Eliciting Beliefs
Theory and Experiments

Andersen, Steffen; Fountain, John; Harrison, Glenn W.; Rutström, E. Elisabet

Department of Economics
Copenhagen Business School

Working paper 3-2009

ELICITING BELIEFS: THEORY AND EXPERIMENTS

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

Department of Economics -Porcelænshaven 16A, 1.fl. - DK-2000 Frederiksberg

Eliciting Beliefs: Theory and Experiments

by

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

October 2007

Working Paper 07-08, Department of Economics,
College of Business Administration, University of Central Florida, 2007

ABSTRACT. Subjective beliefs play a role in many economic decisions. There is a large theoretical literature on the elicitation of beliefs, 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 beliefs, but stresses the need to make strong auxiliary assumptions or "calibrating adjustments" to elicited reports in order to recover the latent belief. With some notable exceptions, the empirical literature seems intent on either making those strong assumptions or ignoring the need for calibration. We make three contributions to bridge this gulf. First, we offer a *general theoretical framework* in which the belief elicitation task can be viewed as an exchange of state-dependent commodities between two traders. Second, we provide a *specific elicitation procedure* which has clear counterparts in field betting environments, and that is directly motivated by our theoretical framework. Finally, we illustrate how one can *jointly estimate risk attitudes and subjective beliefs* using structural maximum likelihood methods. This allows the observer to make inferences about the latent subjective belief, calibrating for virtually any well-specified model of choice under uncertainty. We demonstrate our procedures with an experiment in which we elicit subjective probabilities over three future events and one fact.

† Department of Economics and Centre for Economic and Business Research, Copenhagen Business School, Copenhagen, Denmark (Andersen); Department of Economics, University of Canterbury, Christchurch, New Zealand (Fountain); and Department of Economics, College of Business Administration, University of Central Florida, USA (Harrison and Rutström). Harrison is also affiliated with the Durham Business School, Durham University, UK. E-mail contacts: sa.eco@cbs.dk, john.fountain@canterbury.ac.nz, gharrison@research.bus.ucf.edu and erutstrom@bus.ucf.edu. We thank the U.S. National Science Foundation for research support under grants NSF/IIS 9817518, NSF/HSD 0527675 and NSF/SES 0616746, and the Danish Social Science Research Council for research support under project #24-02-0124. We are grateful to Peter Wakker for many helpful discussions. All instructions and statistical code are available at the *ExLab* Digital Library at http://exlab.bus.ucf.edu.

Table of Contents

Subjective beliefs play a role in many economic decisions. In individual choice settings, any time the agent faces a decision in the presence of risk there is a role for subjective probabilities. Or we might be interested in knowing the forecast an individual makes of some event, since the event is important to others. In strategic settings, virtually all solution concepts hinge on assumptions about consistency of beliefs.

There is a large theoretical literature on the elicitation of beliefs, 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 beliefs, but stresses the need to make auxiliary assumptions or "calibrating adjustments" to elicited reports in order to recover the latent belief. One such assumption is risk neutrality, but there are many others that play a role in practically important elicitation settings as we see below. Unfortunately, none are obviously "weak" assumptions. With some notable exceptions, the empirical literature seems intent on either making those strong assumptions or ignoring the need for calibration.

We make three contributions to bridge this gulf. First, we offer a *general theoretical framework* in which the belief elicitation task can be viewed as an exchange of state-dependent commodities between two traders. This framework is implicit in some of the literature, but serves the important role of unifying the analysis of several important elicitation procedures, such as the Quadratic Scoring Rule (QSR) for individuals and "prediction markets" for groups. It also motivates our second contribution, a *specific elicitation procedure* which has clear counterparts in field betting environments. The procedure also bears a family resemblance to applications of "multiple price list" methods that have become widely adopted in experimental economics for the elicitation of risk attitudes, discount rates and product valuations. We also use the general theoretical framework to point to extensions of the specific procedures we adopt, to better understand differences between different elicitation procedures. Finally, we illustrate how one can *jointly estimate risk attitudes and* *subjective beliefs* using structural maximum likelihood methods. This allows the observer to make inferences about the latent subjective belief, calibrating for virtually any well-specified model of choice under uncertainty. Thus one is not restricted to analytically tractable functional forms, or even to Expected Utility Theory (EUT).[1]

In section 1 we explain the basic elicitation procedure and some variants. In section 2 we provide a formal theoretical framework to understand the procedure, and how it connects with other elicitation procedures that have been proposed. In section 3 we develop an experimental design to examine the behavioral properties of the specific procedure we illustrate. We apply it to elicit subjective beliefs in mid-2006 that Donald Rumsfeld would resign by the end of the year (he did), Tiger Woods would win the next U.S. Open golf championship (he did not), that Brazil would win the World Cup soccer tournament (they did not), and that Madrid was the capital of Portugal (no, but close). In section 4 we present the results of some laboratory experiments testing the procedure, focusing on the joint econometric estimation of subjective beliefs and a model of choice under uncertainty.

We find that inferences about subjective beliefs are fragile, in the constructive sense that one has to commit to some model of decision making under risk before recovering latent subjective beliefs. The need for such corrections has been well-known for decades, but ignored in practical applications. We believe that the time is nigh for a re-balancing of the experimental and econometric procedures used in belief elicitation to better address the theoretical issues raised long ago. These results also have serious implications for ongoing debates over whether open prediction markets recover aggregate beliefs that represent anything useful.

---

[1] When we use the expression "risk attitudes," it should be understood that it is only synonymous with properties of the curvature of the utility function if one assumes EUT. In popular alternative models it is also affected by the possibility of probability weighting and loss aversion, and our approach allows for these possibilities if they are present. In fact, we undertake explicit corrections for probability weighting in §3.

Several appendices provide detailed discussion on select topics. Appendix A presents a formal mathematical statement of the general approach proposed in section 2, as well as additional implications of that approach. Appendix B contains the complete instructions provided to experimental subjects. Appendix C discusses extensions to our approach, building on the specific experimental procedure we use and the insights from the general theoretical framework. We view these extensions as mapping out a research program that can also build a bridge between the elicitation of individual beliefs in experimental settings and the elicitation of aggregate beliefs in market settings. Appendix D draws additional connections with the vast literature in this area, connecting our approach to antecedents.

The reader interested in experimental procedures might skip section 2 initially, and return to it to see where our procedures fit into the broader approaches to belief elicitation. The reader interested in adjustments to elicited beliefs to "correct" for risk attitudes might skip sections 2 and 3 initially. The econometric methodology for recovering latent subjective beliefs is equally applicable to alternative methods for eliciting beliefs, such as scoring rule procedures.

Our approach brings together ideas that have been presented in different parts of several literatures. Appendix D documents these connections in more detail.

The notion that 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 by bookies and parimutuel markets (Epstein [1977; p. 298ff.]). The central insight is that a subjective probability is a marginal rate of substitution between contingent claims, where the contingency is the event that the probability refers to. There are then a myriad of ways in which one can operationalize this notion of a marginal rate of substitution.[2]

The formal link between scoring rules and optimizing decisions by agents is also familiar, particularly in Savage [1971], Kadane and Winkler [1987][1988] and Hanson [2003]. Jose, Nau, and Winkler [2007] stress the interpretation of several popular scoring rules from the perspective of an expected utility maximizing agent with preferences derived from familiar utility functions. Their approach may be viewed as complementary to ours: if one knows the utility function of the agent, they show which scoring rule is incentive compatible. 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.

Finally, the constructive task of "recovering" subjective beliefs from observed betting behavior or responses to scoring rules has been directly attacked by Offerman, Sonnemans, van de Kuilen and Wakker [2007; §6]. Like us, they consider the recovery of true subjective beliefs when the agent may be risk averse in the narrow sense of EUT, as well as the broader sense implied by an allowance for probability weighting. Their approach has a reduced form simplicity, and is actually agnostic about which structural model of decision making under risk one uses. Our approach is explicitly structural, and generates inferences about subjective beliefs that are conditional on the assumed model of decision making under risk. We see these as complementary approaches, and both have strengths and weaknesses.

Our statistical approach uses the notion of joint estimation of preference parameters from several complementary experimental tasks. Intuitively, a series of lottery tasks with objective

---

[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 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 problems. But the basic idea of the method could also be applied to elicit probability intervals, in much the same manner that we elicit probability intervals.

probabilities allow us to estimate the parameters of a latent model of decision making under risk, which then allows us to identify subjective beliefs from a belief betting task. These parameters and subjective beliefs are estimated using joint maximum likelihood methods, such as applied in other experimental settings by Andersen, Harrison, Lau and Rutström [2005]. One insight from this joint estimation approach is that uncertainty about the core parameters of utility functions and/or probability weighting functions affects inferences that can be made about subjective beliefs, as they should. The estimation of subjective beliefs is not a simple matter.

1. An Elicitation Procedure

A. The Basic Idea

Assume for the moment that the subject is risk neutral and does not distort subjective probabilities when evaluating lotteries.[3] The subject is asked to state their belief that event A will occur instead of event B, where A and B are mutually exclusive. Assume that the subject has no stake in whether A or B occurs other than the bets being made with these bookies.[4] Let the subject be told that there are nine bookies willing to take a bet at stated odds. Table 1 shows the house probabilities of each outcome and the associated house odds. The odds are stated in columns 3 and 4 in the *form* that they are naturally stated in the field: what is the amount that the subject would get for a $1 bet if the indicated event occurred? In columns 1 and 2 the same odds are represented in

---

[3] Depending on how the subject perceives the task, one might also want to adjust choices for loss aversion. If the subject bets on the wrong horse, they lose their stake. Thus in Table 1 there is always one negative amount in the "A wins" and "B wins" columns of earnings. It is arguable that when the subject if forced to bet on each bookie, and the stake is provided by the experimenter, that there is no loss involved, since the subject simply does not earn a positive amount. But when we allow the subject the option of not betting on a particular bookie, and keeping the stake, the prospect of a loss may trigger loss aversion. In effect this is a simple extensions of the type of calibration required for risk aversion or probability weighting.

[4] This corresponds to the "no stakes" condition stressed by Kadane and Winkler [1986][1988]. It is redundant when subjects are risk neutral, but essential when one deviates from that special case.

terms of an implied probability of the event. Thus the odds of A occurring are simply the reciprocal of the probability of A, so a probability of 0.1 implies odds of $10: a $1 bet would return a gross payout of $10 and a net payout of $9. Some people (academics) find probabilities easier to understand, and some people (everyone else) seem to find odds easier to understand, so we provide both.[5]

In our base design[6] the subject is simply asked to decide how they want to bet with each of the nine bookies: do they want to bet on A or B? Their "switch point," over the nine bookies, is then used to infer their subjective belief. Our basic procedure has a striking familiarity with "multiple price list" procedures for the elicitation of risk attitudes (Holt and Laury [2002]), discount rates (Harrison, Lau and Williams [2002]) and valuations for goods (Andersen et el. [2007]).

Consider a subject that has a personal belief that A will occur with probability ¾. Assume that the subject *has* to place a bet with each bookie, knowing that only one of these books will actually be played out. The odds offered by a particular bookie are shown on a given row, so

---

[5] One difference between our approach and the field counterparts is that we only state "competitive odds" in which the betting house "take" is zero if there are equal numbers of bettors on each side of the market. To see the parallel to field betting odds, it is worthwhile translating field odds that include a house take into implicit beliefs. Essentially, one has to define the exhaustive set of events, calculate the implied house probability of all events, and normalize the probabilities by that sum. Consider the odds quoted in May 2006 by one betting house, William Hill, for the winner of Group B of the 2006 World Cap. These were England at 1.57, Sweden at 3.00, Paraguay at 8.00 and poor old Trinidad & Tobago at 34.00. The reciprocal of these odds on wagers of 1 unit are 0.64, 0.33, 0.13 and 0.03, which sums to 1.125. So the house take here is 12.5%: if it ended up with a "balanced book" in which wagers were placed on each outcome in inverse proportion to odds, it would earn $1.125 for every $1 bet *with no risk*. We can take the house take away, and re-normalize the implied probabilities of each outcome to be 0.57 (= 0.64 ÷ 1.125), 0.30, 0.11 and 0.03. This adjustment does not much help the chances of the hapless Trinidad & Tobago team, but it does make a noticeable change for the England and Sweden outcomes. With this adjustment, our experimental instrument can be extended to reflect practice in the field even more closely, and we can still infer house probabilities.

[6] We discuss variants throughout, and in particular in Appendix C. In one variant they can also decide what fraction of a $10 stake they would like to bet, and then whether they want to bet on A or B. In a more complex variant we elicit the subject's true belief about the event A and, using this information, provide the subject with their expected income. This information is designed to show the subject some of the consequences of their choice, so that we are not jointly eliciting beliefs and testing cognitive prowess at such calculations.

different rows correspond to different bookies. The subject would rationally bet on A for every bookie offering odds that corresponded to a lower house probability of A winning than ¾, and then switch over to bet on B for every bookie offering odds that corresponded to a higher house probability of A winning than ¾. These bets are shown in Table 1, and imply net earnings of $9 or -$1 with the first bookie, $4.00 or -$1 with the second bookie, and so on.[7] The expected net earnings from each bookie can then be calculated using the subjective belief of ¾ that the subject stated. Hence the expected net earnings from the first bookie are (¾ × $9) + (¼ × -$1) = $6.50, and so on for the other bookies. Average expected income from bets optimally laid is $1.58 = $14.20 ÷ 9.

This logic is illustrated in panel A of Figure 1. This shows the certainty equivalent of the optimal bet for this risk neutral subject for each of the house odds listed on the horizontal axes. The odds and probabilities listed across the horizontal axis correspond to the odds and probabilities listed down the rows of Table 1. One line shows the outcome if there is a bet on A, and the other line shows the outcome if there is a bet on B. At the probability ¾ these two lines cross, reflecting the fact that the subject is indifferent between betting on A and B when visiting the bookie offering exactly this house probability. For lower house probabilities the subject optimally bets on A instead of B, and for higher house probabilities the subject optimally bets on B instead of A. Thus we can directly infer the subjective probability from the point at which the subject switches bets.

The flat line at $1 in panel A of Figure 1 corresponds to the certainty-equivalent of the stake. For now it is simply a reference point, to see that the subject can expect to increase his expected income over the stake by placing the optimal bet. Since we force the subject to place bets on each and every bookie, it does not represent a choice for the subject. Since the subject is assumed to be risk neutral, it would not affect betting behavior even if it was an option: as long as the experimental

[7] The net earnings are the payoffs that our experimental subjects would earn, since their $1 stake is given to them by the experimenter. In the field the better would receive the gross payoffs.

subject makes has a positive expected net earnings from a bet in each row, the subject would place a bet even if they had the opportunity not to.

When we allow for risk attitudes or probability distortion, we need to adjust the observed response to correct for the effect that these behavioral characteristics imply on optimal betting behavior that differs from the risk neutral prediction. In effect, we have to make a joint inference about the subjective probability, risk attitude, and probability weighting of the subject.[8]

B. Correcting for Risk Aversion

Consider a subject who is risk averse, again holding a subjective belief of ¾ that A will occur. Assume further that subject has a Constant Relative Risk Aversion (CRRA) utility function defined by

$$U(y) = (y^{1-r})/(1-r), \tag{1.1}$$

where r is the CRRA coefficient and is set equal to 0.57 for this illustration.[9] Assume further, for now, that the subject behaves consistently with EUT; our approach to calibration generalizes to non-EUT models, as illustrated later. Armed with these parametric assumptions we can recalculate the payoffs in certainty equivalent terms for binary choice the subject can take: bet on A or bet on B.

[8] The same methodological point about the need for joint inference arises in other settings. Harrison and Rutström [2008; §3.1] demonstrate that statements about the curvature of the utility function in the multiple price list task of Holt and Laury [2002] depend on what is assumed about the extent of probability weighting. If it is assumed away, then one can infer that curvature directly from observed behavior, as illustrated by the CRRA intervals reported by Holt and Laury [2002; Table 3, p.1649]. Andersen et al. [2005] demonstrate that statements about the discount rate over utility streams in the multiple price list task of Harrison, Lau and Williams [2002] depend on what is assumed about the curvature of the utility function. If it is assumed to be linear, then one can infer discount rates directly from observed behavior (with some additional assumptions), as illustrated by Coller and Williams [2002; Table 1, p.1610].

[9] The value of 0.57 is the average predicted value from experiments reported in Harrison, Johnson, McInnes and Rutström [2005] for U.S. college students, and is consistent with estimates obtained by many other experimenters with comparable samples and stakes (e.g., Holt and Laury [2002]). Unless otherwise noted, references to "risk attitudes" are synonymous with references to the concavity or convexity of the utility function.

The new certainty equivalents that allow for risk aversion are shown in panel B of Figure 1. It is clear that this risk averse subject will bet on A at all house probabilities in the discrete range 0.1,0.2,......0.9 when she is forced to bet her $1 stake on either A or on B, since the certainty equivalent of the risky $1 bet on A is always higher than the certainty equivalent for the risky $1 bet on B for these house probabilities. Thus the *observed behavior of a subject that is risk neutral is very different from the observed behavior of a subject that holds exactly the same subjective belief but is risk averse.* One cannot recover the subjective belief without correcting observed behavior for the risk attitudes of the subject. Since we have forced bets, this risk averse agent bets in a manner that would be incorrectly interpreted as implying a *higher* subjective probability than ¾ *if we incorrectly assumed he was risk neutral.*

Panel B of Figure 1 also makes it clear that these forced risky bets are worse than not betting at all whenever house probabilities are in the range 0.52 to 0.94. The ability not to bet, or at least not to bet very much, matters to this risk averse subject. As would be expected for a risk averse subject, and unlike the case for a risk neutral subject, when the house probabilities are close to the subject's beliefs, not much betting will take place when the subject has the option not to.

The absence of betting, when it is an option for the subject, is worth pausing over since it is a key feature for all prediction markets open to the public.[10] In our elicitation mechanism we are still able to recover the true belief of the subject, since that was needed to identify which bookies were worth betting with, and assuming some relevant variation in the odds offered by different bookies. But if we simply observed the bets of the subject, what would we infer? We could infer that the subject believed that A would occur with probability greater than 0.5 and less than 0.9, which is not

[10] Although rare, it is perfectly possible to conduct prediction markets for a "closed" set of participants who have no option but to place bets with funds provided by the experimenter solely for that purpose: see Gruca, Berg and Cipriano [2003], for example.

particularly informative.[11] It is not a wrong inference, since the true belief is ¾ by assumption, but it is not very precise. Moreover, the sample selection issues that occur as one moves from "forced betting" to "free range betting" are suggestive of the deeper difficulty of making inferences about true beliefs from market odds based solely on realized bets, since this subject would not have had any impact on the market *within* this wide range.[12]

An additional implication of allowing for the absence of betting concerns the temporal difference between the investment in the bet (today) and the expected payoff (tomorrow). When subjects place a bet *today* they give up the stake: virtually every betting house requires that bets be covered immediately with transfers of funds. But the payout is defined in terms of *future* dollars. So any bettor must discount the future payouts, conditional on outcome, to compare with the current outlay. This implies that individual discount rates, and their heterogeneity across traders, must play a role in inferring individual beliefs from naturally occurring bets.

C. Correcting for Probability Weighting

If the subject engages in probability weighting, the correction is a simple one. If we know the probability weighting function $w(\pi)$ that the subject uses, then we infer $w(\pi)$ and then simply apply the inverse function $w^{-1}(\pi)$ to recover π. Of course, as with the correction for risk aversion due to curvature of the utility function, we need to account for the fact that we will only be able to obtain estimates of $w(\pi)$, and that the standard errors in that estimation will propagate and generate uncertainty in our final inference about π. Nonetheless, this is just an extension of the notion that we have to recover the latent subjective belief from observed behavior, and need to make some

[11] One could imagine a longer list of bookies, arrayed in finer increments of odds, but this would not provide significant improvement on the possible inferences from the observed bets alone.

[12] If the free range option is not all or none, but allows partial bets out of the stake, then there would be some limited market activity in the range of house prices between 0.5 to 0.9.

theoretical and statistics assumptions to do so.

To anticipate an important formal issue, we stress that we are focusing here on decision makers that are "probabilistically sophisticated" in the sense of Machina and Schmeidler [1992; p.747]. Such decision makers are assumed to behave as if they employ subjective probabilities of events and utilities of outcomes that are independent of the assignment of outcomes to events. One important contribution of non-EUT models of decision making under risk is to relax that independence, but it is not the only contribution. We focus on the Rank-Dependant Utility (RDU) model as an alternative to EUT to illustrate the key ideas.

To be specific, one can state a probabilistically sophisticated version of RDU in which individuals have subjective beliefs π and then form decision weights from those beliefs using the probability weighting function $w(\pi)$ and some assumption about rank-dependent weighting of outcomes (e.g., Quiggin [1982]). But one can also state less restrictive versions of RDU in which decisions makers violate probabilistic sophistication (e.g., Schmeidler [1989]); Quiggin [1993; §5.7] reviews the differences. To be clear, we do not know how to meaningfully recover subjective beliefs when decision makers are *not* probabilistically sophisticated, but we are subjectively confident nobody else does as well.

D. Precision of Inferences about Beliefs

Recovering beliefs from observed behavior entails some imprecision to the extent that we only have estimates of the utility function and/or the probability weighting function of the agent. Even if we assume EUT and calculate a point estimate that the subject is risk neutral (r=0, as parameterized here), there will be *some* standard error around that point estimate and hence *some* imprecision about true beliefs. We just have to live with this imprecision, and do our best to recover what we can. The econometric implications of this point are spelled out in §3 with an illustrative application.

But there is another source of imprecision about subjective beliefs, due to the discrete nature of the opportunity set being presented to subjects. To see this, assume that we know with absolute precision that the agent is risk averse and has a CRRA utility function with r=0.57, and does not engage in probability weighting. Our house probabilities vary in increments of 0.1 between 0.1 and 0.9, and we assumed an agent with true beliefs ¾. This agent would bet on event A with every bookie. He would only switch to betting on event B if a bookie offered a house probability of A of 0.93 or better. But in our coarse grid we do not have a house probability greater than 0.9, so we would not be able to detect such a switch. Nor would we be able to tell this agent, with an assumed belief of ¾, apart from an agent with subjective beliefs between 0.73 and 1. In other words, agents with beliefs in the interval [0.73, 1] would be *observationally equivalent*. Hence, from the observation that the agent selected A from all bookies, and even with the absolutely precise knowledge that his r=0.57, we can only infer his true beliefs as coming from some interval.

The solution to this source of imprecision is to explore different opportunity sets for betting. Intuitively, one obvious solution is to concavify the opportunity set, which we current have as linear.[13] Another solution, perhaps less obviously, is to allow "free range" betting so that we obtain more information from the observed betting behavior about risk attitudes and hence inferred subjective beliefs. Another solution, even less obvious, is to allow agents to place bets on more than one bookie for payment. The evaluation of these solutions requires a more general framework, presented in §2, and connects our approach to the traditional literature on scoring rules. We also

[13] 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 linear scoring rules, and that 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.

show there how one can arbitrarily approximate the general opportunity sets with our simple "linear" experimental frame.

2. A Formal Model of Belief Elicitation as a Betting Game

This elicitation framework can be formally viewed from the perspective of a trading game between two agents. This framework provides several insights. First, traditional scoring rules can be interpreted consistently with each other, allowing comparisons of theoretical performance characteristics. Second, one can compare scoring rule procedures with prediction markets in a consistent framework, allowing insights from the former to inform debates on the proper interpretation of outcomes from the latter. Third, one can provide a theoretical micro-foundation for field counterparts of belief elicitation games that take the form of betting markets, as well as our laboratory implementation of those field procedures. Finally, the role of risk aversion and probability weighting can be identified systematically, providing a latent theoretical structure for our econometric estimation procedures in §4.

Our theoretical framework can be explained best by a series of visual examples.[14] We provide a more formal mathematical statement in Appendix A. We then present the special case implemented in our experiments. The relationship to the literature is discussed at some length in §6, apart from references to specific results in the literature along the way.

A. Basic Logic

Consider an agent that is given an endowment of tickets. Each ticket can be used to purchase state-contingent claim that pay $1 if either of states A or B occur, where B is the

[14] Alternative geometrical treatments of different scoring rules are provided by de Finetti [1962], Murphy and Staël von Holstein [1975] and Staël von Holstein and Murphy [1978].

complement of A. The standard quadratic scoring rule can be defined as follows. The subject receives a fixed amount Ω>0, from which is subtracted an amount equal to the square of error in the report. This amount is deflated by some scalar ω before being subtracted from Ω. If r_1 denotes the reported probability of state A occurring, the penalty is just $(1-r_1)^2$ if state A occurs and r_1^2 if state B occurs.[15]

Figure 2 displays an Edgeworth Box that represents the opportunity set that this subject faces if we confront him with a standard quadratic scoring rule penalty function. The vertical axis from the bottom-left origin shows the wealth that the agent ends up with if state B obtains, and the horizontal axis shows the wealth that the agent ends up with if state A obtains. In effect, the Edgeworth Box construct reminds us that *this elicitation task is essentially a trading game between the experimenter and the agent.* The size of this box represents the stake that the experimenter puts on the table, which we have normalized here to $1 in either state, but which could obviously vary up or down.

Now consider the beliefs and preferences of the agent. Start with an agent that is known to risk-neutral, loss-neutral, and does not distort probabilities. Further assume that our agent has true subjective beliefs that lead him to expect state A to occur with a 60% chance. Because the agent holds these beliefs, he is willing *at the margin* of his initial endowment of tickets to trade off state B tickets for state A tickets at a rate of 60-to-40, or 1½-to-1. And because the agent is assumed to be risk neutral, he only cares about the expected value of his wealth after the state is realized, so he does not care about the variance in his expected income. Thus his indifference curves, shown in Figure 3 on the same Edgeworth Box, are linear with a slope given by the ratio of his subjective beliefs. Each

[15] In general, if r_i denotes the reported probability of state *i*, and q_i denotes the true probability of state *i*, then the penalty is $(q_i - r_i)^2$. Staël von Holstein and Murphy [1978] provide a general characterization of the family of quadratic scoring rules, showing that many popular scoring rules can be seen as special cases.

indifference curve can be indexed by the expected value of wealth, where the expectation is taken with respect to the agent's true beliefs.

Faced with the opportunity set in Figure 2, and with these subjective beliefs and risk preferences, the optimal behavior for the agent is to report that state A will occur with 60% chance. We illustrate this result in Figure 4. The optimal outcome for the agent is shown as point x*, which generates an expected income of 64 cents if state B obtains and 84 cents if state A obtains. With the subjective probabilities of these being 0.6 and 0.4, the expected income to the agent is therefore 76 cents (= $\$0.64 \times 0.4 + \$0.84 \times 0.6$). This expected income maximizes expected utility for this agent. Any other report would be sub-optimal. For example, a report that state A would occur with 50% chance, when the agent truly believed that it would occur with 60% chance, would shift x* to the no-risk 45° line, which has an expected income of only 75 cents as shown with the "magnification glass" in Figure 5. Any false report would generate a lower expected income than the true report.

So far we have only provided an alternative, and familiar, characterization of properties of the quadratic scoring rule that are well known. But this characterization generalizes in nice ways, discussed later.

Recovering Beliefs from Elicited Reports

Now consider the effect of knowing that the agent has a particular risk attitude. Using the quadratic scoring rule, Figures 6, 7 and 8 overlay the outcome for an agent that has a CRRA coefficient equal to 0.57 for this illustration. The agent has the same beliefs, but optimally reports something other than those true beliefs. That is the bad news, although not disastrously so as we shall see, since we can calibrate the true beliefs if we have information about the subject's risk attitude. The good news is that the agent is more motivated than the risk neutral agent to provide that report. One way to see this last result is to observe in Figure 8 that the payoff function for the

CRRA agent is more concave than the corresponding payoff function for the risk neutral agent, where payoffs are measured in terms of certainty equivalent income.

The risk averse agent tends to report probabilities under the QSR that are closer to ½ than their true beliefs. The intuition here is simple enough, since the agent is putting less of their eggs in the two extreme event-outcome baskets. Thus, if we know that the agent is risk averse, we should take the observed reports are recover true, subjective beliefs as some probability that is more extreme than the report. This is the *opposite* of the adjustment needed in the procedure we introduced in §1, where we assumed that agents were forced to bet on one or other event outcome. The reconciliation comes from recognizing that the analysis of the QSR is in effect allowing the agent to pick any point on the opportunity set in Figures 6, 7 and 8, rather than the extreme end-points of each opportunity set. Our forced-fed treatment in §1 forces the agent to the end-points of (linear) opportunity sets, as we illustrate below in §2B.

We can generalize the approach to belief elicitation to virtually any well-behaved utility function that might be used to characterize risk attitudes. Furthermore, we can extend the analysis numerically to consider alternatives to EUT, such as RDU and CPT. This extension is best illustrated in §3 when we have more structure on the specific application we provide.

Improving Incentives to Report Accurately

This general framework allows one to easily "see" how to improve the strength of the incentives for reporting one's beliefs accurately. The opportunity set for the quadratic scoring rule is simply a particular specification of a CES indifference curve in a 2-person pure exchange game. The formal proof of this assertion is in Appendix A, but the basic idea is straightforward enough. A risk averse CRRA agent is assumed to hold beliefs b(s) about a finite list of n states $s \in S$, such that $b(s)>0$, $\sum_s b(s)=1$. The agent has a certainty equivalent (CE) wealth function:

$$CE = [\ \textstyle\sum_s b(s)\ x(s)^{(1-r)/r}\ ]^{1/(1-r)} \qquad (2.1)$$

This expression for the CRRA agent's CE is in standard CES form, where the elasticity of substitution $\eta = 1/r$ is the reciprocal of the Arrow-Pratt relative risk aversion parameter r, and is a measure of risk tolerance.

Consider the CRRA agent's optimal choice problem in a competitive contingent claims market where unit s-contingent claims can be bought or sold at prices $p(s)>0$, $\sum_s p(s)=1$. A unit of s-contingent wealth pays \$1 if state s occurs and \$0 otherwise, for all states s=1,2....n. The expenditure minimization problem for a CRRA agent in such a market is to minimize $m=\sum_s p(s)x(s)$ subject to a *certainty equivalent* constraint. This yields Hicksian compensated demand functions $x^c(s)$ for wealth in state s:

$$x^c(s) = CE \times \{[b(s)/(p(s)] \ / \ [\ \textstyle\sum_{\check{s}} [b(\check{s})/x(\check{s})]^{1-r/r}\ ]^{1/1-r}\ \}^{1/r} = CE \times \kappa \qquad (2.2)$$

for $\check{s} \in S$. Embed these compensated demands in an Edgeworth Box of size $\omega(s)$ so that the opportunity set faced by the subject is:

$$\omega(s) - x^c(s) = \omega(s) - (CE \times \kappa) \qquad (2.3)$$

The quadratic scoring rule is a special case of this CES form in which n=2 (so s={A, B}), b(s)=½ ∀s, CE=¼, r=½, $\omega(s)=1$ ∀s, and the subject's reports are the market prices p(s); in this case, b(s) is interpreted as a parameter rather than as true beliefs.

The generality of this CES-in-an-Edgeworth-box perspective provides the experimenter with great control over the shape and position of the incentives faced by the subject. There is, of course, no reason to restrict the analysis to a CES form: in principle, the compensated demands from any indirect utility function will suffice. An important question for further research is to identify flexible functional forms for indifference curves that provide better or worse incentives for particular kinds of experimental subjects.

Restating and Comparing Other Elicitation Procedures

It is possible to re-state other popular scoring rules in the same general framework, and thereby compare their properties. This also provides insights into the design of experiments that can allow one to evaluate alternative procedures, such as the claims that competitive "prediction markets" generate prices that reflect aggregate beliefs in some sense.

