Estimating Subjective Probabilities

by

Steffen Andersen, John Fountain, Glenn W. Harrison & E. Elisabet Rutström[†]

June 2010

ABSTRACT. Subjective probabilities play a role in many economic decisions. There is a large theoretical literature on the elicitation of subjective probabilities, and an equally large empirical literature. However, there is a gulf between the two. The theoretical literature proposes a range of procedures that can be used to recover subjective probabilities, but stresses the need to make strong auxiliary assumptions or "calibrating adjustments" to elicited reports in order to recover the latent probability. With some notable exceptions, the empirical literature seems intent on either making those strong assumptions or ignoring the need for calibration. We illustrate how the *joint estimation of risk attitudes and subjective probabilities* using structural maximum likelihood methods can provide the calibration adjustments that theory calls for. This allows the observer to make inferences about the latent subjective probability, calibrating for virtually any well-specified model of choice under uncertainty. We demonstrate our procedures with experiments in which we elicit subjective probabilities. We calibrate the estimates of subjective beliefs assuming that choices are made consistently with expected utility theory or rank-dependent utility theory. Inferred subjective probabilities are significantly different when calibrated according to either theory, thus showing the importance of undertaking such exercises. Our findings also have implications for the interpretation of probabilities inferred from prediction markets.

[†] Department of Economics, Copenhagen Business School, Copenhagen, Denmark (Andersen); Department of Economics, University of Canterbury, Christchurch, New Zealand (Fountain); Department of Risk Management & Insurance and CEAR, Robinson College of Business, Georgia State University, USA (Harrison); and Robinson College of Business, Georgia State University, USA (Rutström). E-mail contacts: sa.eco@cbs.dk, john.fountain@canterbury.ac.nz, gharrison@gsu.edu and erutstrom@gmail.com. We thank the U.S. National Science Foundation for research support under grants NSF/HSD 0527675 and NSF/SES 0616746. We are grateful to John Duffy, Peter Wakker, Randall Walsh and Nathaniel Wilcox for discussions.

Subjective probabilities about some event are operationally defined as those probabilities that lead an agent to make certain choices over outcomes that depend on that event. These choices could be as natural as placing a bet on a horse race, or as structured as responding to the payoffs provided by some scoring rule. In order to infer subjective probabilities from observed choices of this kind, however, one either has to make some strong assumptions about risk attitudes or jointly estimate risk attitudes and subjective probabilities. We show how the latter can be implemented by pairing several tasks together, some of which identify risk attitudes and some of which identify the interplay between risk attitudes and subjective probabilities. Joint estimation of a structural model of choice across these two types of tasks allows one to then make inferences about subjective probabilities.

If inferred subjective probabilities are conditioned on knowing risk attitudes, then any statistical uncertainty in the estimation of risk attitudes would be expected to "propagate" into some additional uncertainty about subjective probabilities. Joint estimation allows these effects to occur, providing more reliable estimates of subjective probabilities, even if those estimates have large standard errors. In other words, it is possible that the choice task for eliciting subjective probabilities generates a point response that appears to be quite precise by itself, but which is actually not a very precise estimate of the latent subjective probability when one properly accounts for uncertainty over risk attitudes.

An important example is the response to a proper scoring rule, such as the quadratic scoring rule (QSR). A respondent might select 67% when faced with a QSR, and that would be the exact subjective probability if the subject were known to be *exactly* risk neutral. But if the subject was estimated to have risk attitudes in some interval, there would be a corresponding *range of subjective probabilities* to be inferred from that 67% *response* to the scoring rule task. If the estimate of the subject's risk attitudes spanned the possibility of risk neutrality, then the inferred subjective probability would include 67%, but would also include subjective probabilities on either side. If the

estimate of the subject's risk attitudes revealed him to be clearly risk averse, with no statistically significant chance of being risk neutral, then the risk-adjusted subjective beliefs would actually be strictly higher than 67%. The intuition is clear: a response of 50% in the standard QSR removes all uncertainty over payoffs, and is a magnet for risk averse subjects. Sufficiently risk averse subjects would report close to 50% even if their latent subjective probability was 67%. So if we observe the subjects responding at 67%, and we know that they are risk averse to some degree, then we can infer that they must have actually held a latent subjective probability greater than 67% and responded to the pull of the magic 50% magnet.

Formalizing this intuition requires that one obtain estimates of risk attitudes from a task with *objective* probabilities as well as from a task whose outcomes depend on some *subjective* probability, and *then* untangle the effects of risk attitudes and subjective probability with a structural model of choice. We do so for the QSR, and also for a linear scoring rule (LSR).

In section 1 we briefly state the theory underlying our approach and relate this to the literature on belief elicitation. The properties of the QSR and LSR, and the fact that responses to these are affected by risk attitudes, are well known. We assume throughout that the agent is acting in what is called a "probabilistically sophisticated" manner, although we do not restrict the characterization of risk attitudes to expected utility theory (EUT). We also consider the inference of subjective probabilities for subjects who are assumed to make decisions according to the rank-dependent utility (RDU) model. This extension is particularly appropriate in the case of eliciting subjective probabilities, because it involves allowing for probability weighting and non-additive decision weights on the utility of final outcomes. Given that one of the probabilities to be weighted is the subjective probability being estimated, one might expect estimates of the subjective probability to be even more sensitive to the correct specification of the model of risk attitudes employed.

In section 2 we describe the experimental task we posed to 140 subjects, split roughly equally

across the QSR and LSR alternatives. Our subjects made choices over a number of standard lotteries, characterized by objective uncertainty over monetary outcomes between $0 and $100. They also gave responses to either a QSR or LSR choice task. The prizes on each of these scoring rule tasks also spanned $0 and $100, so that we were able to infer risk attitudes over the same prize domain as the scoring rule responses.

Section 3 formally sets out the econometric model used for estimating subjective probabilities, spelling out the manner in which we undertake joint estimation over all tasks in order to identify subjective probabilities. Section 4 then presents our estimates of the inferred subjective probabilities from these scoring rules, after adjusting for risk. Our primary result is that subjective probabilities inferred under the assumption of risk neutrality are very different to the subjective probabilities inferred when one allows the data to say how risk averse the subjects were. This finding has immediate implications for the practical use of scoring rules,[1] which is the focus here; it also has implications for inferences that can be drawn from prediction markets since experimental findings consistently shows that people are risk averse on average. We offer a methodology for jointly estimating risk attitudes and subjective probabilities. Section 5 draws conclusions.

1. Scoring Rules

The notion that subjective probabilities can be usefully viewed as prices at which one might trade has been a common one in statistics, and is associated with de Finetti [1937][1970] and Savage [1971]. It is also clear, of course, in the vast literature on gambling, particularly on the setting of odds

---

[1] There is a large, practical literature on the "normative" elicitation of subjective probabilities, reviewed by O'Hagen et al. [2006] and illustrated well by Shephard and Kirkwood [1994]. One characteristic of those elicitation procedures is that they involve considerable real-time, one-on-one feedback between the elicitor and the elicitee, often including results from proper scoring rules. Our approach is to better model inferences from more static, impersonal applications of those scoring rules, as a prelude to the design and evaluation of normative elicitation procedures with incentives.

by bookies and parimutuel markets (Epstein [1977; p. 298ff.]). The central insight is that subjective

probabilities of events are marginal rate of substitution between contingent claims, where the

contingencies are events that the probabilities refer to. There are then a myriad of ways in which one

can operationalize this notion of a marginal rate of substitution.[2]

Scoring rules are procedures that convert a "report" by an individual into a lottery defined

over the outcome of some event. The formal link between scoring rules and optimizing decisions

by agents is also familiar, particularly in Savage [1971], Kadane and Winkler [1987][1988], Holt

[1986] and Hanson [2003]. Jose, Nau, and Winkler [2008] explore the relationship between expected

scores, expected utility, and generalized information/entropy measures for several popular scoring

rules and the HARA class of utility functions. We start with an arbitrary utility function and belief

betting game, which can be viewed as derived from a particular scoring rule, and draw statistical

inferences about subjective beliefs.

Alternatives to scoring rules include procedures like those developed for utility elicitation by

Becker, DeGroot and Marschak [1964] and adapted to eliciting probabilities. Karni [2009], Grether

[1992], Köszegi and Rabin [2008], and Holt and Smith [2009] are examples of this alternative. Let

there be two prizes, x>y. The subject reports a probability ξ, and a random number ζ is selected

from the unit interval. If $\zeta \leq \xi$ the subject gets the lottery that pays off x if the event occurs, and y

otherwise; if $\zeta > \xi$ the subject gets the lottery that pays off x with probability ζ and y with

probability 1-ζ. If the agent is not satiated, and is probabilistically sophisticated, it is a dominant

strategy to report the true subjective probability as ξ. The main problem for all of these classes of

procedures (both scoring rules, lottery procedures) is the poor incentive properties in practice

---

[2] For example, one could elicit the p that makes the subject indifferent between a lottery paying M
with probability p and m with probability (1-p), for M>m, and a lottery paying M if the event occurs and m if
it does not (Marschak [1964; p. 107ff.]). This method formally requires that one elicit indifference, which
raises procedural issues.

(Harrison [1992]). Thus, precision in the elicited probabilities require substantial payoffs.

For simplicity we assume throughout that the events in question only have two outcomes. A scoring rule asks the subject to make some report θ, and then defines how an elicitor pays a subject depending on their report and the outcome of the event. This framework for eliciting subjective probabilities can be formally viewed from the perspective of a trading game between two agents: you give me a report, and I agree to pay you \$X if one outcome occurs and \$Y if the other outcome occurs. The scoring rule defines the terms of the exchange quantitatively, explaining how the elicitor converts the report from the subject into a lottery. We use the terminology "report" because we want to view this formally as a mechanism, and do not want to presume that the report is in fact the subjective probability π of the subject. In general, it is not.

For example, consider the log scoring rule proposed by Good [1952; p.112] and extended by Bernardo [1979]. In its simplest version, this scoring rule assigns a number, $\log(\theta)$, to the report θ for some event, constraining θ to lie in the unit interval. Hence we can refer to this as a reported probability. Since $0 \le \theta \le 1$, the simple log score is a penalty score, a negative number. Also, since the log function is increasing, higher reported probabilities θ for an event receive higher scores (lower penalties), with a maximum score (minimum penalty) of zero when the reported probability for the event that occurs is 1, i.e, $\theta=1$.

The popular QSR is defined in terms of two positive parameters, α and β that determine a fixed reward the subject gets and a penalty for error. Assume that the possible outcomes are A or B, where B is the complement of A, that θ is the reported probability for A, and that Θ is the true binary-valued outcome for A. Hence $\Theta=1$ if A occurs, and $\Theta=0$ if it does not occur (and thus B occurs instead). The subject is paid $S(\theta \,|\, A \text{ occurs}) = \alpha - \beta(\Theta-\theta)^2 = \alpha - \beta(1-\theta)^2$ if event A occurs and $S(\theta \,|\, B \text{ occurs}) = \alpha - \beta(\Theta-\theta)^2 = \alpha - \beta(0-\theta)^2$ if B occurs. In effect, the score or payment penalizes the subject by the squared deviation of the report from the true binary-valued outcome, Θ, which is 1

and 0 respectively for A and B occurring. An omniscient seer would obviously set $\theta = \Theta$. The fixed reward is a convenience to ensure that subjects are willing to play this trading game, and the penalty function simply accentuates the penalty from not being an omniscient seer. In our experiments $\alpha = \beta = \$100$, so subjects could earn up to $100 or as little as $0. If they reported 1 they earned $100 if event A occurred or $0 if event B occurred; if they reported ¾ they earned $93.75 or $43.75; and if they reported ½ they earned $75 no matter what event occurred.

It is intuitively obvious, and also well known in the literature (e.g., Winkler and Murphy [1970] and Kadane and Winkler [1988]), that risk attitudes will affect the incentive to report one's subjective probability "truthfully" in the QSR.[3] A sufficiently risk averse agent is clearly going to be drawn to a report of ½, and varying degrees of risk aversion will cause varying distortions in reports from subjective probabilities. If we knew the form of the (well-behaved) utility function of the subjects, and their degree of risk aversion, we could infer back from any report what subjective probability they must have had. Indeed, this is exactly what we do below, recognizing that we only ever have estimates of their degree of risk aversion.

The LSR is also defined in terms of two positive parameters, γ and λ, that serve as fixed rewards and penalties, respectively. The subject is paid a fixed reward less some multiple of the absolute difference between their report and what actually happened, which is also what an omniscient seer would have reported. Thus the payment is $S(\theta | A \text{ occurs}) = \gamma - \lambda(1-\theta)$ if event A occurs and $S(\theta | B \text{ occurs}) = \gamma - \lambda(\theta-0)$ if B occurs. We again set $\gamma = \lambda = \$100$, generating payoffs of $100 or $0 for a report of 1; $75 and $25 for a report of ¾; and $50 no matter what the outcome

---

[3] There exist mechanisms that will elicit subjective probabilities without requiring that one correct for risk attitudes, such as the procedures proposed by Köszegi and Rabin [2008; p.199], Karni [2009], Grether [1992], Holt and Smith [2009], and Offerman, Sonnemans, van de Kuilen and Wakker [2009], discussed further below. The latter three employ these mechanisms in an experimental evaluation. Special requirements on the experimental design limits the ability to use these procedures in policy-relevant settings in which it may not be possible to perform the required pairwise tasks on a within-subject basis.

for a report of ½. The LSR is not a favorite of decision theorists, since a risk neutral subject would jump to corner-solution reports of 1 or 0 whenever their true beliefs were either side of ½. But when the subject is (even modestly) risk averse, an interior solution is obtained, and we face the same issues of inference as with the QSR. The LSR is a favorite of experimental economists because of the simplicity of explaining the rule: the score for a report and an event is linear in the reported probability report, so there is no need for elaborate tables showing cryptic payoff scores for discrete reports.[4]

2. Experimental Design

A. Previous Experimental Designs

Experimental economists have used several of the popular scoring rules, but with one notable exception discussed below, none have corrected for any deviation from risk neutrality.

The QSR was apparently first used by McKelvey and Page [1990], and later by Offerman, Sonnemans and Schram [1996], McDaniel and Rutström [2001], Nyarko and Schotter [2002], Schotter and Sopher [2003], Costa-Gomes and Weizsäcker [2008] and Rutström and Wilcox [2009].[5] In each case the subject is implicitly or explicitly assumed to be risk-neutral. Scoring rules that are linear in the absolute deviation of the estimate have been used by Dufwenberg and Gneezy [2000] and Haruvy, Lahav and Noussair [2007]. Croson [2000] and Hurley and Shogren [2005] used scoring rules that are linear in the absolute deviation as well as providing a bonus for an exactly correct

---

[4] Hanson [1996; p. 1224] provides a useful reminder that discrete implementations of proper scoring rules can also engender piecewise linear opportunity sets. He points out that certain regions of the QSR implemented by McKelvey and Page [1990] were actually LSR, and that *risk-neutral* subjects would then rationally report a probability at the extremes of that linear region, and not at the discrete alternative closest to their true belief.

