. Lindsey, R. Hertwig, and G. Gigerenzer (2000): “Communicating Statistical Information,” *Science*, 290, 2261–2262.
- Holt, Charles, A and Angela M. Smith (2009): “An Updating on Bayesian Updating,” *Journal of Economic Behavior and Organization*, 69, 125-134
- 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), 357-363
- 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 (2008): "Auctions: A Survey of Experimental Research, 1995 – 2008," to appear in the *Handbook of Experimental Economics, Vol. 2*.
- 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
- Kagel, John (1995): "Auctions: A Survey of Experimental Research," in (J. Kagel and A.E. Roth, eds.), *Handbook of Experimental Economics*: Princeton University Press, 501-586
- Kahneman, Daniel and Amos Tversky (1979): "An Analysis of Decision under Risk," *Econometrica*, 47(2), 263-291
- Karni, Edi (1999): "Elicitation of Subjective Probabilities when Preferences are State-Dependent," *International Economic Review*, 40(2), 479-486
- Jaffray, Jean-Yves and Edi Karni (1999): "Elicitation of Subjective Probabilities when the Initial Endowment is Unobservable," *Journal of Risk and Uncertainty*, 8, 5-20
- Karni, Edi (2009): "A Mechanism for Eliciting Probabilities," *Econometrica*, 77(2), 603-606
- Karni, Edi (2008): "A Mechanism Design for Probability Elicitation," Unpublished Manuscript, Johns Hopkins University
- Köszegi, Botond and Matthew Rabin (2008): "Revealed Mistakes and Revealed Preferences," in A. Caplin and A. Schotter (eds.), *The Foundations of Positive and Normative Economics: A Handbook* (New York: Oxford University Press, 2008)
- Machina, Mark J. (2004): "Almost-objective uncertainty," *Economic Theory*, 24, 1-54
- Machina, Mark J. and David Schmeidler (1992): "A more robust definition of subjective probability," *Econometrica*, 60, 745-780
- Machina, Mark J. (2009): "Risk, Ambiguity, and the Rank-Dependence Axioms," *American Economic Review*, 99(1), 385-392
- McKelvey, Richard D. and Talbot Page (1990): "Public and Private Information: An Experimental Study of Information Pooling," *Econometrica*, 58(6), 1321-1339
- Möbius, Markus, Muriel Niederle, Paul Niehaus and Tanya Rosenblat (2007): "Gender Differences in Incorporating Performance Feedback," working paper: [http://trosenblat.surveyor.nber.org/papers/Files/WP/gender\\_v4.pdf](http://trosenblat.surveyor.nber.org/papers/Files/WP/gender_v4.pdf)
- 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), 1461-1489
- 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 Bargaining,” *Psychological Review*, 86, 574-594
- Roth, Alvin E. and J. Keith Murningham (1982): “The role of Information in Bargaining: An Experimental Study”, *Econometrica*, 50(5), 1123-1142
- Roth, Alvin E. and Francoise Schoumaker (1983): “Expectations and Reputations in Bargaining: An Experimental Study,” *American Economic Review*, 73(3), 362-372
- Rutström, E. Elisabet (1998): “Home-grown Values and Incentive Compatible Auction Design,” *International Journal of Game Theory*, 27, 427-441
- Rutström, E. Elisabet and Nathaniel T. Wilcox (2009): “Stated Beliefs versus Inferred Beliefs: A Methodological Inquiry and Experimental Test,” *Games and Economic Behavior*, 67, 616-632
- Savage Leonard J. (1971): “Elicitation of Personal Probabilities and Expectations,” *Journal of American Statistical Association*, 66(336), 783-801
- Schlag, Karl H. and Joël van der Weele (2009): “Eliciting Probabilities, Means, Medians, Variances and Covariances without Assuming Risk-Neutrality,” working paper, Universitat Pompeu Fabra, Barcelona
- 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, 211-249
- Smith, Vernon (1994): “Economics in the Laboratory”, *Journal of Economic Perspectives*, 8 (1), 113-131
- Winkler, Robert L. and Allan H. Murphy (1968): “ ‘Good’ Probability Assessors,” *Journal of Applied Meteorology*, 7, 751-758

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.

**Overview:**

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.