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(p), to a probability report p for some event. Since $0 \leq p \leq 1$, the simple log score is a penalty score, a negative number. Also, since the log function is increasing, higher reported probabilities p for an event receive higher scores (lower penalties), with a maximum score (minimum penalty) of zero when the reported probability p=1. The more general version of the log scoring rule (Bernardo and Smith [1994; p. 73]) is a state-dependent linear transformation of the simple Log score: $k(s) + \tau \times \log [b(s)/p(s)]$, where b(s) is a probability report for state S=s from a discrete set of states and a probability mass function $b=[b(1), b(2),....b(n)]$ such that $b(s)>0$, $\sum_s b(s)=1$; p(s) can be thought of as a reference probability report such that $p(s)>0$, $\sum_s p(s)=1$; and k(s) and $\tau>0$ are constants.

With an appropriate specification of parameters this general log scoring rule describes an indifference curve for a CARA agent in contingent commodity space. The logic here is the same as for a CRRA agent. Consider a CARA agent's optimal choice problem in a competitive contingent claims market where unit s-contingent claims can be bought or sold at prices p(s). The expenditure minimization problem for this agent is to minimize $m= \sum_s p(s)\, x(s)$ subject to a *certainty equivalent* constraint that $-\exp[-CE/t] = -\sum_s b(s) \exp[-x(2)/t]$. This yields the Hicksian compensated demand functions $x^c(s)$ for wealth in state s: $x^c(s)= CE + \tau \times \log [b(s)/p(s)]$. Embedding these compensated demands in an Edgeworth Box of size w(s), the opportunity set faced by the subject is

$$\omega(s) - x^c(s)= w(s) - \{CE + \tau \times \log [b(s)/p(s)]\}, \tag{2.4}$$

which is clearly of the general log scoring rule form. Holding $\omega(s)$, CE, τ and b(s) constant, and varying market prices p(s), we trace out a CARA indifference curve in contingent commodity space.

If a scoring rule is just an opportunity set defined by the experimenter and described by an indifference curve in an Edgeworth box, then the optimization problem facing an experimental subject with a well defined utility function is a trading game. To maximize her own preferences, the subject will want to select an allocation of the available aggregate contingent wealth w(s) where her highest indifference curve is just tangent to the indifference curve defined by the scoring rule opportunity set. She can implement this allocation in the experiment by reporting suitably normalized general competitive equilibrium prices for contingent commodities as her probability reports. This is worth repeating: optimal reports for a scoring rule elicitation task are general competitive equilibrium prices in a 2-person game of pure exchange.

Figure 7 shows how the subject optimally reports (0.573, 0.427) in the trading game between a CRRA agent with beliefs (0.6, 0.4) and risk aversion parameter r=0.57 when facing the standard quadratic scoring rule. The reports (0.573, 0.427) are general competitive equilibrium prices for the two contingent wealth commodities in this 2 person game: relative prices are defined by the slope of the tangency line, 1.342 (= 0.573÷0.427), between the two indifference curves. These competitive market prices do not directly reveal the beliefs of the trading agent. If competitive equilibrium prices in a controlled experimental setting do not directly reveal beliefs, then how likely is it that contingent claims prices in real world prediction markets directly reveal beliefs of the participating agents? Moreover, if knowledge of the subject's utility functions in an experiment, along with other features of the trading game such as the aggregate wealth at stake and the shape of the opportunity set (indifference curves) of the other trader, are needed to uncover latent beliefs for one experimental subject, how can we claim to be recovering aggregate beliefs without comparable information on the participants in a real world prediction market? In effect, these questions generalize the concerns expressed by Manski [2006] about the extent to which prediction markets reveal aggregate beliefs in any meaningful sense (e.g., the mean of the distribution of individual beliefs across agents active in the market).

To be more precise in a simple model where we can be explicit, set up a trading game with the log scoring rule, where the opportunity set is described by an indifference curve for a CARA agent with beliefs b(s) about states, b(s)>0, $\sum_s$ b(s)=1, risk tolerance $\grave{\eta}$, aggregate contingent wealth is w(s), and suppose we are trying to elicit beliefs from a *second* CARA agent with beliefs a(s) about states, a(s)>0, $\sum_s$ a(s)=1, and risk tolerance $\acute{\eta}$. The general competitive equilibrium relative prices p(s) turn out to have an analytically tractable form with p(s) proportional to

$$a(s)^{\acute{\eta}/(\acute{\eta}+\grave{\eta})}\ b(s)^{\grave{\eta}/(\acute{\eta}+\grave{\eta})}\ e^{-w(s)/(\acute{\eta}+\grave{\eta})},$$

where the final factor of proportionality makes all prices sum to unity (Fountain [2003]). Thus, general equilibrium contingent claims relative prices are a weighted geometric mean of the individual agents' relative beliefs, with each agent's beliefs a(s) and b(s) about state s weighted by their relative risk tolerances, all adjusted for an aggregate wealth risk scaled back by the aggregate risk tolerance of the traders. Of course, with homogeneous beliefs and no aggregate wealth risk (w(s)= w(t) all s,t) market prices will directly reflect underlying beliefs, but no trading will take place in this case. Once there is some diversity in beliefs and risk attitudes, market prices will diverge from the arithmetic average of participants' beliefs, and increasingly so the more heterogeneity there is.

In Appendix A we present more general results on the extent to which prediction markets can recover aggregate beliefs. We allow for CRRA and log utility specifications, and show that while prediction markets may predict, they do not generally recover aggregate beliefs.

B. Our Elicitation Procedure as a Special Case

The elicitation procedure we implement in §3 flows from our general perspective on the belief elicitation task as an exchange game. We know that subjects have difficulty learning and understanding what a (proper) scoring rule is, which makes it difficult for them to behave sensibly or optimally when faced with one. Why not present a scoring rule opportunity set to them in terms that are likely to be familiar to them, such as choosing points along a simple budget set or setting prices in a betting market? Our economic reinterpretation of scoring rules as indifference-curves-in-a-box lets us do precisely that.

From duality theory we know that the points along any indifference curve can be traced out in two ways. One direct way is in terms of contingent wealth claims. The indirect way is by specifying tangent lines, with the interpretation of relative market prices, that would make any particular point on the indifference curve an optimal choice in the budget set implicitly described by the tangent line to the indifference curve.

Our experimental design exploits this duality by having the subject take on one of two potential roles.

In the first role, as outlined in §1, she can reveal her beliefs directly by her choices of contingent wealth claims from a finite list of bets. The bets we chose there earlier in fact lie along a linear opportunity set defined in contingent commodity space by the \$1 stake and the house prices. This line can be viewed as the limit of a CES indifference curve in a 2-event contingent commodity space through an endowment point (1,1). In a "forced fed" design we restrict the subjects' choices to the end-points or intercepts of that budget line, corresponding to spending all of the stake on either one or other of the two contingent commodities. In a "free range" design we increase the number of discrete points along the budget line, and implicitly include a "no-bet" option. In either case, when these choices are analyzed jointly with other choices designed to elicit risk attitudes, the experimenter can draw inferences about the subjects' beliefs. Conditional on knowing something about the subject's risk attitude we can infer bounds on her beliefs from her choices.

We view this role as the most natural one for subjects to adopt, since they essentially place bets at given house odds.

Alternatively, our experimental design permits the subject to move along an opportunity set indirectly, by taking the role of the brokerage house and setting contingent claims prices. In this representation beliefs are revealed by price setting behavior, or odds setting behavior in the language of naturally occurring betting markets. In this case the opportunity set defined by the CES Hicksian compensated demands for contingent claims can be viewed as the demands of a computer-generated subject, "RoboPunters." When reporting competitive market odds, or relative prices, to a particular RoboPunter, the subject reveals information which can be used to infer her beliefs after calibration for risk attitudes. Varying the opportunity set facing the subject can be viewed as providing her with a heterogeneous range of punters. Both "forced fed" and "free range" treatments are also feasible in this frame. We do not view this role as the most natural one for typical subjects to take, but might be more natural for expert decision-makers in some settings.

Returning to the experimental frame we adopted, Figure 11 illustrates the opportunity sets presented to subjects in our experiments. In panel A we show the opportunity set presented by one bookie, based on a house probability of ½. The budget line facing the subject is linear, but otherwise the logic is the same as our general framework. The dashed lines show the indifference curves over the two state-contingent wealth claims, and since this is a risk neutral agent they are linear. The subject is assumed to hold a subjective belief π=0.65, so the slope of the indifference curves reflects that rate of subjective transformation between wealth when A is the outcome and wealth when B is the outcome. This subject would optimally select A to win in this instance, maximizing expected income.

Panel B of Figure 11 extends this environment to allow the subject to face 4 bookies instead of 1. The solid lines show the budget sets presented by each bookie, and the associated house probabilities are shown beside each budget line. These bookies are selecting odds based on house probabilities of A winning of 0.5, 0.6, 0.7 and 0.8. The risk neutral agent on the left of panel B picks bet A from the bookies offering house probabilities of 0.5 and 0.6, but then optimally switches to bet B from the bookies offering house probabilities of 0.7 and 0.8.[16] Thus the elicitation procedure correctly identifies this subject as having a true belief in the closed interval [0.6, 0.7].

The risk averse agent on the right of panel B of Figure 11 faces the same opportunity set defined by these 4 bookies, but chooses differently. In this case the indifference curves are non-linear, reflecting a CRRA parameter of 0.45. The agent switches from betting on A to betting on B only when the house probability is between 0.7 and 0.8, even though his subjective belief is still just 0.65.[17] Thus one would need to recover the correct belief by adjusting for risk attitudes in this instance, and the true belief is not directly observable from the choices the subject makes.

C. Approximating the General Case

It is an easy matter to see how the special case used in our experimental frame can be used to approximate arbitrarily well the general, concave opportunity sets discussed in §A. The basic idea is to use that frame to take piecewise linear approximations of the concave opportunity set. There are two approaches that can be used.

---

[16] The house probabilities of 0.5, 0.6, 0.7 and 0.8 generate an EV for a bet on A of $1.30, $1.08, $0.93 and $0.81, and an EV for a bet on B of $0.70, $0.88, $1.17 and $1.75, respectively.

[17] The house probabilities of 0.5, 0.6, 0.7 and 0.8 generate a CE for a bet on A of $0.914, $0.762, $0.652 and $0.572, and an EV for a bet on B of $0.296, $0.37, $0.494 and $0.742, respectively.

Approximation When Only One Bookie is Used for Payment

The outer envelope of panel B of Figure 11 traces out a convex opportunity set. In this case we know by revealed preference that no better would want to place a bet in the interior of this envelope. So with enough betting houses we can approximate a convex opportunity set in this manner.

The same idea can be extended to a concave opportunity set, which is more interesting for belief elicitation. The construction is based on Figure 11 again. Allow the bettors facing house probabilities of 0.5, in one betting house, to have a higher stake to wager than the bettors in the bookies offering house probabilities greater than 0.5. This raises the budget set for this bookie up, and/or the budget sets for the other bookies down: the relative shift is all that is needed in this construction. Then we have to place a cap on how much one can bet (or earn) from any one bookie, to ensure that the outer envelope only consists of portions of the budget set from each bookie. So the bookie offering house probabilities of 0.5 is only willing to take bets up to a certain amount, which means that the relevant budget set is truncated, and only contributes the "top left" part of the concave opportunity set. Similarly for the other bookies.

The upshot is that the same frame that we use for our experiments can be adapted to approximate any opportunity set, such as the QSR. The subjects are again told that there are a finite number of bookies offering house probabilities, but now we (a) vary the stake they have to bet with each bookie, and (b) place minimum and maximum bets that can be placed with each bookie. Of course, in this manner we are free to approximate any concave opportunity set, and not just the one implied by the QSR.

Approximation When All Bookies Are Used for Payment

An even simpler way of using the experimental frame to approximate an arbitrary concave opportunity set arises if we let the subject be paid for all bets on all bookies. In effect, we let the subject create a portfolio of bets across all bookies. The subject is then assumed to maximize over the set of portfolios available.

To see this construction, assume we have two betting houses, α and β, respectively offering house probabilities of 0.75 and 0.4 that event A will occur. For the moment assume our forced fed environment, where the agent has to bet all on A or all on B in each betting house. Then the possible portfolio consists of 4 combinations. Betting on A with both houses generates ($4, $1.67), betting on B with both houses generates ($1.33, $2.50), betting on A with α and B with β generates ($4, $2.50), and betting on B with α and A with β generates ($1.33, $1.67), where the amounts in brackets refer to outcomes from each betting house (the first element is the amount earned from bookie α if the event occurs, and the second element is the amount earned from bookie β if the event occurs). Hence the implied state-contingent wealth outcomes are {$5.67, $0}, {$0, $3.83}, {$4, $2.50} and {$1.67, $1.33}, respectively, where the elements in curly brackets refer to state-contingent outcomes from these combinations of bets (the first element is the payoff if event A occurs, and the second element is the payoff if event B occurs). The last of these 4 wealth outcomes is dominated, but the other three trace out a discrete set of 3 final wealth outcomes. The convex hull of these 3 outcomes traces out a weakly concave opportunity set.

If we extend this construction to allowing free range betting with any of the bookies, we obtain a piecewise linear concave opportunity set. Although this construction is Frankenstein-like for just two bookies, it can obviously be made to approximate a smoother surface with more bookies. The attractive feature of this construction is that it utilizes the standard experimental frame, but only differs by allowing subjects to be paid for the bets that they place on all bookies instead of one picked at random.[18]

D. Point Beliefs and Belief Distributions

We recover the subjective belief π of an event, but we do not recover the subjective belief distribution. The belief π is a *predictive probability* for a specific event, characterizing the subject's uncertainty about the event (Roberts [1965], Winkler [2003; §3.7]). But what about his uncertainty about the probability used to predict this event? The belief π that we recover is consistent with many supporting belief distributions, and is the mean of those distributions. To see this well-known result, initially assume that we elicit the belief π_A=0.65, and that the subject has degenerate beliefs with all mass at that point. Under EUT the subject has the following EU for a bet on A occurring

$$\begin{aligned} EU_A = \;& \pi_A \times U(\text{payout if A occurs} \mid \text{bet on A}) + \\ & (1-\pi_A) \times U(\text{payout if B occurs} \mid \text{bet on A}) \end{aligned} \tag{2.5}$$

and the following EU for a bet on B occurring

$$\begin{aligned} EU_B = \;& \pi_A \times U(\text{payout if A occurs} \mid \text{bet on B}) + \\ & (1-\pi_A) \times U(\text{payout if B occurs} \mid \text{bet on B}). \end{aligned} \tag{2.6}$$

Now assume that the subject actually had a 2-point distribution with density $f(\pi_A^1)$=½ at 0.60 and $f(\pi_A^2)$=½ at 0.70. Define EU_A^1 by substituting π_A^1 for π_A in (2.5), define EU_A^2 by substituting π_A^2 for π_A in (2.5), and similarly define EU_B^1 and EU_B^2 by corresponding substitutions in (2.6). Then the EU for a bet on A is now the compound lottery

$$EU_A = [\, f(\pi_A^1) \times EU_A^1 \,] + [\, f(\pi_A^2) \times EU_A^2 \,] \tag{2.7}$$

and the EU for a bet on B is similarly defined. Since the outcomes in the conditional lotteries are the same, one can collect terms and see that (2.7) is identical to

[18] Experimental economists worry a lot about the "contaminating effects" of allowing subjects to choose a portfolio of lotteries, as discussed by Harrison and Rutström [2008; §2.6] in the context of inferring risk attitudes. In this instance there are no such difficulties, as long as one considers the complete set of possible portfolios.

$$\mu(\pi)_A \times U(\text{payout if A occurs} \mid \text{bet on A}) + (1-\mu(\pi_A)) \times U(\text{payout if B occurs} \mid \text{bet on A}) \quad (2.8)$$

where $\mu(\pi)_A \equiv [\, f(\pi_A^1) \times \pi_A^1 \,] + [\, f(\pi_A^2) \times \pi_A^2 \,]$, the mean of the 2-point distribution. Thus the report provides us with the mean of the distribution. It is immediate that this result generalizes to asymmetric distributions with more than 2 mass points, and indeed to continuous distributions.

Thus we can claim, at least under EUT, that we elicit the mean belief. This also means that we cannot identify the underlying distribution of beliefs, if there is one. The same result generalizes immediately to RDU if one maintains the Reduction of Compound Lottery axiom, as is standard.[19] Of course, in the case of RDU it is the average *weighted* probability that is elicited, and an additional step is needed to recover the subjective probability itself.

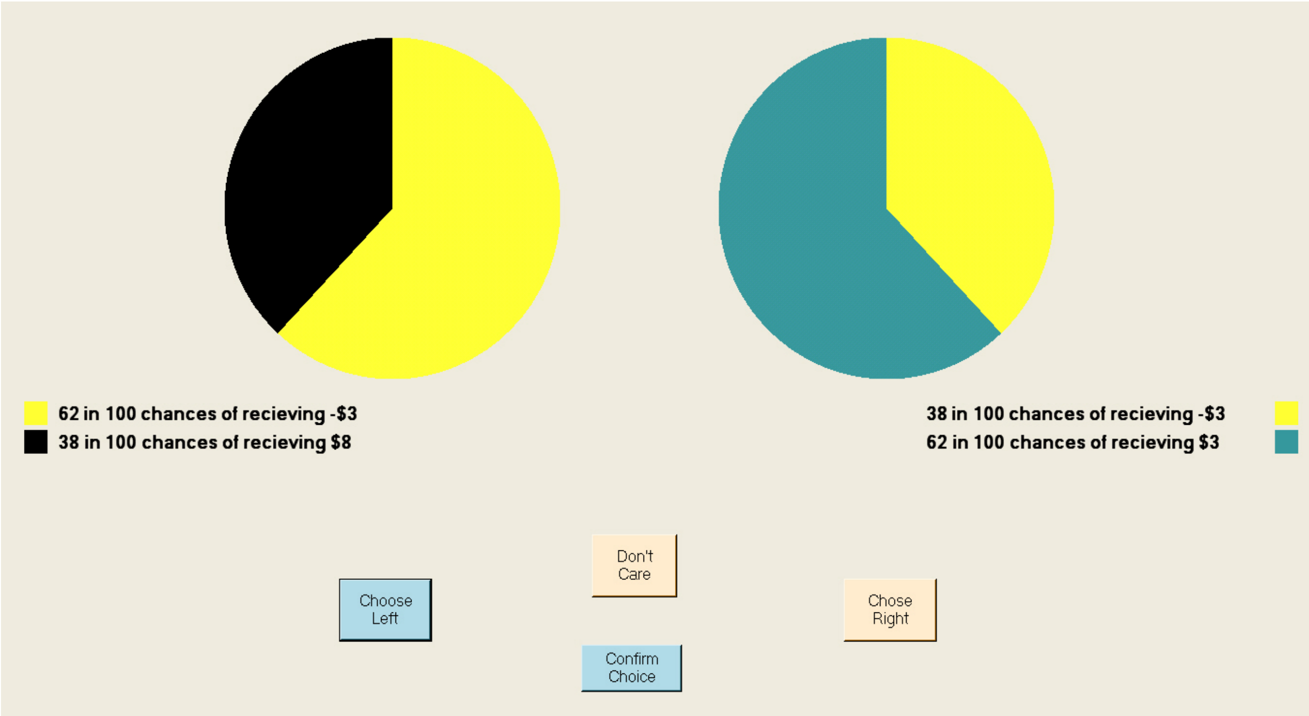
3. Experimental Design

Our overall design is to embed the belief elicitation procedure in tasks that allow us to identify some characteristics of the choice behavior of the subject (their risk attitudes, loss aversion, and probability weighting), to be trained in the use of the elicitation procedure, and to apply the procedure to some naturally occurring events. We pair the belief elicitation task with another task that allows us to identify attitudes to risk, loss aversion and probability weighting from the sample, since we are jointly eliciting these and beliefs.

A. Characterizing Attitudes to Lotteries

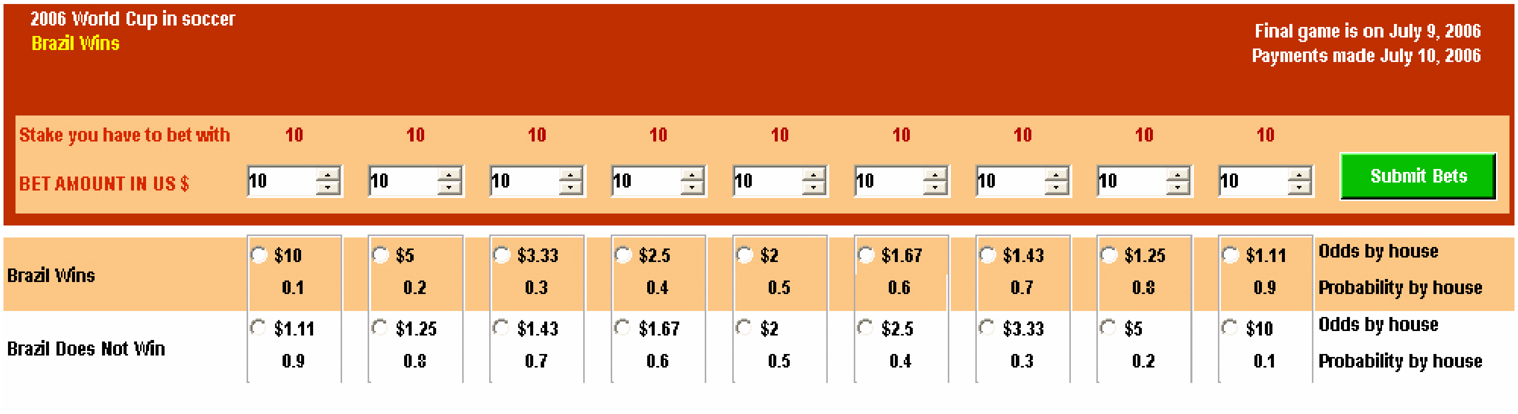
The first series of tasks are designed to tell us if the subject is better characterized as

[19] The axiom of EUT that is typically relaxed is the Independence Axiom, and the Reduction of Compound Lotteries axiom is almost always retained (e.g., Quiggin [1993; p. 19, 134, 154]). It also has considerable normative appeal in a-temporal settings, such as we have in mind here. However, Segal [1987] [1990] illustrates some implications of relaxing the Reduction of Compound Lotteries axiom for RDU. To reiterate, we are focusing solely on probabilistically sophisticated versions of RDU in which the subjective belief can be recovered.



following expected utility theory or prospect theory, and then to identify the specific attitudes to risk or losses that the subject has. This will be useful when we evaluate performance in the elicitation task, as explained later. For this purpose we use a series of binary choice tasks patterned after Hey and Orme [1992], but with the "mixed frame" implementation of Harrison and Rutström [2005]. Each lottery consisted of 2 or more monetary prizes consisting of $8, $3, $0, -$3 or -$8, with varying probabilities. Each subject received an endowment at the start of the task to cover any losses. There were 60 choices, and three were to be chosen at random for payment, so the endowment was $24. Hence, if the subject ended up losing $8 three times, net earnings from these choices would be $0. Some lotteries consisted of non-negative prizes, some of non-positive prizes, but most had a mixture of positive and negative prizes. The screen shot displayed above illustrates a typical choice task. Appendix B shows the instructions provided to subjects, who also had a practice session of 6 hypothetical choices.

These choices allow one to estimate a parametric model of behavior that allows expected utility theory (EUT) or prospect theory (PT) to explain behavior. Conditional on the choices being

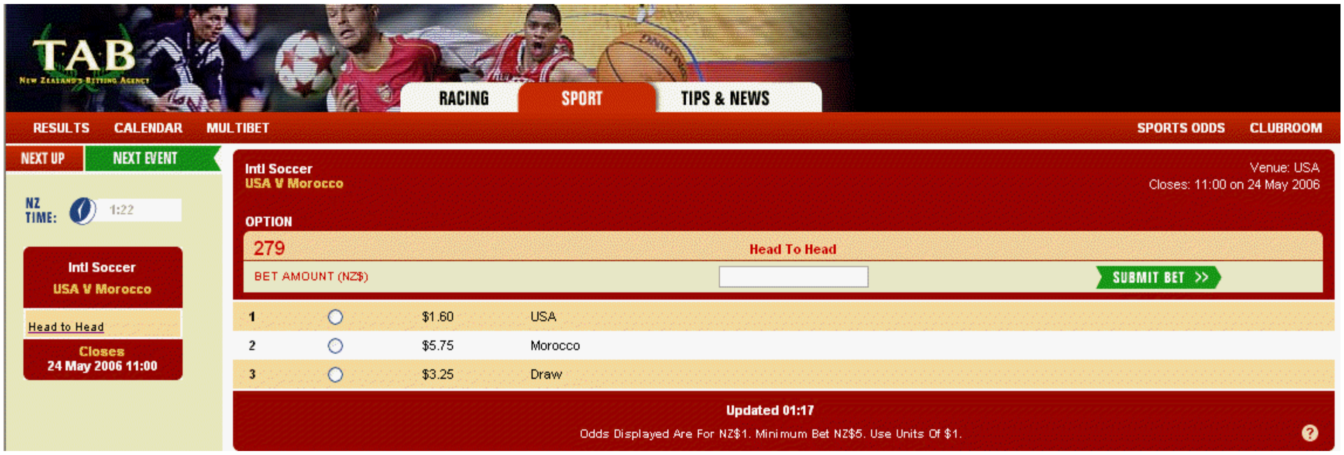


generated by EUT we estimate a coefficient of relative risk aversion. Conditional on the choices being generated by PT we estimate risk aversion coefficients for gains and losses, a loss aversion coefficient, and a coefficient reflecting the extent of probability weighting.

B. Belief Elicitation With Odds

The second task is the elicitation task trainer. We use a simple event that can be resolved in the laboratory to train subjects on the procedure. The instructions are presented in written form, and read out aloud; they are provided in full in Appendix B. Rewards for the trainer are hypothetical, but our experience is that a hands-on trainer typically provides more information than only providing written instructions.

The third task is the elicitation task of interest. We use a naturally occurring event as the object of the elicitation, such as a football game. This event was one that was to be resolved in the future. The basic task is illustrated in the screen shot shown above, which is one of the actual tasks given to subjects. The event is whether Brazil will win the 2006 World Cup. There are 9 betting houses offering different odds that Brazil would win. Prior to this interface the subject is provided information on 3 betting houses that offer odds on events ranging from the winner of the U.S. Presidential race in 2008, the winner of the *American Idol* television show in 2006, an Australian Rules football match, and a soccer match between the USA and Morocco. The laboratory interface was



patterned after the field interface used by the New Zealand Totalisater Agency Board (NZ TAB) shown above, so that the subjects would realize that there are numerous field referents to the interface we were providing. The example used in the instructions was actually the USA *v.* Morocco soccer match referenced in the NZ TAB screen shot, to help subjects see how our lab interface corresponded to the field interface actually used.

The use of 4 field bets from 3 betting houses also served to show the subjects that there are many betting houses available, offering different odds for the same event. To reinforce the fact that different betting houses in the field might offer different odds for the same outcome, we also provided subjects with a screen shot from a web site that simply collects these. This display is shown above, and used the winner of the 2006 U.S. National Basketball Association championship as the exemplar. Odds for the top 3 teams are displayed from 14 active betting houses, and the range of odds for the same team pointed out to subjects.

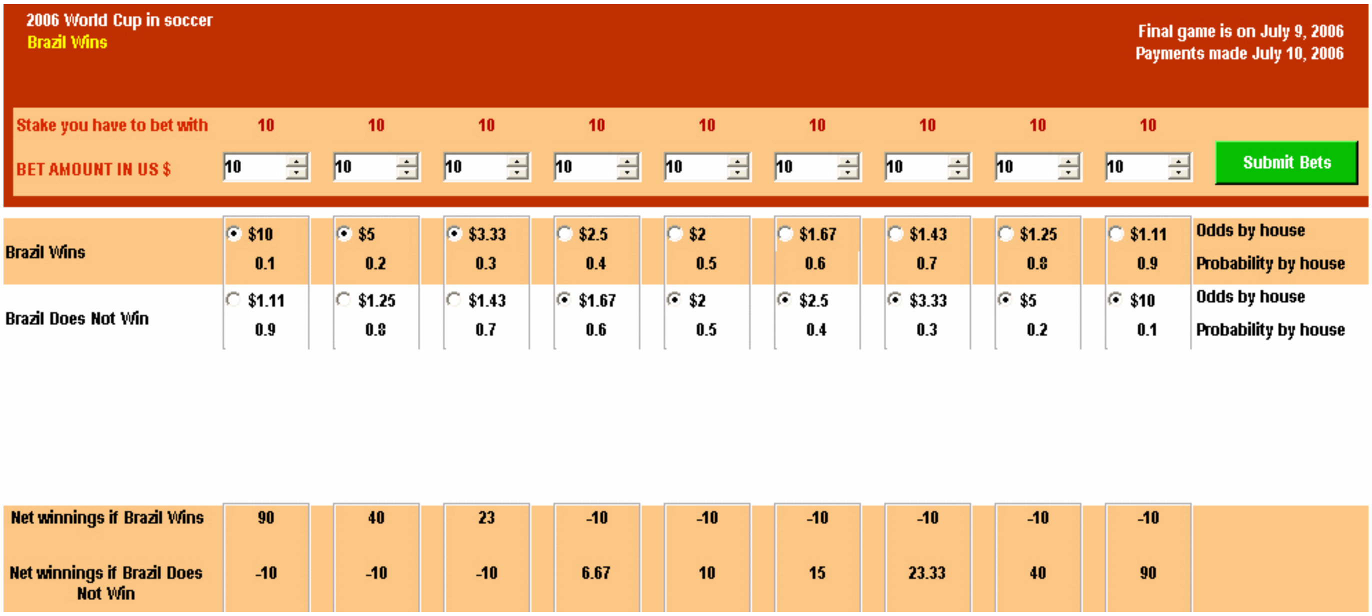
Our instructions are deliberate in using these graphics to provide subjects with field referents to the nature of the task, without of course providing information on prevailing market odds for any of the bets that were to be used to reward subjects. Harrison and List [2004] argue that there are many features of field experiments compared to conventional lab experiments, and that the use of field referents in instructions is one component, among many, that might encourage subject behavior in an experiment that is closer to naturally-occurring behavior.

Returning to our lab interface, the subject is given a $10 stake with which to bet in each betting house. The $10 from one house is not transferable to other houses, and one of the bets will be selected to be actually played out (selected using a die that the subject throws). We "force fed" the subjects by requiring that they place a bet with each house, and did not allow them to change the $10 stakes: all the subject can do is decide if they want to bet on event A or event B in each house.[20]

The interface updated the display showing net winnings conditional on each event occurring, as illustrated below. This illustration shows a subject that was willing to bet on Brazil winning when the odds being quoted were favorable enough, but who switches to betting on Brazil not winning between odds of $3.33 to $1.43 and $2.50 to $1.67. This switch point corresponds to house probabilities of Brazil winning between 0.3 and 0.4, which are also displayed to the subjects. The subject is told in the instructions that their net winnings are "net" of the stake that we gave them to bet with.

The actual elicitation tasks involved a hypothetical trainer, one fact which could be resolved

[20] An alternative "free range" design also allows the subject to vary the stake that is bet in each house, from $0 up to $10 in $1 increments, and is discussed in §5 as an important extension that builds a bridge between our elicitation task and those found in open prediction markets.



during the session, and three natural events occurring in the future. The hypothetical trainer was whether Mercury was the closest planet to the sun. This question was chosen since there is some survey evidence that U.S. college seniors do not universally know that this is true, and it can be verified with no time delay. Thus subjects could see the effect of their choices when compared to an actual outcome. The fact used for real rewards was either whether Madrid was the capital of Portugal. The natural events used for real rewards were (i) whether Brazil would win the 2006 World Cup outright, (ii) whether Donald Rumsfeld would resign by the end of 2006, and (iii) whether Tiger Woods would win the 2006 U.S. Golf Championship. The outcome of the two sporting events were to be known in late June and early July, and subjects were to be paid mid-July. The political event was to be known by the end of 2006, and subjects were to be paid at the beginning of 2007.