[5] Hanson [1996] contains some important corrections to some of the claims about QSR elicitation in McKelvey and Page [1990].

prediction.

Scoring rules that provide a positive reward for an "exact" prediction and zero otherwise have been used by Charness and Dufwenberg [2006] and Dufwenberg, Gächter and Hennig-Schmidt [2007]. In each case the inferential objective has been to test hypotheses drawn from "psychological game theory," which rest entirely on making operational the beliefs of players in strategic games. For this purpose, and also in other applications, the elicitation of beliefs is combined with some other experimental task. Other applications include testing the hypothesis that the belief elicitation task will encourage players in a game to think more strategically (Croson [2000], Costa-Gomes and Weizsäcker [2008], Rutström and Wilcox [2009]). Of course, combining tasks in this way violates the "no stakes condition" required for the QSR to elicit beliefs reliably unless one assumes that the subject is risk neutral (Kadane and Winkler [1988], Karni and Safra [1995] and Karni [1999]). Only one study employs a "spectator" treatment in which players are asked to provide beliefs but do not take part in the constituent game determining the event outcome: study #2 of Offerman, Sonnemans and Schram [1996].

In order to avoid portfolio effects from the combined choices rewards in these scoring rule task are sometimes very, very small. For example, Nyarko and Schotter [2002] and Rutström and Wilcox [2009] gave each subject an endowment of 10 cents, from which their penalties are to be deducted. In the latter study this does not present a problem since the focus is not on analyzing the elicited beliefs but on analyzing the game play as a function of exposing subjects to a belief elicitation process. Nevertheless, one has to worry about the incentive properties of the elicitation method once the interest is on analyzing the beliefs themselves.

The need to calibrate or control for risk aversion is often not made explicit, or is claimed to be of marginal concern. Schotter and Sopher [2003; p. 504] recognize the role of risk aversion, but appear to argue that it is not a factor behaviorally:

It can easily be demonstrated that this reward function provides an incentive for subjects to reveal their true beliefs about the actions of their opponents. Telling the truth is optimal; however, this is true only if the subjects are risk neutral. Risk aversion can lead subjects to make a "secure" prediction and place a .50 probability of each strategy. We see no evidence of this type of behavior.

Of course, evidence of subjects selecting the probability report of ½ only shows that the subject has *extreme* risk aversion. The absence of that extreme evidence says nothing about the role that risk aversion might play in general.

Only one study attempts to explicitly calibrate the beliefs for non-linear utility functions and/or probability weighting: Offerman, Sonnemans, van de Kuilen and Wakker [2009; §6].[6] Their elegant approach has a reduced form simplicity, and is agnostic about which structural model of decision making under risk one uses. The maintained assumption is that any deviation in reports from probabilities in a QSR based on objective and known probabilities transfer directly to a task where probabilities are subjective, less precise and not known with certainty. While precise and agnostic about structural assumptions, it is limited in applicability since it requires a within-subject approach and the employment of identical tasks and incentives across the natural lottery for which one wants to elicit subjective beliefs and a calibration task with known objective probabilities. Our approach is less precise but more general in the sense that it does not require a within-subject approach or two identical instruments for eliciting the beliefs and the belief calibration function. We introduce an experimental method where the decision model can be estimated on separate tasks for the same subjects, or even on a different pool of subjects drawn from the same population, and then

---

[6] The need for some correction is also recognized by Offerman, Sonnemans and Schram [1996; p.824, fn.8] and Rutström and Wilcox [2009; p.11, fn.8]. Although not explicitly performing calibration, Köszegi and Rabin [2008; p.199], Karni [2009], Grether [1992] and Holt and Smith [2009] all use procedures based on the same logic of comparing choices under objective probabilities to those under unobserved, subjective probabilities. However, all use the valuation methods due to Becker DeGroot and Marshak [1964] instead of the QSR. Qiu and Steiger [2010] extend the approach of Offerman et al. [2009] to identify the role of probability weighting on the correction of raw responses to the QSR.

statistically integrated with the belief elicitation task while transparently allowing error terms to propagate.

B. Our Experimental Designs

Figure 1 illustrates the lottery choice that our subjects were given. Each subject faced 45 such choices, where prizes spanned the domain $0 up to $100. One choice was selected to be paid out at random after all choices had been entered. Choices of indifference were resolved by rolling a die and picking one lottery, as had been explained to subject. This interface builds on the classic design of Hey and Orme [1994], and is discussed in greater detail in Harrison and Rutström [2008; Appendix B]. The lotteries were presented sequentially in 3 blocks of 15, where each block had prizes in one of three intervals between $0 and some higher level. One level was between $0 and $1, the other level was between $0 and $10, and the third level was between $0 and $100. We presented the lotteries sequentially so that the subject could see that all of the lotteries in one block were for a given scale. The sequence of blocks was randomized across subjects. Complete instructions are provided in appendix A.

The beliefs tasks were presented to subjects with a novel interface that has many attractive features.[7] Figure 2 shows the interface for the QSR as it was presented to subjects on a computer screen and in printed instructions. The interface was explained with these instructions, which used a trusty old bingo cage to illustrate one underlying random process:

> You place your bets by sliding the bar at the bottom of the screen until you are happy with your choice of a probability report. The computer will start at some point on this bar at random: in the above screen it started at 71%, but you are free to change this as much as you like. In fact, you should slide this bar and see the effects on earnings, until you are happy to confirm your choice.

---

[7] This interface immediately extends to other tasks with a cardinal scale that experimental economists use to elicit risk preferences, discount rates, or social preferences.

In this hypothetical example the maximum payoff you can earn is $1,000. In the actual tasks the maximum payoff will be lower than that, and we will tell you what it is when we come to those tasks. But the layout of the screen will be the same.

In this demonstration, the event you are betting on is whether a Ping Pong ball drawn from a bingo cage will be Orange or White. We have a bingo cage here, and we have 20 ping-pong balls. 15 of the balls are white, and 5 of the balls are orange. We will give the cage a few turns to scramble them up, and then select one ball at random.

What we want you to do is place a bet on which color will be picked. At the top of your screen we tell you what the event is: in this case, it is **Picking a Ping-Pong Ball**, and you need to bet on whether you think it will be **Orange or White**.

Your possible earnings are displayed in the two bars in the main part of the screen, and also in numbers at the bottom of each bar. For example, if you choose to report 71% you can see that you would earn $915.90 if the Ping Pong Ball was ORANGE, and $495.90 if the Ping Pong Ball was WHITE.

The subject was then taken through displays of their payoffs if they chose to report 0% or 100%.[8]

We then concluded the main instructions in this manner:

Summarizing, then, there are two important points for you to keep in mind when placing your bets:

1. **Your belief about the chances of each outcome is a personal judgment that depends on information you have about the different events.** In this practice example, the information you have consists of the total number of Orange balls and White balls.

2. **Your choices might also depend on your willingness to take risks or to gamble.** There is no right choice for everyone. For example, in a horse race you might want to bet on the longshot since it will bring you more money if it wins. On the other hand, you might want to bet on the favorite since it is more likely to win something.

For each task, your choice will depend on two things: your judgment about how likely it is that each outcome will occur, and how much you like to gamble or take risks.

Each subject then participated in an unpaid training choice, in which they were told the number of

---

[8] The display of the probability on the right side of the slider always show an integer percentage, and the earnings were always calculated on that value, thus making the continuous probability scale into an integer one. In the statistical analysis we treat the probability as taking on 101 integer values in the range 0 to 100.

orange balls in the bingo cage that was on public display, and asked to make a report and confirm it. We deliberately adopted an extremely high scale of a maximum $1000 payoff to ensure that the subjects understood that this was to be a trainer.

Each subject then participated in 7 belief elicitation tasks, knowing that one would be selected for payment. The first 3 were repetitions of the training task with orange and white ping pong balls: subjects were told that there were 60 balls in all in a publicly visible, but initially covered, bingo cage, but were not told the number of orange or white balls. The urn was uncovered and spun for 10 rotations, and then the subject had to make a report that a ball drawn at random would be orange. We do not consider these events here. The fourth task was based on the outcomes of a test in psychology for empathy known as The Eyes Test (e.g., Baron-Cohen [2003]). All subjects had completed this test at the outset of the session, and the event they were asked about was whether the score that a randomly chosen man got on the Eyes Test was equal to or greater than the score that a randomly chosen woman would get. The final three tasks were based on the 2008 U.S. Presidential Election, which was to be held about one week after the session. One task was whether the outright winner of the Presidency would be a Democrat or a Republican, one task was whether the winning share of the popular vote would be at least 5 percentage points greater than the losing share, and the final task was whether the winning share of the popular vote would be at least 10 percentage points greater than the losing share.[9] Our own *a priori* expectations for these subjective probabilities were just below 50%, around 80%, around 65% and less than 10%, respectively.

The exact phrasing of these events was explained in written instructions, which were also read out loud. The event based on the Eyes Test was explained as follows:

---

[9] These events compare to similar events employed in popular prediction markets, inspired by Forsythe, Nelson, Neumann and Wright [1992]. See http://www.biz.uiowa.edu/iem/ for the current version of this market, and the contracts traded in the 2008 Presidential Election.

The Eyes Test

At the beginning of today's experimental session we asked you to answer 36 questions, called **<u>The Eyes Test</u>**. These questions were designed by psychologists to measure a person's ability to "read someone else's mind" by just looking at their eyes.

Each and everyone of you were given the same 36 Eyes Test questions in today's experiment and a total score was recorded for each and every one of you in this experiment.

Now we come to the outcome we want you to place bets on in this portion of the experiment. We will pick one man and one woman in the room. Do you think the man who is selected will have a higher score on the Eyes Test than the woman who is selected?

After everyone in the experiment has made their bets for this event we will randomly select one man, and randomly select one woman from this experimental session. We will use the cards we collected, and sort then into one pile for men and one pile for women. Each pile will be shuffled, and one card drawn from each pile. We will then compare the score for the man that is drawn with the score for the woman that is drawn, and write these scores up on the board for you to see.

We therefore pose the following outcome for you to bet on now:

**That the man we select at random will have a higher score
on the Eyes Test than the woman we select at random.**

Do you have any questions?

After subjects had completed their bets on the bingo cage and Eyes Test tasks, the final three events

were explained as follows:

2008 Presidential Elections

We want you to place bets on some questions about the U.S. Presidential Elections being held in a few weeks:

1.   Will the next President of the United States be a Democrat or a Republican?

2.   Will the popular vote for the winning candidate be 5 or more percentage points greater than the popular vote for the losing candidate?

3.   Will the popular vote for the winning candidate be 10 or more percentage points greater than the popular vote for the losing candidate?

It is important that you understand that the first question is about the outcome of the Electoral College vote, and not the popular vote. The popular vote is just the sum of all votes across the United States. We are only referring to the Presidential Election, and not to any other elections that might occur on the same day.

For the second and third question, we are asking if you think that the winner of the popular vote will beat the loser by 5 or 10 percentage points or more. For example, if the winner of the popular vote gets 51% of the vote and the loser gets 49%, then this is a 2 percentage point difference. If the winner gets 53% and the loser gets 47%, then this is a 6 percentage point difference.

The election will be on Tuesday, November 4, 2008. To use a widely respected public source for the outcome, we will use the New York Times of Friday, November 7, 2008 as the official source used to determine your payoffs. In the event that there is a drawn out determination of the next President, such as in the 2000 election, we will delay payments until Inauguration Day, which is on January 20, 2009.

You will be paid for your bets in checks that will be mailed out on Monday, November 10, assuming we know who the next President will be at that time.

Please go ahead now and place your bets for this event, unless you have any questions.

The subjects then completed their belief elicitation tasks for these Presidential election events, and went on to the lottery choice tasks described earlier.

We recruited 140 subjects from the student population of the University of Central Florida, split equally across the QSR and LSR treatments. The experiments were conducted in the week of Monday October 27 through Friday October 31, in the week prior to the 2008 election. Our 140 subjects earned about $90 on average for the two paid tasks. Each session lasted around 2 hours. There was considerable variation in earnings, with one subject taking home $3 and another subject taking home $205.

3. Econometric Model

We develop the econometric model to be estimated in three stages. First we present the specification of risk attitudes assuming an EUT model of latent choice, where the focus is entirely

on the concavity of the estimated utility function. Second, we present the specification of risk

attitudes assuming a RDU model of latent choice, so that risk attitudes are determined by the

interplay of concave utility functions and non-linear probability weighting.[10] Third, we consider the

joint estimation of risk attitudes and subjective probability, using either the EUT or the RDU

specification.

A. Risk Attitudes under Expected Utility Theory

We assume an Expo-Power (EP) utility function originally proposed by Saha [1993].

Following Holt and Laury [2002], the EP function is defined as

$$u(y) = [1-\exp(-\alpha y^{1-r})]/\alpha, \tag{1}$$

where α and r are parameters to be estimated, and y is income from the experimental choice. The

EP function can exhibit increasing or decreasing relative risk aversion (RRA), depending on the

parameter α: RRA is defined by $r + \alpha(1-r)y^{1-r}$, so RRA varies with income if $\alpha \neq 0$ and the estimate

of r defines RRA at a zero income. This function nests CRRA (as $\alpha \to 0$) and CARA (as $r \to 0$).