**Details:**

**Bag A** has 2 white chips and 8 black chips for a total of 10.

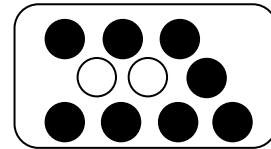
**Bag B** also has 10 chips, some white, some black, but

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.

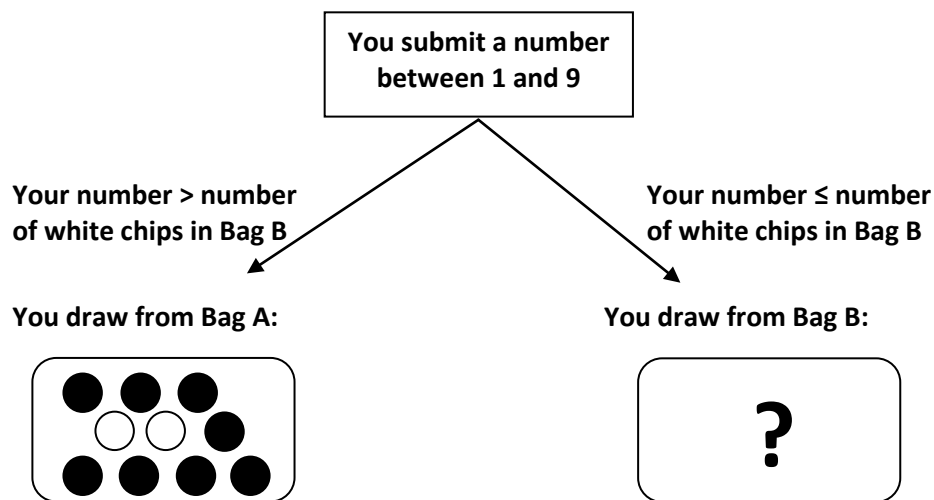
**Bag A:**



**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**.



**Your payment:** 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:**

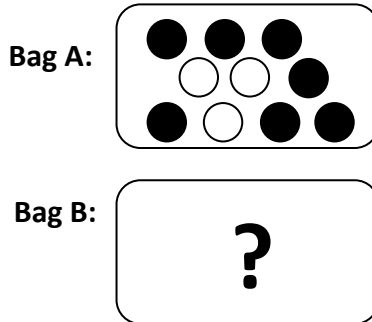
Welcome to this experiment! In addition to the \$5 for showing up on time, you will be paid in cash based on your decisions in the experiment. Please note that no other participant's decisions in this experiment will affect your earnings, and vice versa. Please read these instructions carefully. Raise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

**Overview:**

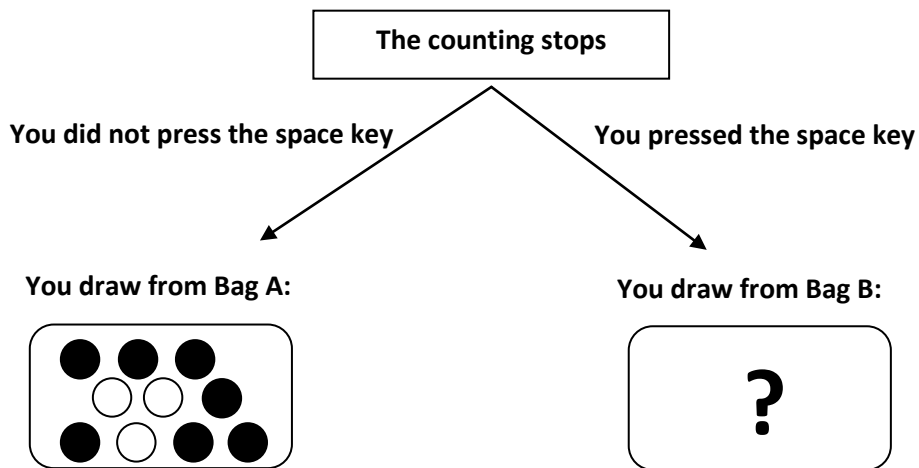
The procedure is simple. You will first participate in an exercise, and then you will draw a chip from one of two bags. If you draw a white chip you will earn \$10, if it is black you will earn \$1.

**Details:**

**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.



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**.



**Your payment:** If you draw a white chip you earn \$10; a black chip earns you \$1.