For simplicity we assume that the risk attitudes we elicit using lotteries defined over contemporaneous payouts can be used to recover the subjective beliefs inherent in bets defined over future payouts. In effect, we assume that risk attitudes "today" apply "in the future." In principle one could elicit risk attitudes applicable to future payouts, and there are good reasons to eventually do that; nonetheless, we avoid procedural complexity by making this assumption.[21]

We recruited 25 subjects from the general student population of the University of Central Florida. Subjects were recruited for this specific experiment using the computerized *ExLab* interface (http://exlab.bus.ucf.edu), after being solicited in general terms to register with *Exlab* for paid experiments. All subjects received a $5 show-up fee, and participated in the same session. Apart from the choice tasks described above, all subjects completed a survey of demographic characteristics. Payments for the belief elicitation task totaled $1,094, and all deferred payments were cashed.

4. Results

We present the results of our experiments in three stages. First, we examine the raw responses to our belief elicitation task, which can be interpreted directly as elicited beliefs under the assumption that subjects are risk neutral. Second, we specify the estimation framework and define a joint likelihood over the responses to the lottery tasks and the belief tasks, allowing us to generate maximum likelihood estimates of all parameters. We specify one framework assuming EUT, and one framework assuming RDU. Third, we examine the evidence for the strong assumption of risk neutrality – it will not come as a shock if we reject it. Fourth, we show how inferences about beliefs are adjusted or calibrated for the "distortions" generated by deviations from the risk-neutrality assumption.

---

[21] One reason is that risk attitudes could be state-dependent, and subjects might expect some changes in some state in the future. Methods for eliciting state-dependent preferences are evaluated in Andersen et al. [2008]. Another reason is that "dual-self" models of decision making under risk posit differences in the preferences that are brought to bear when making choices over money today and money in the future: see Fudenberg and Levine [2006] for a theoretical exposition and Andersen et al. [2005] for an empirical application to eliciting preferences. Finally, there is the possibility that preferences over bets might depend on the timing of the resolution of the event, as proposed by Kreps and Porteus [1978].

A. Elicited Beliefs from the Raw Data

The raw data from our belief elicitation task can be represented in terms of the fraction of subjects that picked option A at each given house probability. To understand the nature of the choices at the individual level, Figure 12 displays a binary indicator of the bets of the first 4 subjects. A value of 1 on the vertical axis indicates that the subject bet on the event or state occurring, and a value of 0 indicates otherwise. Thus subject #6001 was certain that Rumsfeld would resign, and would not bet against this outcome no matter what odds were offered. The next three subjects were less certain. If we assume that subjects behave consistently[22] and are risk neutral, we can infer subjective beliefs from these choices: subject #6001 has a degenerate probability of 1, subject #6002 has a belief in the interval [0.5, 0.6], subject #6003 has a belief in the interval [0.4, 0.5], and subject #6004 has a belief in the interval [0.3, 0.4].

Figure 13 displays comparable data pooled over all subjects. These lines represent the average over all subjects, so the 1 and 0 values on the vertical axes represent the same outcome as the binary indicator in Figure 12. But as we pool over subjects we have averages that are between 0 and 1, of course. Hence the panel for Rumsfeld Resigning in Figure 13 is generated by taking averages over the panels illustrated in Figure 12, although for all 25 subjects.

From Figure 13 we observe that this sample is virtually certain that Madrid is not the capital of Portugal, which is re-assuring.[23] A majority of subjects held subjective beliefs that Rumsfeld would resign with probability 0.4 or less, that Tiger would win the next U.S. Open with probability 0.4, and that Brazil would win the World Cup with probability 0.5.

[22] The term "consistently" anticipates the possibility that subjects might make errors, as formally allowed for in our econometric framework.

[23] As it happens, two subjects did switch their bets as the odds varied, but almost perfectly offset each other in terms of the direction of their switches.

B. Joint Estimates of Beliefs and Preferences Towards Risk

Expected Utility Theory

We assume a CRRA utility function defined over the gain domain of lottery prizes y as

$$U(y) = (y^{1-r})/(1-r) \quad (4.1)$$

where r is the CRRA coefficient and $y \geq 0$. With this specification, and for $y \geq 0$, $r = 0$ denotes risk neutral behavior, $r > 0$ denotes risk aversion, and $r < 0$ denotes risk loving. All arguments of utility are defined as the *gross* earnings from each bet, as defined in Table 1, so they are non-negative.

The utility function (4.1) can be estimated using maximum likelihood and a latent structural model of choice, such as EUT. 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:

$$EU_i = \sum_{k=1,K} [\, p_k \times u_k \,]. \quad (4.2)$$

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

$$\nabla EU = EU_R - EU_L \quad (4.3)$$

calculated, where EU_L is the "left" lottery and EU_R is the "right" lottery. This latent index, based on latent preferences, is then linked to the observed choices using a standard cumulative normal distribution function $\Phi(\nabla EU)$. This "probit" function takes any argument between $\pm\infty$ and transforms it into a number between 0 and 1. Thus we have the probit link function,

$$\text{prob(choose lottery R)} = \Phi(\nabla EU) \quad (4.4)$$

The latent index defined by (4.3) is linked to the observed choices by specifying that the R lottery is chosen when $\nabla EU > ½$, which is implied by (4.4).

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

$$\ln L(r; y, \mathbf{X}) = \sum_i [(\ln \Phi(\nabla EU) \mid y_i = 1) + (\ln \Phi(1-\nabla EU) \mid y_i = -1)] \quad (4.5)$$

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

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 by recognizing that such choices implied a 50:50 mixture of the likelihood of choosing either lottery:

$$\ln L(r; y, \mathbf{X}) = \sum_i [(\ln \Phi(\nabla EU) \mid y_i = 1) + (\ln \Phi(1-\nabla EU) \mid y_i = -1) + (\ln (\tfrac{1}{2} \Phi(\nabla EU) + \tfrac{1}{2} \Phi(1-\nabla EU)) \mid y_i = 0)] \quad (4.5')$$

where $y_i = 0$ denotes the choice of indifference. In our experience very few subjects choose the indifference option, but this formal statistical extension accommodates those responses.[24]

To allow for subject heterogeneity with respect to risk attitudes, the parameter r is modeled as a linear function 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 (in years) and Female (0 for males, and 1 for females). Then we would estimate the coefficients α, β and η in $r = \alpha + \beta \times \text{Age} +$

[24] Our treatment of indifferent responses uses the specification developed by Papke and Wooldridge [1996; equation 5, p.621] for fractional dependant variables. Alternatively, one could follow Hey and Orme [1994; p.1302] and introduce a new parameter τ to capture the idea that certain subjects state indifference when the latent index showing how much they prefer one lottery over another falls below some threshold τ in absolute value. This is a natural assumption to make, particularly for the experiments they ran in which the subjects were told that expressions of indifference would be resolved by the experimenter, but not told how the experimenter would do that (p.1295, footnote 4). It adds one more parameter to estimate, but for good cause.

$\eta \times$ Female. Therefore, each subject would have a different estimated r, $\hat{r}$, for a given set of estimates of α, β and η 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 8 individual characteristics in our model. Apart from age and sex, these include binary indicators for subjects that self-declare their ethnicity as Asian, declare themselves to be Hispanic, a Business major, a U.S. Citizen, a Graduate student, and having a low GPA (below 3.25). Age is measured as years in excess of 19, so it shows the effect of age increments.

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 (4.4) that the probability of choosing a lottery is not 1 when the EU of that lottery exceeds the EU of the other lottery.[25] By varying the shape of the link function implicit in (4.4), 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 error specification originally due to Fechner [1860], and popularized by Becker, DeGroot and Marschak [1963] and Hey and Orme [1994]. It posits the latent index

$$\nabla EU = (EU_R - EU_L)/\mu \tag{4.3'}$$

instead of (4.3), where $\mu>0$ is a structural "noise parameter" used to allow some errors from the perspective of the deterministic EUT model. As $\mu \rightarrow \infty$ this specification collapses ∇EU to 0, 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 (4.4). This is just

[25] This assumption is clear in the use of a link function between the latent index ∇EU and the probability of picking one or other lottery; in the case of the normal CDF, this link function is $\Phi(\nabla EU)$. If the subject exhibited no errors from the perspective of EUT, this function would be a step function: zero for all values of $\nabla EU<0$, anywhere between 0 and 1 for $\nabla EU=0$, and 1 for all values of $\nabla EU>0$.

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,

$$\ln L(r, \mu; y, \mathbf{X}) = \sum_i [(\ln \Phi(\nabla EU) \mid y_i = 1) + (\ln \Phi(1-\nabla EU) \mid y_i = -1) + (\ln (½ \Phi(\nabla EU) + ½ \Phi(1-\nabla EU)) \mid y_i = 0)] \tag{4.5″}$$

and estimate r and μ using maximum likelihood, given observations on y and **X**. 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].

Rank Dependent Utility

The RDU model extends the EUT model by allowing for decision weights on lottery outcomes. Instead of (4.1) we have

$$U(y) = (y^{1-\rho})/(1-\rho) \tag{4.6}$$

for $y \geq 0$, and where ρ is the coefficient for the curvature of the utility function. To calculate decision weights under RDU one replaces expected utility defined by (4.2) with RDU

$$RDU_i = \sum_{k=1, K} [w_k \times u_k] \tag{4.7}$$

where

$$w_i = \omega(p_i + ... + p_n) - \omega(p_{i+1} + ... + p_n) \tag{4.8a}$$

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

$$w_i = \omega(p_i) \tag{4.8b}$$

for i=n, the subscript indicating outcomes ranked from worst to best, and where $\omega(p)$ is some probability weighting function. We adopt the popular probability weighting function employed by Tversky and Kahneman [1992], with curvature parameter γ. It is assumed to have well-behaved endpoints such that $\omega(0)=0$ and $\omega(1)=1$ and to imply weights

$$\omega(p) = p^\gamma / [\, p^\gamma + (1-p)^\gamma \,]^{1/\gamma} \tag{4.9}$$

for $0<p<1$. So $\gamma<1$ implies the usual inverse-S shaped function, but $\gamma>1$ has been observed in some studies.

The same Fechner noise specification is used as in the EUT model, and defined over the difference in the RDU calculated from (4.7):

$$\nabla RDU = (RDU_R - RDU_L)/\mu \tag{4.10}$$

The likelihood specification for the RDU model is therefore

$$\ln L(\rho, \gamma, \mu; y, \mathbf{X}) = \sum_i [\, (\ln \Phi(\nabla RDU) \mid y_i = 1) + (\ln \Phi(1-\nabla RDU) \mid y_i = -1) + (\ln (\tfrac{1}{2}\, \Phi(\nabla RDU) + \tfrac{1}{2}\, \Phi(1-\nabla RDU)) \mid y_i = 0) \,] \tag{4.11}$$

and entails the estimate of ρ, γ and μ using maximum likelihood. Individual heterogeneity is allowed for by estimating the parameters ρ and γ as linear functions of the 8 demographic characteristics defined earlier.

Figure 14 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 $\gamma=0.7<1$ in this illustration, the probability weighting function $\omega(p)$ has the standard inverse-S shaped form. 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 $p=½$; with 3 prizes, each prize has $p=⅓$; and with 4 prizes, each prize has $p=¼$. For the 3-prize and 4-prize lottery we see the standard result, that the decision weights on the extreme prizes are relatively greater than the true probability, and the decision weights on the interior prizes are relatively smaller than the true probability. For $\gamma>1$ the interior-ranked prizes receive greater decision weight.

Each panel in Figure 14 is important for our analysis. For the purposes of estimating γ from the observed lottery choices we only need the decision weights in the right panel of Figure 14. But for the purposes of recovering subjective beliefs subject to probability weighting, we only need the probability weighting function. In fact, we need it's inverse function, since it is the p in the $\omega(p)$ function that we are seeking to recover. We do not directly observe $\omega(p)$, but we can estimate it as part of the latent structure generating the observed choices in the belief task. Once we have $\omega(p)$ we can then recover p by directly applying the estimated probability weighting function shown, for a typical γ, in the left panel of Figure 14.

C. Elicited Beliefs of the Sample after Correcting for Preferences Towards Risk

The responses to the belief elicitation task can be used to draw estimates about the belief that each subject holds if we are willing to assume something about how they make decisions under risk.

If they are assumed to be risk neutral, then we can directly infer intervals that contain that point estimate, as our discussion of Figure 12 illustrated. If we do that for all subjects and all tasks, we obtain interval estimates that can then be used to make inferences about beliefs using interval regression methods. Figure 15 displays the risk neutral estimates that we obtain in this manner from the pooled data, after controlling for observable demographics. We infer that the sample holds beliefs that Rumsfeld will resign of 0.54, that Tiger will win of 0.50, and that Brazil will win of 0.46. There is also considerable diversity in beliefs, with standard deviations of 0.13, 0.09 and 0.14, respectively.

Moving to the models that allow for varying risk attitudes, we jointly estimate the subjective probability and the parameters of the core model. Using the schema in Table 1, the subject that selects event A from a given bookie b receives EU

$$\begin{aligned} EU_A = \quad & \pi_A \times U(\text{payout if A occurs} \mid \text{bet on A}) + \\ & (1-\pi_A) \times U(\text{payout if B occurs} \mid \text{bet on A}) \end{aligned} \tag{4.12}$$

where π_A is the subjective probability that A will occur; this is the same expression introduced earlier as (2.5). The payouts that enter the utility function are defined by the odds that each bookie offers, and are illustrated in Table 1. For the bet offered by the first bookie, for example, these payouts are \$10 and \$0, so we have

$$EU_A = \pi_A \times U(\$10) + (1-\pi_A) \times U(\$0) \tag{4.12'}$$

We similarly define the EU received from a bet on event B, restating (2.6), as the complement of event A:

$$\begin{aligned} EU_B = \quad & \pi_A \times U(\text{payout if A occurs} \mid \text{bet on B}) + \\ & (1-\pi_A) \times U(\text{payout if B occurs} \mid \text{bet on B}). \end{aligned} \tag{4.13}$$

and this translates for the first bookie in Table 1 into payouts of \$0 and \$1.11, so we have

$$EU_B = \pi_A \times U(\$0) + (1-\pi_A) \times U(\$1.11) \tag{4.13'}$$

for this particular bookie and bet. We observe the bet made by the subject, so we can calculate the likelihood of that choice given values of r, π_A and μ.

We need r to evaluate the utility function in (4.12) and (4.13), we need π_A to calculate the EU in (4.12) and (4.13) once we know the utility values, and we need μ to calculate the latent indices (4.3′) or (4.10) that generate the probability of observing the choice of event A or event B 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. In effect, the lottery task allow us to identify r under EUT, and ρ 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 parameters, one for each belief task, as linear functions of the demographic characteristics defined earlier.

Table 3 and 4 collect the detailed estimates for each model, and Figure 16 displays densities of the predicted parameter estimates across individuals. These estimates serve as a reminder that the heterogeneity in beliefs here derives from variations in demographic characteristics across individuals in the sample. This variation is reflected not just in the estimates of the beliefs themselves, but in the estimates of the core parameters r, ρ and γ, which in turn affect the estimates of the beliefs indirectly. Figure 16 shows that the estimates of the curvature of the utility function under EUT and RDU each suggest a predominance of risk loving behavior ($r<0$ and $\rho<0$), with a mode at risk neutrality. In general the probability weighting parameter γ indicates the conventional inverse-S shaped probability weighting ($\gamma<1$), although there is a smaller mode corresponding to subjects that exhibit S-shaped probability weighting.[26]

Since the focus of the estimation is on the subjective probability, it is important to ensure that the estimates of π lie in the closed interval [0,1]. It is straightforward to constrain the estimates to the open unit interval, which is sufficient for our purposes, by estimating the log odds transform κ of the parameters we are really interested in, and converting using $\kappa = 1/(1+e^{\pi})$. This transform is one of several used by statisticians; for example, see Rabe-Hesketh and Everitt [2004; p. 244]. These commands use the Delta Method from statistics to calculate the standard errors correctly: see Oehlert [1992] for an exposition and historical note.

Figures 17, 18 and 19 show the effect of jointly estimating subjective beliefs conditional on a specific model of choice under uncertainty other than risk neutrality. The solid density in each panel is the baseline distribution assuming risk neutrality, for reference; these just repeat, for each task, the corresponding density from Figure 15. In the left panel we see the effect of calibrating subjective beliefs using an EUT model that allows for risk attitudes, and in the right panel we see the effect of calibrating subjective beliefs using an RDU model that allows for utility curvature and probability

[26] The log-likelihood for the RDU model is lower than the log-likelihood for the EUT model, as one would expect since the EUT model is nested in the RDU model. However, we do not view this as a legitimate basis for declaring RDU to be the sole "true" model to describe behavior of this sample. In general we need to allow for behavior to be explained by a mixture of such latent data-generating processes, and the evidence suggests that such mixtures apply even when the aggregate log-likelihood of one model dominates the other (Harrison and Rutström [2005]).

weighting.

Consider Figure 17, and beliefs about the resignation of Rumsfeld. The effect of the "salvage operation" is to recover a distribution of subjective beliefs that is strikingly different than the one implied by assuming that subjects are risk neutral. Figures 18 and 19 show the estimated densities for the belief about whether Tiger will win, and that Brazil will win. The same qualitative pattern emerges.

D. Why are the Recovered Beliefs so Fragile?

The pattern of adjustment is relatively easy to account for given the evidence of risk loving behavior. Figure 20 shows the difference between the true belief and the reported switch point of a risk averse CRRA agent with r=0.57, the example used in §2, as well as a risk loving CRRA agent with r= -0.57. For each type of agent we show the switch points, in terms of house probabilities, at which the agent changes from betting on A to betting on B. The paradigm here is our forced feed environment, in which the agent faces bookies at each house probability and has to bet a given stake all on A or all on B. Thus it corresponds to the experimental frame we used, and Figure 11. The risk neutral agent switches at house probabilities equal to her true belief.

In panel A of Figure 20 we see that the risk averse agent, however, switches at lower probabilities when true beliefs are below ½, and switches at higher probabilities when true beliefs are above ½. Thus the *risk averse agent will appear to be more certain, when incorrectly interpreted under the assumption of risk neutrality, than she really is.* The risk loving agent in panel B of Figure 20 exhibits the reverse qualitative behavior: he switches at house probabilities greater than his true beliefs when those true beliefs are less than ½, and switches at house probabilities less than his true beliefs otherwise. Thus the *risk loving agent will appear to be less certain, from a risk neutral interpretation of observed switching behavior, than he really is.*

Figure 21 uses the risk averse agent to show the extent of the salvage operation involved in recovering true beliefs. For an agent with a true belief of ¾, we would see a switch at a house probability of 0.93 if bookies with house probabilities incremented in 0.01 amounts were available. Thus the error from interpreting her behavior as being generated by a risk neutral response would be 18 percentage points, as shown.

Figure 22 provides a warning about the difficulty of the salvage operation involved in inferring true beliefs from observed behavior, and another hint as to why inferences about subjective beliefs might be fragile. The same point was made less formally in §1D, couched there in terms of the "observational equivalence" of a range of subjective beliefs. *If we only have house probabilities incremented in 0.1 amounts*, as in our experiments, we would not be able to differentiate observationally beliefs in the heavily shaded region to the right of panels A and B in Figure 22, as marked by the arrow. In panel A, with a risk averse agent, every such belief, from 0.73 up to 1, would result in the agent choosing A for all house probabilities. This result does not depend on the agent picking all A or all B, as panel B illustrates. Here we depict, in the heavily shaded region, the range of true beliefs consistent with the agent picking A for all house probabilities up to and including 0.7, and then picking B. This range is between 0.64 up to 0.72.

Of course, if we are uncertain about the CRRA (and probability weighting) coefficient for this agent, these ranges would be even wider. And they are, as strikingly demonstrated in Figures 16, 17 and 18.

5. Conclusions

Our approach brings together theoretical tools, experimental procedures, and econometric techniques that allow us to elicit subjective beliefs. The theoretical tools are based on many ideas found in the literature, and show that scoring rule procedures and betting procedures can be viewed from the same formal perspective: they both set up a game of exchange between two parties, defined over state-contingent wealth claims. This perspective allows one to see how this game can be applied in experiments, and also how the game itself can be modified to encourage truthful and natural decisions. But the most important insight is that we can directly apply alternative models of decision making under risk to recover subjective beliefs in a rigorous manner. We show that those corrections can be substantial, depending on the model of decision making adopted.

One theme of our results is that inferences about subjective beliefs are fragile. One cannot simply read off the subjective beliefs of an individual from some intelligently designed task. The subjects often find these tasks artefactual, and it is only blind faith that leads one to believe that they perceive or believe the dominant strategy of making a truthful response. Moreover, that strategy is often not a dominant strategy when subjects are risk averse or risk loving, so the claim itself is often simply irrelevant to subjects. In fact, the subjects do still respond truthfully, but they also take into account the risk consequences of the decisions they are being asked to make. The challenge is for the observer to recover the subjective beliefs embedded in those decisions.

The complementary, and constructive, theme of our results is that we do have the tools to elicit subjective beliefs, if we are prepared to make enough structural assumptions. The structural assumptions we use are stronger than those that are necessary, but our objective was to demonstrate the range of assumptions necessary and to adopt popular functional forms. To recover subjective beliefs we need to know the risk attitudes of individuals, and we can define risk attitudes in terms of several popular models. We used EUT and RDU, but the approach extends to Prospect Theory, Regret Theory, and the myriad variants in the literature. We prefer to recover latent subjective beliefs using explicit structural models, so we know what forces are causing the subject to deviate from the choices implied by a risk neutral agent. If we are going to claim that subject #6003, from Figure 12, has some belief that Rumsfeld will resign that is *not* between 0.4 and 0.5, we want to be able to tell a sensible story to justify the claim.

Three extensions merit mention, with a longer list in Appendix C. One is to allow subjects to engage in "free range" betting, in which they can decide to hold back some of their stake and not place a bet. This option connects our approach to the beliefs elicited on open prediction markets, where subjects are not forced to place a bet.[27] Free range betting also allows a more precise joint determination of latent risk attitudes and subjective beliefs. A second extension would be to allow subjects to place (free range) bets on all bookies and be paid for all bets, so that they optimize with respect to the portfolio available to them. This would concavify the opportunity set, as explained in §2C, and should result in significantly less fragile statistical inferences about beliefs. A third extension is to compare our results to experimental frames that present traditional scoring rules, such as the QSR, to see the effect of our "natural betting" frame compared to the complicated scoring rules that have been popular.

[27] A related issue is allowing subjects to bet with their own money, earned in some prior task, rather than with house money.

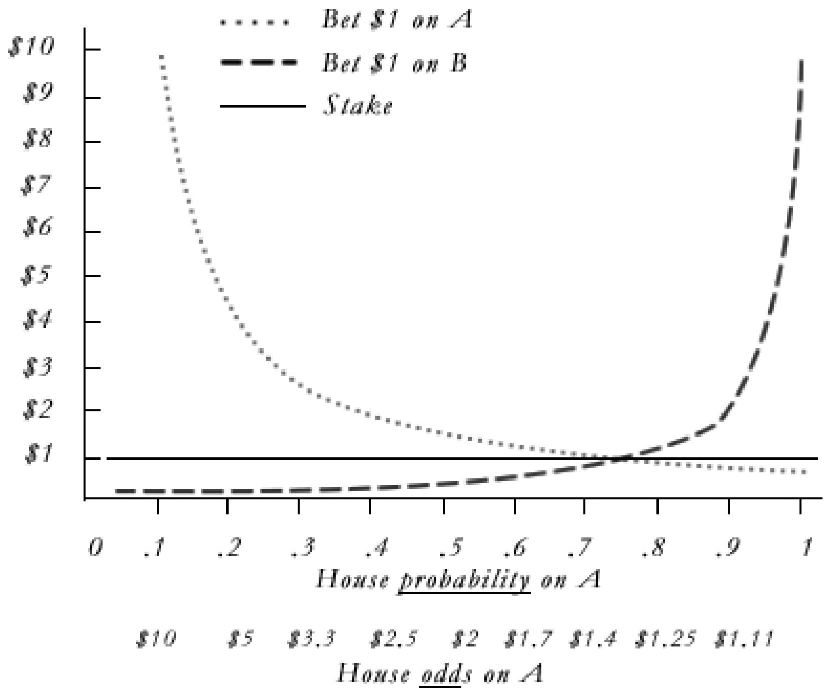
Table 1: Betting Choices Facing Subject Forced to Place a Bet

Assume subject has a personal belief that A will occur with probability ¾.

House Probabilities		Odds		Stake			Net Earnings			Gross Earnings		
									Expected			Expected
A	B	A	B	to Bet	Bet on A	Bet on B	A wins	B wins	from bets	A wins	B wins	from bets
0.1	0.9	$10.00	$1.11	$1.00	1	0	$9.00	$-1.00	$6.50	$10.00	$0.00	$7.50
0.2	0.8	$5.00	$1.25	$1.00	1	0	$4.00	$-1.00	$2.75	$5.00	$0.00	$3.75
0.3	0.7	$3.33	$1.43	$1.00	1	0	$2.33	$-1.00	$1.50	$3.33	$0.00	$2.50
0.4	0.6	$2.50	$1.67	$1.00	1	0	$1.50	$-1.00	$0.88	$2.50	$0.00	$1.88
0.5	0.5	$2.00	$2.00	$1.00	1	0	$1.00	$-1.00	$0.50	$2.00	$0.00	$1.50
0.6	0.4	$1.67	$2.50	$1.00	1	0	$0.67	$-1.00	$0.25	$1.67	$0.00	$1.25
0.7	0.3	$1.43	$3.33	$1.00	1	0	$0.43	$-1.00	$0.07	$1.43	$0.00	$1.07
0.8	0.2	$1.25	$5.00	$1.00	0	1	$-1.00	$4.00	$0.25	$0.00	$5.00	$1.25
0.9	0.1	$1.11	$10.00	$1.00	0	1	$-1.00	$9.00	$1.50	$0.00	$10.00	$2.50
									$14.20			$23.20

Figure 1: Certainty Equivalents of Optimal Bets Across Betting Houses

A. Risk Neutral agent with subjective belief = ¾



B. Risk averse agent with subjective belief = ¾

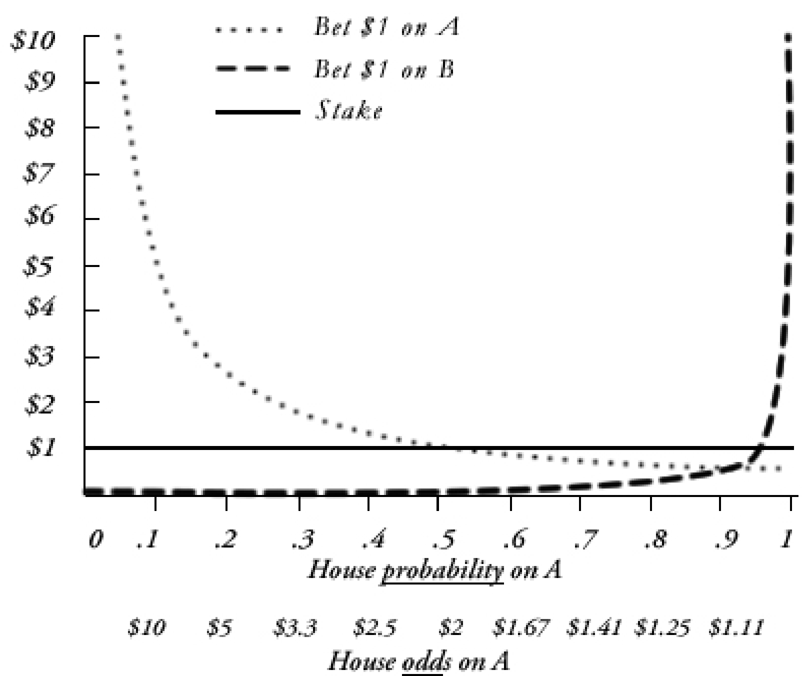
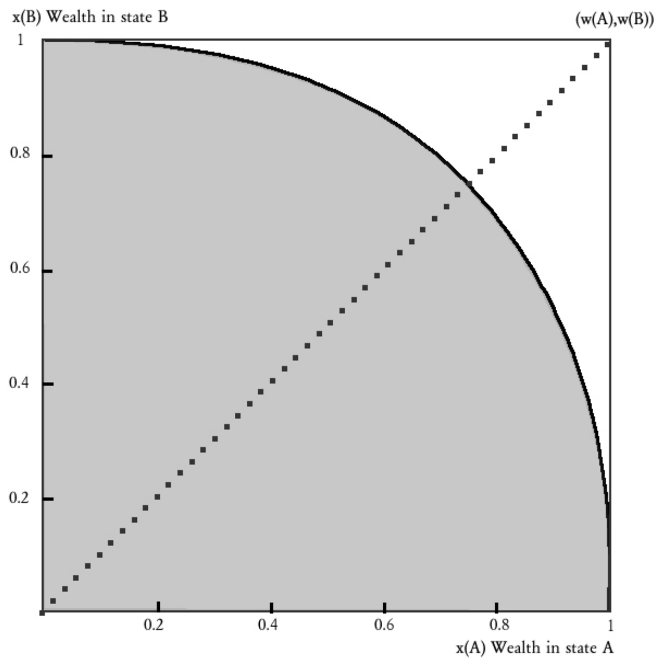
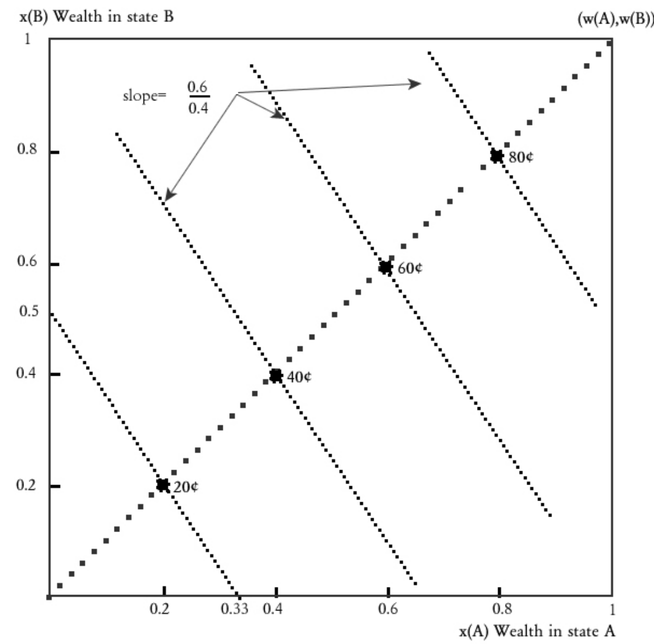


Figure 2: The Quadratic Scoring Rule as an Opportunity Set



Opportunity set for quadratic score [$1-(1-r)^2$, $1-r^2$],
where r is the reported probability for state A , $0 \leq r \leq 1$

Figure 3: Indifference Curves for a Risk Neutral Agent



Indifference curves reflect beliefs (p(A),p(B))=(0.6,0.4}

Figure 4: Optimal Behaviour for a Risk Neutral Agent with the Quadratic Scoring Rule