The utility function (1) can be estimated using maximum likelihood and a latent EUT

structural model of choice. Let there be K possible outcomes in a lottery; in our lottery choice task

$K \leq 4$. Under EUT the probabilities for each outcome k in the lottery choice task, p_k, are those that

are induced by the experimenter, so expected utility is simply the probability weighted utility of each

outcome in each lottery i:

$$EU_i = \sum_{k=1,K} [p_k \times u_k]. \tag{2}$$

The EU for each lottery pair is calculated for a candidate estimate of r and α, and the index

---

[10] We could just develop an RDU model and test if the estimated probability weighting is the identity function, in which case the RDU model collapses to an EUT model. However, the exposition is, in our view, simpler if one develops the models separately because of the familiarity of EUT to most economists.

$$\nabla EU = eu_R/(eu_L + eu_R) \tag{3}$$

calculated, where

$$eu_i = exp(EU_i), \tag{4}$$

for i= {R, L}, and where EU_L is the "left" lottery and EU_R is the "right" lottery, as displayed to the subject and illustrated in Figure 1. This latent index, based on latent preferences, is already in the form of a probability.[11]

Thus the likelihood of the observed responses, conditional on the EUT and EP specifications being true, depends on the estimates of r and α given the above statistical specification and the observed choices. If we ignore responses that reflect indifference[12] the log-likelihood is then

$$\ln L(r, \alpha; y, \mathbf{X}) = \sum_i [(\ln \nabla EU \times \mathbf{I}(y_i = 1)) + (\ln (1-\nabla EU) \times \mathbf{I}(y_i = -1))] \tag{5}$$

where $\mathbf{I}(\cdot)$ is the indicator function, $y_i =1(-1)$ denotes the choice of the Option R (L) lottery in risk aversion task i, and $\mathbf{X}$ is a vector of individual characteristics reflecting age, sex, race, and so on. When we pool responses over subjects the $\mathbf{X}$ vector will play an important role to allow for some heterogeneity of preferences.

To allow for subject heterogeneity with respect to risk attitudes, the parameters r and α are each modeled as linear functions of observed individual characteristics of the subject. For example, assume that we only had information on the age and sex of the subject, denoted Age (years of age) and Female (0 for males, and 1 for females). Then we would estimate the coefficients r_0, r_1 and r_2 in

---

[11] It is well known, but useful to note, that (3) is equivalent to $\Lambda(EU_R - EU_L)$ where $\Lambda(\cdot)$ is the logistic cumulative density function. Thus (3) embodies a statistical "link function" between the difference in the EU of the two lotteries and the probability of the observed choice.

[12] In our lottery experiments the subjects are told at the outset that any expression of indifference would mean that the experimenter would toss a fair coin to make the decision for them if that choice was selected to be played out. Hence one can modify the likelihood to take these responses into account either by recognizing this is a third option, the compound lottery of the two lotteries, or alternatively that such choices implies a 50:50 mixture of the likelihood of choosing either lottery, as illustrated by Harrison and Rutström [2008; p.71]. We do not consider indifference here because it was an extremely rare event.

$r = r_0 + r_1 \times \text{Age} + r_2 \times \text{Female}$. Therefore, each subject would have a different estimated r, $\hat{r}$, for a given set of estimates of r_0, r_1 and r_2 to the extent that the subject had distinct individual characteristics. So if there were two subjects with the same sex and age, to use the above example, they would literally have the same $\hat{r}$, but if they differed in sex and/or age they would generally have distinct $\hat{r}$. In fact, we use four individual characteristics to model heterogeneity in our estimated risk preferences. Apart from a dummy for peopled aged over 22 years and gender, these include binary indicators for subjects that self-declare as having a high GPA (over 3.75) and self-declare as graduate students.

An important extension of the core model is to allow for subjects to make some errors. The notion of error is one that has already been encountered in the form of the statistical assumption (4) that the probability of choosing a lottery is not 1 when the EU of that lottery exceeds the EU of the other lottery.[13] By varying the shape of the link function implicit in (4), one can informally imagine subjects that are more sensitive to a given difference in the index ∇EU and subjects that are not so sensitive. We use the contextual error specification proposed by Wilcox [2010]. It posits the latent index

$$eu_i = \exp[(EU_i/\nu)/\mu], \tag{4'}$$

instead of (4), where ν is a normalizing term for each lottery pair L and R, and $\mu > 0$ is a structural "noise parameter" used to allow some errors from the perspective of the deterministic EUT model. The normalizing term ν is defined as the maximum utility over all prizes in this lottery pair minus the minimum utility over all prizes in this lottery pair, and ensures that the *normalized* EU difference

---

[13] This assumption is clear in the use of a link function from the difference between the EU of each option to the probability of picking one or other lottery; in the case of the logistic CDF that is implied by our approach in (3), this link function is $\Lambda(EU_R - EU_L)$. If the subject exhibited no errors from the perspective of EUT, this link function would instead be a step function: zero for all values of $(EU_R - EU_L) < 0$, anywhere between 0 and 1 for $(EU_R - EU_L) = 0$, and 1 for all values of $(EU_R - EU_L) > 0$. Harrison [2008; p.326] illustrates the implied CDF, referring to it as the CDF of a "Hardnose Theorist."

$[(EU_R - EU_L)/\nu]$ remains in the unit interval. As $\mu \to \infty$ this specification collapses ∇EU to 0 for any

values of EU_R and EU_L, so the probability of either choice converges to ½. So a larger μ means that

the difference in the EU of the two lotteries, conditional on the estimate of r, has less predictive

effect on choices. Thus μ can be viewed as a parameter that flattens out, or "sharpens," the link

functions implicit in (4). This is just one of several different types of error story that could be used,

and Wilcox [2008] provides a masterful review of the implications of the strengths and weaknesses

of the major alternatives.

Thus we extend the likelihood specification to include the noise parameter μ and maximize

$\ln L(r, \alpha, \mu; y, \mathbf{X})$ by estimating r, α and μ, given observations on y and $\mathbf{X}$.[14] Additional details of the

estimation methods used, including corrections for "clustered" errors when we pool choices over

subjects and tasks, is provided by Harrison and Rutström [2008; p.69ff].

B. Risk Attitudes under Rank Dependent Utility Theory

The RDU model extends the EUT model by allowing for decision weights on lottery

outcomes. Instead of (1) we have

$$u(y) = [1-\exp(-\acute{\alpha}y^{1-\acute{r}})]/\acute{\alpha}, \tag{6}$$

where $\acute{r}$ and $\acute{\alpha}$ are the coefficient for the curvature of the utility function, comparable to r and α in

(1). To calculate decision weights under RDU one replaces expected utility defined by (2) with RDU

$$RDU_i = \sum_{k=1, K} [w_k \times u_k] \tag{7}$$

where

$$w_i = \omega(p_i + ... + p_n) - \omega(p_{i+1} + ... + p_n) \tag{8}$$

for i=1,... , n-1, and

---

[14] The normalizing term ν is given by the value of r and the lottery parameters, which are part of $\mathbf{X}$.

$$w_i = \omega(p_i) \tag{9}$$

for i=n, the subscript indicating outcomes ranked from worst to best, and where $\omega(p)$ is some probability weighting function. We adopt the simple "power" probability weighting function proposed by Quiggin [1982], with curvature parameter γ:

$$\omega(p) = p^{\gamma} \tag{10}$$

So $\gamma \neq 1$ is consistent with a deviation from the conventional EUT representation.

The same behavioral noise specification is used as in the EUT model, replacing EU_i with RDU_i, leading to the definition of a comparable latent index ∇RDU. The likelihood specification for the RDU model is therefore

$$\ln L(\acute{r}, \acute{\alpha}, \gamma, \mu; y, \mathbf{X}) = \sum_i [(\ln \Phi(\nabla RDU) \times \mathbf{I}(y_i=1)) + (\ln (1-\Phi(\nabla RDU)) \times \mathbf{I}(y_i=-1))] \tag{11}$$

and entails the estimate of $\acute{r}$, $\acute{\alpha}$, γ and μ using maximum likelihood. Individual heterogeneity is allowed for by estimating the parameters $\acute{r}$, $\acute{\alpha}$ and γ as linear functions of the demographic characteristics defined earlier.

Figure 3 shows the manner in which the parameter γ characterizes the probability weighting function and the decisions weights used to evaluate lottery choices. Since we assume $\gamma=0.77<1$ in this illustration, to anticipate our estimates, the probability weighting function $\omega(p)$ is concave. For simplicity here we assume lotteries with 2, 3 or 4 prizes that are equally likely when we generate the decision weights. So for the case of 2 prizes, each prize has $p=\frac{1}{2}$; with 3 prizes, each prize has $p=\frac{1}{3}$; and with 4 prizes, each prize has $p=\frac{1}{4}$. For the 3-prize and 4-prize lottery we see the standard result, that the decision weights on the largest prizes are relatively greater than the true probability, and the decision weights on the smallest prizes are relatively smaller than the true probability. In the belief elicitation task there were only 2 prizes per lottery, so this value of the parameter γ puts greater

decision weight on the higher prize.[15]

Each panel in Figure 3 is important for our analysis. For the purposes of estimating γ from the observed lottery choices with known probabilities we only need the decision weights in the right panel of Figure 3. But for the purposes of recovering a subjective probability π subject to probability weighting, as distinct from p, the induced objective probability, we instead only need the probability weighting function. In fact, we need it's inverse function, since it is the π in the $\omega(\pi)$ function that we are seeking to recover in that case. We do not directly observe $\omega(p)$ or $\omega(\pi)$, but we can estimate $\omega(\cdot)$ as part of the latent structure generating the observed choices in the two types of task, implicitly assuming that $\omega(p) = \omega(\pi)$. To anticipate slightly the exposition below: once we have $\omega(\cdot)$ we can then recover π by directly applying the estimated probability weighting function, such as the one shown, for a typical γ, in the left panel of Figure 3.[16]

C. Estimating the Subjective Probability

To estimate the subjective probability π that each subject holds from LSR and QSR responses we have to assume something about how they make decisions under risk. This is obvious

---

[15] Hence a concave probability weighting function, as in the left panel of Figure 3, implies risk-seeking behavior (*ceteris paribus* the curvature of the utility function). As explained by Wakker [2010; §7.1], "The implications of probability weighting for risk attitude can be compared to those of utility. Concave utility curves enhance risk aversion. Similarly, *convex* weighting curves enhance pessimism and risk aversion. [...] Convex utility curves enhance risk seeking and so do *concave* weighting curves." This can be confused with the original exposition, as in Quiggin [1982; p.335][1993; p.61], which uses a cumulative distribution in (8) instead of the decumulative version the subsequent literature uses. There is no formal reason to use one or the other, as long as the interpretation of the implications of concave or convex probability weighting for risk attitudes is consistent.

[16] There is a serious caveat to this intuition when subjective probabilities are close to 0.5, due to a jump discontinuity in decision weights as reports under the QSR vary around 0.5. At the point where the report is 0.5 the rank ordering of the choice options reverses. For reports less than 0.5 the A option implies higher monetary rewards, but for reports greater than 0.5 it is the B option that offers the higher reward. For sufficiently large probability weighting this jump discontinuity can wreak havoc with the ability to infer latent subjective probabilities. In our applications the probability weighting is not severe, and we can initialize the maximum likelihood estimation at the solution values obtained by assuming EUT (so that directional derivatives do not need to be numerically evaluated in the problematic regions of γ and π).

in theory, and the only issue then is how to operationalize that property of these scoring rules.

If they are assumed to be risk neutral, then we can directly infer the subjective probability from the report of the subject.[17] This result is immediate under the QSR, but raises a problem of interpretation under the LSR if the reports are not at the corner solutions of 0% and 100%. On the other hand, any minimal level of risk aversion will suffice, under the LSR, to generate interior responses, so we assume that the subjects indeed have some minimal level of risk aversion when we report "risk neutral subjective beliefs" for the LSR.

Moving to the models that allow for varying risk attitudes, we jointly estimate the subjective probability and the parameters of the core model. Assume for the moment that we have an EUT specification. The subject that selects report θ from a given scoring rule receives the following EU

$$EU_\theta = \quad \pi_A \times u(\text{payout if A occurs} \mid \text{report } \theta) + \\ (1-\pi_A) \times u(\text{payout if B occurs} \mid \text{report } \theta) \tag{12}$$

where π_A is the subjective probability that A will occur. The payouts that enter the utility function are defined by the scoring rule and of course the specific report θ, and span the interval [\$0, \$100]. For the QSR and a report of 75%, for example, we have

$$EU_{75\%} = \pi_A \times u(\$93.75) + (1-\pi_A) \times u(\$43.75) \tag{12'}$$

For the LSR, and the same report, we have:

$$EU_{75\%} = \pi_A \times u(\$75) + (1-\pi_A) \times u(\$25) \tag{12''}$$

and so on for other possible reports. We observe the report made by the subject for QSR or LSR. This report can take 101 different integer values defined over percentage points. Then we can calculate the likelihood of that choice given values of r, π_A and μ, where the likelihood is the multinomial analogue of the binary logit specification used for lottery choices. We define

---

[17] The expression "risk neutral" here should be understood to include the curvature of the utility function *and* the curvature of the probability weighting function. So it is not just a statement about the former, unless one assumes EUT.

$$eu_\Theta = \exp[(EU_\Theta/\nu)/\mu] \tag{13}$$

for any report Θ, analogously to (4′), and then

$$\nabla EU = eu_\theta/(eu_{0\%} + eu_{1\%} + ... + eu_{100\%}) \tag{14}$$

for the specific report θ observed, analogously to (3).

We need r and α to evaluate the utility function in (12), we need π_A to calculate the EU_θ in (12) for each *possible* report Θ in {0%, 1%, 2%, ..., 100% }once we know the utility values, and we need μ to calculate the latent indices (13) and (14) that generate the subjective probability of observing the choice of specific report θ when we allow for some noise in that process. The *joint* maximum likelihood problem is to find the values of these parameters that best explain observed choices in the belief elicitation tasks as well as observed choices in the lottery tasks.

Exactly the same logic extends to the model in which we assume an RDU latent structure instead of an EUT latent structure. In effect, the lottery task allow us to identify r and α under EUT, and ŕ, ά and γ under RDU, since π_A plays no direct role in explaining the choices in that task. Individual heterogeneity is allowed for by estimating the π parameter as a linear functions of the demographic characteristics defined earlier, as well as allowing risk attitudes to vary with demographic characteristics.

4. Results

A. Raw Elicited Beliefs

Figure 4 displays the raw responses from each of the scoring rules for each event, in the form of kernel densities, and Table 1 shows summary statistics of the elicited responses. The four events are the "Eyes Test," the "President," "win by 5%," and "win by 10%." The summary statistics suggest that the QSR and LSR provided roughly the same responses, but the densities in Figure 4 do have some differences in shape. In part this simply alerts us to be aware of the non-

Gaussian shape of these distributions.

The general location of the densities corresponds with our qualitative priors on the subjective beliefs that were to be expected for these events. Recall that the Eyes Test asked for the probability that a typical male would score better than a typical female. For the Eyes Test it appears, from the observation that the modal response is around 0.4 for both LSR and QSR treatments, that the sample did not expect the male score to exceed the female score, but that there was a wide variability around this modal belief. The sample appeared confident that Barack Obama would indeed win the election outright, but displayed a healthy sense of perspective on what the winning margin would be. For our purposes, the choices of a 5% and 10% threshold for the popular vote could not have worked out better, with a majority believing that a 5% margin would be attained but that a 10% margin would not. These results show, at a minimum, that responses were at least correlated with what we believe to be reasonably coherent subjective beliefs for these events.