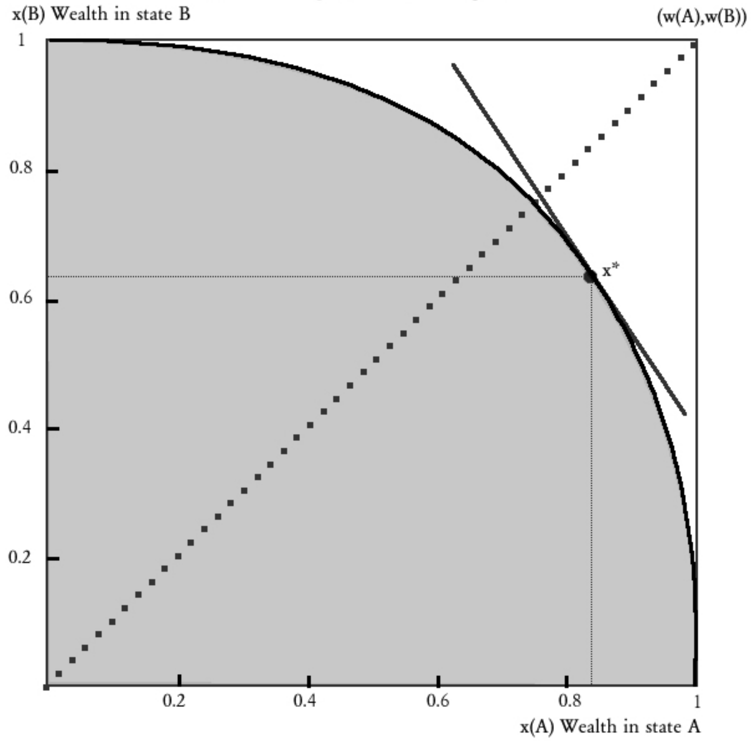


Figure 5: Getting out the Magnifying Glass

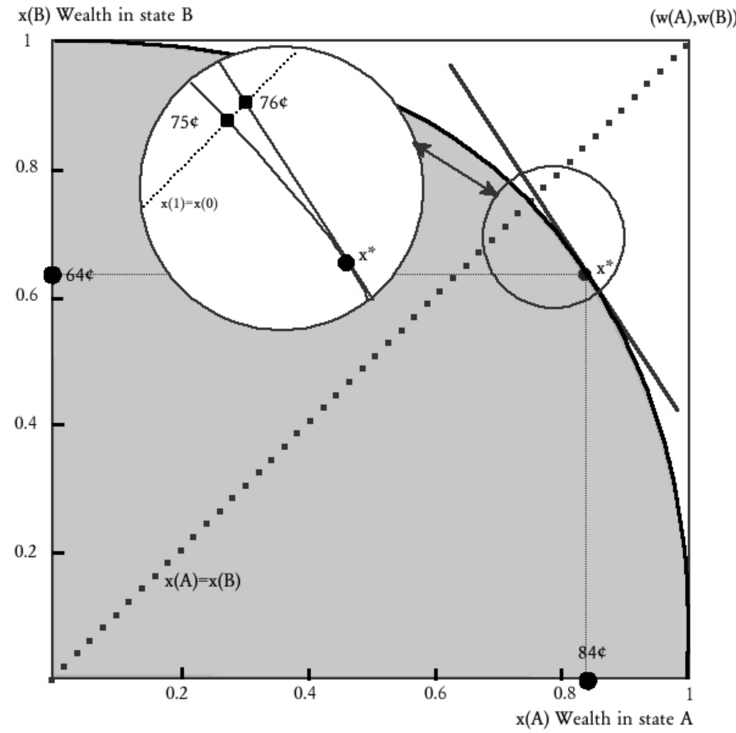
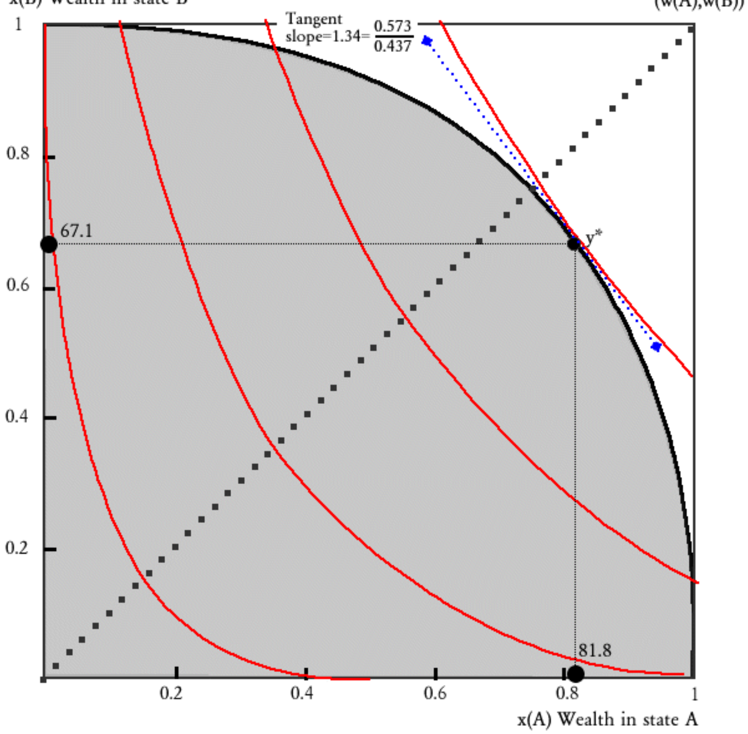
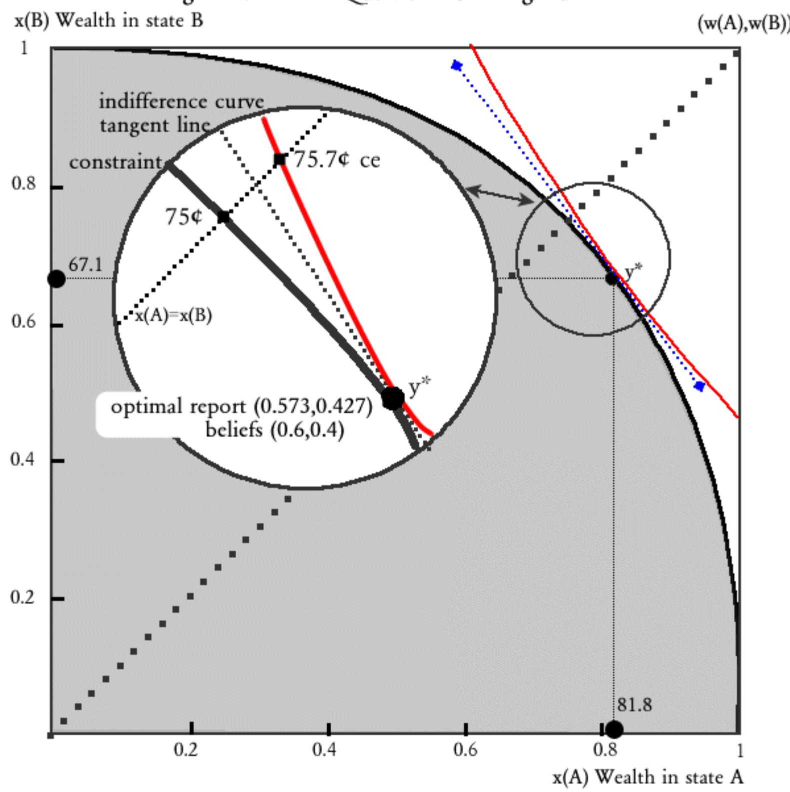


Figure 6: Optimal Behaviour for a CRRA(0.57) Agent with the Quadratic Scoring Rule



Indifference curves reflect beliefs (p(A),p(B))=(0.6,0.4}

Figure 7: Magnifying Glass of Optimal Behaviour for a CRRA(0.57) Agent with the Quadratic Scoring Rule



Indifference curves reflect beliefs (p(A),p(B))=(0.6,0.4)

Figure 8: Certainty Equivalent Wealth for Risk Neutral and CRRA(0.57) with the Quadratic Scoring Rule

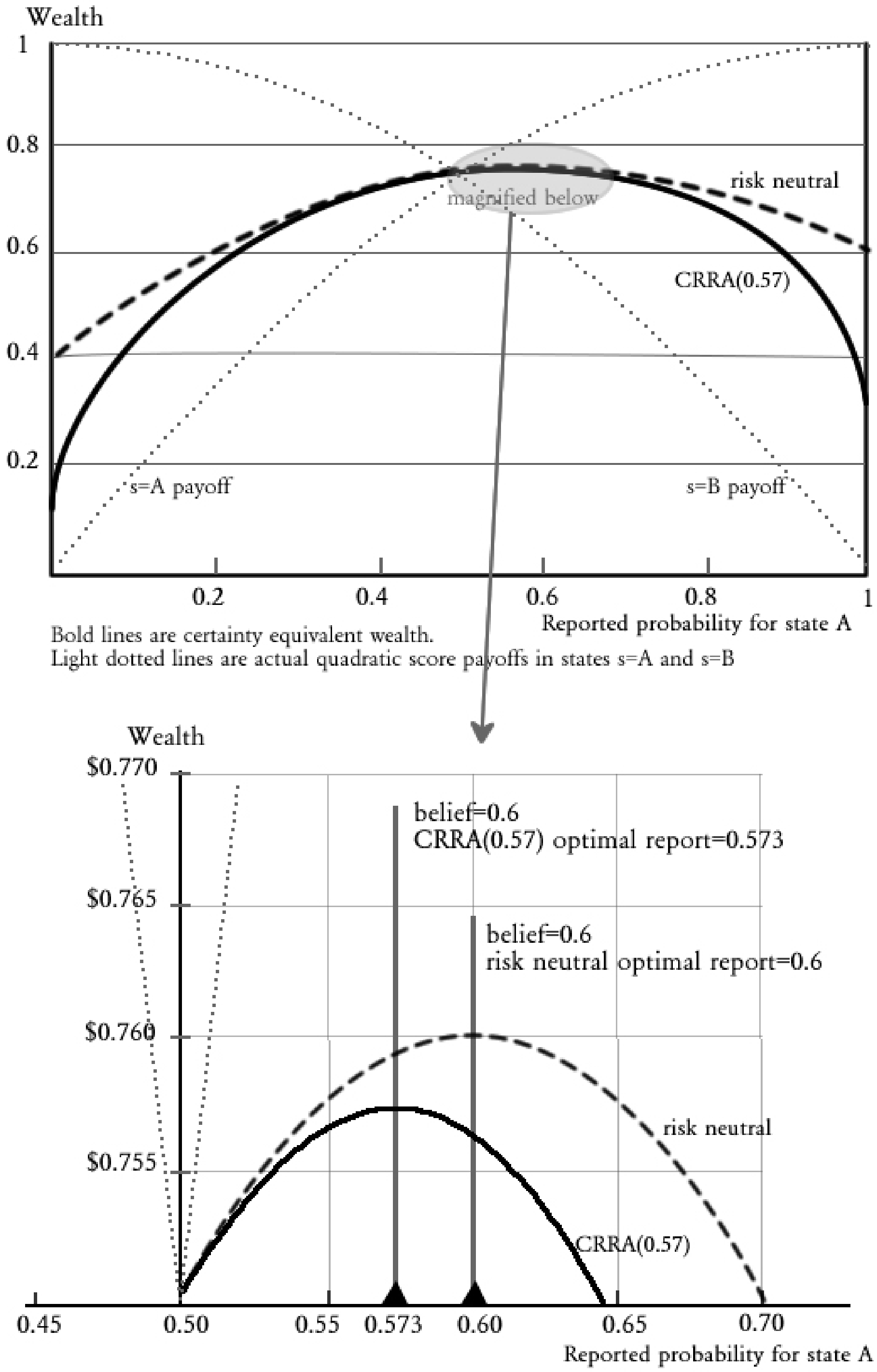


Figure 9: Opportunity Sets for Modified Scoring Rules

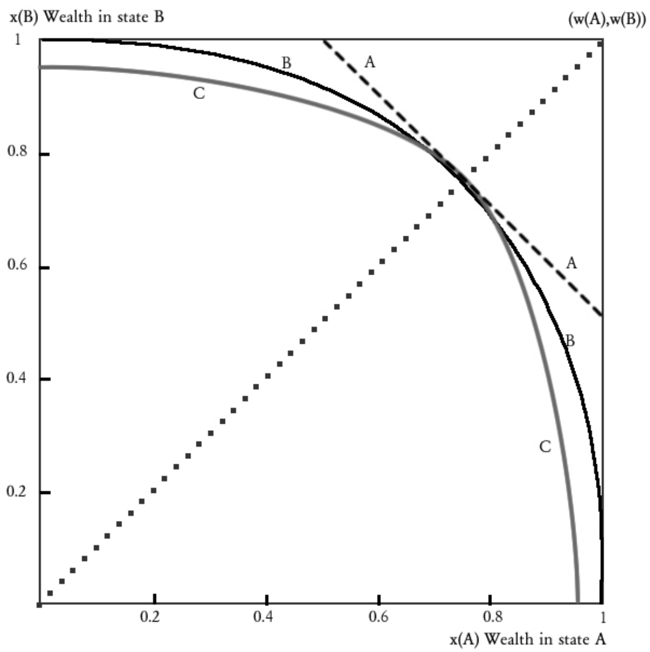
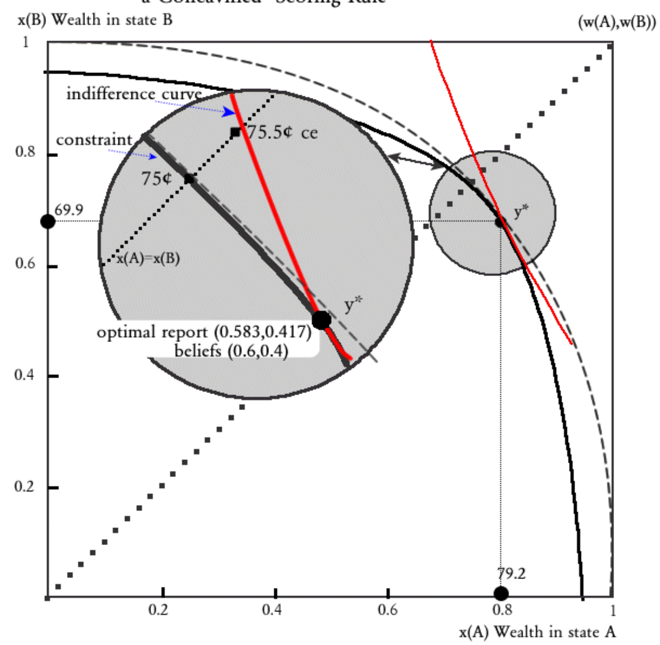


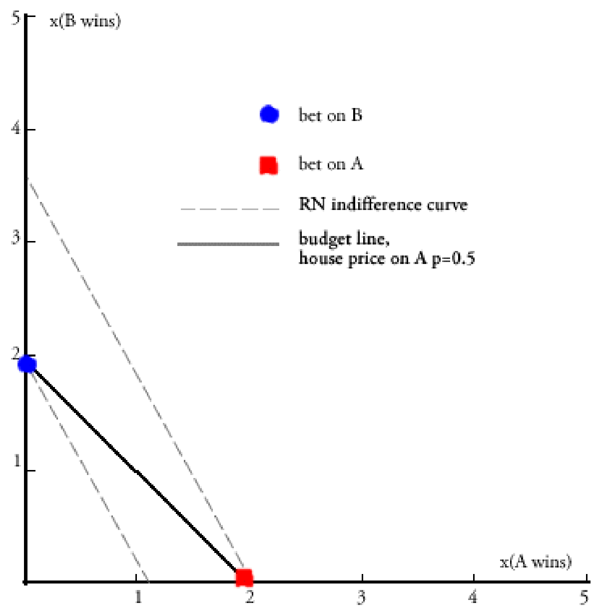
Figure 10: Optimal Behaviour for a CRRA(0.57) Agent with a Concavified Scoring Rule



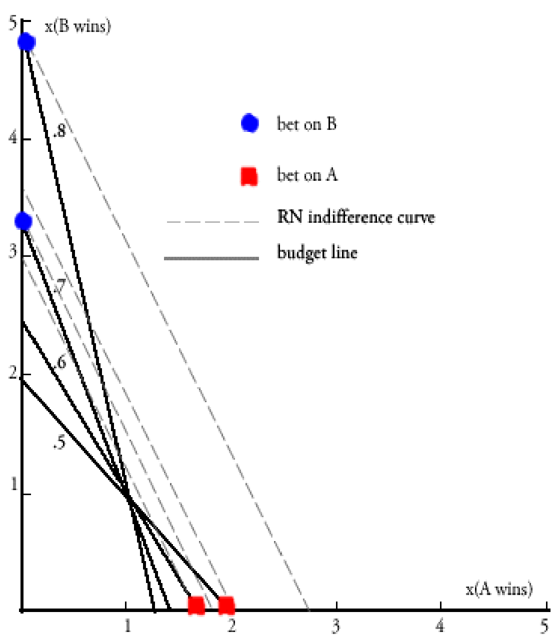
Indifference curves reflect beliefs (p(A),p(B))=(0.6,0.4)
Quadratic Scoring Rule is the dashed opportunity set.
Concavified Scoring Rule is the bold opportunity set.

Figure 11: Opportunity Sets for Experimental Design

A. Risk Neutral agent with subjective belief = 0.65, facing one bookie



B. Risk neutral and risk averse agents facing multiple bookies



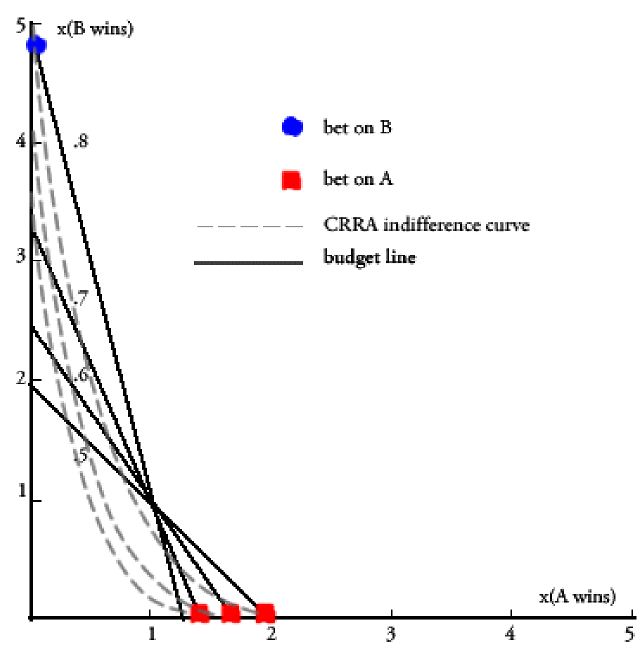


Figure 12: Elicited Bets by Four Subjects on Rumsfeld Resigning

Binary indicator of subject betting on the state or event

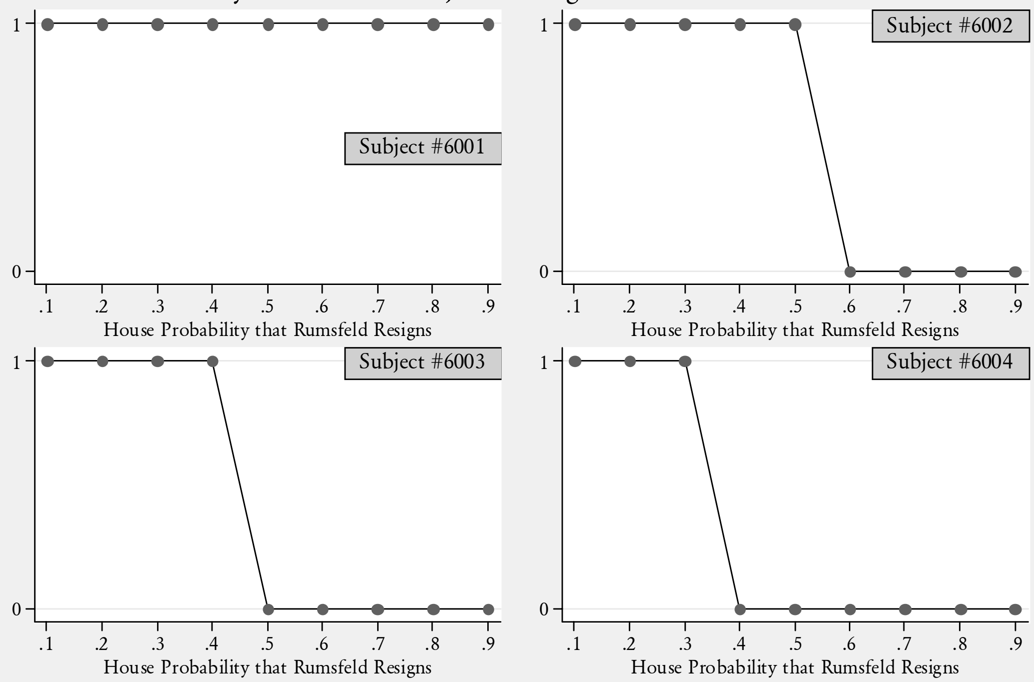


Figure 13: Elicited Bets

Fraction of subjects betting on the state or event

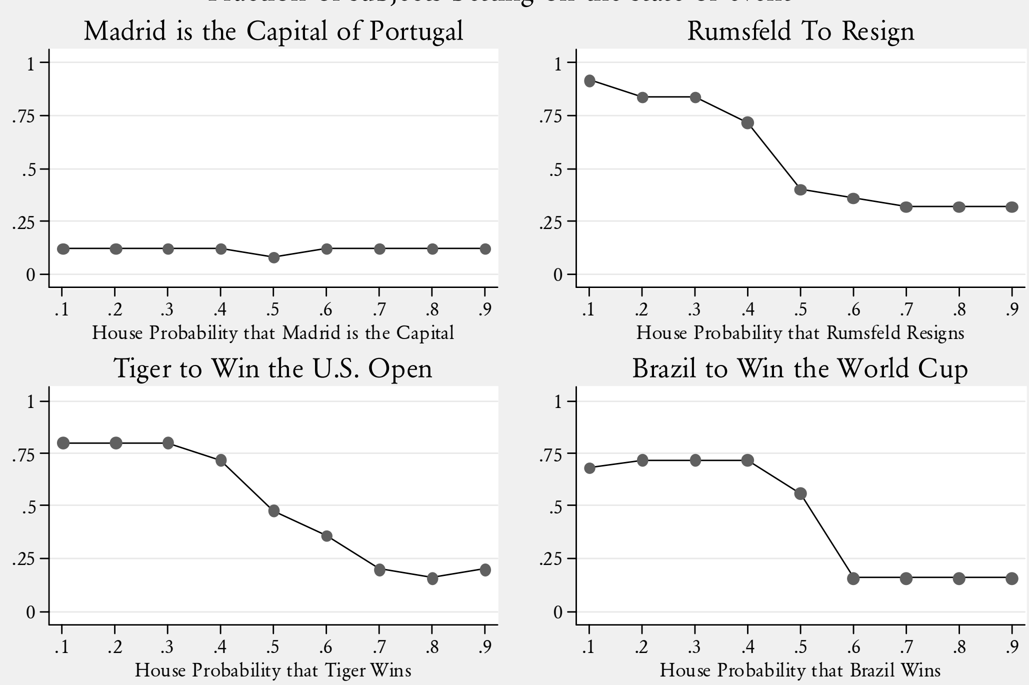


Figure 14: Probability Weighting and Decision Weights

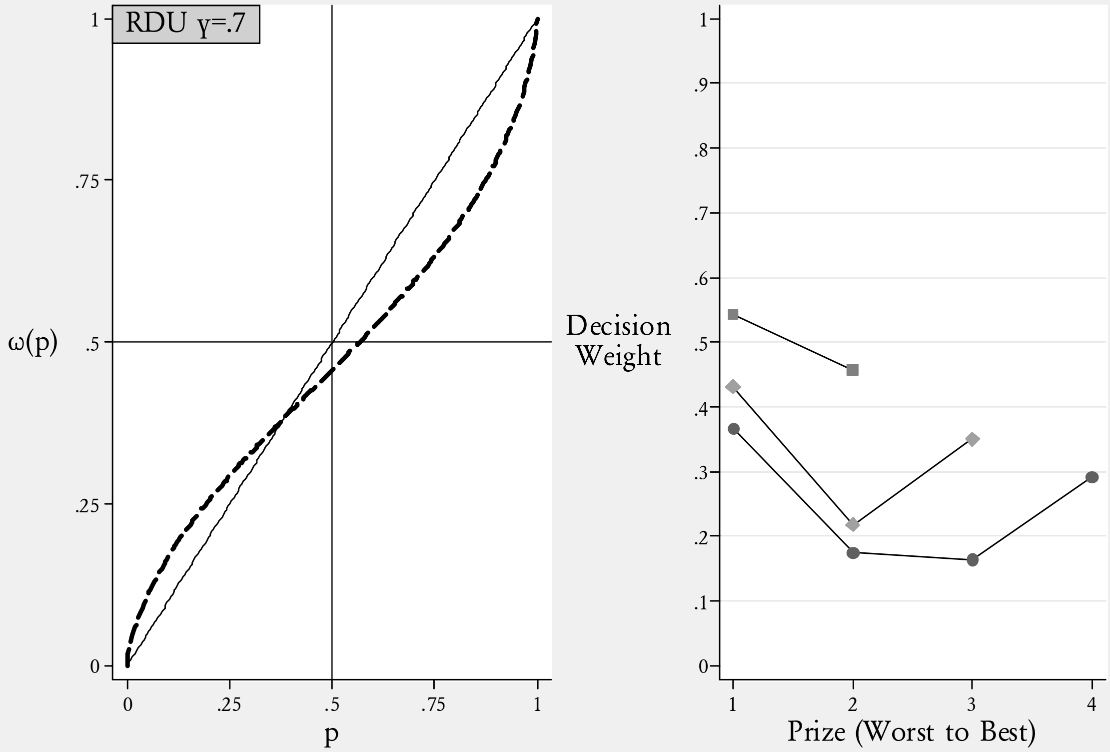


Figure 15: Elicited Risk Neutral Beliefs

Inferred from intervals that subjects switched at.
Assumes risk neutrality.

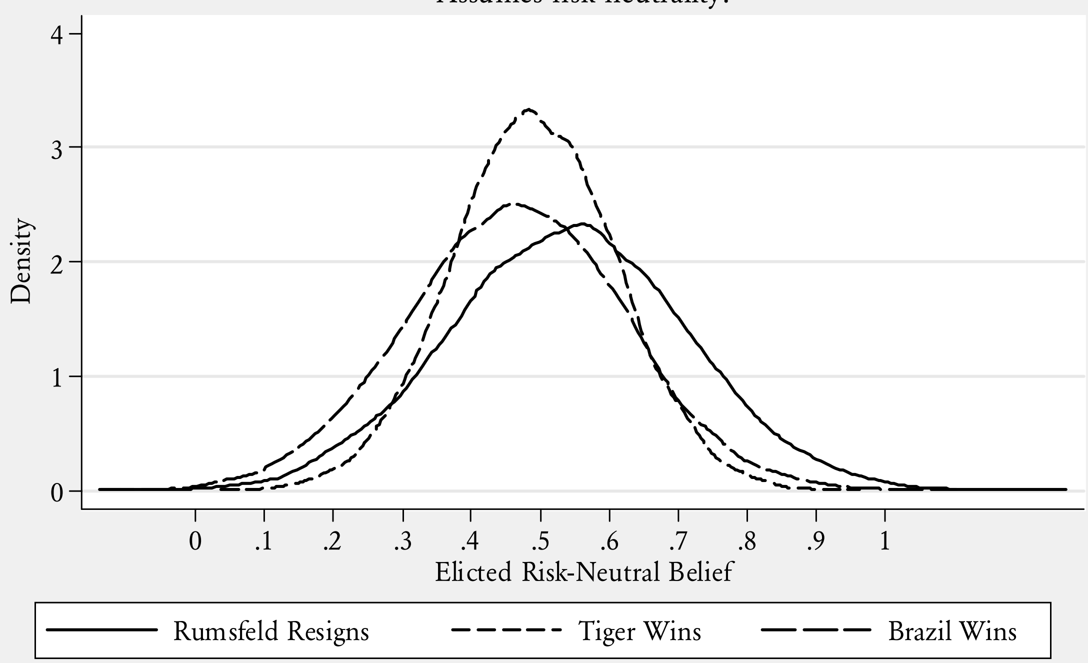


Table 2: Maximum Likelihood Estimates of Risk Neutral Subjective Beliefs

Interval regression estimates of elicited probability. N=25 subjects.

Variable	Description	Estimate	Standard Error	*p*-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
			A. Rumsfeld to Resign			
Constant		-0.10	0.18	0.59	-0.46	0.26
female	Female	-0.08	0.09	0.37	-0.24	0.09
Asian	Asian heritage	0.06	0.11	0.55	-0.15	0.27
Hispanic	Hispanic heritage	0.00	0.10	0.97	-0.21	0.20
Citizen	US citizen	0.45	0.15	0.00	0.16	0.74
Age	Years in age over 17	0.03	0.01	0.03	0.00	0.05
Business	Business	-0.21	0.09	0.01	-0.38	-0.04
Graduate	Graduate Student	0.33	0.12	0.01	0.09	0.56
GPAlow	GPA score lower than 3.24	0.33	0.10	0.00	0.14	0.52
			B. Tiger To Win			
Constant		0.53	0.23	0.02	0.08	0.99
female	Female	0.07	0.11	0.54	-0.15	0.28
Asian	Asian heritage	-0.24	0.14	0.08	-0.50	0.03
Hispanic	Hispanic heritage	-0.21	0.13	0.11	-0.46	0.05
Citizen	US citizen	0.02	0.19	0.93	-0.35	0.38
Age	Years in age over 17	0.01	0.01	0.26	-0.01	0.04
Business	Business	-0.22	0.11	0.04	-0.44	-0.01
Graduate	Graduate Student	-0.04	0.15	0.79	-0.33	0.25
GPAlow	GPA score lower than 3.24	0.13	0.12	0.26	-0.10	0.36
			C. Brazil To Win			
Constant		0.62	0.16	0.00	0.31	0.94
female	Female	0.09	0.08	0.22	-0.06	0.24
Asian	Asian heritage	-0.17	0.10	0.08	-0.37	0.02
Hispanic	Hispanic heritage	-0.20	0.09	0.02	-0.37	-0.03
Citizen	US citizen	0.01	0.13	0.92	-0.24	0.26
Age	Years in age over 17	0.01	0.01	0.38	-0.01	0.02
Business	Business	-0.32	0.08	0.00	-0.48	-0.17
Graduate	Graduate Student	0.12	0.10	0.22	-0.07	0.31
GPAlow	GPA score lower than 3.24	-0.03	0.08	0.71	-0.19	0.13

Figure 16: Distribution of Estimated EUT and RDU Parameters

Solid line is RDU, dashed line is EUT

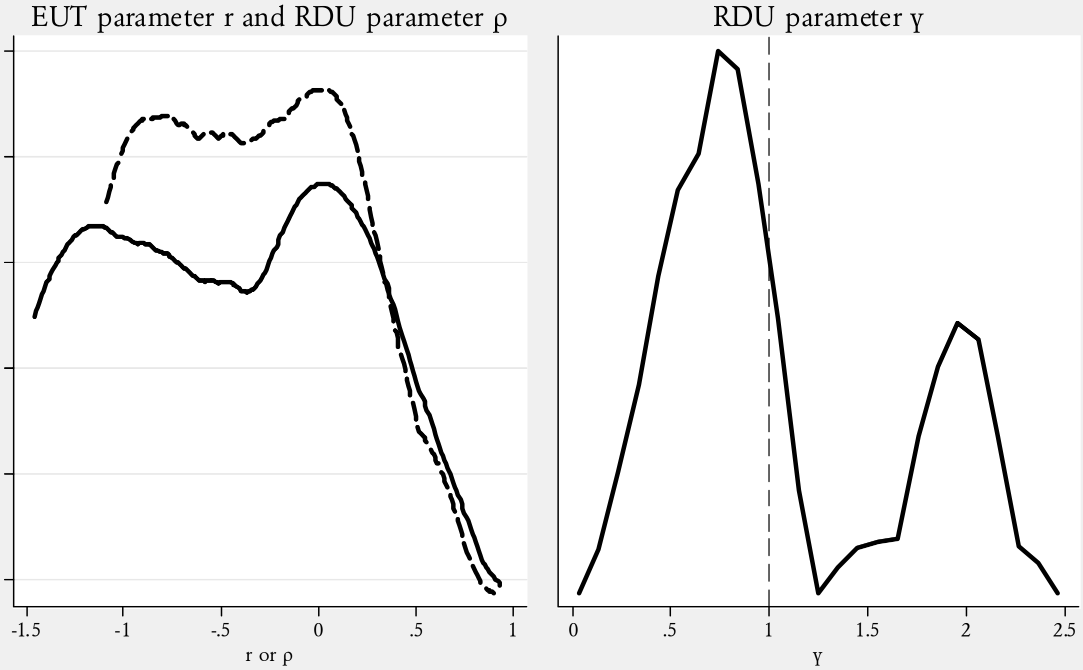


Figure 17: Effect of Calibration on Elicited Belief that Rumsfeld Will Resign

Solid line reflects calibration with Risk Neutral model

Dashed line reflects calibration with EUT or RDU model

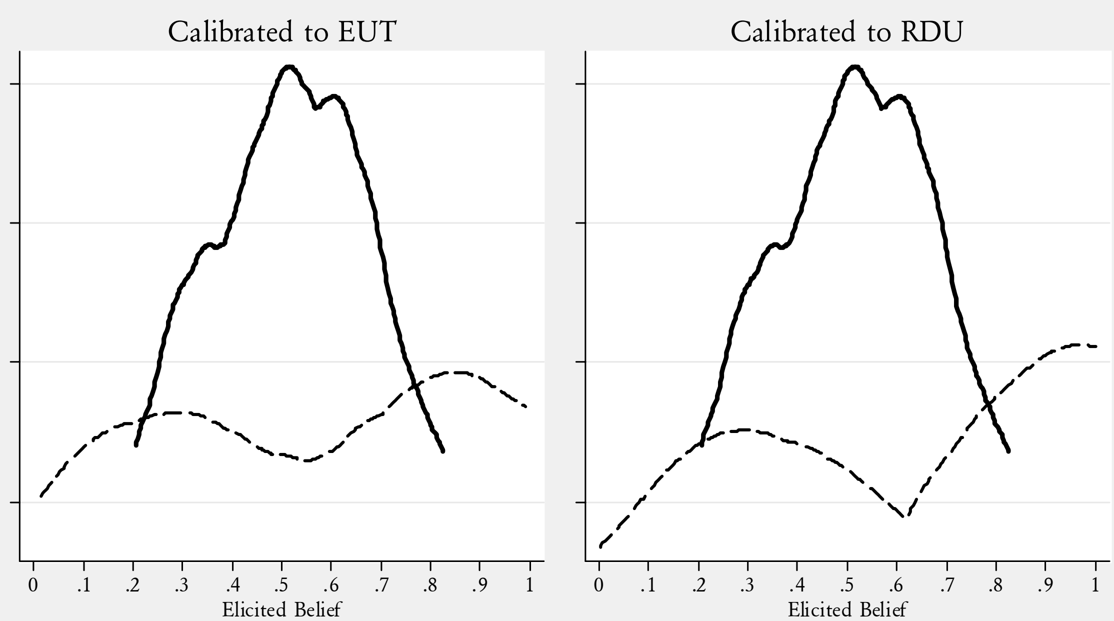


Figure 18: Effect of Calibration on Elicited Belief
that Tiger Will Win

Solid line reflects calibration with Risk Neutral model
Dashed line reflects calibration with EUT or RDU model

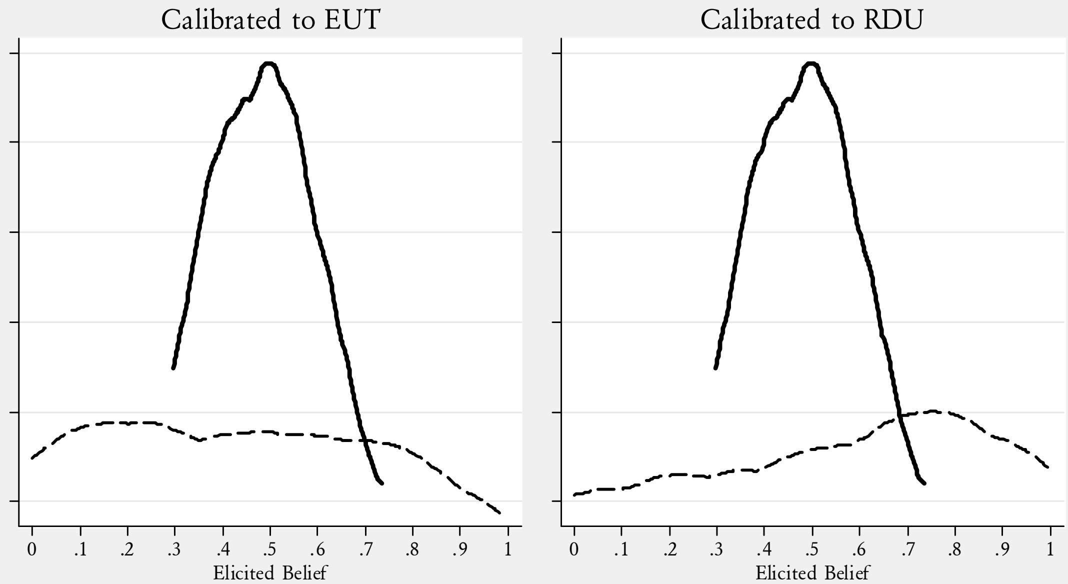


Figure 19: Effect of Calibration on Elicited Belief
that Brazil Will Win

Solid line reflects calibration with Risk Neutral model
Dashed line reflects calibration with EUT or RDU model

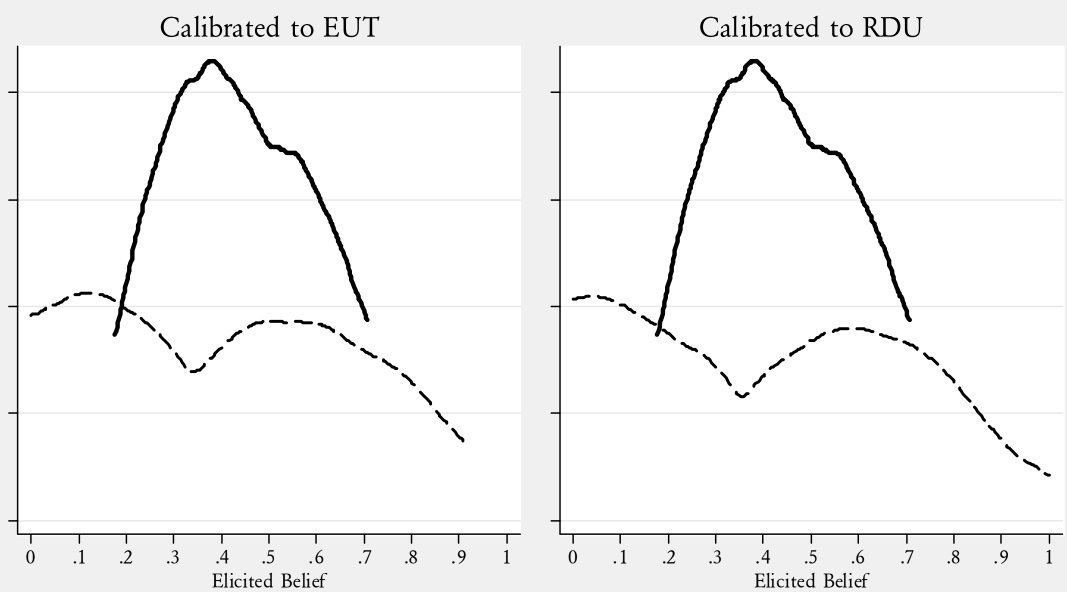


Table 3: Maximum Likelihood Estimates of EUT Subjective Beliefs

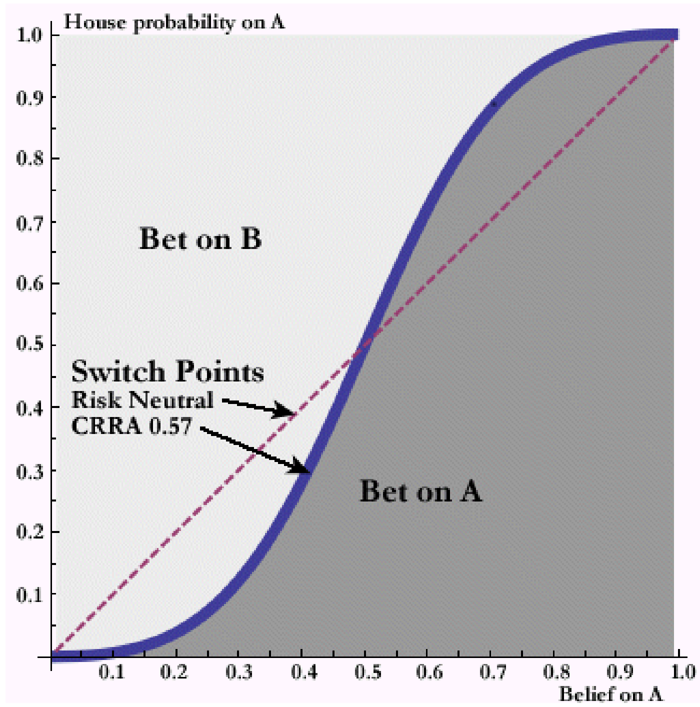
Variable	Description	Estimate	Standard Error	*p*-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
r						
Constant		-0.59	0.63	0.36	-1.85	0.68
female	Female	0.09	0.19	0.65	-0.29	0.47
Asian	Asian heritage	0.54	0.28	0.05	-0.01	1.09
Hispanic	Hispanic heritage	-0.74	0.22	0.00	-1.18	-0.30
Citizen	US citizen	0.24	0.48	0.61	-0.72	1.20
Age	Years in age over 17	0.10	0.01	0.00	0.07	0.13
Business	Business	-0.93	0.17	0.00	-1.27	-0.59
Graduate	Graduate Student	0.14	0.33	0.67	-0.51	0.79
GPAlow	GPA score lower than 3.24	-0.01	0.16	0.94	-0.33	0.30
μ	Fechner Error	34.58	18.45	0.07	-2.18	71.34
κ_R						
Constant		4.72	1.40	0.00	1.93	7.51
female	Female	0.34	0.56	0.55	-0.78	1.45
Asian	Asian heritage	-0.71	0.59	0.24	-1.89	0.48
Hispanic	Hispanic heritage	-0.96	1.97	0.63	-4.88	2.97
Citizen	US citizen	-2.20	1.51	0.15	-5.20	0.80
Age	Years in age over 17	-0.36	0.12	0.00	-0.59	-0.13
Business	Business	2.82	1.67	0.10	-0.51	6.15
Graduate	Graduate Student	-3.87	1.12	0.00	-6.10	-1.64
GPAlow	GPA score lower than 3.24	-3.91	0.90	0.00	-5.70	-2.12
κ_R						
Constant		2.12	1.16	0.07	-0.19	4.42
female	Female	-1.92	0.61	0.00	-3.13	-0.71
Asian	Asian heritage	3.56	0.84	0.00	1.89	5.22
Hispanic	Hispanic heritage	-1.02	0.72	0.16	-2.45	0.40
Citizen	US citizen	-1.07	0.92	0.25	-2.91	0.76
Age	Years in age over 17	0.69	0.29	0.02	0.10	1.27
Business	Business	1.43	0.72	0.05	0.00	2.87
Graduate	Graduate Student	-2.52	0.62	0.00	-3.76	-1.29
GPAlow	GPA score lower than 3.24	-4.47	0.65	0.00	-5.76	-3.18
κ_R						
Constant		-1.15	0.78	0.15	-2.69	0.40
female	Female	-1.75	0.57	0.00	-2.88	-0.62
Asian	Asian heritage	6.10	1.15	0.00	3.81	8.40
Hispanic	Hispanic heritage	-0.59	1.47	0.69	-3.53	2.34
Citizen	US citizen	0.94	1.42	0.51	-1.89	3.77
Age	Years in age over 17	1.24	0.27	0.00	0.70	1.77
Business	Business	0.47	1.36	0.73	-2.24	3.17
Graduate	Graduate Student	-2.13	1.32	0.11	-4.77	0.51
GPAlow	GPA score lower than 3.24	-2.85	0.81	0.00	-4.46	-1.24

Table 4: Maximum Likelihood Estimates of RDEU Subjective Beliefs

Variable	Description	Estimate	Standard Error	*p*-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
r						
Constant		0.17	0.64	0.79	-1.10	1.44
female	Female	-0.24	0.17	0.17	-0.58	0.10
Asian	Asian heritage	0.42	0.28	0.13	-0.13	0.97
Hispanic	Hispanic heritage	-0.92	0.17	0.00	-1.26	-0.58
Citizen	US citizen	-0.34	0.50	0.50	-1.34	0.66
Age	Years in age over 17	0.10	0.01	0.00	0.08	0.13
Business	Business	-1.26	0.23	0.00	-1.72	-0.79
Graduate	Graduate Student	-0.20	0.22	0.35	-0.64	0.23
GPAlow	GPA score lower than 3.24	-0.01	0.14	0.92	-0.30	0.27
γ						
Constant		0.67	0.30	0.03	0.09	1.26
female	Female	1.26	0.48	0.01	0.30	2.23
Asian	Asian heritage	-0.19	0.09	0.04	-0.37	-0.01
Hispanic	Hispanic heritage	0.15	0.12	0.23	-0.09	0.39
Citizen	US citizen	0.05	0.26	0.85	-0.47	0.57
Age	Years in age over 17	-0.01	0.02	0.42	-0.05	0.02
Business	Business	0.27	0.11	0.02	0.05	0.49
Graduate	Graduate Student	-0.29	0.18	0.11	-0.63	0.06
GPAlow	GPA score lower than 3.24	-0.02	0.21	0.91	-0.45	0.40
μ	Fechner Error	50.49	32.46	0.12	-14.17	115.14
κ_R						
Constant		-14.53	9.93	0.15	-34.30	5.25
female	Female	0.27	0.31	0.39	-0.34	0.88
Asian	Asian heritage	-0.32	0.45	0.48	-1.22	0.57
Hispanic	Hispanic heritage	0.71	0.99	0.48	-1.27	2.69
Citizen	US citizen	15.63	9.94	0.12	-4.17	35.44
Age	Years in age over 17	-0.57	0.17	0.00	-0.90	-0.23
Business	Business	7.44	2.42	0.00	2.62	12.25
Graduate	Graduate Student	-6.47	2.37	0.01	-11.18	-1.75
GPAlow	GPA score lower than 3.24	-6.77	2.34	0.01	-11.44	-2.10
κ_R						
Constant		-1.54	2.05	0.46	-5.63	2.55
female	Female	-1.90	0.66	0.01	-3.21	-0.59
Asian	Asian heritage	3.75	0.69	0.00	2.39	5.12
Hispanic	Hispanic heritage	-0.55	0.58	0.35	-1.70	0.61
Citizen	US citizen	2.11	2.02	0.30	-1.91	6.13
Age	Years in age over 17	0.67	0.22	0.00	0.23	1.12
Business	Business	3.19	1.32	0.02	0.56	5.82
Graduate	Graduate Student	-4.48	1.38	0.00	-7.22	-1.74
GPAlow	GPA score lower than 3.24	-5.78	1.49	0.00	-8.74	-2.82
κ_R						
Constant		-20.68	8.60	0.02	-37.81	-3.56
female	Female	-2.87	1.20	0.02	-5.27	-0.48
Asian	Asian heritage	10.15	3.16	0.00	3.85	16.45
Hispanic	Hispanic heritage	-1.75	1.11	0.12	-3.96	0.46
Citizen	US citizen	20.98	8.60	0.02	3.85	38.11
Age	Years in age over 17	2.23	1.03	0.03	0.17	4.28
Business	Business	-2.38	1.41	0.10	-5.19	0.42
Graduate	Graduate Student	-3.84	1.79	0.04	-7.40	-0.28
GPAlow	GPA score lower than 3.24	-2.63	0.95	0.01	-4.53	-0.73

Figure 20: Recovering Subjective Beliefs

A. Risk Averse Agent



B. Risk Loving Agent

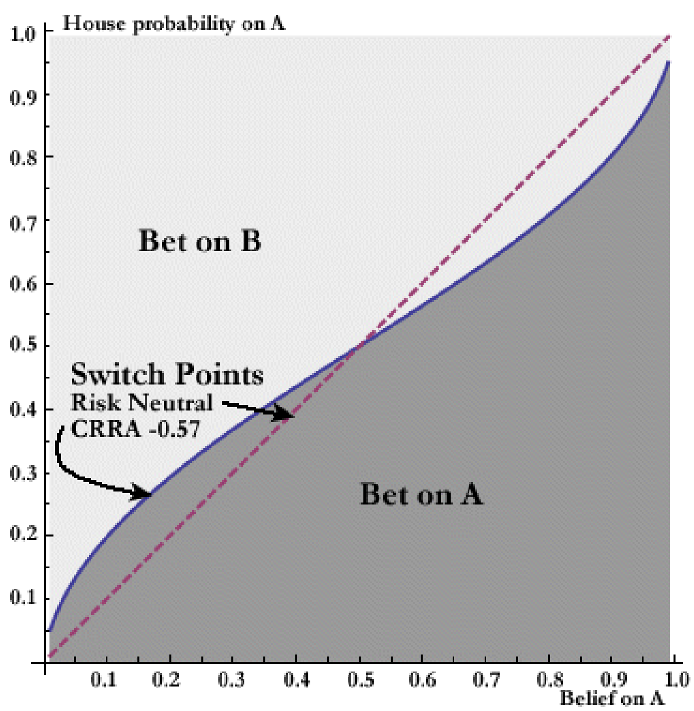


Figure 21: The Extent of the Salvage Operation

Risk averse agent with true belief ¾

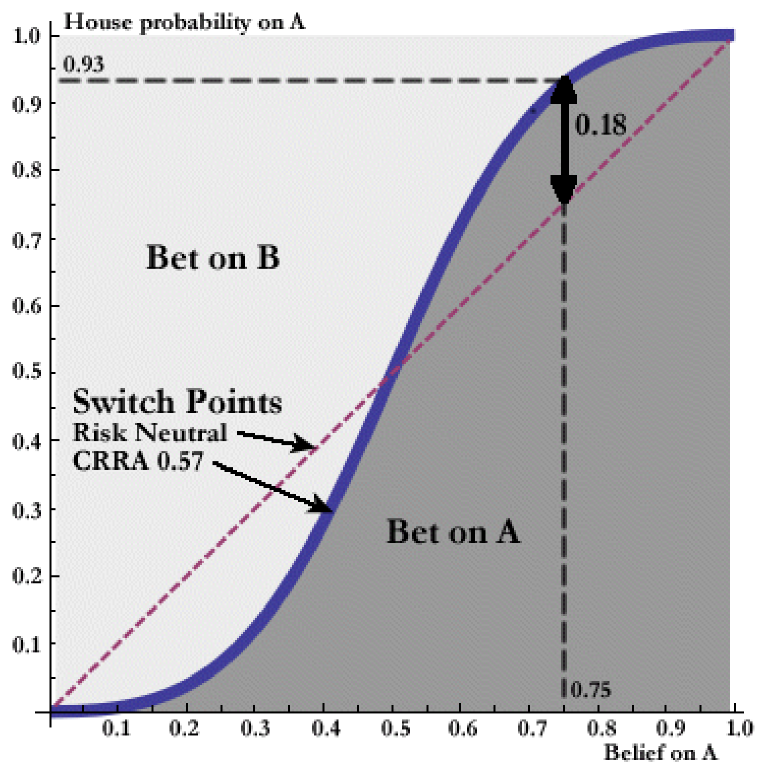
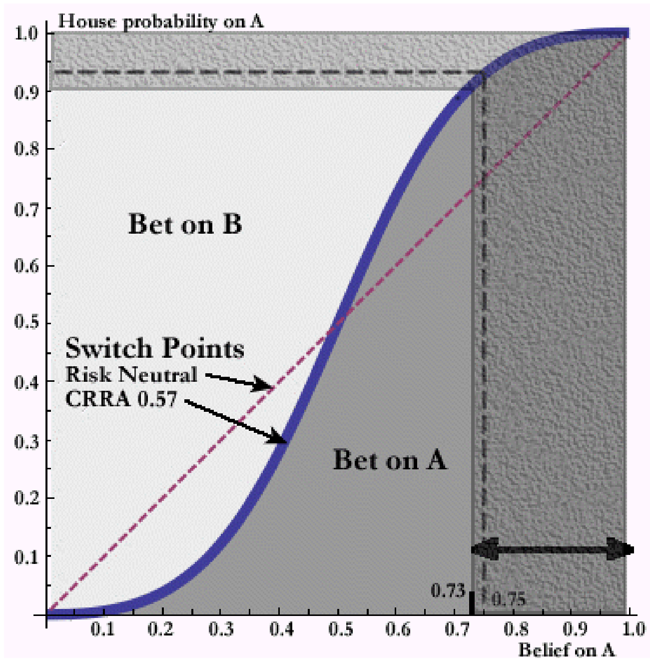
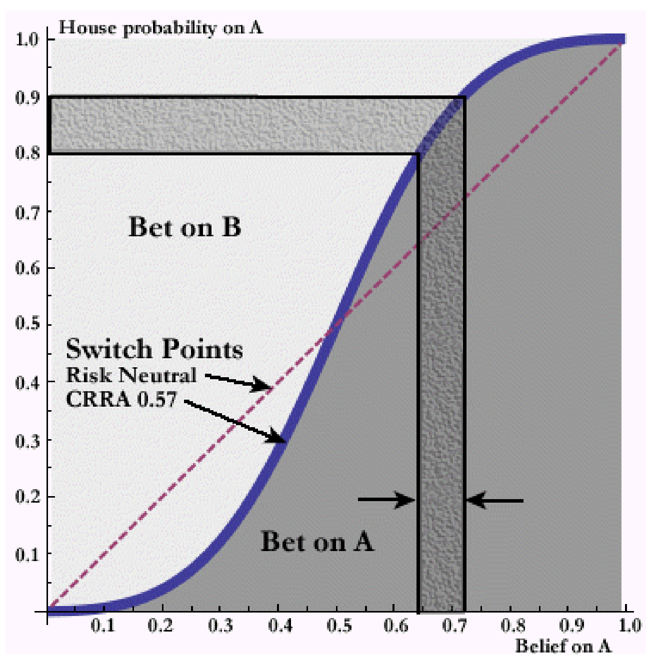


Figure 22: The Difficulty of the Salvage Operation

A. Agent observed to pick A for all house probabilities
0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7. 0.8 and 0.9



B. Agent observed to pick A for house probabilities
0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 and 0.8, and to pick B for house probability 0.9



References

Andersen, Steffen; Harrison, Glenn W.; Lau, Morten Igel, and Rutström, E. Elisabet, "Eliciting Risk and Time Preferences," *Working Paper 05-24*, Department of Economics, College of Business Administration, University of Central Florida, 2005; forthcoming, *Econometrica*.

Andersen, Steffen; Harrison, Glenn W.; Lau, Morten Igel, and Rutström, E. Elisabet, "Elicitation Using Multiple Price Lists," *Experimental Economics*, 9(4), December 2006, 383-405.

Andersen, Steffen; Harrison, Glenn W.; Lau, Morten Igel, and Rutström, E. Elisabet, "Valuation Using Multiple Price List Formats," *Applied Economics*, 39(6), April 2007, 675-682.

Andersen, Steffen; Harrison, Glenn W., Lau, Morten I., and Rutström, E. Elisabet, "Lost in State Space: Are Preferences Stable?" *International Economic Review*, 49, 2008, forthcoming.

Becker, Gordon M.; DeGroot, Morris H., and Marschak, Jacob, "Stochastic Models of Choice Behavior," *Behavioral Science*, 8, 1963, 41-55.

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

Bernardo, José, and Smith, Adrian F. M., *Bayesian Theory* (Chichester: Wiley, 1994).

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, "Does it make sense to speak of 'good probability appraisers'?" in I.J. Good (ed.), *The Scientist Speculates: An Anthology of Partly-Baked Ideas* (New York: Basic Books, 1962, 357-364).

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

Diecidue, Enrico; Wakker, Peter P., and Zeelenberg, Marcel, "Eliciting Decision Weights by Adapting de Finetti's Betting-Odds Method to Prospect Theory," *Journal of Risk and Uncertainty*, 34, 2007, 179-199.

Fechner, Gustav, *Elements of Psychophysics* (New York: Holt, Rinehart & Winston, Volume 1, 1966; originally published 1860).

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, "Measuring Risk Relatively," *Working Paper*, Department of Economics, University of

Christchurch, Canterbury, New Zealand, April 2003.

Fudenberg, Drew, and Levine, David K., "A Dual-Self Model of Impulse Control,"*American Economic Review*, 96(5), December 2006, 1449-1476.

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

Gruca, Thomas S.; Berg, Joyce, and Cipriano, Michael, "The Effects of Electronic Markets on Forecasts of New Product Success," *Information Systems Frontiers*, 5(1), 2003, 95-105.

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.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Risk Aversion and Incentive Effects: Comment," *American Economic Review*, 95(3), June 2005, 897-901.

Harrison, Glenn W.; Lau, Morten Igel, and Williams, Melonie B., "Estimating Individual Discount Rates for Denmark: A Field Experiment," *American Economic Review*, 92(5), December 2002, 1606-1617.

Harrison, Glenn W., and Rutström, E. Elisabet, "Expected Utility Theory *And* Prospect Theory: One Wedding and A Decent Funeral," *Working Paper 05-18*, Department of Economics, College of Business Administration, University of Central Florida, 2005.

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* (Greenwich, CT: JAI Press, Research in Experimental Economics, Volume 12, 2008).

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.

Jose, Victor Richmond R.; Nau, Robert F, and Winkler, Robert L., "Scoring Rules, Generalized Entropy, and Utility Maximization," *Operations Research*, 55, 2007 forthcoming.

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.

Köbberling, Veronika, and Wakker, Peter P., "An Index of Loss Aversion," *Journal of Economic Theory*, 122, 2005, 119-131.

Kreps, David M., and Porteus, Evan L., "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46(1), January 1978, 185-200.

Machina, Mark J., and Schmeidler, David, "A More Robust Definition of Subjective Probability," *Econometrica*, 60(4), July 1992, 745-780.

Manski, Charles F., "Interpreting the Predictions of Prediction Markets," *Economics Letters*, 91(3), 2006, 425-429.

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

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

Murphy, Allan H., and Staël von Holstein, Carl-Axel, "A Geometrical Framework for the Ranked Probability Score," *Monthly Weather Review*, 103, January 1975, 16-20.

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

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," *Working Paper*, Econometric Institute, Erasmus University, April 2007.

Papke, Leslie E., and Wooldridge, Jeffrey M., "Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates," *Journal of Applied Econometrics*, 11, 1996, 619-632.

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).

Rabe-Hesketh, Sophia, and Everitt, Brian, *A Handbook of Statistical Analyses Using Stata* (New York: Chapman & Hall/CRC, Third Edition, 2004).

Roberts, Harry V., "Probabilistic Prediction," *Journal of the American Statistical Association*, 60, March 1965, 50-62.

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

Schmeidler, David, "Subjective Probability and Expected Utility Without Additivity," *Econometrica*, 57(3), May 1989, 571-587.

Segal, Uzi, "The Ellsberg Paradox and Risk Aversion: An Anticipated Utility Approach," *International Economic Review*, 28(1), February 1987, 175-202.

Segal, Uzi, "Two-Stage Lotteries Without the Independence Axiom," *Econometrica*, 58(2), March 1990, 349-377.

Staël von Holstein, Carl-Axel, and Murphy, Allan H., "The Family of Quadratic Scoring Rules," *Monthly Weather Review*, 106(7), July 1978, 917-924.

Wilcox, Nathaniel T., "Predicting Individual Risky Choices Out-of-Context: A Critical Stochastic Modeling Primer and Monte Carlo Study 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* (Greenwich, CT: JAI Press, Research in Experimental Economics, Volume 12, 2008).

Winkler, Robert L., *An Introduction to Bayesian Inference and Decision* (Gainesville, FL: Probabilistic Publishing, 2003, Second Edition).

Appendix A: A Formal Model and Additional Results on Prediction Markets

We provide formal derivations of compensated demands for state contingent commodities, $x^c(s)$, for some common utility functions. In each case we represent beliefs about state $s \in S$ by vectors $b(s)>0$, $\sum_s b(s)=1$ and market prices (reports) by vectors $p(s)>0$, $\sum_s p(s)=1$. A unit of s-contingent wealth pays $1 if state s occurs and $0 otherwise, for all states s=1,2....n. We also define $\check{s} \in S$, to aid the statement of summation operations. We then use these results to generate simulation results on the extent to which prediction markets recover aggregate beliefs.

A. Formal Results

The table below shows four common utility functions (CRRA, CARA, logarithmic and Quasi-Linear), their corresponding compensated demand functions, and the scoring rules they are associated with. These scoring rules, or positive affine transformations of them, are defined in Jose, Nau, and Winkler [2007] and Lad [1996].

Utility function	Compensated demand function	Scoring rule
$u(x) = x^{(1-r)}/(1-r)$	$CE \times \{[b(s)/(p(s)] \,/\, [\sum_{\check{s}} [b(\check{s})/x(\check{s})]^{1-r/r}]^{1/1-r}\}^{1/r}$	Pseudo-spherical
$u(x) = -\exp\{-x/\eta\}$	$CE + \eta \times \ln[(b(s))/p(s)]$	Logarithmic
$u(x) = \ln(x)$	$CE \times \{[b(s)/(p(s)] \,/\, [\prod_{\check{s}} [b(\check{s})/x(\check{s})]^{b(\check{s})}\}$	Reciprocal
$u(c,x) = c+x^{(1-r)}/(1-r)$	$EU + \{[b(s)/(p(s)]^{1/r} - \sum_{\check{s}} b(\check{s})[b(\check{s})/x(\check{s})]^{(1-r)/r}\}/(1-r)$	Power

If w(s) describes the aggregate wealth available, then the scoring rule in each case pays the elicitee $w(s)-x^c(s)$for a coherent probability report p(s). A report p(s) functions as a supporting price p(s) for the compensated demands, selecting contingent commodities along the opportunity set presented by the corresponding indifference curve, subject to staying within the bounds of the available aggregate wealth, w(s) in each state.

<u>CRRA Utility and the Pseudo-Spherical Scoring Rule</u>

A risk averse CRRA agent is assumed to hold beliefs b(s) about a finite list of n states $s \in S$, such that $b(s)>0$, $\sum_s b(s)=1$. The EU for this agent is of the form

$$EU = [\sum_s b(s)\, x(s)^{(1-r)}]/(1-r) \tag{A1}$$

The corresponding certainty equivalent (CE) wealth is a weighted mean wealth of order 1-r:

$$CE = [\sum_s b(s)\, x(s)^{(1-r)/r}]^{1/(1-r)} \tag{A2}$$

This expression for the CRRA agent's CE is in standard CES form, where the elasticity of substitution $\eta = 1/r$ is the reciprocal of the Arrow-Pratt relative risk aversion parameter r, and is a measure of risk tolerance.