The fact that the responses to the LSR are not at "corner" values of 0 or 1 shows that the subjects were not exactly risk neutral. But it does not show much more, because one would observe some interior response even for small amounts of risk aversion, as noted earlier.

With the exception of the bets on the outright winner of the Presidential election, the distribution of responses for the two scoring rules are roughly the same. This conclusion is supported by a Kolmogorov-Smirnov test of the null hypothesis that the two distributions are equal. In the case of the outright winner event, the *p*-value on this test is only 0.011, so we can reject that hypothesis in this instance. This finding provides some support for those that would prefer to use the LSR on the grounds that it is simpler to explain to subjects than the QSR. Of course, the real issue is whether they generate the same estimates of subjective probability when one allows for risk attitudes.

B. Characterizing Risk Attitudes

Looking just at the lottery choices under a maintained hypothesis of EUT for now, we find evidence of modest risk aversion at low stakes (since r=0.3>0, and r defines RRA at y=0), and evidence of slightly increasing relative risk aversion as the prizes climb to $100 (since α=0.03>0). Detailed results are provided in Appendix B, since they are only of indirect interest here. Given these parameter estimates we can calculate RRA at various prize levels: at $25, $50, $75 and $100 the RRA is estimated to be 0.49, 0.61, 0.71 and 0.81, respectively.[18] Thus, despite the relatively low estimate of α the implied risk aversion at higher stakes is much stronger than for the lower stakes.

When we allow for an array of covariates to better characterize the heterogeneity of risk attitudes, we observe females to be significantly more risk averse, with RRA 0.14 higher at the $0 level. Detailed estimates are provided in a Appendix B. As expected, allowing for covariates has no significant average effect on the RRA by prize level.

These results also suggest that one might see somewhat different effects of "risk-conditioning" in the reports for scoring rules depending on the stakes involved. Our stakes are large in relation to the literature: a maximum prize of $100, compared to common implementations in experiments of a maximum prize of less than $1. To be fair, those low-stake implementations are often in the context of the probability elicitation task being paired with another task, such as the choice of a strategy in a game, and the stake for the probability elicitation task is kept small to avoid the subject attempting to construct a portfolio of paired responses across the two tasks. In other applications, however, where the interest is directly on eliciting beliefs, such small stakes will not generally result in precise estimates, so larger stakes are necessary. Our results on risk attitudes for low stakes and high stakes imply that the extent of adjustment for risk attitudes is much greater for

---

[18] We simply use the estimated model to predict the point estimate of the RRA for each subject and each of these prizes, and report the average of those RRA point estimates here.

higher stake elicitations. It is a factor for both, since we estimate RRA to be positive for the lowest stakes, but it is not as serious a factor as when the stakes are higher.

The results from estimating the RDU model are slightly different. Detailed estimates are again reported in Appendix B. The estimates of the utility function parameters $\acute{r}$ and $\acute{\alpha}$ are both larger than their counterparts (r and α) under EUT, implying greater concavity of the utility function. We estimate the probability weighting parameter under RDU to be $\gamma=0.72$ without covariates, and can reject the hypothesis that this is equal to 1 (p-value<0.001). A likelihood ratio test of the hypothesis that the EUT model and the RDU model are the same when there are no covariates has a $\chi^2_1 = 35.47$ (p-value<0.01), so we reject that null. The same conclusion is true when we account for covariates and heterogeneity of responses. There are no strikingly different demographic characteristics, and the small changes across the board do not add up to a statistically significant difference. A likelihood ratio test of the hypothesis that the EUT model and the RDU model are the same when including covariates has a $\chi^2_{14} = 47.74$ (p-value<0.01). Thus, we find support for the hypothesis of probability weighting in this case, although here we are not investigating to what extent this is generally true or just holds for portions of our subjects and/or tasks.[19]

One noteworthy feature of these estimates is that one can reject the CRRA specification for both EUT and RDU models in this case. Of course, one might accept CRRA if estimating risk attitudes over a much smaller income domain, such as between $0 and $10, or when the variation in stakes as a percentage is relatively small.

---

[19] Investigating the breakdown of support for non-linear probability weights could be done by using a finite mixture specification to evaluate the fraction of choices consistent with EUT and the fraction consistent with RDU, following Harrison and Rutström [2009]. For our immediate purposes such a refined characterization of risk attitudes is not needed.

C. Estimating Subjective Probabilities

Table 2 lists the main results from estimating subjective probabilities for each of the four

events considered here, and assuming either an EUT or RDU specification. We pool data over LSR

and QSR, and control for the effect of the scoring rule with a binary dummy variable.

Given that we find evidence of risk aversion in our subjects over the domain of prizes used

in the belief elicitation tasks, our estimated subjective probabilities are all translations of the raw

responses away from the 50% response. In both the LSR and QSR we expect risk averse subjects to

make choices biased towards 50%, so that when we correct for their risk attitudes the inferred

probabilities should move away from 50%. The reason, again, is that risk averse subjects are drawn

to respond toward 50% simply to reduce the uncertainty over payoffs, so evidence of risk aversion

implies that their true, latent, subjective probabilities must be further away from 50% than their raw

responses. Our maximum likelihood estimates simply impose some parametric structure on that

theoretical structure, to be able to quantify the extent of the required translation and the precision of

the resulting inference about the latent subjective probability.

For the Eyes Test, we observe a very small movement in subjective probabilities away from

the raw responses under EUT. For any given risk attitude, the closer the true subjective

probabilities are to 50% the smaller is the absolute bias generated by the tendency for risk aversion

to move reports towards 50%. The effect of probability weighting is statistically significant, further

adjusting the inferred probabilities away from the 50% responses. The estimated value of the

probability weighting parameter in this case is $\gamma=0.72$, which is close to the value used to generate

the illustration in Figure 3. By inspection of the left panel of Figure 3 we observe that probability

values of just less than 50% would appear to be the most significantly affected by this functional

form. This explains why, for this task with probabilities in this range, the RDU model generates

different estimates than the EUT model.

For the three election events, we see more interesting effects of adjusting for risk aversion. In the "win by 5%" event, which is again one in which the raw responses are relatively close to 50%, we again infer a small translation from the raw response average of 59% to 62% under EUT, and from 59% to 65% under RDU. The 95% confidence interval on each of these estimate includes 59%. The other two election events, on the other hand, illustrate when one might expect to see the effect of risk aversion exert a more significant quantitative effect. In the case of the outright winner event we estimate latent subjective probabilities to be 80% (EUT) or 93% (RDU), rather than the raw response average of 71%. Moreover, the 95% confidence interval on these estimates does not include the raw response. Similarly, in the case of the chance of the popular vote for the winner being more than 10% of the popular vote of the loser, we estimate subjective probabilities of 18% (EUT) or 5% (RDU), rather than the raw response average of 28%.

It is possible to estimate the structural model of risk attitudes and subjective beliefs allowing for observable covariates on each parameter. To illustrate, Table B3 in Appendix B lists detailed EUT estimates for the model of the "win by 5%" event when we include covariates.[20] One reason for controlling for covariates is to allow for sample composition differences between treatments.[21] We observe no statistically significant effect from using the QSR or LSR, but there is no reason to expect one *after we properly condition for risk attitudes*. That is, the two scoring rules simply provide subjects with different lotteries with which to place bets about their subjective beliefs. So the same

---

[20] The estimates for the subjective probability π refer to a non-linear transform in which we actually estimate the parameter κ and then convert κ to π using $\pi = 1/(1+\exp(\kappa))$. Thus κ can vary between $\pm\infty$ and π is constrained to the open unit interval. To interpret these coefficients, $\kappa=0$ implies $\pi=\frac{1}{2}$, $\kappa>0$ implies $\pi<\frac{1}{2}$, and $\kappa<0$ implies $\pi>\frac{1}{2}$. The estimated subjective probabilities shown in Figures 5 and 6 have been converted back from κ to π using this non-linear function and the "delta method" to correctly calculate standard errors (Oehlert [1992]). In addition, it is the linear function of κ that is constrained by this transform to be in the unit interval, not each element of that function. Thus, in Table B3 the constant term for π has a statistically significant coefficient of -1.2, which would violate that constraint *if there were no covariates*.

[21] For example, men and women might have different risk attitudes or subjective beliefs, and the mix of men and women could vary from treatment to treatment. This can occur even with randomization to treatment, particularly when considering a wide range of covariates.

subjective belief should be estimated from each treatment, once one has conditioned on risk attitudes. Again, this provides support for those that would use the LSR over the QSR on the grounds that it is easier to explain to subjects.

To put these results into perspective, particularly those for the outright winner, it is important to note that in the week of the experiments the tide of public opinion clearly favored Barack Obama to win in a landslide. Eliciting a raw belief of only a 70% chance of Obama winning is therefore puzzling, and an *a priori* challenge for those that would use the raw results from a QSR or LSR procedure. For example, consider the Iowa Presidential Market, the current version of the prediction market developed by Forsythe, Nelson, Neumann and Wright [1992]. Average daily prices on this market for the month of October 2008, and specifically the week in which our experiments were conducted, implied that the market probability of Obama winning was around 85%. Moreover, this was the prevailing sense of the market for at least 2 weeks prior to our experiments. Indeed, the online Irish betting house, PaddyPower.com, was already paying out on over €1 million in pro-Obama bets as early as October 17! However, calibrating these inferred market probabilities to reflect the fact that most populations are risk averse on average would imply an even higher subjective probability.[22]

The prices from the Iowa Presidential Market on the popular vote share suggested that a 10% difference was possible. On the other hand, one striking feature of the two contracts, the "winner take all" contract and the "popular vote share" contract, is that the latter was poorly traded in comparison to the former as measured by volume. Of course, that could reflect a market that has found an equilibrium price, but it also could reflect a market in which there is too much uncertainty

---

[22] Fountain and Harrison [2010] examine many ways on which averages or medians from prediction markets might not reflect the average or median of the aggregate distribution of beliefs, apart from heterogeneity of risk attitudes.

about the outcome for traders to feel safe making a bet. Additionally, correcting for the risk preferences of the average market participant would imply a much lower probability, consistent again with what we report here.

Finally, we can assess the effect of econometric uncertainty about risk attitudes on inferences about subjective probabilities. To illustrate, focus on the "win by 5%" event and the EUT model of risk attitudes, although nothing essential relies on those choices. The unconstrained estimated subjective probability, conditioning on risk attitudes, is 0.730 with a standard error of 0.079. If we constrain all parameters, other than the subjective probability itself, to equal their unconstrained point estimates, the estimated subjective probability is still 0.730 but the standard error drops to 0.054. This effect might seem quantitatively small, but it is exactly what theory would predict qualitatively; the small quantitative effect is due to the relative precision of our unconstrained estimates. Apart from the standard error of the estimate, reflecting sampling error, it is possible that we do not have a precise point estimate of risk attitudes for some other reason (e.g., women and men might have slightly different risk attitudes, and the point estimate just reflects the average). If we increase the r parameter by 0.05 from its unconstrained point estimate, continuing to hold all other parameters at their unconstrained point estimates, the estimated subjective probability drops to 0.728 with a standard error of 0.054. If we decrease the r parameter by 0.05, the subjective probability increases to 0.735 with a standard error of 0.055. So imprecision in our estimates of risk attitudes translates directly into imprecision in our estimates of subjective probabilities. On the other hand, assuming a precise value of an unknown risk attitude will simply make the belief estimates appear more precise than they really are.

5. Conclusions

We demonstrate how one can make the theory of subjective probability elicitation practical. Our experimental design shows how one can pair different types of choice tasks to allow estimation of risk attitudes, which can then be used to condition the inferences from responses to the scoring rule tasks. Our structural econometric model shows how maximum likelihood methods can then be used to estimate subjective probabilities while properly controlling for both the heterogeneous risk attitudes and for the statistical errors of the latter. We applied this approach to elicit subjective probabilities over four naturally occurring events from a sample of 140 subjects. As an added advantage, our novel experimental interface allows subjects to see the tradeoff between alternative reports and payoff consequences, without the need for long lists of payoff tables to be printed out. Once we correct for risk attitudes using our joint estimation approach we find no significant difference between the Quadratic and Linear Scoring rules. This makes the Linear Scoring rule appealing due to the simplicity of explaining it to subjects.

Our results show that one has to be sensitive to the risk attitudes of subjects before drawing inferences about subjective probabilities from responses to scoring rules. More accurately, we show that one cannot just directly treat the responses to the scoring rule as if it is a subjective probability, unless one is willing *a priori* to make striking assumptions about risk attitudes. Those assumptions are rejected in our data.

Quite apart from inferring the correct point estimate of subjective probability, uncertainty about risk attitudes affects the confidence that one should have in any point estimate. Even if subjects are "approximately risk neutral" on average, and the QSR is used, uncertainty about the precise level of risk attitudes should be properly reflected in uncertainty about inferences over subjective probabilities. Our analysis has demonstrated how to combine theory and econometrics to do just that. The choice task for eliciting subjective probabilities generates a point response that

might appear to be quite precise by itself, but which is actually not a very precise estimate of the latent subjective probability when one properly accounts for uncertainty over risk attitudes.

Of course, although the estimation of subjective probabilities is an important objective in itself, the issue of how to best characterize subjective uncertainty, and attitudes towards it, involve deeper issues. Our estimation approach is clearly within the conventional Bayesian subjective probability framework of Savage [1971][1972]. That framework provides one obvious point of departure for criticisms of EUT based on ideas that subjective *beliefs* should not be represented by subjective *probabilities*. Exactly how one then models subjective beliefs is an open and important area of research (e.g., Smith [1969], Gilboa and Schmeidler [1989], Ghirardoto, Maccheroni and Marinacci [2004], Klibanoff, Marinacci and Mukerji [2005], Nau [2007] and Gilboa, Postlewaite and Schmeidler [2008]). Further, it seems plausible that subjects exhibit some degree of "uncertainty aversion" in addition to traditional risk aversion when faced with making decisions about events that involve subjective probabilities rather than objective probabilities.
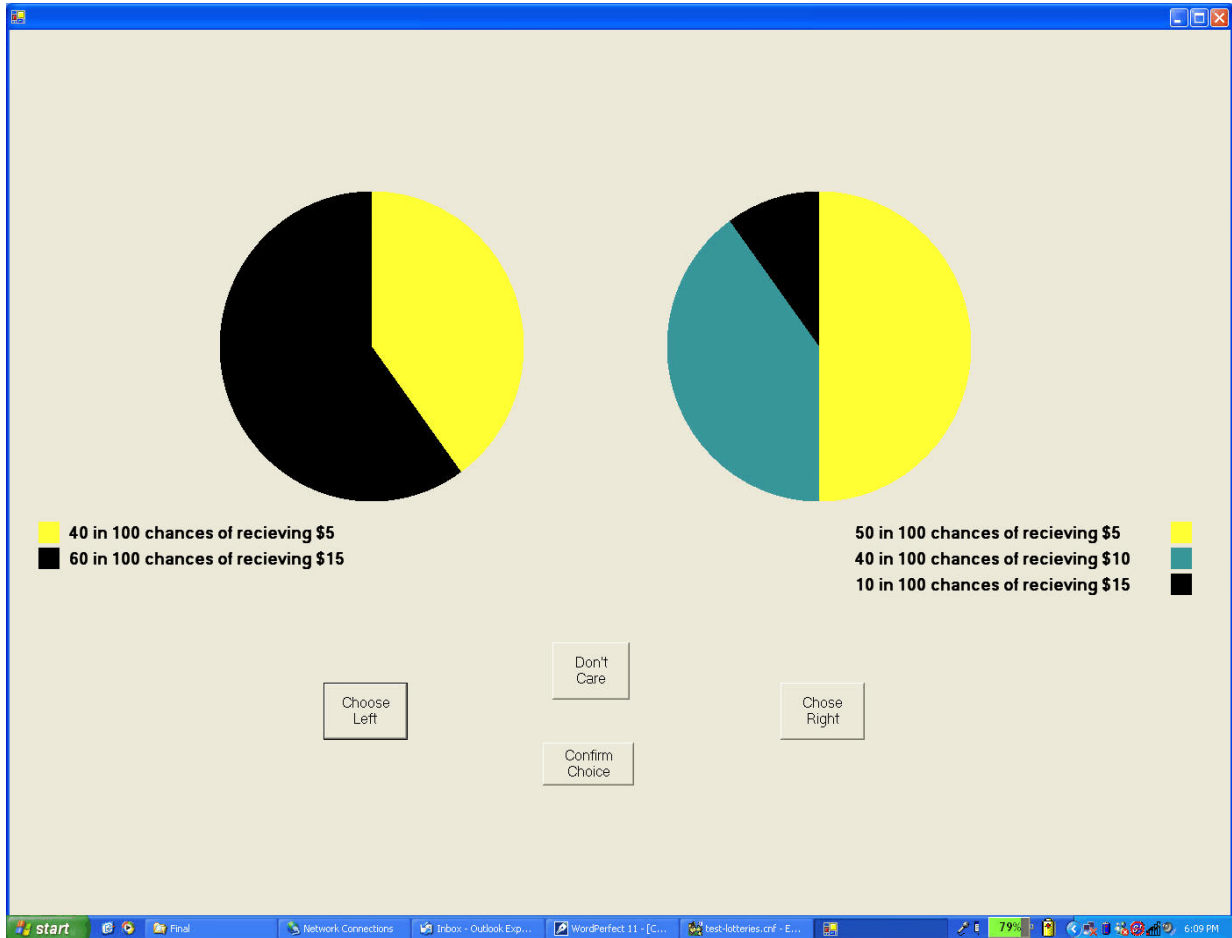
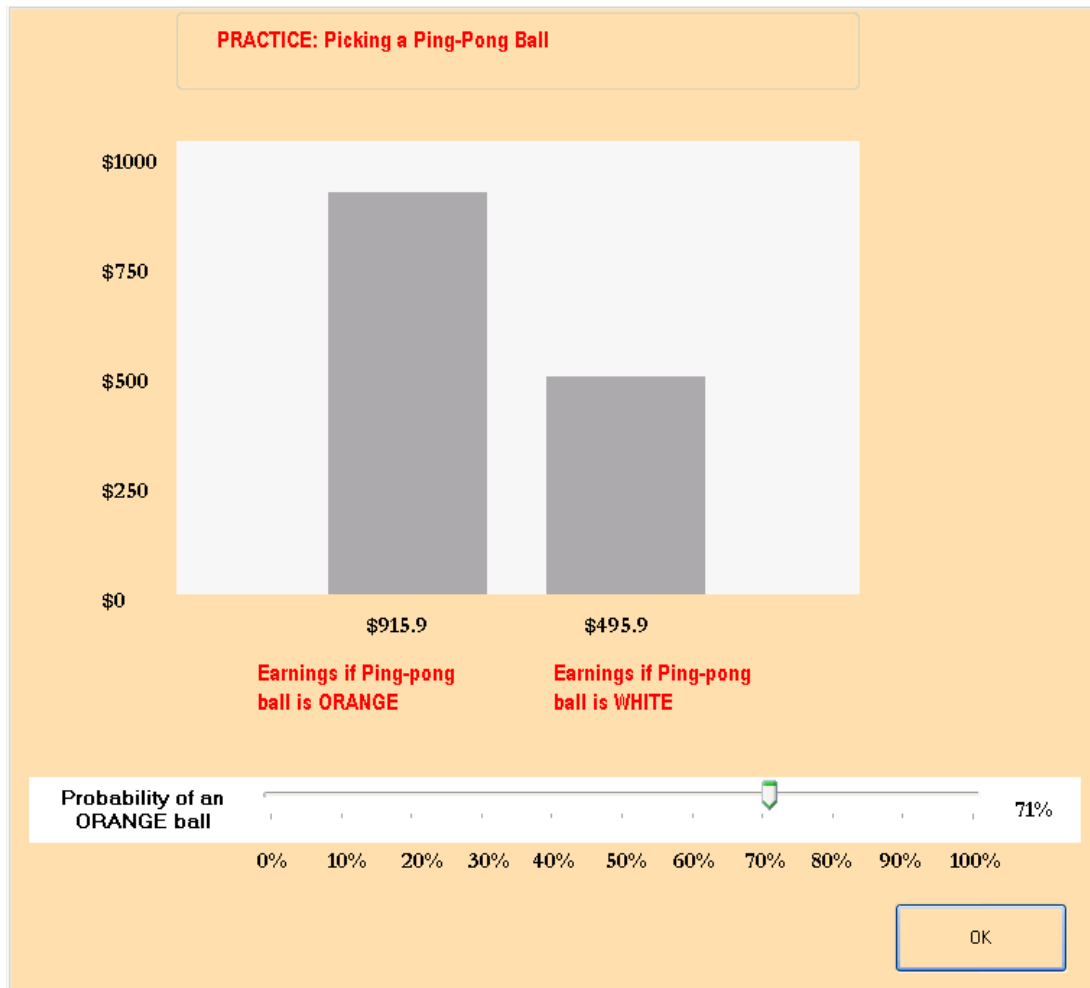**Figure 1**: Illustrative Lottery Choice

Figure 2: Illustrative Quadratic Scoring Rule Interface

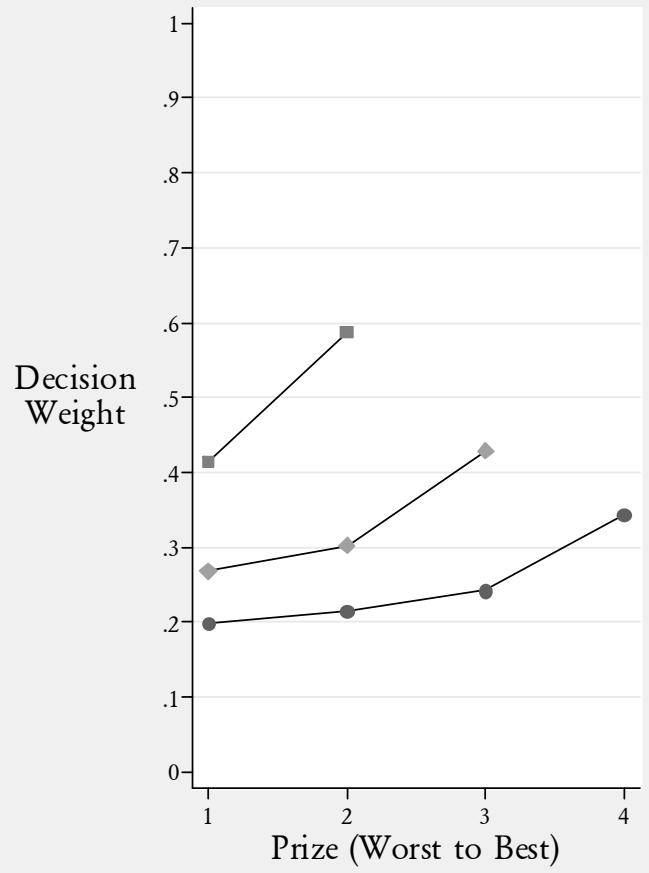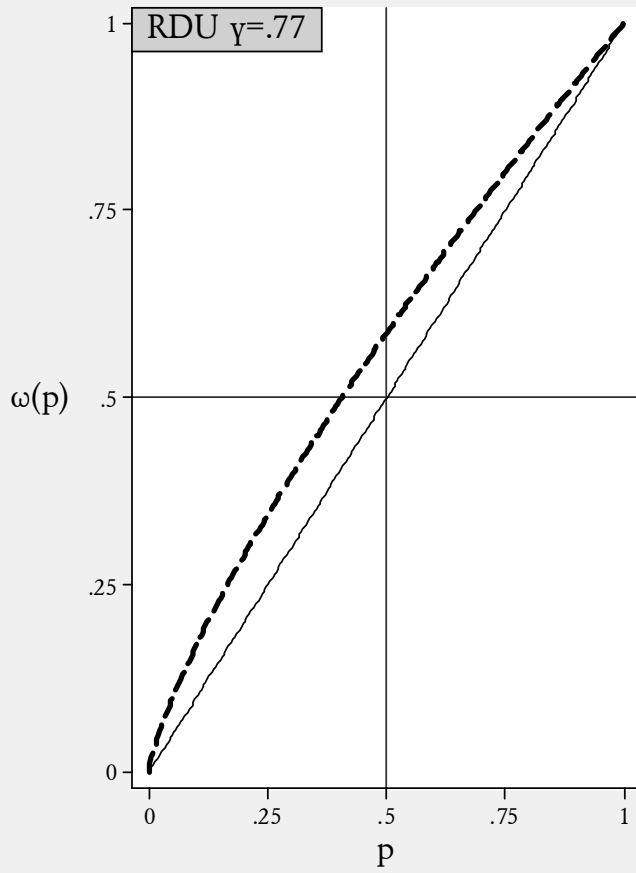Figure 3: Probability Weighting and Decision Weights

Table 1: Descriptive Statistics for Scoring Rule Responses

Event	Scoring Rule	Mean	Median	Standard Deviation
Eyes Test	Quadratic	0.43	0.4	0.19
	Linear	0.43	0.4	0.19
	Both	0.43	0.4	0.19
President	Quadratic	0.69	0.7	0.2
	Linear	0.74	0.75	0.22
	Both	0.71	0.7	0.21
Win by 5%	Quadratic	0.59	0.6	0.23
	Linear	0.58	0.6	0.26
	Both	0.59	0.6	0.24
Win by 10%	Quadratic	0.28	0.3	0.2
	Linear	0.29	0.26	0.21
	Both	0.28	0.3	0.21

Figure 4: Elicited Responses from QSR and LSR

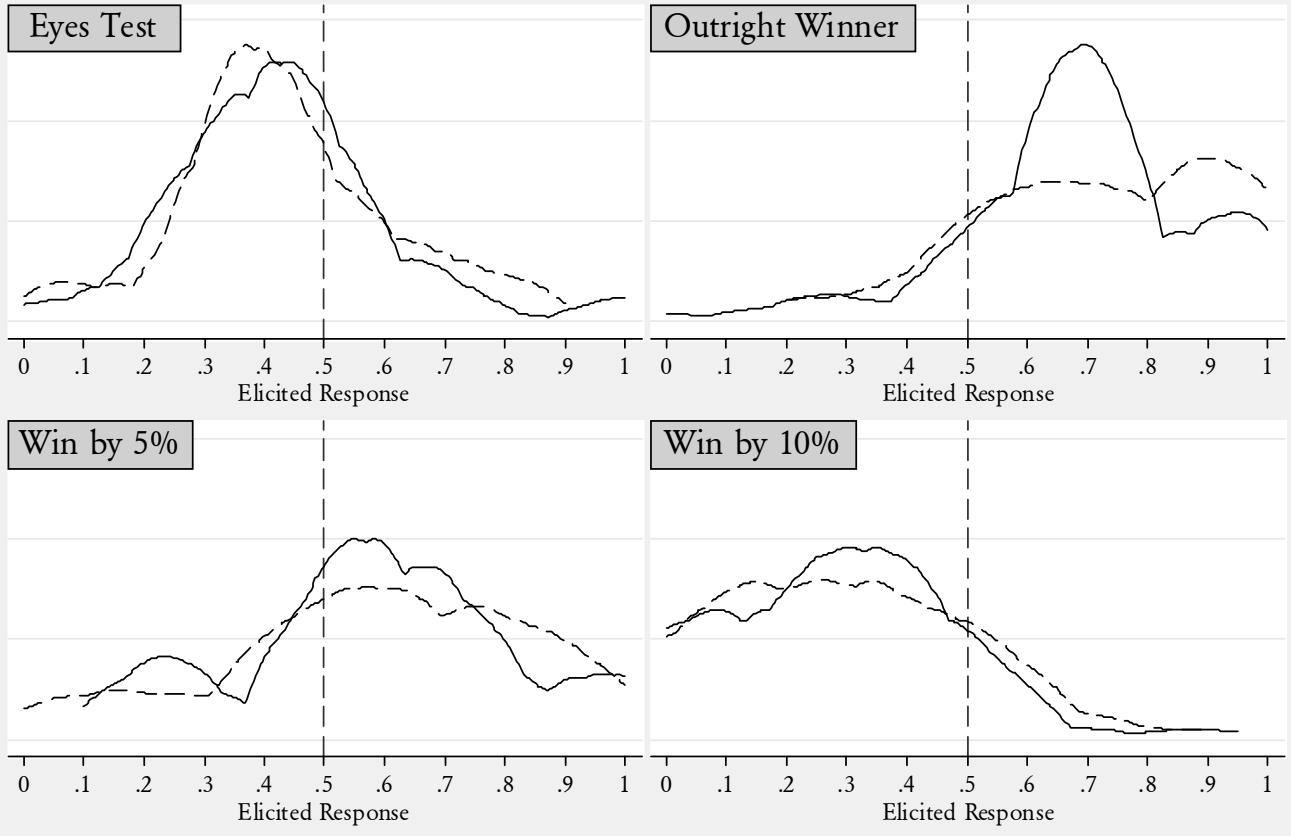Kernel density for QSR is solid line, and LSR is dashed line

Table 2: Estimated Subjective Probabilities

Event	Specification (Log-Likelihood)	Point Estimate	Standard Error or Standard Deviation	95% Confidence Interval
Eyes Test	EUT (-4563.2)	0.42	0.022	0.38/0.46
	RDU (-4553.7)	0.42	0.027	0.36/0.47
	Raw Responses	0.43	0.19	0/0.95
President	EUT (-4558.1)	0.80	0.033	0.74/0.87
	RDU (-4541.2)	0.93	0.094	0.75/‡
	Raw Responses	0.71	0.21	0.20/1
Win by 5%	EUT (-4593.8)	0.62	0.023	0.57/0.67
	RDU (-4475.4)	0.65	0.041	0.57/0.73
	Raw Responses	0.59	0.24	0.10/1
Win by 10%	EUT (-4560.9)	0.18	0.038	0.11/0.25
	RDU (-4538.9)	0.053	0.092	‡/0.23
	Raw Responses	0.28	0.2	0/0.77

Notes: ‡ Due to numerical unreliability of the subjective probability at extreme "corner" values, the standard errors and confidence intervals could not be reliably estimated.