Consider the CRRA agent's optimal choice problem as an expenditure minimization problem:

minimize $m=\sum_s p(s)x(s)$ subject to a *certainty equivalent* constraint

$$CE^{1-r} = \sum_s b(s)\ x(s)^{(1-r)} \tag{A3}$$

The first-order conditions are, for an appropriate Lagrange multiplier λ,

$$p(s) = \lambda(1-r)\ b(s)\ (x(s)^{-r}) \tag{A4}$$

We may identify λ by using (A4) with the constraint (A3), from the following expression:

$$CE^r = \lambda(1-r)\ \{ \sum_{\check{s}} b(\check{s})\ [b(\check{s})/p(\check{s})]^{1-r/r} \}^{r/1-r} \tag{A5}$$

Substituting (A5) back into (A4) yields Hicksian compensated demand functions $x^c(s)$ for wealth in state s:

$$x^c(s) = CE \times \{[b(s)/(p(s)] \ /\ [\sum_{\check{s}} [b(\check{s})/x(\check{s})]^{1-r/r}]^{1/1-r} \}^{1/r} \tag{A6}$$

The denominator of (A6) is a risk-attitude-adjusted expected rate of return, a power mean of order (1-r)/r of individual contingent asset expected rates of return b(s)/p(s) for this CRRA agent.

As noted in the text, the quadratic scoring rule is a special case of this form in which there are only two states, so $s=\{1,2\}$, $b(s)=½\ \forall s$, $CE=¼$, $r=½$, $\omega(s)=1\ \forall s$, and the subject's reports are the market prices p(s). In this case b(s) is interpreted as a parameter, and not as true beliefs.

CARA Utility and the Logarithmic Scoring Rule

A CARA EU function, based on an increasing, concave utility of wealth function $u(x) = -\exp\{-x/\eta\}$, for $\eta>0$ and beliefs b(s), has form:

$$EU = - \sum_s b(s)\ \exp\{-x(s)/\eta\} \tag{A7}$$

The corresponding CE is

$$CE=-\eta\ \ln(-EU) \tag{A8}$$

From the first order conditions for expenditure minimization subject to an EU constraint we have:

$$p(s) = (\lambda/\eta)\ b(s)\ \exp\{-x(s)/\eta\} \tag{A9}$$

for an appropriate Lagrange multiplier λ. Solving for $x^c(s)$ by taking logs and re-arranging:

$$x^c(s) = \eta\ \ln(\lambda/\eta) + \eta\ \ln[b(s)/p(s)] \tag{A10}$$

To eliminate λ, sum (A9) over all s, use the constraint (A7) and the definition of CE in (A8), to obtain:

$$x^c(s)=CE + \eta \times \ln[(b(s))/p(s)] \tag{A11}$$

Logarithmic Utility and the Reciprocal Scoring Rule

A Logarithmic EU function, based on an increasing, concave utility of wealth function $u(x)=\ln(x)$, has form:

$$EU = \sum_s b(s) \ln(x(s)) \tag{A12}$$

The corresponding CE is defined by

$$EU = \ln(CE) \tag{A13}$$

From the first order conditions for expenditure minimization subject to an EU constraint we have:

$$p(s) = \lambda\, b(s)\, (1/x(s)) \tag{A14}$$

for an appropriate Lagrange multiplier λ. Solving for $x^c(s)$ we obtain

$$x^c(s) = \lambda\, [b(s)/p(s)] \tag{A15}$$

To eliminate λ, take the log of both sides of (A15) and substitute back into the constraint (A12) :

$$\ln(CE) = \ln(\lambda) + \sum_s b(s) \ln[(b(s))/p(s)] \tag{A16}$$

Exponentiating both sides to eliminate λ, and substituting back into (A15):

$$x^c(s) = CE \times \{[b(s)/(p(s)] \,/\, [\prod_{\check{s}} [b(\check{s})/x(\check{s})]^{b(\check{s})}\} \tag{A17}$$

Quasi-Linear Utility and the Power Scoring Rule

Consider a quasi-linear utility function $u(c, x) = c + x^{(1-r)}/1\text{-}r$ in two commodities, corn c and wealth x. Assuming corn consumption is non-contingent, so that $c(s) = c\ \forall\ s$, the expected utility function is

$$EU = c + (1/r) \sum_s b(s)\, x(s)^{(1-r)} \tag{A18}$$

For convenience let the price of corn be \$1 for sure, so that the budget constraint can be written $m=c+\sum_s p(s)x(s)$. Then the expenditure minimization problem is to minimize m subject to an EU constraint in the form of (A18). The first order conditions are, for an appropriate Lagrangian multiplier λ, that $\lambda=1$ (derivative of the Lagrangian w.r.t. c) and that $p(s) = \lambda\, b(s)\, x(s)^{-r}$ (derivatives of the Lagrangian w.r.t. each x(s)), which simplifies rather easily to

$$p(s) = b(s)\, x(s)^{-r} \tag{A19}$$

Solving for the optimal compensated demands $x^c(s)$ and c^c, we have

$$x^c(s) = \sum_s [b(s)/p(s)]^{1/r} \tag{A20}$$

$$c^c = EU + (1/r) \sum_s b(s)[b(s)/p(s)]^{(1-r)/r} \tag{A21}$$

From the quasi-linearity of utility we have u(c, x(s)) = u(c+x(s), 0). Let n denote the number of possible states. Hence as prices p(s) vary and EU from (A18) remains constant, an indifference curve (c^c, $x^c(1)$, ..., $x^c(n)$) traced out in n+1 dimensions also traces out an EU indifference curve in n-dimensional contingent commodity space ($c^c+x^c(1)/(1-r)$, ..., $c^c+x^c(n)/(1-r)$) according to the following compensated demand functions $y^c(s)$, where $y^c(s) = c^c + x^c(s)/(1-r)$ is the money-metric, equivalent contingent wealth for this agent. Substituting in (A20) and (A21) we obtain

$$y^c(s) = EU + \{[b(s)/(p(s)]^{1/r} - \sum_{\check{s}} b(\check{s})[b(\check{s})/x(\check{s})]^{(1-r)/r}\}/(1-r) \qquad (A23)$$

This has the form of a generalized power scoring rule.

B. Prediction Markets

Experimental economists were the first to develop prediction markets, originally defined over presidential elections in the United States: see Forsythe, Nelson, Neumann and Wright [1992]. There has been some controversy over the claim that these markets elicit "aggregate beliefs," normally understood to mean the average of beliefs for the population. There is no debate over whether these markets generate good predictions.[28] Instead the debate has been over the additional claim that the prices in these markets recover the mean of aggregate beliefs. Our formal derivations can be used to explore this controversy, beyond the analytical result presented in §2A.

Manski [2006] presented a simple, formal model in which a theoretical market did not elicit average beliefs. He built in asymmetry on the buying and selling side of the market in terms of point-mass beliefs, and assumed risk neutral agents with the same wealth (or finite bet constraint, which amounts to the same thing in terms of the effect on market behavior).

Gjerstad [2005] and Wolfers and Zitzewitz [2005] present models in which they relax the assumption of risk neutrality, and assume uni-modal beliefs, and find that markets can generate prices that reflect average beliefs.[29] However, their models build in homogeneity on both sides of the market, which is the key criticism of Manski [2006]. Every agent is risk averse, but has the same attitude to risk. And every agent has the same wealth level, or faces the same maximal bet constraint. So each agent chooses to bet the same amount, on both sides of the market, and one recovers a homogenous bidder model. One common feature of their models is that the log utility specification emerges as the "poster boy" of aggregate belief elicitation: deviations from log utility are associated with deviations from belief elicitation, although they each argue that these deviations are small for a wide range of belief distributions.

Consider the log utility model initially. Assume 1,000 simulated agents with beliefs distributed according to some parametric form. Initially we focus on a unimodal Normal distribution with mean 0.30 and standard deviation 0.10, and truncate at 0 the very few random draws that are negative. Let

---

[28] A separate issue is the ability of prediction markets to reliably elicit beliefs in "informationally complex" environments: see Healy, Ledyard, Linardi and Lowery [2007]. As the state-space of events grows, markets defined over each possible event-combination would be expected to become thin, resulting in unreliable information aggregation.

[29] In private correspondence Gjerstad shows that symmetric, bi-modal beliefs will also lead to near-perfect aggregate belief elicitation in his model.

wealth be distributed uniformly between 1 and 100. Initially we assume that there is no discounting, even though a key feature of prediction markets is that one places bets today for payoff in the future. We also allow for free range betting, so that agents can rationally decide if they want to place a bet or not; in the absence of discounting they always want to place a bet, but when discounting enters they may not.

Given their preferences, wealth, and beliefs, each agent evaluates their ordinary demand and decides on an optimal bet on the event or its complement. We undertake this evaluation for each agent, and then select the equilibrium price as the (unique) price at which demand equals supply. We evaluate all prices between 0.001 and 0.999 in increments of 0.001. Figure A1 shows the happy result that in this homogenous world prediction markets predict well and indeed recover the mean of aggregate beliefs perfectly.

Figure A2 relaxes some of these assumptions. The top left panel is a baseline, that repeats the assumptions from Figure A1 for reference. The top right panel allows for discounting behavior, and free-range betting. Thus agents can determine if the *present value* of the certainty-equivalent of their optimal bet is less than their current wealth, which is non-stochastic, and decide not to participate in the market. We select discount rates uniformly between 1% and 25%. Again we find that prediction markets do just fine.

But then we introduce some bias, in the bottom left panel of Figure A2, in favor of the "optimistic" side of the market. We define the optimistic side as those agents that have beliefs which are greater than the average belief. In this case we allow them to have greater wealth, and to be more patient, than the baseline assumptions for them.[30] These assumptions move the equilibrium price in favor of their beliefs, and away from the average of beliefs. The bottom right panel combines discounting and bias, for the same qualitative result. The difference between equilibrium price and average belief is not quantitatively large, but it is clear.

Figure A3 repeats this exercise, but substituting CRRA utility for the log utility specification. We draw CRRA values from a uniform distribution defined over the open interval (0, 1), to correspond to modest amounts of risk aversion; log utility, of course, is the special case of CRRA when the CRRA parameter tends to 1, and risk neutrality is the case of the CRRA parameter being 0. The results are virtually the same as for log utility in Figure A2, but the inaccuracy of the equilibrium price is increasing compared to the log utility case.

Figure A4 repeats the simulation from Figure A2, using log utility, but with an asymmetric, bi-modal distribution generated as a linear average of two normal distributions. We now see even larger differences between the equilibrium price and the average belief. Figure A5 then considers a completely diffuse distribution of beliefs, generated uniformly between 0.01 and 0.99, and shows a similarly difference between equilibrium price and average belief.

Taking all these results together, and in conjunction with those of the earlier literature, we conclude that prediction markets can be expected to do a good job recovering the average of aggregate beliefs under certain circumstances: unimodal distributions of beliefs, with no *a priori* reason to expect

[30] Specifically, the have 100 units of wealth more than the baseline, and their discount rates are cut in half.

heterogeneity on either side of the market. Indeed, this environment might characterize many interesting settings, such as political elections or closed prediction markets in which there is minimal sample selection into the market (on the basis of beliefs, preferences and endowments). But the result is not general, and it is easy to construct examples in which prediction markets do a predictably poor job of recovering average beliefs.

Additional Reference

Gjerstad, Steven, "Risk Aversion, Beliefs, and Prediction Market Equilibrium," *Unpublished Manuscript*, Economic Science Laboratory, University of Arizona, January 2005.

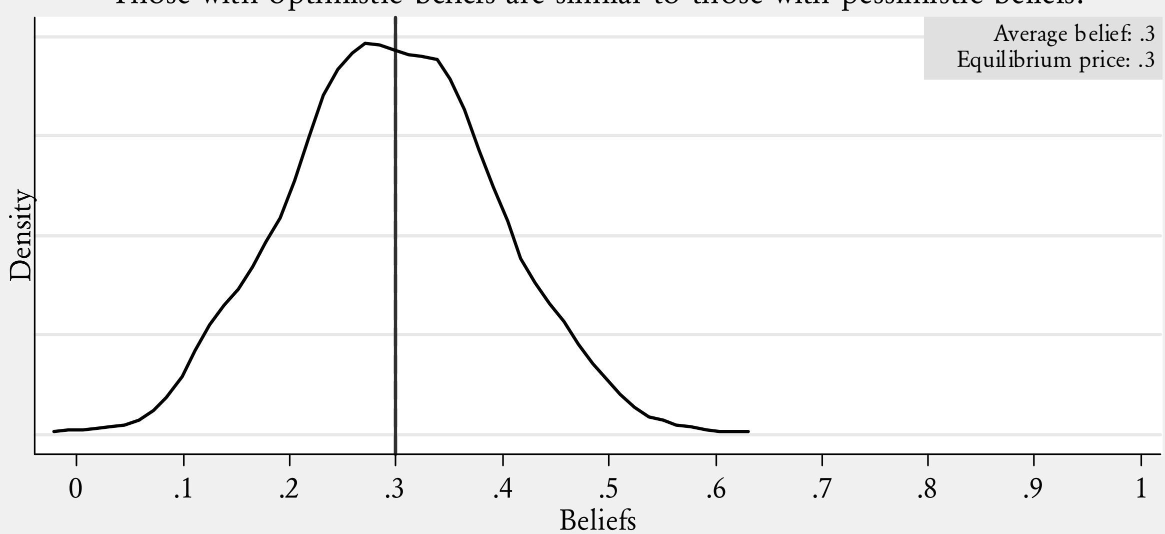
Healy, Paul J.; Ledyard, John O.; Linardi, Sera, and Lowery, J. Richard, "Information Aggregation Mechanisms and Environmental Complexity," *Working Paper*, Department of Economics, Ohio State University, October 2007.

Lad, Frank, *Operational Subjective Statistical Methods: A Mathematical, Philosophical and Historical Introduction* (New York: Wiley, 1996).

Wolfers, Justin, and Zitzewitz, Eric, "Interpreting Prediction Market Prices as Probabilities," *Unpublished Manuscript*, The Wharton School, University of Pennsylvania, February 2005.

Figure A1: Log Utility Agents in a Homogeneous Prediction Market

Solid vertical line is mean of true beliefs. Dashed vertical line is equilibrium price.
Simulations with 1000 agents and assuming free range betting.
Fraction of agents with CRRA utility is 0; fraction with log utility is 1.
Those with optimistic beliefs are similar to those with pessimistic beliefs.



Parameterization for all agents: wealth distributed uniformly between 1 and 100.
risk attitudes derived from log utility function;
no discounting of future payoffs assumed.

Figure A2: Prediction Markets with Log Utility

Solid vertical line is mean of true beliefs. Dashed vertical line is equilibrium price.

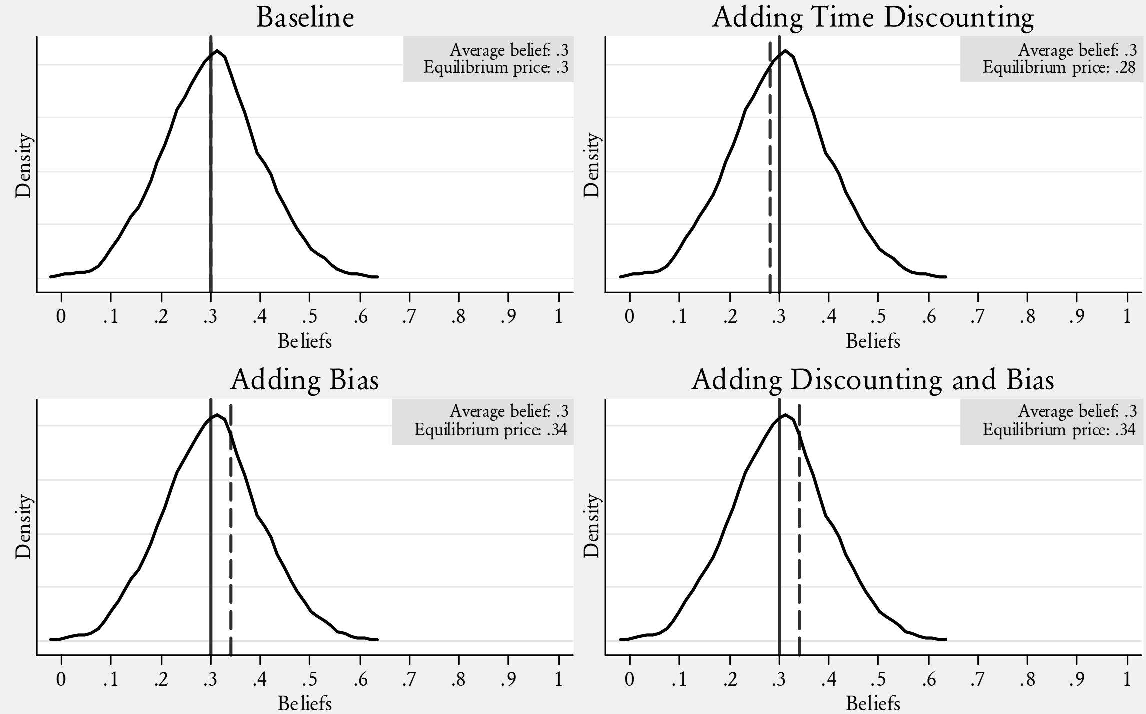


Figure A3: Prediction Markets with CRRA Utility

Solid vertical line is mean of true beliefs. Dashed vertical line is equilibrium price.

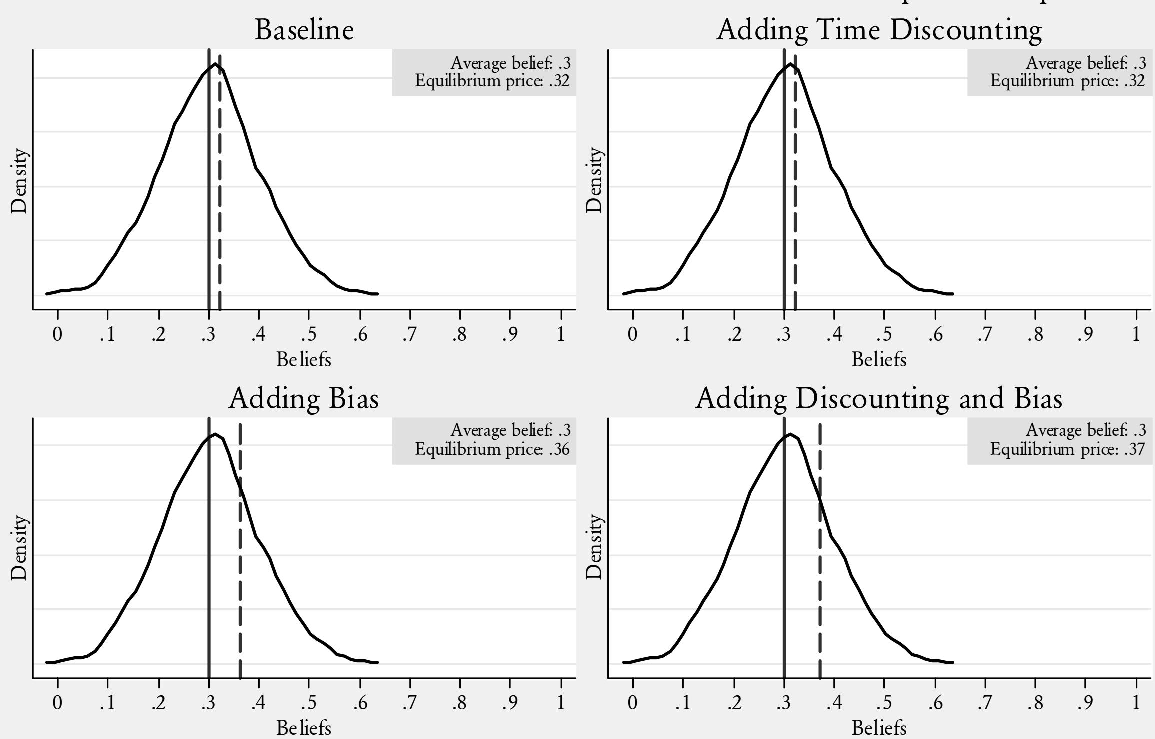


Figure A4: Prediction Markets with Log Utility And Asymmetric, Bi-Modal Aggregate Beliefs

Solid vertical line is mean of true beliefs. Dashed vertical line is equilibrium price.

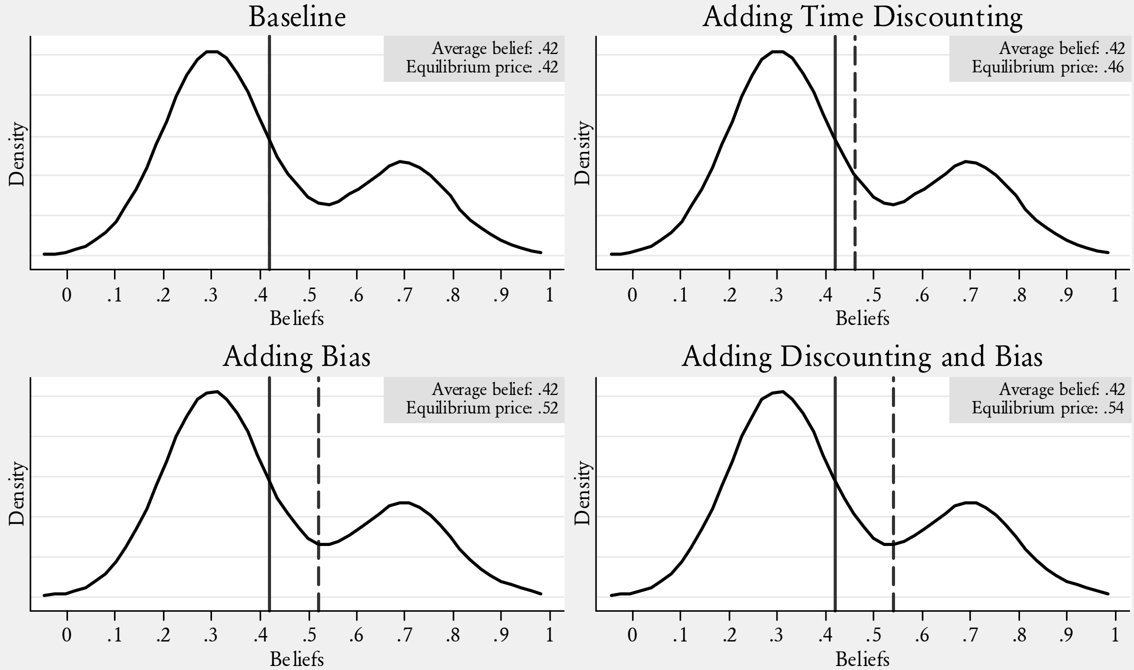
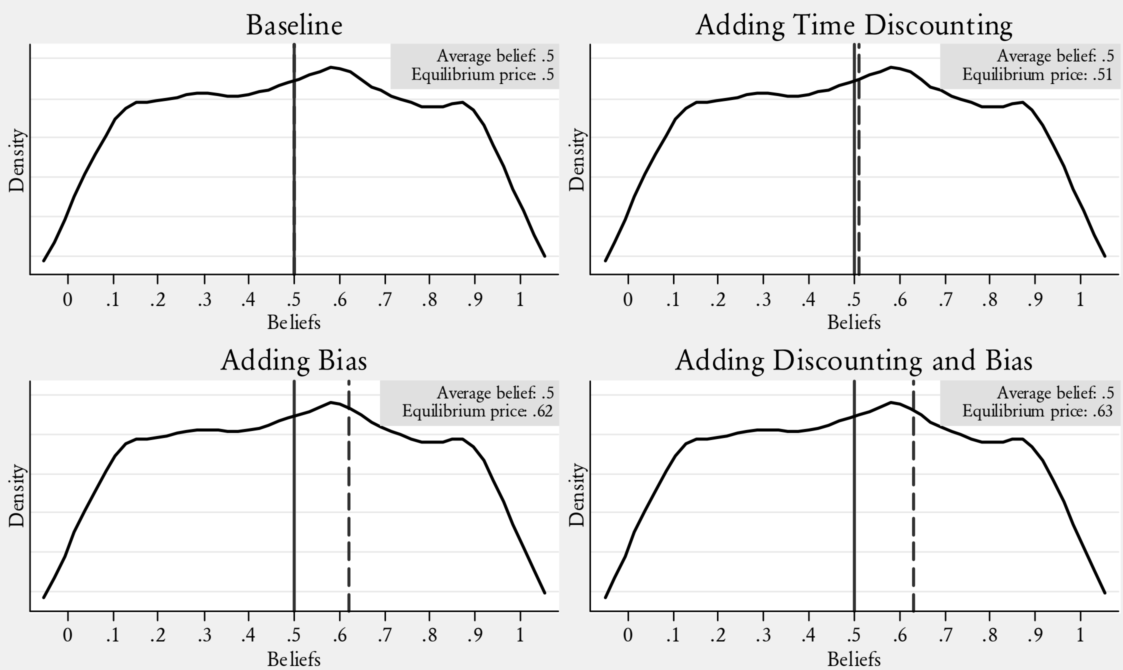


Figure A5: Prediction Markets with Log Utility And Diffuse Aggregate Beliefs

Solid vertical line is mean of true beliefs. Dashed vertical line is equilibrium price.



Appendix B: Experimental Instructions

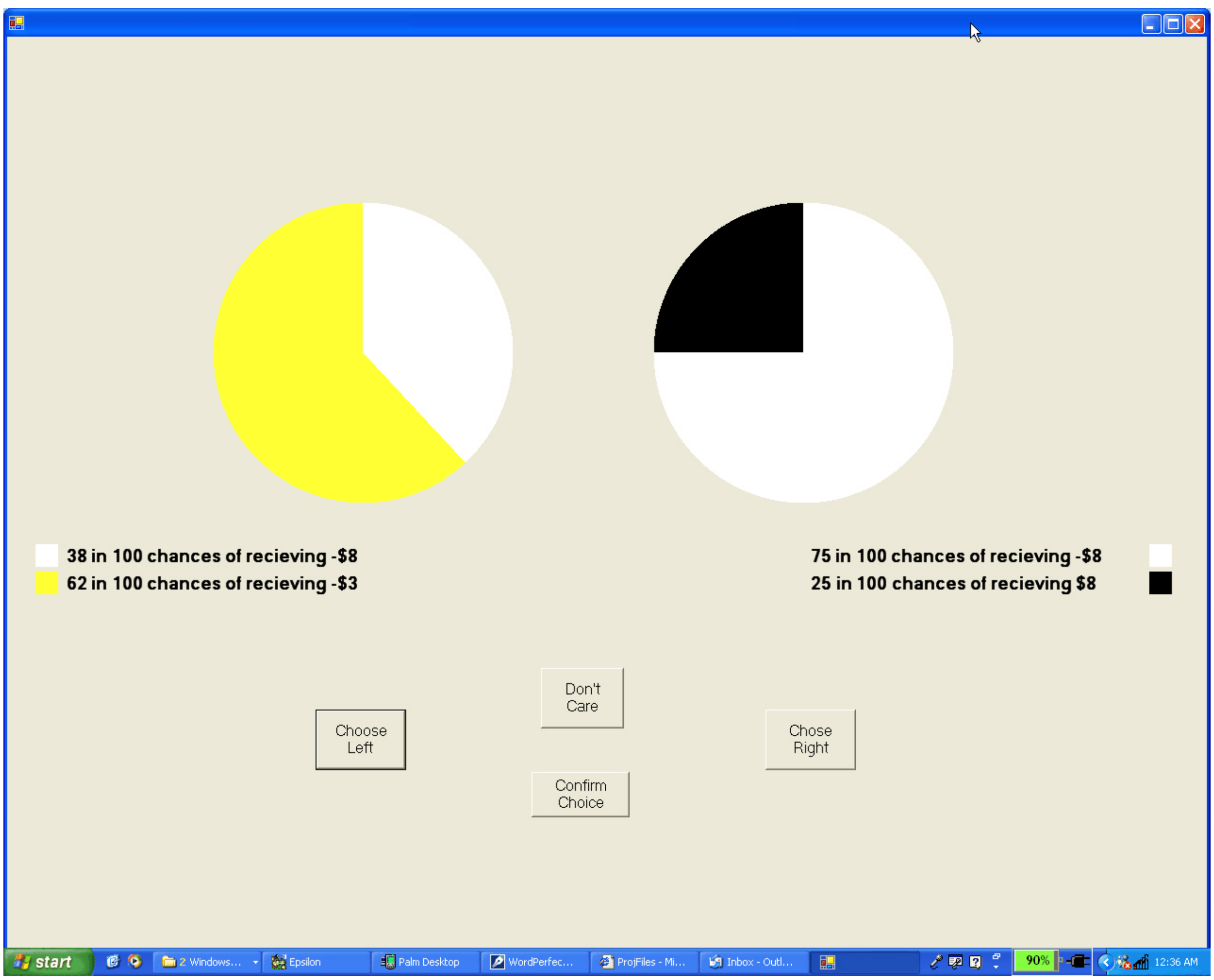
A. Binary Lottery Choice Task

YOUR INSTRUCTIONS

This is an experiment about choosing 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 60 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 three of the lotteries you choose, and will be paid according to the outcome of those lotteries, so you should think carefully about which lotteries you prefer.

You will be given $24 before you make a choice. Any losses will be deducted from that $24. You cannot lose more than $24, no matter what choice you make in this experiment.

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 a special die that has all the numbers 1 through 100 on it.

In the above example the left lottery pays minus eight dollars (-$8) if the random number on the ball drawn is between 1 and 38, and pays minus three dollars (-$3) if the number is between 39 and 100. The yellow color in the pie chart corresponds to 62% of the area and illustrates the chances that the ball drawn will be between 39 and 100 and your payoff will be -$3. The white area in the pie chart corresponds to 38% of the area and illustrates the chances that the ball drawn will be between 1 and 38 and your payoff will be -$8.

We have selected colors for the pie charts such that a lighter color indicates a lower prize. White will be used when the prize is minus eight dollars (-$8).

Now look at the pie in the chart on the right. It pays minus eight dollars (-$10) if the number drawn is between 1 and 75, and eight dollars ($8) if the number is between 76 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 -$8 pie slice is 75% 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 one of the three 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 standard six sided die, where the numbers 1-3 correspond to the left lottery and the numbers 4-6 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 a 20 sided die three times to determine which pairs of lotteries that will be played out. One lottery pair from the first 20 pairs will be selected, one from the next 20 pairs, and finally one from the last 20 pairs. If you picked DON'T CARE for one of those pairs, you will use the six sided die to decide which one you will play. Finally, you will roll the 100 sided die 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 was 37, you would lose $8; if it was 93, you would lose $3. If you picked the lottery on the right and drew the number 37, you would lose $8; if it was 93, you would gain $8.

Therefore, your payoff is determined by three things:

- by which three lottery pairs that are chosen to be played out in the series of 60 such pairs using the 20 sided die;
- by which lottery you selected, the left or the right, for each of these three pairs; and
- by the outcome of that lottery when you roll the 100 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 difference lotteries, 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. In addition to these earnings you will also receive a randomly determined amount of money before we start. This can be any amount between $1 and $10 and will be determined by rolling the 10 sided die. You will also receive $24 to cover any losses.

When you have finished reading these instructions, please raise your hand. We will then come to you, answer any questions you may have, let you roll the 10 sided die for the additional money, and start you on a short series of practice choices. The practice series will not be for payment, and consists of only 6 lottery pairs. After you complete the practice we will start you on the actual series for which you will be paid.

As soon as you have finished the actual series, and after you have rolled the necessary dice, you will be paid in cash and are free to leave.