References

Baron-Cohen, Simon, *The Essential Difference: The Truth About the Male and Female Brain* (New York: Basic Books, 2003).

Becker, Gordon M.; DeGroot, Morris H., and Marschak, Jacob., "Measuring Utility By A Single-Response Sequential Method," *Behavioral Science*, 9, July 1964, 226-232.

Bernardo, José, "Expected Information as Expected Utility," *Annals of Statistics*, 7(3), 1979, 686-690.

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

Costa-Gomes, Miguel A., and Weizsäcker, Georg, "Stated Beliefs and Play in Normal-Form Games," *Review of Economic Studies*, 75(3), 2008, 729-762.

Croson, Rachel, "Thinking like a game theorist: Factors affecting the frequency of equilibrium play," *Journal of Economic Behavior and Organization*, 41(3), 2000, 299-314.

de Finetti, Bruno, "La Prévision: Ses Lois Logiques, Ses Sources Subjectives," *Annales de l'Institut Henri Poincairé*, 7, 1937, 1-68; English translation as "Foresight: Its Logical Laws, Its Subjective Sources," in H.E. Kyburg and H.E. Smokler (eds.), *Studies in Subjective Probability* (Huntington, NY: Robert E. Krieger, 1980, Second Edition).

de Finetti, Bruno, "Logical Foundations and Measurement of Subjective Probability," *Acta Psychologica*, 34, December 1970, 129-145.

Dufwenberg, Martin; Gächter, Simon, and Hennig-Schmidt, Heike, "The Framing of Games and the Psychology of Play," *Working Paper*, Department of Economics, University of Arizona, May 2007.

Dufwenberg, Martin, and Gneezy, Uri, "Measuring beliefs in an experimental lost wallet game," *Games & Economic Behavior*, 30, 2000, 163-182.

Epstein, Richard A., *The Theory of Gambling and Statistical Logic* (San Diego: Academic Press, 1977).

Forsythe, Robert; Nelson, Forrest; Neumann, George R., and Wright, Jack, "Anatomy of an Experimental Political Stock Market," *American Economic Review*, 82(5), December 1992, 1142-1161.

Fountain, John, and Harrison, Glenn W., "What Do Prediction Markets Predict?" *Applied Economics Letters*, 17, 2010 forthcoming.

Ghirardoto, Paolo; Maccheroni, Fabio, and Marinacci, Massimo, "Differentiating Ambiguity and Ambiguity Attitude," *Journal of Economic Theory*, 118, 2004, 133-173.

Gilboa, Itzhak, and Schmeidler, David, "Maxmin Expected Utility with a Non-Unique Prior," *Journal of Mathematical Economics*, 18, 1989, 141-153.

Gilboa, Itzhak; Postlewaite, Andrew P., and Schmeidler, David, "Probability and Uncertainty in Economic Modeling," *Journal of Economic Perspectives*, 22(3), Summer 2008, 173-188.

Good, I.J., "Rational Decisions," *Journal of the Royal Statistical Society*, 14(1), 1952, 107-114.

Grether, David M., "Testing Bayes Rule and the Representativeness Heuristic: Some Experimental Evidence," *Journal of Economic Behavior & Organization*, 17, 1992, 31-57.

Hanson, Robin, "Correction to McKelvey and Page, 'Public and Private Information: An Experimental Study of Information Pooling'," *Econometrica*, 64(5), September 1996, 1223-1224.

Hanson, Robin, "Combinatorial Information Market Design," *Information Systems Frontiers*, 5(1), 2003, 107-119.

Harrison, Glenn W., "Theory and Misbehavior of First-Price Auctions: Reply," *American Economic Review*, 82, December 1992, 1426-1443.

Harrison, Glenn W., "Neuroeconomics: A Critical Reconsideration," *Economics & Philosophy*, 24(3), 2008, 303-344.

Harrison, Glenn W., and Rutström, E. Elisabet, "Risk Aversion in the Laboratory," in J.C. Cox and G.W. Harrison (eds.), *Risk Aversion in Experiments* (Bingley, UK: Emerald, Research in Experimental Economics, Volume 12, 2008).

Harrison, Glenn W., and Rutström, E. Elisabet, "Representative Agents in Lottery Choice Experiments: One Wedding and A Decent Funeral," *Experimental Economics*, 12, 2009, 133-158.

Haruvy, Ernan; Lahav, Yaron, and Noussair, Charles, "Traders' Expectations in Asset Markets: Experimental Evidence," *American Economic Review*, 97(5), December 2007, 1901-1920.

Holt, Charles A, "Scoring Rule Procedures for Eliciting Subjective Probability and Utility Functions," in P.K. Goel and A. Zellner (eds.), *Bayesian Inference and Decision Theory: Essays in Honor of Bruno de Finetti* (Amsterdam: North-Holland, 1986).

Hey, John D., and Orme, Chris, "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62(6), November 1994, 1291-1326.

Holt, Charles A., and Laury, Susan K., "Risk Aversion and Incentive Effects," *American Economic Review*, 92(5), December 2002, 1644-1655.

Holt, Charles A., and Smith, Angela M., "An Update on Bayesian Updating," *Journal of Economic Behavior & Organization*, 69, 2009, 125-134.

Hurley, Terrance M., and Shogren, Jason F., "An Experimental Comparison of Induced and Elicited Beliefs," *Journal of Risk and Uncertainty*, 30(2), 2005,169-188.

Jose, Victor Richmond R.; Nau, Robert F, and Winkler, Robert L., "Scoring Rules, Generalized Entropy, and Utility Maximization," *Operations Research*, 56 (5), 2008, 1146-1157

Kadane, Joseph B., and Winkler, Robert L., "de Finetti's Method of Elicitation," in R. Viertl (ed.), *Probability and Bayesian Statistics* (New York: Plenum, 1987).

Kadane, Joseph B., and Winkler, Robert L., "Separating Probability Elicitation from Utilities," *Journal of the American Statistical Association*, 83(402), June 1988, 357-363.

Karni, Edi, "Elicitation of Subjective Probabilities when Preferences are State-Dependent," *International Economic Review*, 40(2), May 1999, 479-486.

Karni, Edi, "A Mechanism for Eliciting Probabilities," *Econometrica*, 77(2), March 2009, 603-606.

Karni, Edi, and Safra, Zvi, "The Impossibility of Experimental Elicitation of Subjective Probabilities," *Theory and Decision*, 38, 1995, 313-320.

Klibanoff, Peter; Marinacci, Massimo, and Mukerji, Sujoy, "A Smooth Model of Decision Making Under Ambiguity," *Econometrica*, 73(6), November 2005, 1849-1892.

Köszegi, Botond, and Rabin, Matthew, "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).

Marschak, Jacob, "Actual Versus Consistent Decision Behavior," *Behavioral Science*, 9, April 1964, 103-110.

McDaniel, Tanga M., and Rutström, E. Elisabet, "Decision Making Costs and Problem Solving Performance," *Experimental Economics*, 4(2), October 2001, 145-161.

McKelvey, Richard D., and Page, Talbot, "Public and Private Information: An Experimental Study of Information Pooling," *Econometrica*, 58(6), November 1990, 1321-1339.

Nau, Robert F., "Extensions of the Subjective Expected Utility Model," in Ward Edwards, Ralph Miles, Jr., and Detlof von Winterfeldt (eds.), *Advances in Decision Analysis: From Foundations to Applications* (New York: Cambridge University Press, 2007).

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

Oehlert, Gary W., "A Note on the Delta Method," *The American Statistician*, 46(1), February 1992, 27-29.

O'Hagen, Anthony; Buck, Caitlin E.; Daneshkhah, Alireza; Eiser, J. Richard; Garthwaite, Paul H.; Jenkinson, David J.; Oakley, Jeremy E., and Rakow, Tim, *Uncertain Judgements: Eliciting Experts' Probabilities* (Hoboken, NJ: Wiley, 2006).

Offerman, Theo; Sonnemans, Joep, and Schram, Arthur, "Value Orientations, Expectations and Voluntary Contributions in Public Goods," *Economic Journal*, 106, 1996, 817-845.

Offerman, Theo; Sonnemans, Joep; van de Kuilen, Gijs, and Wakker, Peter P., "A Truth-Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes," *Review of Economic Studies*, 76(4), 2009, 1461-1489.

Qiu, Jianying, and Steiger, Eva-Maria, "On the Impact of Risk Attitudes on Belief Updating: An Experimental Study," *Working Paper*, Department of Economics, University of Innsbruck, March 2010.

Quiggin, John, "A Theory of Anticipated Utility," *Journal of Economic Behavior & Organization*, 3(4), 1982, 323-343.

Quiggin, John, *Generalized Expected Utility Theory: The Rank-Dependent Model* (Norwell, MA: Kluwer Academic, 1993).

Rutström, E. Elisabet, and Wilcox, Nathaniel T., "Stated Beliefs Versus Empirical Beliefs: A Methodological Inquiry and Experimental Test," *Games and Economic Behavior*, 67(2), November 2009, 616-632.

Saha, Atanu, "Expo-Power Utility: A Flexible Form for Absolute and Relative Risk Aversion," *American Journal of Agricultural Economics*, 75(4), November 1993, 905-913.

Savage, Leonard J., "Elicitation of Personal Probabilities and Expectations," *Journal of American Statistical Association*, 66, December 1971, 783-801.

Savage, Leonard J., *The Foundations of Statistics* (New York: Dover Publications, 1972; Second Edition).

Schotter, Andrew, and Sopher, Barry, "Social Learning and Coordination Conventions in Intergenerational Games: An Experimental Study," *Journal of Political Economy*, 2003, 111(3), 498-529.

Shephard, Glenn G., and Kirkwood, Craig W., "Managing the Judgemental Probability Elicitation Process: A Case Study of Analyst/Manager Interaction," *IEEE Transactions on Engineering Management*, 41(4), November 1994, 414-425.

Smith, Vernon L., "Measuring Nonmonetary Utilities in Uncertain Choices: the Ellsberg Urn," *Quarterly Journal of Economics*, 83(2), May 1969, 324-329.

Wakker, Peter, *Prospect Theory: For Risk and Ambiguity* (New York: Cambridge University Press, 2010).

Wilcox, Nathaniel T., "Stochastic Models for Binary Discrete Choice Under Risk: A Critical Primer and Econometric Comparison," in J. Cox and G.W. Harrison (eds.), *Risk Aversion in Experiments* (Bingley, UK: Emerald, Research in Experimental Economics, Volume 12, 2008).

Wilcox, Nathaniel T., "'Stochastically More Risk Averse:' A Contextual Theory of Stochastic Discrete Choice Under Risk," *Journal of Econometrics*, 2010 forthcoming; available as doi:10.1016/j.jeconom.2009.10.012.

Winkler, Robert L., and Murphy, Allan H., "Nonlinear Utility and the Probability Score," *Journal of Applied Meteorology*, 9, 1970, 143-148.

Appendix A: Instructions (NOT FOR PUBLICATION)

We provide complete instructions for the introduction to the session (marked I in the top, right corner of the first page of instructions, the quadratic scoring rule task trainer (marked q), the linear scoring rule trainer (marked L), the actual belief elicitation tasks for which subjects are paid (marked sr100), and then the lottery choice tasks (marked LOT). Copies of the exact instructions, which were printed in color, are available on request. Each subject received either the QSR or the LSR instructions.

Welcome! Today we will be asking you several types of questions. Some of these will earn you cash, which we will pay you today. And some may earn you cash which we will pay you in a few weeks. You have already earned $5 for showing up and agreeing to participate.

There are basically four stages today:

1.   We will ask you a series of questions about yourself, such as some basic information about your age. The computer will prompt you for these questions, and you should just work through them at your own pace when we log you in.

2.   We will then pause, and provide some instructions on the next task, which involves you making some judgements about what someone in a picture is thinking. We will explain that task when we come to it.

3.   We will then pause, and provide more instructions on the next task, which will involve you placing some bets on things that have yet to happen. In this stage we will take small breaks between the bets you place, so that we may explain the next specific thing that you are to bet on. These choices will directly affect your earnings. Nothing comes out of your own pocket.

4.   We will then pause, and provide more instructions on some choices you are to make over different amounts of money that have different chances of occurring. These choices will also directly affect your earnings.

The instructions for the second, third and fourth stage will provide more information on the type of choices you are being asked to make.

The experimenters will then collate all of your earnings and pay you for the money you have earned, as well as provide a receipt for any earnings that will be paid in the future.

Your choices are private, and will only be associated with an ID that we will enter when we log you in to the computer. So your name, address and SSN will not be linked to any choices you make. We will pay you privately, one at a time, at the end to keep your earnings private.

Are there any questions? If not, go ahead and answer the questions until the computer pauses and asks for a password. When everyone is finished this stage we will announce the password and we can go on to the second stage. There is no hurry, so take your time.
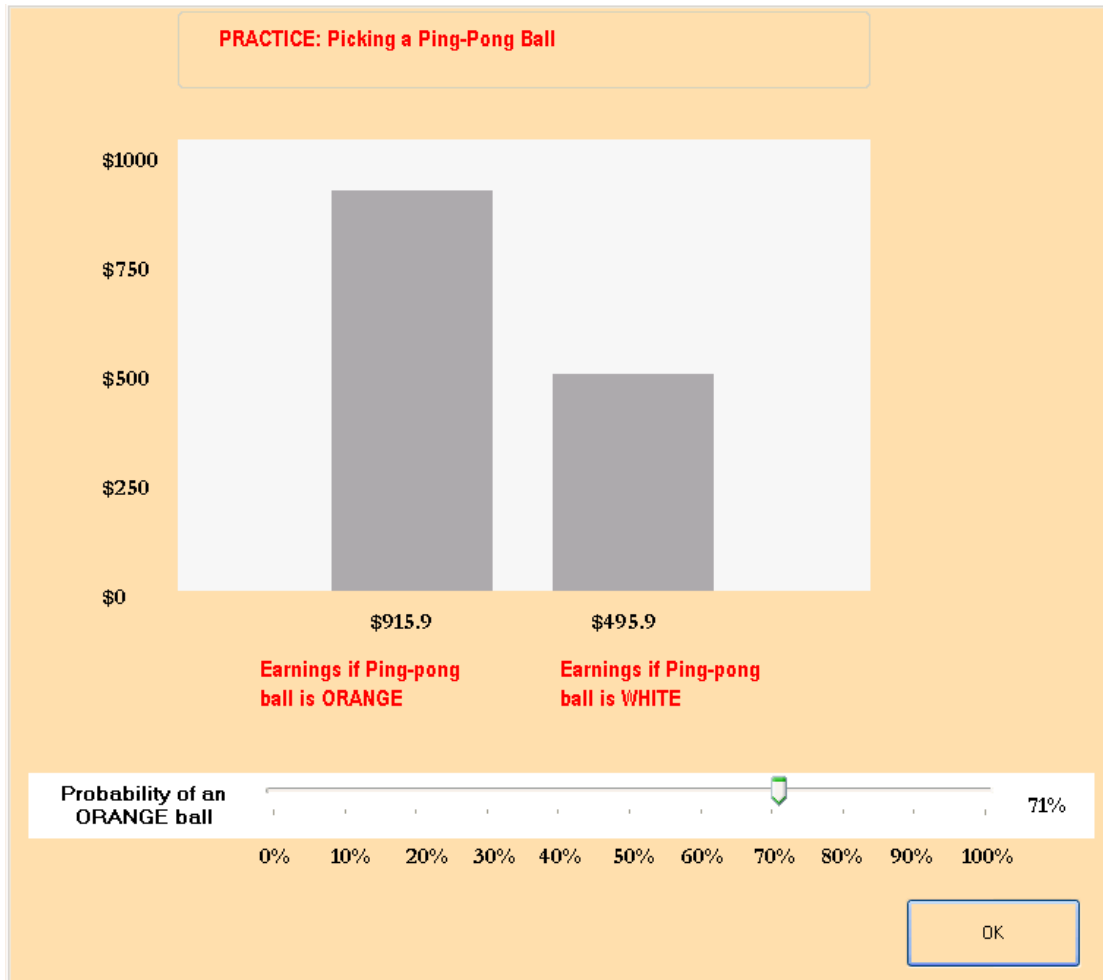
YOU MAY NOW PROCEED WITH THE FIRST STAGE.

In this stage we will give you tasks where you will place bets on the outcome of events that will happen today or in the future. For example, who will be the next U.S. President? You can make more money the more accurately you can predict these outcomes.

You place these bets on a screen like the one below. In a moment we will let you practice with this screen on your computer. Remember, any betting you do today is with our money, not your money.

You place your bets by sliding the bar at the bottom of the screen until you are happy with your choice of a probability report. The computer will start at some point on this bar at random: in the above screen it started at 71%, but you are free to change this as much as you like. In fact, you should slide this bar and see the effects on earnings, until you are happy to confirm your choice.

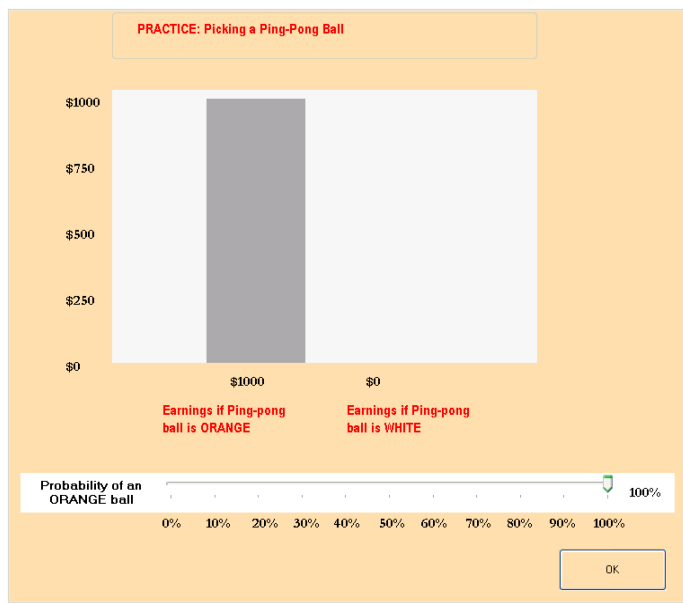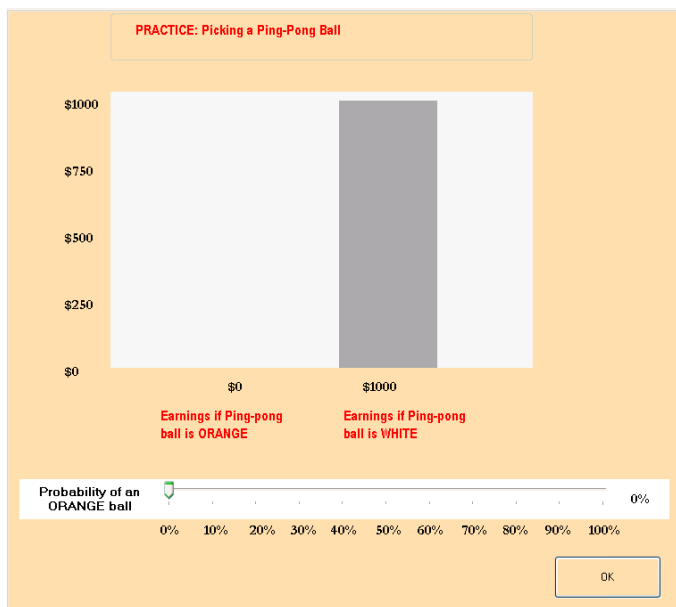ENTER THE PASSWORD THAT IS BEING ANNOUNCED NOW.

In this hypothetical example the maximum payoff you can earn is $1,000. In the actual tasks the maximum payoff will be lower than that, and we will tell you what it is when we come to those tasks. But the layout of the screen will be the same.

In this demonstration, the event you are betting on is whether a Ping Pong ball drawn from a bingo cage will be Orange or White. We have a bingo cage here, and we have 20 ping-pong balls. 15 of the balls are white, and 5 of the balls are orange. We will give the cage a few turns to scramble them up, and then select one ball at random.

What we want you to do is place a bet on which color will be picked. At the top of your screen we tell you what the event is: in this case, it is **Picking a Ping-Pong Ball**, and you need to bet on whether you think it will be **Orange or White**.

Your possible earnings are displayed in the two bars in the main part of the screen, and also in numbers at the bottom of each bar. For example, if you choose to report 71% you can see that you would earn $915.90 if the Ping Pong Ball was ORANGE, and $495.90 if the Ping Pong Ball was WHITE.

Lets see what happens if you make different reports. If you chose to report 0% or 100% here is what you would see, and earn:
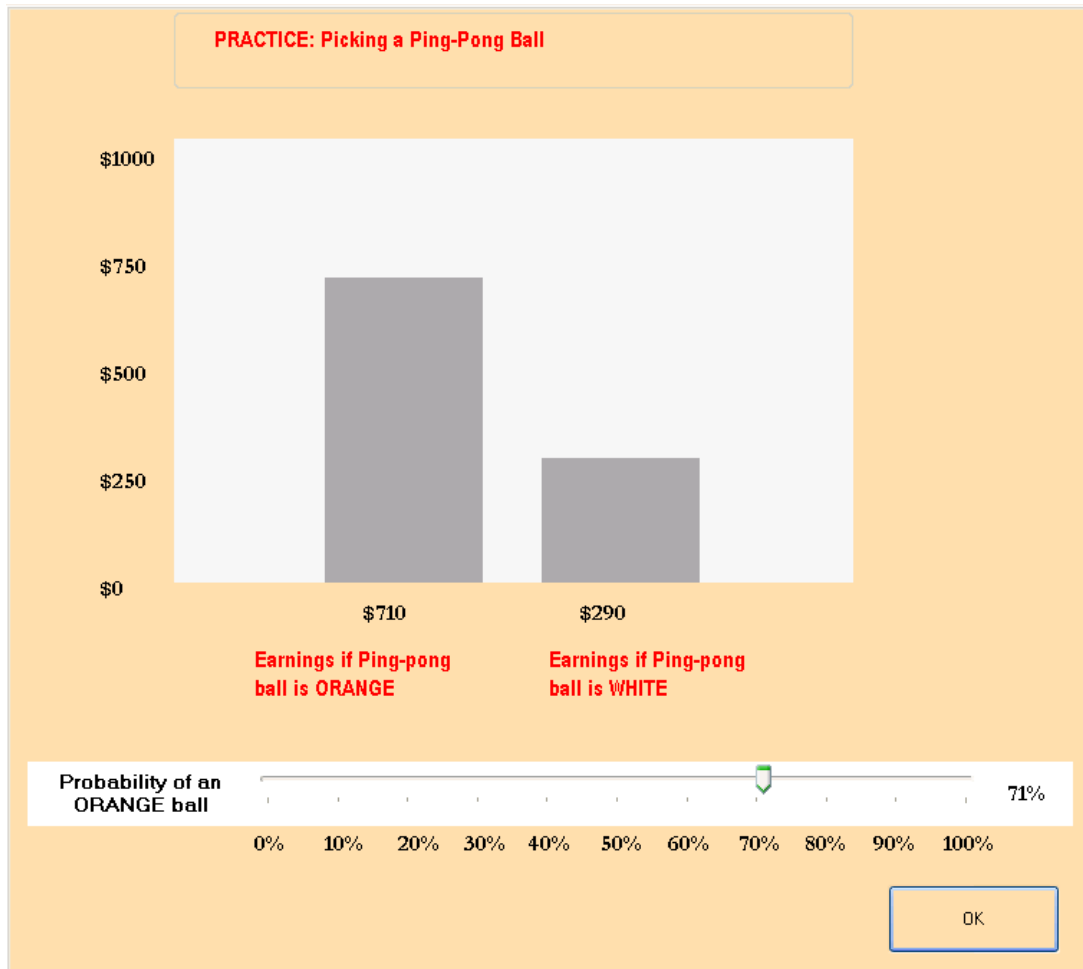

These screens are a little small, but you can see that these two reports lead to extreme payoffs. The "good news" is the possible $1,000 payoff, but the "bad news" is the possible $0 payoff. In between the reports of 0% and 100% you will have some positive payoff no matter what happens, but it will vary, as you can see from the report of 71%.

In this stage we will give you tasks where you will place bets on the outcome of events that will happen today or in the future. For example, who will be the next U.S. President? You can make more money the more accurately you can predict these outcomes.

You place these bets on a screen like the one below. In a moment we will let you practice with this screen on your computer. Remember, any betting you do today is with our money, not your money.

PRACTICE: Picking a Ping-Pong Ball

$710 — Earnings if Ping-pong ball is ORANGE

$290 — Earnings if Ping-pong ball is WHITE

Probability of an ORANGE ball — 71%

You place your bets by sliding the bar at the bottom of the screen until you are happy with your choice of a probability report. The computer will start at some point on this bar at random: in the above screen it started at 71%, but you are free to change this as much as you like. In fact, you should slide this bar and see the effects on earnings, until you are happy to confirm your choice.

ENTER THE PASSWORD THAT IS BEING ANNOUNCED NOW.

In this hypothetical example the maximum payoff you can earn is $1,000. In the actual tasks the maximum payoff will be lower than that, and we will tell you what it is when we come to those tasks. But the layout of the screen will be the same.

In this demonstration, the event you are betting on is whether a Ping Pong ball drawn from a bingo cage will be Orange or White. We have a bingo cage here, and we have 20 ping-pong balls. 15 of the balls are white, and 5 of the balls are orange. We will give the cage a few turns to scramble them up, and then select one ball at random.

What we want you to do is place a bet on which color will be picked. At the top of your screen we tell you what the event is: in this case, it is **Picking a Ping-Pong Ball**, and you need to bet on whether you think it will be **Orange or White**.

Your possible earnings are displayed in the two bars in the main part of the screen, and also in numbers at the bottom of each bar. For example, if you choose to report 71% you can see that you would earn $710 if the Ping Pong Ball was ORANGE, and $290 if the Ping Pong Ball was WHITE.

Lets see what happens if you make different reports. If you chose to report 0% or 100% here is what you would see, and earn:
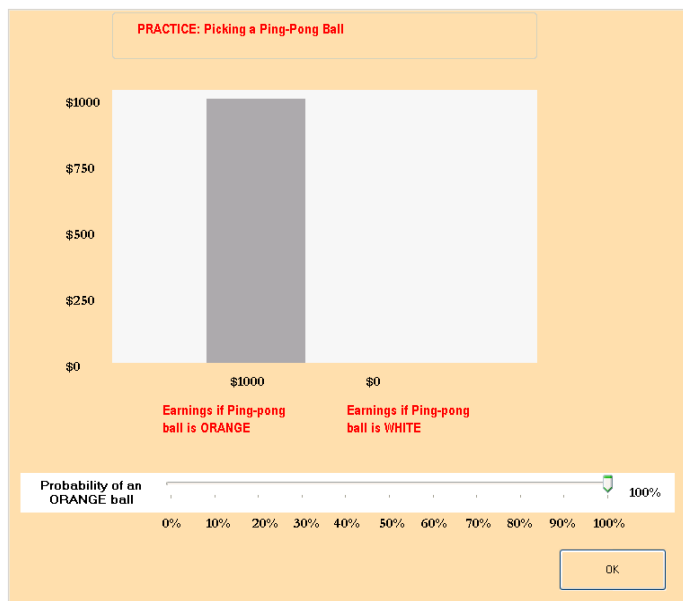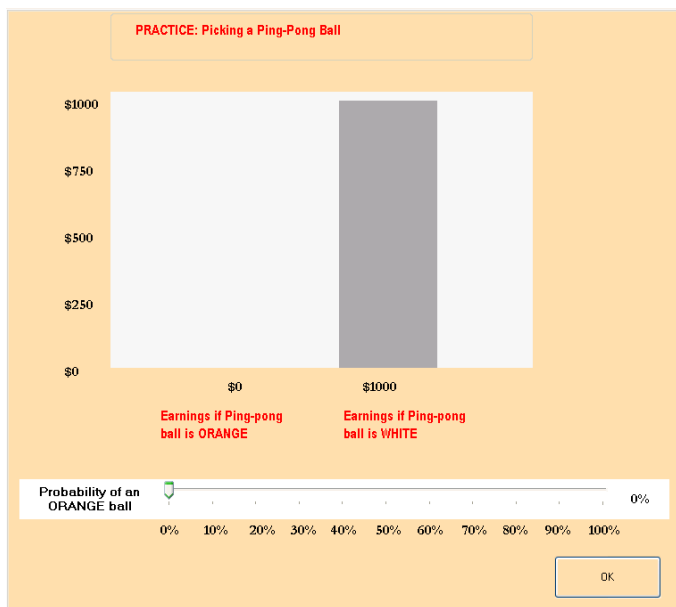
These screens are a little small, but you can see that these two reports lead to extreme payoffs. The "good news" is the possible $1,000 payoff, but the "bad news" is the possible $0 payoff. In between the reports of 0% and 100% you will have some positive payoff no matter what happens, but it will vary, as you can see from the report of 71%.