Practice Session

In this practice session you will be asked to make 6 choices, just like the ones that you will be asked to make after the practice. You may take as long as you like to make your choice on each lottery.

You will not be paid your earnings from the practice session, or incur any losses, so feel free to use this practice to become familiar with the task. When you have finished the practice you should raise your hand so that we can start you on the tasks for which you will be paid.

YOUR INSTRUCTIONS

This is an experiment that allows you to place bets on the outcome of events that will happen in the future. For example, who will be the next U.S. President? Or who will win the next U.S. Masters Golf Championship? Or will interest rates in the United States go up? You can make more money if you have better estimates about these outcomes.

Our bets today will be similar to those that you can find in any number of betting houses around the world. So we begin by showing you how some of these houses offer their bets. *William Hill* is one of the world's largest betting houses, and is based in the United Kingdom. They offer legal off-course betting on a wide range of events, illustrated on their web site at http://www.willhill.com. For example, here are the odds that they had on May 23, 2006, for the first question posed above: who will be the next U.S. President?

Competitor	Price	Unit Stake
Hilary Clinton	3.00	
John Mccain	6.00	
John Edwards	8.00	
Mark Warner	11.00	
G Allen	11.00	
Rudolph Giuliani	15.00	
C Rice	15.00	
Bill Frist	26.00	
Mitt Romney	34.00	
Evan Bayh	34.00	
Arnold Schwarzenegger	34.00	
Sam Brownback	41.00	
Russ Feingold	41.00	
John Kerry	41.00	
Jeb Bush	41.00	

You see that the odds that they quote imply that Hilary Clinton is the favorite, followed by John McCain and then John Edwards. Clinton is listed as having a price of 3.00. This means that for every $1 that you bet on Clinton winning, you will get back $3 if she does win – but you lose your $1 if she does not win. If you think that Clinton is virtually certain to win, this is a good bet for you to take on. But if you think that Clinton has no hope of winning, you would not take this bet.

The odds quoted here suggest that it is less likely that McCain or Edwards will win. If McCain wins you would get $6 for every $1 bet, compared to $3 if you bet on Clinton and she won. Similarly, if Condaleeza Rice wins you would get $15 for every $1 bet.

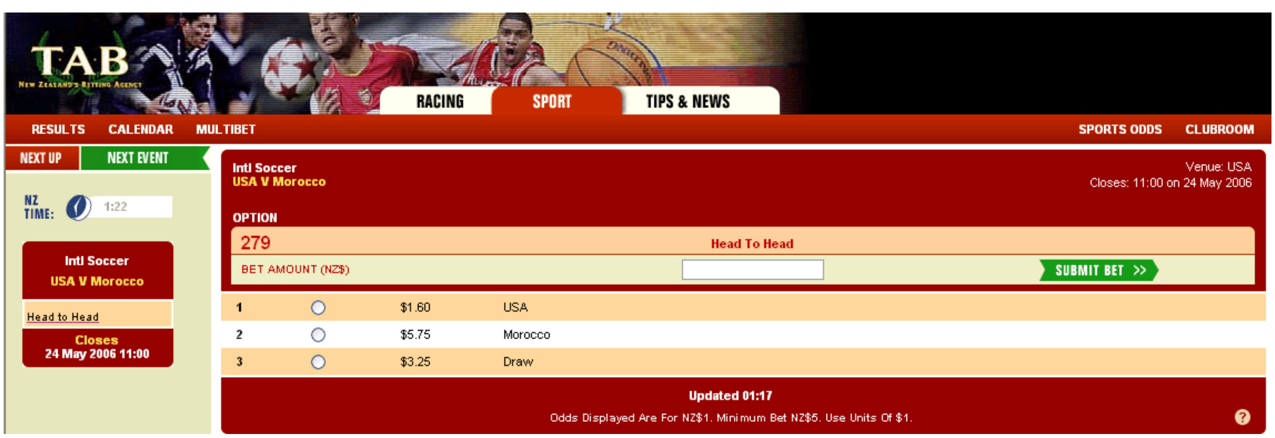
You can also see on the left hand side that they offer bets in a wide range of sports, including blockbuster events such as Hurling and Darts! They even have some special events in the entertainment field. Here are the odds that Taylor Hicks and Katherine McPhee will win the *American Idol* competition on TV:

Specials : Outright			
	Bet Until : 22:00 23/05/2006		
	Competitor	**Price**	**Unit Stake**
	Taylor Hicks	1.40	
	Katherine Mc Phee	2.75	
Win Only			Bet Now

So Hicks was the favorite a few hours before the show aired. If you bet $10 on Hicks and he won you would earn $14.00 in winnings. If you bet $10 on McPhee and she won you would earn $27.50. Of course, this final is now over and you probably know that Hicks won, but these were the odds being offered before the show aired. Do you believe that these were good odds to bet on or not? That is what our experiment asks you to decide for several events yet to occur.

Another popular betting site is *CentreBet*, at http://sports.centrebet.com/. This is a large betting house based in Australia. Here you can bet on a similar range of sporting events from around the world, as well as some local Australian sports. For example, the screen shot below shows their odds for the round #9 match between Hawthorn and Sydney in the Australian Football League. The odds are $3.50 for Hawthorn to win and $1.30 for Sydney to win, each on a wager of $1. The orange box on the right shows the bet and payout for each outcome.

One last example is the *New Zealand TAB* (Totalisater Agency Board), which is in New Zealand of course. Their betting web sit is at http://www.tab.co.nz/. They are offering the following odds on a soccer match between the United States and Morocco, a warm-up for the World Cup later this year:



So these odds suggest that the *NZ TAB* believes that the US is more likely to win than Morocco is, and that a draw is more likely than Morocco winning.

As you can see, there are many betting houses around the world. Each one offers odds on a wide range of outcomes, including local events and events in other countries. It is common to see some differences in the odds quoted by different betting houses.

To illustrate, consider bets on which U.S. NBA basketball team is going to win the NBA Championship for 2006. This event attracts odds-makers from many betting houses. The web site http://www.oddschecker.com/ lists odds for the same event from most of the big online betting houses. Here are the odds being quoted for the top 3 teams. The teams are listed in each row, and each betting house is listed in each column.

We see that there are differences in the odds quoted by different betting houses. For comparability, these odds are all taken for the same day and time: May 24 at 5pm. They are also taken for the same level of wager, shown here as £100 (about $187), since the odds can vary with the size of the bet. The odds for the Dallas Mavericks range from $2.49 up to $3.04 for every $1 wagered. The odds for the Detroit Pistons range from $1.98 up to $2.88 for every $1 wagered. And the odds for the Miami Heat range from $3.50 up to $5.20 for every $1 wagered.

selections	sportingodds	totalbet	William Hill	skybet	Ladbrokes	Coral	vc bet	Paddy Power	ukbetting	bluesq.com	bet365	betdirect	Pinnacle	Premier Bet	totesport	Stan James	Betfred	expekt.com	extrabet	betfair	BetDAQ	Mansion	iBetX	Sporting Index	Bet Basket
Dallas Mavericks	2.5	2.5			2.75	2.63			2.5		2.7		2.49		2.5	2.75	2.63		**2.77**	3.04	2.78	2.55			
Detroit Pistons	2.75	2.63			**2.88**	2.75			2.63		2.7		2.65		2.63	2.75	2.63		2.5	2.08	1.98	**2.7**			
Miami Heat	4.3	4.3			3.5	4			4.3		4		**5.2**		4.3	3.5	4		4.7	7.5	<6.23	5.3			
Phoenix Suns	9	9			7	7			9		8		**10.9**		10	8	9		7.6	10	<8.76	**11**			
each-way terms	1 1/1										1 1/1														

The betting odds we show come directly from the online bookmakers. Whilst every effort is made to ensure that the betting odds are correct, it is your responsibility to check before you place a bet.

Lets go through an example of the betting tasks we will ask you to do. These are patterned after the display shown above for the *NZ TAB*. All of the screen shots below were taken from the betting screens you will use. The choices were ones we made just to illustrate. Your computer will provide better screen shots that are easier to read.

Our betting tasks will differ in several ways:

- **First**, we will provide you with odds from several betting houses. There may just be one betting house, or there may be several. Think of these as just alternative places you can place your bet. You need to decide what to bet for each betting house.

- **Second**, you will have a stake of $10 to bet with for each betting house. In some cases you will have to use the whole $10 to place a bet. In other cases you can decide what fraction of the $10 to bet, or bet none of it. The computer will tell you which case applies to you today. You cannot use your stake in one betting house and apply it to a bet in another betting house.

- **Third**, if there are several betting houses offering you odds, we will pick one of them to actually pay you. You will be asked to "place your bets," or "not bet," for each house. Then we will pause and come over and you can roll a die to decide which betting house we will actually use.

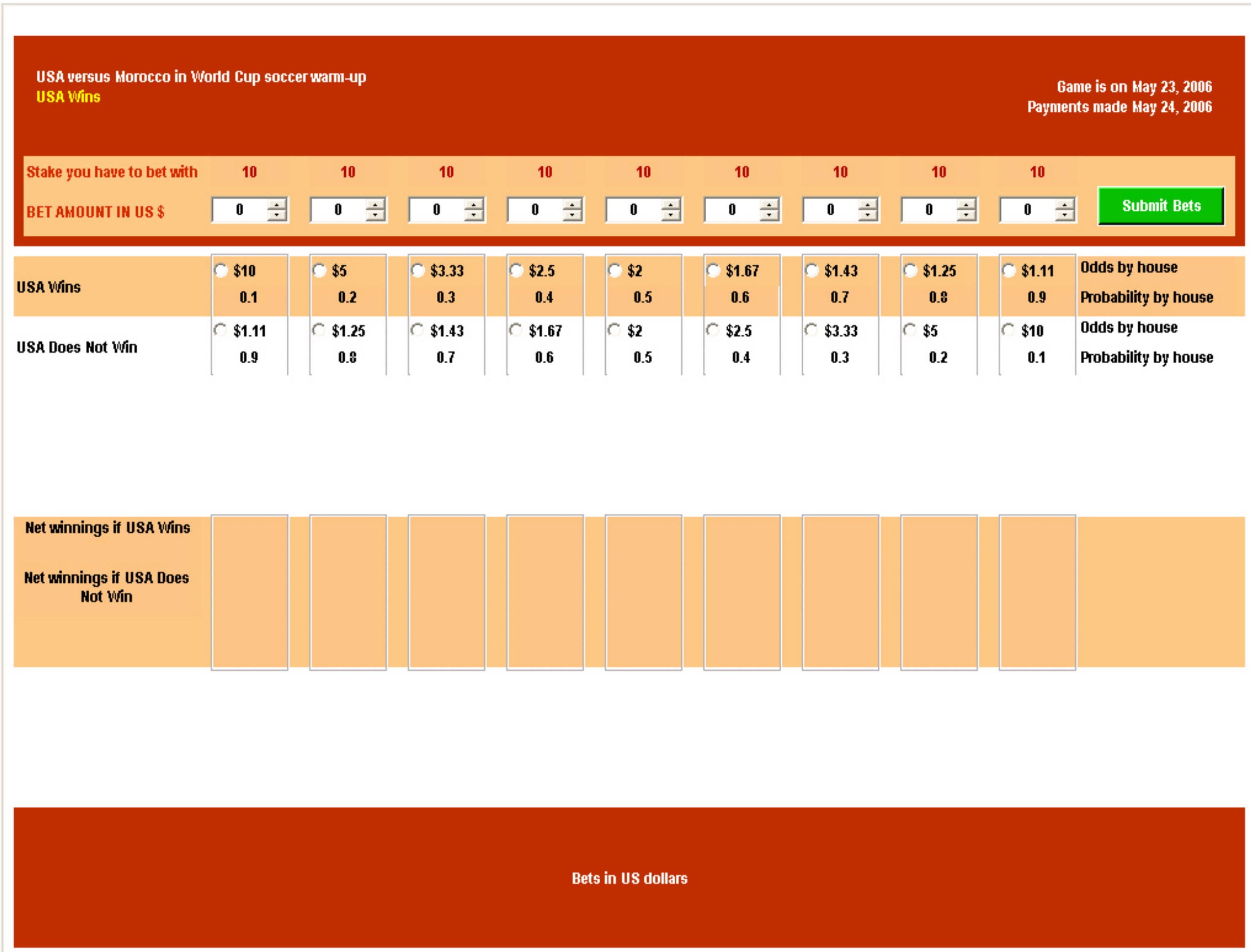
- **Finally**, all payments will be made on a date specified on the betting screen. Some events are over in a few weeks, and some are over in a few months. You will then be paid all earnings, including any money you decide not to bet (if you have that choice).

Here is the opening page, telling you the setup for your choices. This setup will apply to all of your choices today. **The information on your screen may differ from this display, so be sure to read it before making any choices.**

In this case you are told that you will be quoted odds from 9 betting houses. You have a $10 stake for each betting house, and you can decide how much to bet for each house. Click on Begin, in the bottom of the screen (not shown below) to start...

To illustrate we will use the USA versus Morocco soccer match as an example. The event that you will be asked to bet on will be whether the USA wins or not. Note that you are betting either that the USA wins or that the USA does not win. Thus, if there is a draw, that is the same as the USA not winning. This screen provides you with the betting form you will use...

On this screen you will first see the **event** listed in the top left hand corner. We also



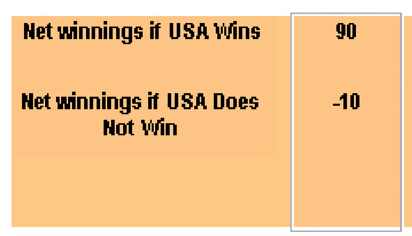
list **the bet**, in this case that the USA wins. In the top right hand corner you see when the event is to occur, and when we would pay out any earnings. We will obtain an address and/or e-mail contact for you at the end of the experiment, so that we can send you a check for your earnings. Of course, you will receive all other earnings from today's session in cash, paid out at the end of the session.

You will next see the **stake** amounts you can bet at each of the 9 betting houses. You choose how much to bet, or if you prefer not to bet at all. Then you choose whether you want to bet on USA winning or on USA not winning.

You will see that each betting house offers different **odds**. The first betting house, for example, pays $10 for every $1 that you bet on the USA winning, if in fact the USA wins, or $1.11 for every $1 that you bet on the USA not winning.

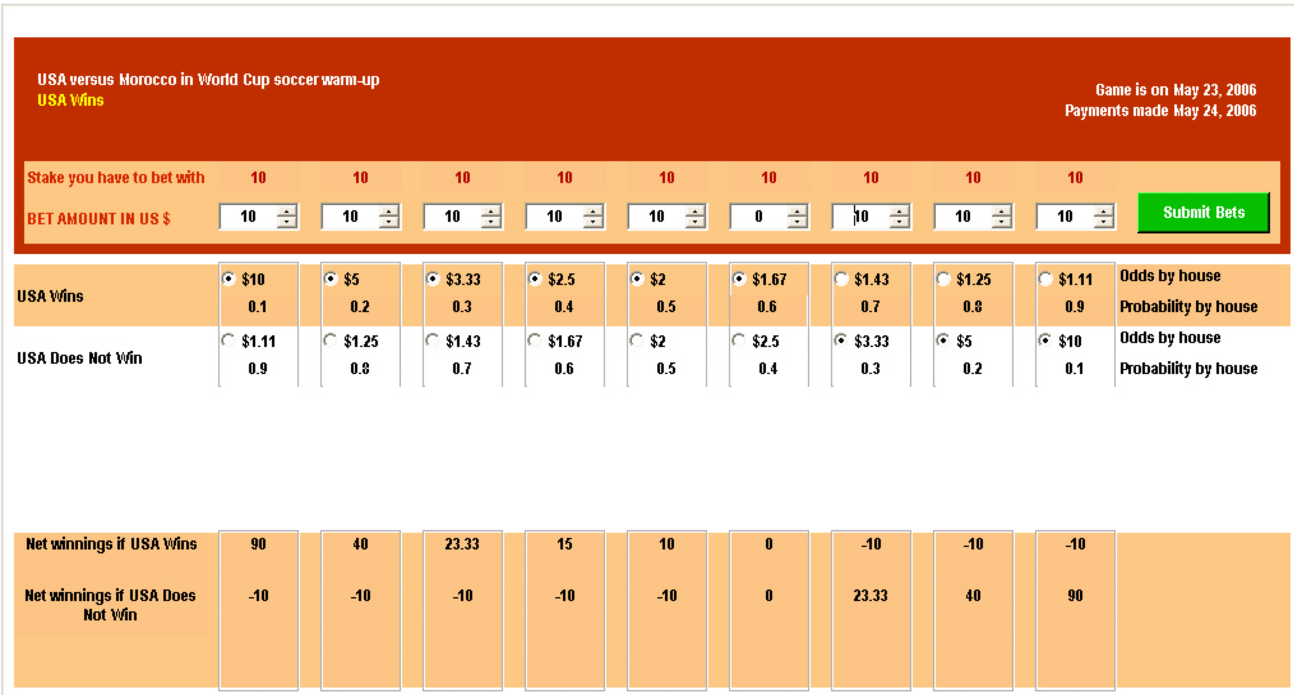
Beside each set of **odds**, we also display the **probability** that each event will occur. This is just another way of thinking about the odds. Some people understand odds better, and some people understand probabilities better. The probabilities are just the inverse of the odds. So if the odds say that you will be paid $5 for every $1 bet if the USA wins, as in betting house 2, then this implies a probability for the USA winning of $1 ÷ $5 = 0.2. This is the same things as there being a 20% chance of the USA winning. We have rounded some of the probabilities to make the screen easier to read, and you will be paid according to the odds.

Here is an example of how you place a bet, using betting house 1. In this case we chose to bet all of the $10 stake, and we bet that the USA would win. The betting screen then tells us what our net winnings would be if either outcome occurs. If the USA wins we get $10 for every $1 we bet, and we bet $10, so we earn $100. But that cost us our $10 bet, so the net winnings in this case are $90. But if the USA does not win, we lose our $10 bet.



In this case we would place this bet if we were confident that the USA has a better chance of winning than the 10% implied by the odds that this betting house is quoting. If we thought that the USA had less than a 10% chance of winning, we could (a) bet $10 that the USA does not win, (b) bet nothing, or (c) bet something between $1 and $10 that the USA does not win.

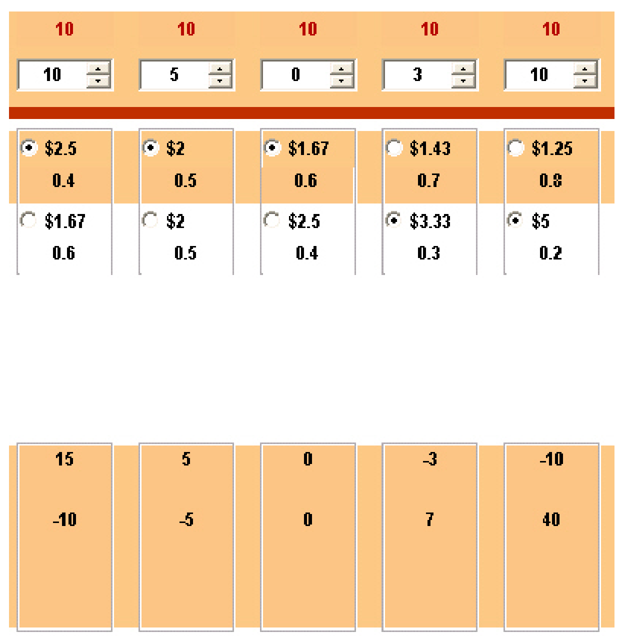
In the screen below we have filled in bets for every one of the 9 betting houses, to illustrate how you fill in the complete sheet. You do not have to place the same bet for every betting house, or bet the same amount:



USA versus Morocco in World Cup soccer warm-up
USA Wins

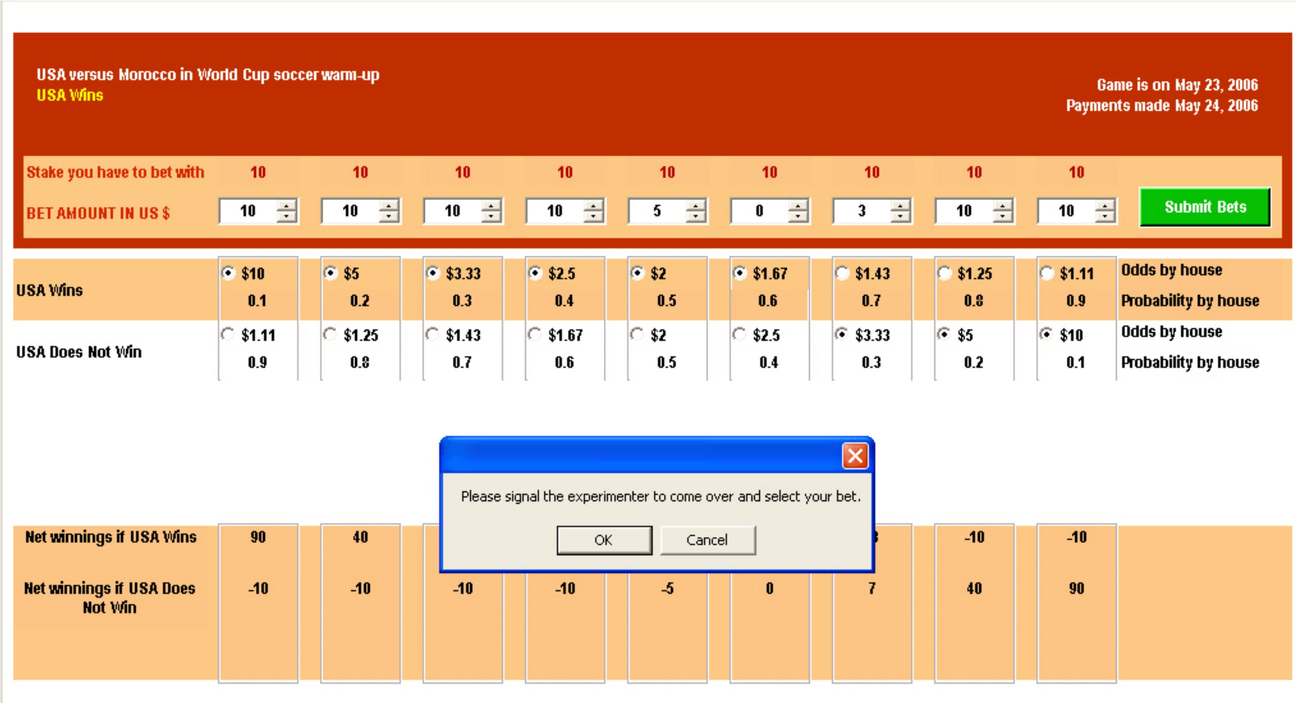
Game is on May 23, 2006
Payments made May 24, 2006

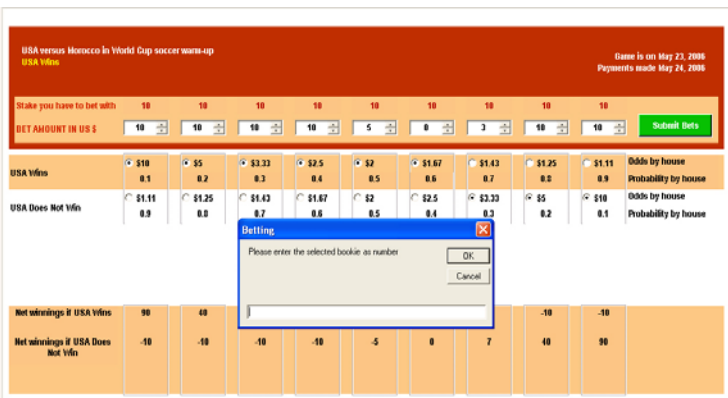
Stake you have to bet with	10	10	10	10	10	10	10	10	10	
BET AMOUNT IN US $	10	10	10	10	10	0	10	10	10	Submit Bets
USA Wins	(•) $10	(•) $5	(•) $3.33	(•) $2.5	(•) $2	(•) $1.67	( ) $1.43	( ) $1.25	( ) $1.11	Odds by house
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Probability by house
USA Does Not Win	( ) $1.11	( ) $1.25	( ) $1.43	( ) $1.67	( ) $2	( ) $2.5	(•) $3.33	(•) $5	(•) $10	Odds by house
	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	Probability by house
Net winnings if USA Wins	90	40	23.33	15	10	0	-10	-10	-10	
Net winnings if USA Does Not Win	-10	-10	-10	-10	-10	0	23.33	40	90	



10	10	10	10	10
10	5	0	3	10
(•) $2.5	(•) $2	(•) $1.67	( ) $1.43	( ) $1.25
0.4	0.5	0.6	0.7	0.8
( ) $1.67	( ) $2	( ) $2.5	(•) $3.33	(•) $5
0.6	0.5	0.4	0.3	0.2
15	5	0	-3	-10
-10	-5	0	7	40

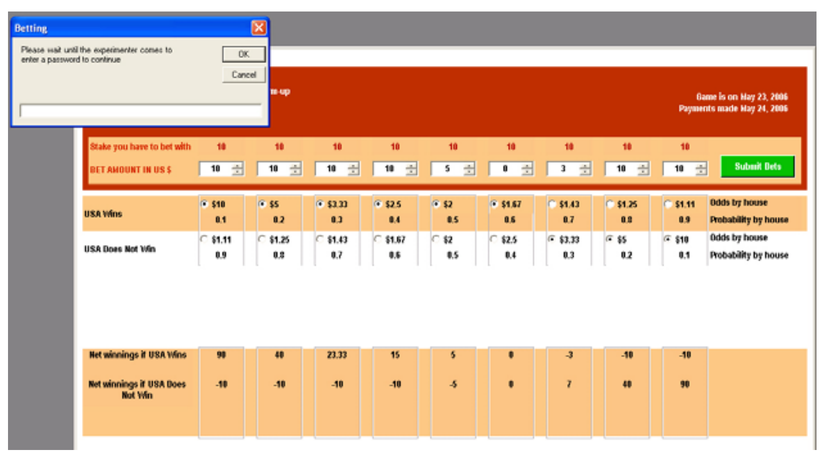
If we wanted to bet less than $10 for any betting house, we simply click on the amount we want to bet. The display changes to show the change in net winnings. In this case we bet $5 with one betting house, $0 with another, $3 with another, and $10 with the remaining houses.





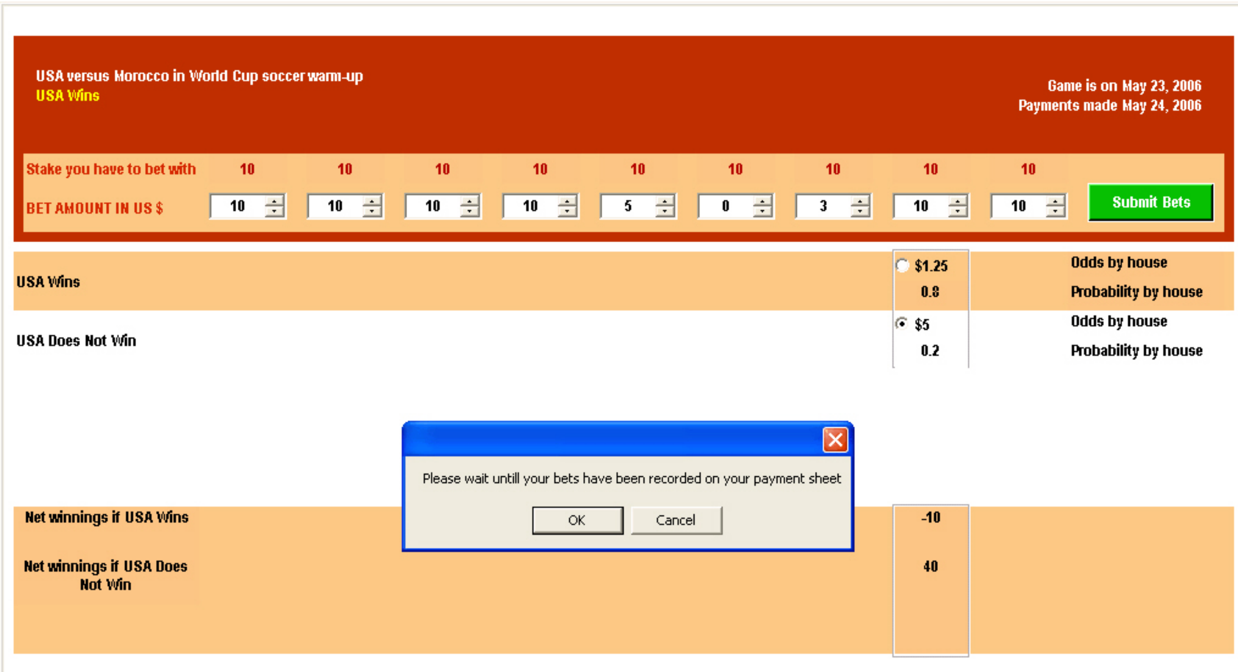
When you have entered all of your bets, and you are happy with your choices, you click on the green SUBMIT BETS tab, and you will be asked to signal the experimenter to come over. We will then roll a die to determine which betting house you will actually place your bet with.

After you click OK, a special box will come up which the experimenter uses to unlock the screen with a super-secret password. Please do not type anything here. The experimenter will do that when he or she comes over.



The experimenter will then have you roll a die to determine the betting house you will place your bet with. The experimenter will enter this, and then another password.

In this case we selected betting house 8, so the screen displays that bet. This information will then be recorded on your payment sheet.



You will be asked to place bets on several events. When you finish one event you will be given a new betting sheet for the other event. In a moment we will provide you with a sheet listing the events that you will be asked to place bets on.

Remember that we will pay you all earnings from these bets on the dates specified. We will also then pay you any money that you choose not to wager, if you have that option.

At the very end we will have a series of questions about you. Your responses will be confidential, and will not be connected with your name or any other identifying ID such as a SSN.

There are no right or wrong choices. Which choices you make depends on your personal preferences and your beliefs about what the chances are that each event will actually occur. The people next to you will have different tasks, and may have different preferences or beliefs, so their responses should not matter to you. Nor do their choices affect your earnings in any way. Please work silently, and make your choices by thinking carefully about the odds being offered in relation to how likely you believe each outcome is.

Do you have any questions?

Appendix C: Extensions

Our application illustrates all of the steps in the belief elicitation procedure we propose. But we realize that there are many possible extensions. Some are procedural, in the sense that they test features of the experimental design that might have influenced behavior; some are conceptual, in the sense that they reflect insights from our general theoretical framework; and some are statistical, in the sense that they modify some of the steps undertaken in the econometric recovery of latent subjective beliefs.

A. Procedural Extensions

Framing

The belief elicitation task is another variant on the "multiple price list" formats used in the elicitation of many other subjective values (e.g., risk attitudes, discount rates, valuations). It is subject to one major behavioral concern, the effect that the frame of the task has on responses.[31] It is possible that subjects will be drawn to respond on the middle of the frame, inferring that the experimenter has picked the end-points on the basis of some extra information that will help them make a better choice. This inference is false, but can nonetheless be a behavioral anchor, even though probabilities do have a natural frame between 0 and 1 as a matter of definition.

One solution is to frame the task differently to subjects, and determine if it systematically affects their responses. There is some evidence that there are such anchoring effects, but that they are not large in relation to the uncertainty over individual responses, and can be measured and hence controlled for (Andersen et al. [2006]). For example, for roughly half of our subjects one would use a frame that spans probabilities of each event from 0.1 up to 0.9 in increments of 0.1, and for the other half use a frame that spans probabilities of 0.1 up to 0.5 in increments of 0.05. If there is an anchoring effect biasing responses to the mid-point of the frame we would expect to see lower elicited probabilities in the "narrow" frame compared to the "wide frame."