Summarizing, then, there are two important points for you to keep in mind when placing your bets:

1. **Your belief about the chances of each outcome is a personal judgment that depends on information you have about the different events.** In this practice example, the information you have consists of the total number of Orange balls and White balls.

2. **Your choices might also depend on your willingness to take risks or to gamble.** There is no right choice for everyone. For example, in a horse race you might want to bet on the longshot since it will bring you more money if it wins. On the other hand, you might want to bet on the favorite since it is more likely to win something.

For each task, your choice will depend on two things: your judgment about how likely it is that each outcome will occur, and how much you like to gamble or take risks.

       You will now make your report in this practice round. When you have chosen the report, confirm your bet by clicking on the OK tab.

       After you click OK, a special box will come up which causes the program to pause. We will do this after every series of bets, and then explain what the next few bets are about. We will tell you what the password is when we are all ready to proceed. There is plenty of time, so there is no need to rush.

       When everyone has placed their bets we will pick the ball and you will see what your earnings would have been if this had been for money. After that we will go on with the bets for which you can earn real money.

       Does anyone have any questions?

       ENTER THE PASSWORD THAT IS BEING ANNOUNCED NOW.

We are now ready to begin the choices for which you will be paid. There will be several sets of choices. In each case we will describe the bet to you, and then you can place your bets. Then we will explain the next couple of bets to you, and you place those bets.

Some of these bets will be about outcomes we know today, here and now, and some will be about outcomes we will only know in a few weeks. There will be 7 bets in all. We will pay you for one of these 7 bets. We will pick this bet at random after all bets are made, and tell you which one will be paid. You should view each bet as if it could be the one to determine your payoffs, since one of them actually will.

The maximum payoff for your bets today will be $100.

Ping Pong Balls Again

We will now repeat the task with Ping Pong balls a few times.

We have a number of ping pong balls in each of three bingo cages, which we have labeled Cage A, Cage B and Cage C. Some of the ping pong balls are Orange and some are White. We will roll each bingo cage and you can decide for yourself what fraction of Orange balls you think are in the cage. Of course, the balls will be rolling around, and you may not be able to tell exactly how many Orange balls are in the cage. You will be asked to bet on the color of one ping pong ball, selected at random after you all place your bets. For example, if there are 20 Orange balls and 80 White balls, the chance of an Orange ball being picked at random is 20 ÷ 100, or 20%.

We will do this task 3 times, with 3 different bingo cages. Just be sure that you check which cage you are placing a bet on. You can see this listed in the top left corner of your screen, where it refers to Cage A, Cage B or Cage C. We will show you each cage one at a time, and allow you to place your bets after we show it to you.

Do you have any questions?

The Eyes Test

At the beginning of today's experimental session we asked you to answer 36 questions, called **The Eyes Test**. These questions were designed by psychologists to measure a person's ability to "read someone else's mind" by just looking at their eyes.

Each and everyone of you were given the same 36 Eyes Test questions in today's experiment and a total score was recorded for each and every one of you in this experiment.

Now we come to the outcome we want you to place bets on in this portion of the experiment. We will pick one man and one woman in the room. Do you think the man who is selected will have a higher score on the Eyes Test than the woman who is selected?

After everyone in the experiment has made their bets for this event we will randomly select one man, and randomly select one woman from this experimental session. We will use the cards we collected, and sort then into one pile for men and one pile for women. Each pile will be shuffled, and one card drawn from each pile. We will then compare the score for the man that is drawn with the score for the woman that is drawn, and write these scores up on the board for you to see.

We therefore pose the following outcome for you to bet on now:

That the man we select at random will have a higher score

on the Eyes Test than the woman we select at random.

Do you have any questions?

2008 Presidential Elections

We want you to place bets on some questions about the U.S. Presidential Elections being held in a few weeks:

5.   Will the next President of the United States be a Democrat or a Republican?

6.   Will the popular vote for the winning candidate be 5 or more percentage points greater than the popular vote for the losing candidate?

7.   Will the popular vote for the winning candidate be 10 or more percentage points greater than the popular vote for the losing candidate?

It is important that you understand that the first question is about the outcome of the Electoral College vote, and not the popular vote. The popular vote is just the sum of all votes across the United States. We are only referring to the Presidential Election, and not to any other elections that might occur on the same day.

For the second and third question, we are asking if you think that the winner of the popular vote will beat the loser by 5 or 10 percentage points or more. For example, if the winner of the popular vote gets 51% of the vote and the loser gets 49%, then this

is a 2 percentage point difference. If the winner gets 53% and the loser gets 47%, then this is a 6 percentage point difference.

The election will be on Tuesday, November 4, 2008. To use a widely respected public source for the outcome, we will use the New York Times of Friday, November 7, 2008 as the official source used to determine your payoffs. In the event that there is a drawn out determination of the next President, such as in the 2000 election, we will delay payments until Inauguration Day, which is on January 20, 2009.

You will be paid for your bets in checks that will be mailed out on Monday, November 10, assuming we know who the next President will be at that time.

Please go ahead now and place your bets for this event, unless you have any questions.

This is the final stage of today's experiment. You will be asked to choose between lotteries with varying prizes and chances of winning. You will be presented with a series of lotteries where you will make choices between pairs of them. There are 45 pairs in the series. For each pair of lotteries, you should indicate which of the two lotteries you prefer to play. You will actually get the chance to play one of the lotteries you choose, and will be paid according to the outcome of that lottery, so you should think carefully about which lotteries you prefer.

Here is an example of what the computer display of such a pair of lotteries will look like. The display on your screen will be bigger and easier to read.

The outcome of the lotteries will be determined by the draw of a random number between 1 and 100. Each number between (and including) 1 and 100 is equally likely to occur. In fact, you will be able to draw the number yourself using two 10-sided dice.

In the above example the left lottery pays five dollars ($5) if the number on the dice that is rolled is between 1 and 40, and pays fifteen dollars ($15) if the number is between 41 and 100. The yellow color in the pie chart corresponds to 40% of the area and illustrates the chances that the number on the dice rolled will be between 1 and 40 and your prize will be $5. The black area in the pie chart corresponds to 60% of the area and illustrates the chances that the number on the dice rolled will be between 41 and 100 and your prize will be $15.

We have selected colors for the pie charts such that a darker color indicates a higher prize. White will be used when the prize is zero dollars ($0).

Now look at the pie in the chart on the right. It pays five dollars ($5) if the number on the dice rolled is between 1 and 50, ten dollars ($10) if the number is between 51 and 90, and fifteen dollars ($15) if the number is between 91 and 100. As with the lottery on the left, the pie slices represent the fraction of the possible numbers which yield each payoff. For example, the size of the $15 pie slice is 10% of the total pie.

Each pair of lotteries is shown on a separate screen on the computer. On each screen, you should indicate which of the lotteries you prefer to play by clicking on one of the three boxes beneath the lotteries. You should click the LEFT box if you prefer the lottery on the left, the RIGHT box if you prefer the lottery on the right, and the DON'T CARE box if you do not prefer one or the other.

You should approach each pair of lotteries as if it is the one out of the 45 that you will play out. If you chose DON'T CARE in the lottery pair that we play out, you will pick one using a 10-sided die, where the numbers 1-5 correspond to the left lottery and the numbers 6-10 to the right lottery.

After you have worked through all of the pairs of lotteries, raise your hand and an experimenter will come over. You will then roll two 10-sided die to determine which pair of lotteries that will be played out. You roll the die until a number between 1 and 45 comes up, and that is the lottery pair to be played. If you picked DON'T CARE for that pair, you will use the 10-sided die to decide which one you will play. Finally, you will roll the two 10-sided dice to determine the outcome of the lottery you chose.

For instance, suppose you picked the lottery on the left in the above example. If the random number you rolled was 37, you would win $5; if it was 93, you would get $15. If you picked the lottery on the right and drew the number 37, you would get $5; if it was 93, you would get $15.

Therefore, your payoff is determined by three things:

- by which lottery pair is chosen to be played out in the series of 45 such pairs using the two 10-sided die;
- by which lottery you selected, the left or the right, for that pair; and
- by the outcome of that lottery when you roll the two 10-sided die.

This is not a test of whether you can pick the best lottery in each pair, because none of the lotteries are necessarily better than the others. Which lotteries you prefer is a matter of personal taste. The people next to you will have different lotteries in front of them when you make your choices, and may have different tastes, so their responses should not matter to you. Please work silently, and make your choices by thinking carefully about each lottery.

All payoffs are in cash, and are in addition to the $5 show-up fee that you receive just for being here and any earnings from the previous stage.

We will now come around to your computer and get you started. When you are finished, please signal someone to come around to play out your lottery and record your earnings. As soon as you have finished the actual series, and after you have rolled the necessary dice, you will be asked to check with someone in the next room to make sure that your earnings sheet is complete, and then you will be paid. You are then free to go. Thanks again for your participation today!

Appendix B. Additional Statistical Results (NOT FOR PUBLICATION)

Table B1: Estimates of the Utility Function Under EUT

Variable	Description	Estimate	Standard Error	p-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
A. No Covariates						
r		0.30	0.03	<0.01	0.23	0.36
α		0.03	0.01	<0.01	0.02	0.04
μ		0.07	0.00	<0.01	0.07	0.08
B. Adding Covariates						
r						
Constant		0.23	0.06	<0.01	0.12	0.35
female	Female	0.14	0.07	0.04	0.01	0.28
Over22	Over 22 Years of age	0.09	0.10	0.37	-0.11	0.29
GPAhi	GPA great than 3.75	0.05	0.09	0.57	-0.13	0.23
Graduate	Graduate Student	-0.16	0.12	0.17	-0.40	0.07
α						
Constant		0.02	0.01	0.02	0.00	0.03
female	Female	0.02	0.01	0.18	-0.01	0.05
Over22	Over 22 Years of age	0.01	0.01	0.39	-0.02	0.04
GPAhi	GPA great than 3.75	0.00	0.02	0.96	-0.03	0.03
Graduate	Graduate Student	0.00	0.02	0.95	-0.03	0.03
μ		0.07	0.00	<0.01	0.07	0.08

Table B2: Estimates of the Utility Function Under RDU

Variable	Description	Estimate	Standard Error	p-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
A. No Covariates						
ŕ		0.49	0.05	<0.01	0.39	0.58
ά		0.07	0.02	<0.01	0.03	0.12
γ		0.72	0.06	<0.01	0.60	0.83
μ		0.08	0.00	<0.01	0.07	0.09
B. Adding Covariates						
ŕ						
Constant		0.45	0.07	<0.01	0.31	0.58
female	Female	0.08	0.10	0.93	-0.18	0.20
Over22	Over 22 Years of age	0.02	0.15	0.87	-0.28	0.33
GPAhi	GPA great than 3.75	0.07	0.10	0.48	-0.12	0.26
Graduate	Graduate Student	0.18	0.25	0.46	-0.30	0.67
ά						
Constant		0.05	0.02	0.04	0.00	0.09
female	Female	-0.01	0.04	0.82	-0.08	0.06
Over22	Over 22 Years of age	0.03	0.08	0.72	-0.12	0.18
GPAhi	GPA great than 3.75	0.07	0.07	0.30	-0.06	0.21
Graduate	Graduate Student	0.29	0.37	0.42	-0.42	1.01
γ						
Constant		0.72	0.07	<0.01	0.59	0.86
female	Female	0.15	0.10	0.15	-0.05	0.35
Over22	Over 22 Years of age	0.05	0.25	0.85	-0.45	0.54
GPAhi	GPA great than 3.75	-0.09	0.13	0.46	-0.34	0.16
Graduate	Graduate Student	-0.47	0.41	0.25	-1.28	0.33
μ		0.07	0.01	<0.01	0.06	0.08

Table B3: Detailed EUT Estimates for Bets on Win Margin Being 5% or More

Variable	Description	Estimate	Standard Error	p-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
r						
Constant		0.23	0.06	<0.01	0.12	0.35
female	Female	0.14	0.07	0.04	0.01	0.28
Over22	Over 22 Years of age	0.09	0.10	0.36	-0.11	0.30
GPAhi	GPA great than 3.75	0.05	0.09	0.59	-0.13	0.23
Graduate	Graduate Student	-0.17	0.12	0.17	-0.40	0.07
α						
Constant		0.02	0.01	0.02	0.00	0.03
female	Female	0.02	0.01	0.18	-0.01	0.05
Over22	Over 22 Years of age	0.01	0.01	0.41	-0.02	0.04
GPAhi	GPA great than 3.75	0.00	0.02	1.00	-0.03	0.03
Graduate	Graduate Student	0.00	0.02	0.97	-0.03	0.03
μ_{RA}	Fechner Error on Risk Tasks	0.07	0.00	<0.01	0.07	0.08
π						
Constant		-1.20	0.65	0.07	-2.47	0.08
female	Female	0.76	0.79	0.34	-0.80	2.31
Over22	Over 22 Years of age	-0.28	0.86	0.75	-1.97	1.42
GPAhi	GPA great than 3.75	0.32	0.46	0.49	-0.58	1.23
Graduate	Graduate Student	-0.48	1.55	0.76	-3.52	2.56
LSR	Linear Scoring Rule tasks	0.23	0.83	0.79	-1.40	1.85
μ_{BE}	Fechner Error on Belief Tasks	0.23	0.10	0.03	0.03	0.43