Another solution might be to operationalize the manner in which the house odds are chosen for each subject, to make it clear that there is no experimenter role. One could mark 20 bingo balls with odds reflecting 0.05 increments in probability between 0.05 and 0.95, put them in a bingo cage that is vigorously shuffled, then have 9 selected without replacement. The risk is that one might draw several odds that are close to each other, and leave some large gaps, but over a larger sample of subjects this lumpiness would not be a major problem.

Cognitive Burden

Another direct extension to our elicitation procedure would be to provide feedback to subjects on the expected income implications of each possible bet, and the manner in which their subjective

[31] Two additional concerns are also sometimes expressed: (a) that there may be a large fraction of inconsistent decisions where subjects switch back and forth between A and B, implying a violation of monotonicity of preference, and (b) that it only elicits "interval responses," and therefore is not as accurate as open-ended procedures that elicit "point responses." These concerns are addressed in Harrison, Lau, Rutström and Sullivan [2005] and Andersen et al. [2006].

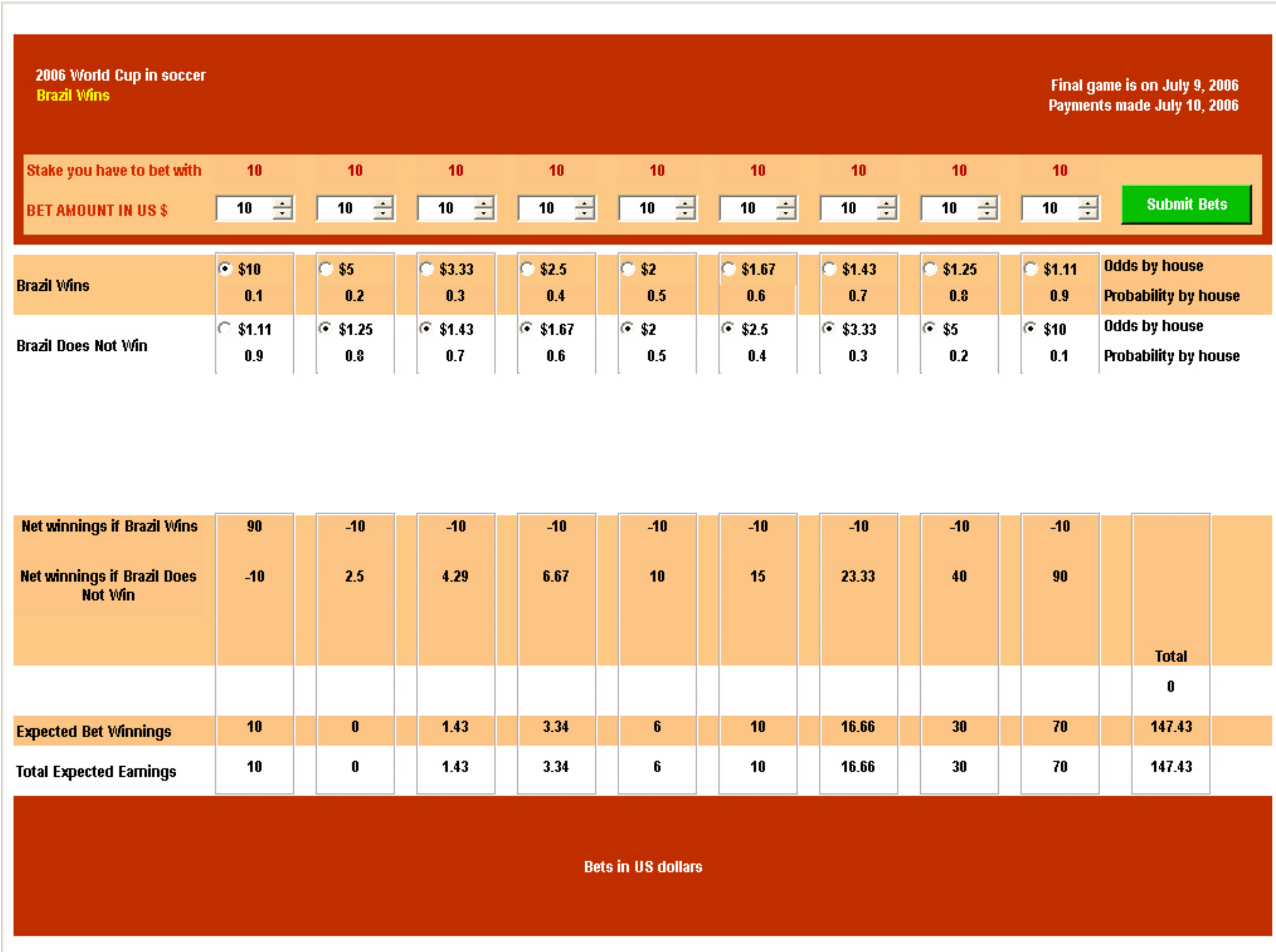
beliefs affect expected income. We could use the computerized interface to show the subject the implications of rational bets conditional on a stated subjective belief. The subject would move his beliefs, denominated in terms of the probability of one event, on a slider, and the betting form interface changes the expected income implied. This would significantly reduce the computational burden of these calculations, if one relaxes the "as if" assumption and assumes for the moment that subjects evaluate options the way we model them. It might also anchor behavior towards the risk neutral bet, since there is some evidence that providing information on expected income of lotteries reduces elicited risk aversion; on the other hand, such shifts might just reflect true risk attitudes without the confound of the cognitive burden of calculating EV (Harrison and Rutström [2008; §2.7]).

One potentially important treatment would be to initiate the information display at the optimal response for a risk-neutral bettor with a stated subject probability belief. In this version the subject would be initially asked to provide a belief that event A would occur. This is accomplished by providing a slider, illustrated here, that the subject could move between 0% and 100%. A numerical display would then identify the exact percentages implied.

Once the subject had entered this information, and confirmed it, a betting sheet comes up as shown on the next page. This sheet is the same as the one displayed in our experiments, except that *the bets are already initialized and there is a display at the bottom of the expected bet winnings and total expected earnings*. These expectations reflect the belief that the subject had entered immediately prior, which is a 20% chance of Brazil winning. Consider the first bet, in which the subject wins \$90 if Brazil wins and loses the \$10 wager if Brazil does not win. The expected winnings are then \$18 (= \$90 × 0.2) minus \$8 (= -\$10 × 0.8), or \$10 as shown.

If the subject had entered a different belief, the initial bets and numeric display of expected winnings would have changed. The subject can change any bet, and the interface shows the effect on total expected earnings. For example, if the subject chooses to bet *for* Brazil with the 9th betting house, which is offering such extreme odds that Brazil will lose, then total expected earnings drop from \$147.43 to \$69.63, as shown in the extract at the bottom of the next page. If this subject had stated a belief that Brazil would win that was 90% or greater, this might be an optimal bet for a risk-neutral agent to make, but not with the stated 20% belief that Brazil would win.

We stress, and this is explained carefully to the subject, that *these initial bets are optimal only if the subject is neutral with respect to risk*. We actually explain in the instructions that if the subject is extremely averse to any risks, the best thing they could do is to not bet at all and just collect \$10 at some future date. Or, if the subject is forced to bet, to bet as if they thought that Brazil would win with a 50% chance. We also point out that if they are not so risk averse, doing either of these things would reduce their expected earnings.

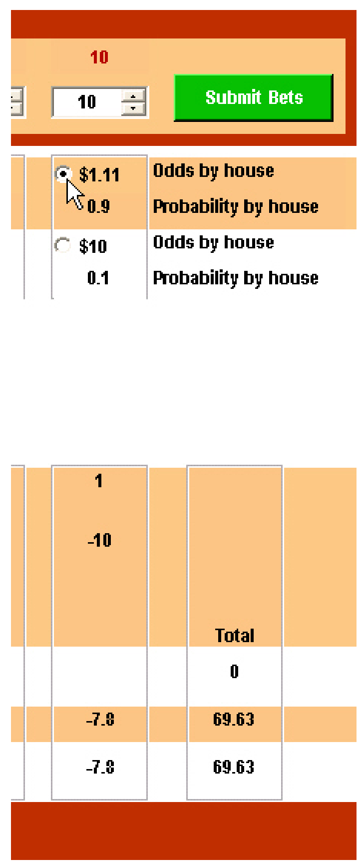


2006 World Cup in soccer
Brazil Wins

Final game is on July 9, 2006
Payments made July 10, 2006

Stake you have to bet with	10	10	10	10	10	10	10	10	10	
BET AMOUNT IN US $	10	10	10	10	10	10	10	10	10	Submit Bets
Brazil Wins	$10	$5	$3.33	$2.5	$2	$1.67	$1.43	$1.25	$1.11	Odds by house
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Probability by house
Brazil Does Not Win	$1.11	$1.25	$1.43	$1.67	$2	$2.5	$3.33	$5	$10	Odds by house
	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	Probability by house
Net winnings if Brazil Wins	90	-10	-10	-10	-10	-10	-10	-10	-10	
Net winnings if Brazil Does Not Win	-10	2.5	4.29	6.67	10	15	23.33	40	90	Total
										0
Expected Bet Winnings	10	0	1.43	3.34	6	10	16.66	30	70	147.43
Total Expected Earnings	10	0	1.43	3.34	6	10	16.66	30	70	147.43

Bets in US dollars



B. Conceptual Extensions

Free Range Betting

We have already indicated throughout that an important extension would be to allow subjects to pick which bookies to place bets on, and to vary the stake that they apply to any bet. This extension is not necessary in the laboratory environment, where we can gloss the "participation constraint" and provide subjects with play money to make bets with. But it provides an important bridge to help us understand the behavior of prediction markets that are open to the public. Our experimental instructions, in Appendix B, illustrate how we explain this free range option to subjects, since we designed them to be general.[32]

Free range betting opens up two additional issues. One is to allow subjects not to bet, but to restrict them to bets of $0 or $10 (or whatever the finite stake is). The ability to select intermediate stakes allows a constructive response to the issues of field censoring in "interval elicitation methods" raised by Cubitt and Read [2007]. The second issue is the effect on behavior of having subjects make choices with stakes that they have earned by some non-trivial effort rather than been endowed with as "play money." There is striking evidence of differences in behavior in social preference settings, and this could be a more general behavioral phenomenon (Rutström and Williams [2000] and Cherry, Frykblom and Shogren [2002]).

Improving Incentives

We could modify the opportunity set for subjects in ways that enhance our ability to recover their latent subject belief. We assumed that our subjects would tend to be risk averse, as explained in §2.C and lamented in §4.C when we discovered that some subjects were in fact risk loving. This assumption was, of course, well justified by the available evidence, but simply failed to reflect the individual preferences of some subjects in our sample. One could, however, use the computerized interface to provide subjects with an *opportunity set constructed specifically for them*. If they are shown to be (weakly) risk averse in some initial lottery task, then use the opportunity set in our experiment. Otherwise, if they are shown to be risk loving, modify the opportunity set as explained in §2.C.

Does one run the risk that subjects might strategically mis-represent their preferences in the lottery task so as to generate a more attractive opportunity set in the belief elicitation task? This is a theoretical possibility, but a practically remote one that could be mitigated by various experimental procedures. Or one could predict the risk attitudes of the belief-elicitation subject from a separate cohort of subjects that had participated in the risk attitude elicitation task.

Changing Places

A more fundamental extension is to change roles in the "trading game" between bookie and

[32] The econometric analysis would be slightly modified to allow for the fact that subjects have at least 3 choices now, even if we ignore the variation in stake: bet all on A, bet all on B, do not bet on either. So one cannot use the binary choice model, but would have to move to well-known generalizations (e.g., Train [2003; ch.2, 3]).

bettor, and have subjects take the part of the bookie rather than the bettor. Just as duality allows one to attain the same optima by utility maximization or cost minimization, when we view the belief elicitation task as an exchange game defined over state-contingent commodities, it is natural to see how one can elicit subjective beliefs by having subjects play either role.

We adopted the frame in our experiment that had the subject take the role of the bettor, since it was more natural for our subjects as consumers to be facing betting odds rather than setting them. But we can envisage subject pools for whom the other role would be more natural. Managers of firms who are deciding on product placement and pricing are in effect making book against the market based on their subjective beliefs.

Uncertainty in Beliefs

We noted earlier that the belief we elicit is what a statistician calls a predictive distribution, and that the elicited belief is consistent with a wide range of supporting beliefs about the underlying probability. The reason for the latter inference is that there are two sources of uncertainty embedded in a predictive distribution, and in general we cannot know from our design which source is more important. One source is uncertainty in the estimates of parameters defining the probability of the event, and the other source is uncertainty over the sample realization. For example, assume that the event is to bet on the number of times a coin toss comes out Heads. If the sample to be used to determine payoffs is 1, then there is some well-defined probability distribution that a fair coin would come out heads. But if the sample is much larger, say 30, then that distribution has the same mean but with a much smaller variance. Does that change in the predictive distribution affect behavior in belief elicitation?

This question also raises deeper conceptual issues about the formal modeling of "second order beliefs," "ambiguity," and "uncertainty" which we choose to side-step for now. Our objective is to make a start with settled theory about decision making under subjective risk, as usually defined. We do not believe that there is currently sufficient structure on these more general models to permit one to recover subjective beliefs, although ongoing research may change that judgement.

C. Econometric Modeling Extensions

Alternative Models of Choice Behavior

It was a simple matter to "swap out" the EUT specification of choice under uncertainty and replace it with the RDU specification. The extension to consider probability weighting is particularly important in this setting, since we are after all eliciting probabilities. But the same logic would allow one to extend the specification to include Cumulative Prospect Theory or any other preferred model of choice under uncertainty. Although we stress that the observer has to take a stand with respect to the model of choice under uncertainty that the subject is assumed to employ, our procedures generalize easily to alternative specifications.

Parametric Assumptions

We concede that we rely heavily on parametric functional forms to illustrate the manner in which one recovers latent subjective beliefs. The functions we adopt are standard and reasonably general, but we appreciate that there are alternatives worth exploring. Indeed, given the importance of the probability weighting function, we illustrated one such extension ourselves.

A related issue is to consider non-parametric approaches. These come of two general varieties: non-parametric specifications of the utility function or probability weighting function, as in Hey and Orme [1994], and/or non-parametric specifications of the statistical specification linking the deterministic choice model to observed behavior. We find these alternatives seductive, but they add needless complexity to our illustration of the basic steps of recovering latent subjective beliefs from observed betting choices.

Individual Heterogeneity

It is, in principle, a simple matter to estimate latent subjective beliefs at the level of the individual, rather than pooling over individuals. Although we do account for a rich set of individual characteristics in all core parameters, there might still be some unobserved individual heterogeneity that can best be addressed by estimating at the level of the individual. In terms of models of decision making under risk, Hey and Orme [1994] remains a classic in this regard. However, one faces inevitable numerical issues with specific individual estimates, and that can become time-consuming to evaluate and report efficiently. For our methodological purposes there is nothing new involved in extending the analysis to the level of the individual, so we avoid it here.

Additional References

Cherry, Todd L.; Frykblom, Peter, and Shogren, Jason F., "Hardnose the Dictator," *American Economic Review*, 92(4), September 2002, 1218-1221.

Cubitt, Robin P., and Read, Daniel, "Can Intertemporal Choice Experiments Elicit Time Preferences for Consumption?" *Experimental Economics*, 10, 2007 forthcoming.

Harrison, Glenn W.; Lau, Morten I.; Rutström, E. Elisabet, and Sullivan, Melonie B., "Eliciting Risk and Time Preferences Using Field Experiments: Some Methodological Issues," in J. Carpenter, G.W. Harrison and J.A. List (eds.), *Field Experiments in Economics* (Greenwich, CT: JAI Press, Research in Experimental Economics, Volume 10, 2005).

Rutström, E. Elisabet, and Williams, Melonie B., "Entitlements and Fairness: An Experimental Study of Distributive Preferences," *Journal of Economic Behavior & Organization*, 43, 2000, 75-89.

Train, Kenneth E., *Discrete Choice Methods with Simulation* (New York: Cambridge University Press, 2003).

Appendix D: Connections to the Literature

It is not appropriate to survey the literature on belief elicitation, not least because of the vast scope of the task. Instead we restrict ourselves to the more important connections to building blocks we rely on, and directly comparable approaches.

A. Theoretical Literature

The scoring rule derivations we provide in §2 and Appendix A have also been developed independently by Jose, Nau and Winkler [2007]. However, their decision-theoretic approach to scoring rules and belief elicitation problems is driven by a desire to generalize and unify the plethora of measures of divergence between probability distributions and measures of the quantity of information in statistics. The methods they use to achieve their objective perhaps obscures a fundamental insight about belief elicitation in their analysis, familiar from the mechanism design literature in economics: if you know (or assume) specific details about the general form of preferences you can always find incentive compatible elicitation mechanisms for beliefs via ordinary demand functions.

Suppose A has an EU utility function of wealth f(x) with beliefs b, {b(1).... b(n)}, and a budget set in contingent commodity space described by $\sum_s$ p(s)x(s)=m. Under standard regularity conditions there will be a unique set of optimal wealth choices x*(s,b,p,m,f), one for each state s, that will depend on A's personal beliefs b. These are A's ordinary demand functions for s-contingent wealth in a competitive market. The choices x*(s, b′, p, m, f) maximize somebody else's utility in the budget set $\sum_s$ p(s)x(s)=m, a somebody perhaps just like A except that their beliefs are b′, and not b. These choices are not EU-maximizing for A – by revealed preference, it cannot be any better, and is probably worse. The s-contingent wealth functions x*(s, r, p, m, f), for varying reports r, can however function as an incentive compatible scoring rule for A. That is, if A is faced with an opportunity set of s-contingent wealth payments, x*(s, r, p, m, f) , all of which satisfy $\sum_s$ p(s)x(s)=m, and A's means of implementing a wealth choice is to choose a vector of numbers r, plug them in to her ordinary demand functions, and obtain the resulting wealth, then A would choose to report her beliefs r=b in order to maximize her EU in this budget set.

While these ordinary demand functions, if we knew them, could function as an incentive compatible scoring rule for A, call it an x-score, they would not typically be a "proper"scoring rule (i.e., one maximizing A's expected x-score of A's contingent wealth along a linear budget line), because A may not be risk neutral. With a mathematical slight of hand we can, however, turn the incentive compatible scoring rules x*(s, r, p, m, f) for A into proper scoring rules. Imagine that A's score-payouts are no longer observable wealth levels x(s) lying along a linear wealth constraint in s-contingent commodity space, but unobservable utility levels u(s)=f(x(s)) lying along a corresponding non-linear constraint in s-contingent commodity space. (The constraint would be linear if f is linear). If we try to maximize A's expected u-score, $\sum_s$ b(s)u(s), constrained to u(s)=f(x*(s, r, p, m, f)), by choosing a "report" r, then r=b will maximize that expected u-score. Hence the u-score is a proper (expected u-score maximizing) scoring rule.

This general characterization of incentive compatible scoring rules, and how to turn them into proper scoring rules, is not trivial. It does reveal that there are as many different scoring rules as there are different utility of wealth functions. Jose, Nau and Winkler [2007] takes this idea seriously, using f(x) as a power function, x^a, and through an appropriate re-scaling of this utility function, $(1/(\beta-1)\{[1+\beta x)^{(\beta-1)}-1\}$, derive ordinary demand functions which, when optimized against a linear budget constraint and transformed using the mathematical sleight of hand noted above, yield generalized forms of many well-knows scoring rules.

Jose, Nau and Winkler [2007] contains excellent deductive logic: if preferences are given, they show that we can find a proper scoring rule in utility units to elicit true beliefs. This is valuable, but does not address the elicitation problem that we are concerned with: given an opportunity set described in wealth space by a scoring rule, what will forecasters of differing risk attitudes choose to report?

B. Experimental Literature

Experimental economists have used several of the popular scoring rules, but with one notable exception discussed in §C 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], Rutström and Wilcox [2006] and Costa-Gomes and Weizsäcker [2007].[33] In each case the subject is implicitly or explicitly assumed to be risk-neutral. 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 selected 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.

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 prediction. It is well-known that linear scoring rules elicit the *median* of the subjective predictive distribution for a risk-neutral agent, and of course this will also be the mean and mode if the distribution is unimodal and symmetric.

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

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

game theory," which rest entirely on making operational the beliefs of players in strategic games. In the former study the "exact" prediction of a probability was defined as an estimate within 5 percentage points of the true outcome; in the latter study the estimates were over 41 finite contribution levels in a public good, so the prediction had to be the correct integer to receive the reward. It is easy to show that this scoring rule elicits the *mode* for a risk-neutral agent. In the case of contributions to a public good this is not at all likely to be a unimodal and symmetric distribution, given the expectation of a significant spike at the zero contribution level implied by perfect free-riding (e.g., see the histograms displayed in Dufwenberg, Gächter and Hennig-Schmidt [2007; Appendix B]). The rationale for this scoring rule, rather than the QSR, is provided by Charness and Dufwenberg [2006; p.1586]: "Overall, we chose our belief-elicitation protocol mainly because it is simple and rather easy to describe in instructions. [...] Our idea is to get a rough-but-meaningful ballpark estimate of participants' degrees of belief." We certainly accept that the QSR can be difficult to explain to subjects, but do not know what the phrase "rough-but-meaningful ballpark estimate" means: either we take the incentives to report beliefs seriously, or we do not.

Most experimental economists embed the elicitation of probabilities in another experimental task that the subject is undertaking. Indeed, one of the hypotheses being studied is whether the effort to elicit beliefs will encourage players in a game to think more strategically (Croson [2000], Rutström and Wilcox [2006], Costa-Gomes and Weizsäcker [2007]). Of course, this violates the "no stakes condition" required for the QSR to elicit beliefs reliably unless one assumes that the subject is risk neutral. 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].

The most serious concern with the experimental implementations of scoring rules is that the rewards are very, very small. For example, Nyarko and Schotter [2002] and Rutström and Wilcox [2006] gave each subject an endowment of 10 cents, from which their penalties are to be deducted. So the effect of the scoring rule is literally defined in terms of fractions of pennies, and the additional rewards are not very substantial for optimal reports as compared to reports near the optimum. Whatever position one takes on the issue of "flat payoff functions" raised by Von Winterfeldt and Edwards [1986] and Harrison [1989], these rewards for accuracy are disappointing.

We have discussed the development of prediction markets by experimental economists in Appendix A. In addition, experimental economists have considered parimutuel betting markets as a means of eliciting aggregate beliefs: see Hurley and McDonough [1995] and Plott, Wit and Yang [2003].

C. Econometric Literature

Only one study attempts to recover elicited beliefs from observed choices, calibrating for non-linear utility functions and/or probability weighting: Offerman, Sonnemans, van de Kuilen and Wakker [2007; §6].[34] They provide a statement of some alternative ways in which this recovery could

[34] The need for some correction is also recognized by Offerman, Sonnemans and Schram [1996; p.824, fn.8] and Rutström and Wilcox [2007; p.11, fn.8].

be undertaken, essentially the method we use, and then propose a new method. Their method can be viewed as an empirical "reduced form" approach to adjusting for risk attitudes.

One method they consider is by estimating or eliciting the functional forms of a model of choice under risk (e.g., EUT, RDU or CPT), then observing beliefs over a natural event in some task, and econometrically recovering the implied subjective beliefs by using the estimated model of choice under risk to "back out" the subjective probability that must have been used in the belief elicitation task. They dismiss this approach, which is the one we follow. They claim, without further discussion, that estimating or eliciting the functional forms is "laborious" and that it involves "complex multi-parameter estimations." It is certainly true that the joint likelihood involves several parameters, but such estimation is standard fare with maximum likelihood modeling, so that is hardly a concern (unless one wants to avoid writing out customized likelihoods). It is not clear in what sense this is a "complex" undertaking. The labor involved depends on how one undertakes the estimation or elicitation. In our case the subjects need to do one task, which consists of 60 binary choices over lotteries, and then all of the labor involved is by the computer estimating maximum likelihood models that have been well-studied for years (e.g., Harrison and Rutström [2008; §2] for a survey).[35]

The empirical method they use instead has an attractive reduced form simplicity. For a given subject, it uses the QSR to elicit reported probabilities for naturally occurring events, and then uses the QSR in a calibration task to elicit a "risk correction function" that allows them to recover the subjective probability that generated the report for the *naturally occurring event*. The risk correction function simply elicits reports for "objective probabilities," such as the chance that a single roll of a 100-sided die will come up between 1 and 25. Assume the subject report 0.30 for this event. Then, if the subject ever reported a 0.30 in the initial task for the naturally occurring event, they would infer that he had a subjective probability of 0.25 underlying it, since that was the objective probability that generated this report using the (same) scoring rule. Thus the difference between the report of 0.30 in the calibration task and the true underlying probability is attributed solely to the effects of non-linear utility and/or probability weighting. By eliciting a risk correction function for a wide range of probabilities, and with a sufficiently fine grid, one can recover any report with some reasonable accuracy.

This approach is attractive, and avoids the need for the researcher to "take a stand" on which model of choice under uncertainty determines betting behavior. To see the key assumption underlying their approach, let ϕ be the *actuarial* probability that the calibration event will occur. For some artefactual events, such as tossing coins and rolling die, ϕ is well defined, but for other events it is not so well defined. Let $\pi(\phi)$ be the function that summarizes the subjective belief that the subject actually holds that the calibration event will occur, and let $R(\pi(\phi))$ be the function transforming $\pi(\phi)$ into a report using the QSR, or any appropriate scoring rule. Offerman et al. [2007] first assume that $\pi(\phi) = \phi$ *in the calibration task*, so that the only reason that $R(\pi(\phi)) \neq \phi$ is that the subject has non-linear utility

[35] On the other hand, if one uses other elicitation procedures, such as the Trade-Off design of Wakker and Deneffe [1996], Fennema and van Assen [1999], Abdellaoui [2000] and Abdellaoui, Bleichrodt and Paraschiv [2007], then the procedures can indeed become laborious for the subject. There are other reasons not to use these methods, the most significant of which is their lack of incentive compatibility as conventionally applied (Harrison and Rutström [2008; §1.5]). But these methods are not needed, and the stated concerns with this approach to recovering subjective beliefs are not substantial.

and/or undertakes probability weighting.[36] Why might $\pi(\phi) \neq \phi$, for such simple tasks? Apart from concerns with loaded die, or certain cultures imbuing randomizing devices or colors on chips with some animist intent, we would be concerned with psychological editing processes based on similarity relations. To take a simplistic example, someone might "round down" to the nearest increment of 0.05 or 0.10 and then decide how to report using this subjectively edited probability $\pi(\phi)$ as the basis for any adjustments due to non-linear utility or probability weighting. Is the actuarial probability ϕ the one we really want to compare $R(\pi(\phi))$ to in such a case, or is it $\pi(\phi)$?

This might seem to be nit-picking when it comes to the rolling of a 100-sided die in the calibration task, and perhaps viewed as part of a latent structural psychological story underlying the idea of probability weighting. But it is surely more significant for naturally occurring events. Here is where the second assumption comes in: that $\pi(\phi) = \phi$ *in the belief elicitation task* where ϕ is defined (or not) over naturally occurring events. Thus, what if we accept that $\pi(\phi) = \phi$ is a reasonable assumption for the calibration task with the artefactual event, but cannot be so sure for the task with the naturally occurring event? Our position is that we are recovering $\pi(\phi)$, "warts and all" in terms of how the subject conceives of the event and defines the probability ϕ. Offerman et al. [2007] would appear to be recovering the "touched up" image of $\pi(\phi)$, ϕ, after the warts have been removed.

---

[36] So there is no allowance for subjects to make decision errors in the calibration task, or the elicitation task for the naturally occurring event for that matter. These errors could be subsumed into some sampling error on estimates of $R(\pi(\phi))$ as an empirical function of ϕ, but then one is relying on the errors being well-behaved statistically. In fact, Offerman et al. [2007; equation (8.3)] do allow for an additive error term which they assume to be truncated normal to ensure that reported probabilities lie between 0 and 1. Their pooled estimates indicate that there is a need for some correction for non-linear utility, but that it is not so clear that probability weighting is an issue (the log-likelihood for row #1 in their Table 9.1, which allows for both effects, is virtually identical to the log-likelihood for row #5, in which no probability weighting is assumed). Although they allow for errors to vary with an incentives treatment applied between-subjects, it would be useful to extend their statistical analysis of the pooled data to allow for correlated errors at the level of the individual subject, rather than implicitly assume homoskedasticity.

Additional References

Abdellaoui, Mohammed, "Parameter-Free Elicitation of Utilities and Probability Weighting Functions," *Management Science*, 46, 2000, 1497-1512.

Abdellaoui, Mohammed; Baillon, Aurélien, and Wakker, Peter P., "Combining Bayesian Beliefs and Willingness to Bet to Analyze Attitudes towards Uncertainty," *Working Paper*, Econometric Institute, Erasmus University, September 2007.

Abdellaoui, Mohammed; Bleichrodt, Han, and Paraschiv, Corina, "Measuring Loss Aversion under Prospect Theory: A Parameter-Free Approach," *Management Science*, forthcoming 2007.

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," *ELSE Working Paper*, University College London, July 2007.

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.

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).

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

Harrison, Glenn W., "Theory and Misbehavior of First-Price Auctions," *American Economic Review*, 79, September 1989, 749-762.

Harrison, Glenn W., "Hypothetical Bias Over Uncertain Outcomes," in J.A. List (ed)., *Using Experimental Methods in Environmental and Resource Economics* (Northampton, MA: Elgar, 2007).

Haruvy, Ernan; Lahav, Yaron, and Noussair, Charles, "Traders' Expectations in Asset Markets: Experimental Evidence," *American Economic Review*, 2007, forthcoming.

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).

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.

Hurley, William, and McDonough, Lawrence, "A Note on the Hayek Hypothesis and the Favorite-Longshot Bias in Parimutuel Betting," *American Economic Review*, 85(4), September 1995, 949-955.

Jaffray, Jean-Yves, and Karni, Edi, "Elicitation of Subjective Probabilities when the Initial Endowment is Unobservable," *Journal of Risk and Uncertainty*, 8, 1999, 5-20.

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

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

Kass, R.E., and Raftery, A.E., "Bayes Factors," *Journal of The American Statistical Association*, 90, June 1995, 773-795.

Lad, Frank, *Operational Subjective Statistical Methods: A Mathematical, Philosophical and Historical Introduction* (New York: Wiley, 1996).

Leamer, Edward E., "Bid-Ask Spreads for Subjective Probabilities," in P.K. Goel and A. Zellner (eds.), *Bayesian Inference and Decision Theory: Essays in Honor of Bruno de Finetti* (Amsterdam: North-Holland, 1986).

Ledyard, John O., "Designing Information Markers for Policy Analysis," in R.W. Hahn and P.C. Tetlock (eds.), *Information Markets: A New Way of Making Decisions* (Washington, DC: AEI Press, 2006).

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

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

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

Plott, Charles R.; Wit, Jorgen, and Yang, Winston C., "Parimutuel Betting Markets as Information Aggregation Devices: Experimental Results," *Economic Theory*, 22, 2003, 311-351.

Rutström, E. Elisabet, and Wilcox, Nathaniel T., "Stated Beliefs Versus Empirical Beliefs: A Methodological Inquiry and Experimental Test," *Working Paper 06-18*, Department of Economic, College of Business Administration, University of Central Florida, 2006.

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.

Von Winterfeldt, Detlof, and Edwards, Ward, *Decision Analysis and Behavioral Research* (New York: Cambridge University Press,1986).

Wakker, Peter P., and Deneffe, Daniel, "Eliciting von Neumann-Morgenstern Utilities when Probabilities Are Distorted or Unknown," *Management Science*, 42, 1996, 1131-1150.

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