Lossed in Translation: An Off-the-Shelf Method to Recover Probabilistic Beliefs from Loss-Averse Agents

Theo Offerman*

Asa B. Palley†‡

November 12, 2013

Abstract: Although strictly proper rules are designed to truthfully elicit subjective probabilistic beliefs, experimental results have shown that risk aversion causes agents to bias their reports towards the rule's baseline probability of 1/2, and a conservative preference for certain outcomes leads agents with moderate beliefs to simply report 1/2. While both of these distortions make recovery of true beliefs difficult, the second effect is particularly pernicious because it leaves the assessor unable to discriminate amongst a broad range of moderate probabilities. Applying a prospect theory model of risk preferences, we show that loss aversion can explain both of these behavioral phenomena. Using the insights of this model, we develop a modified off-the-shelf probability assessment mechanism that corrects these distortions and allows the assessor to recover an accurate estimate of an agent's true beliefs. In an experiment, we demonstrate the effectiveness of this modification in both eliminating uninformative reports and eliciting true probabilistic beliefs.

Keywords: Scoring Rule, Subjective Probability Assessment, Loss Aversion, Prospect Theory

JEL Classification C81, C91, D03, D81

---

*CREED, The University of Amsterdam. Roetersstraat 11, 1018 WB Amsterdam. t.j.s.offerman@uva.nl.

†The Fuqua School of Business, Duke University. 100 Fuqua Drive, Box 90120, Durham, North Carolina 27708. asa.palley@duke.edu.

‡We thank CREED-programmer Jos Theelen for programming the experiment. We are grateful for financial support from the Research Priority Area Behavioral Economics of the University of Amsterdam.

1 Introduction

Accurately obtaining subjective probabilistic information about uncertain future events is an essential step in the decision making process in many different economic and public policy settings. In many cases, rather than trying to build a model to estimate probabilities, the best and most informative assessments come from an agent in the field who has a good amount of relevant experience and can use her collected wisdom to estimate a subjective probability. Eliciting this information presents an important and difficult problem in many fields such as finance and macroeconomics (Diebold and Rudebusch, 1989; Ghysels, 1993), decision analysis (Keeney, 1982), and meteorology and weather forecasting (Murphy and Winkler, 1984). In addition, probability assessments often comprise an important component of economic experiments. Even when the ultimate objective is not to elicit subjective beliefs, obtaining this information may be a critical secondary step in an experimental procedure.

Well-designed scoring rules provide a useful tool for procuring this subjective information by providing an agent with the right incentives to carefully evaluate and quantify her beliefs, and to honestly reveal her subjective assessment of the likelihood of these uncertain future events. The quadratic scoring rule (QSR), a variant of which was first introduced by Brier (1950), is the most commonly used.[1]

The incentive design of scoring rules implicitly assumes, however, that the agent is risk neutral, which contrasts with how people often behave. Winkler and Murphy (1970) examine the effects of nonlinear utility on the optimal report under a proper scoring rule, showing that risk aversion leads an agent to hedge her reports away from categorical forecasts of 0 and 1 and risk seeking leads the agent to bias her reports closer to 0 or 1. This biasing effect of risk preferences can be easily corrected by applying the inverse utility function to the scoring rule (Winkler, 1969). In practice, however, an even more troubling pattern of excessive reports equal to the baseline probability of 1/2 emerges as well, a phenomenon not explained by classical expected utility models. For example, Offerman et al. (2009) tested responses by 93 subjects to a QSR for objective probabilities that ranged from 0.05 to 1 and found that they reported $r = 1/2$ more than three times as often as they should have (15.3% versus 5%). This particular type of conservatism inhibits the assessor's ability to discern among a broad domain of moderate beliefs and conceals a significant amount of useful information.

In this paper, we employ the insights of prospect theory (Kahneman and Tversky, 1979;

[1]References include McKelvey and Page (1990), Offerman, Sonnemans, and Schram (1996), Huck and Weizsäcker (2002), Nyarko and Schotter (2002), Costa-Gomes and Weizsäcker (2008), Armantier and Treich (2013), Offerman et al. (2009), Blanco et al. (2010), Andersen et al. (2010), and Kothiyal, Spinu, and Wakker (2011).

Tversky and Kahneman, 1992) to understand the ways in which an agent will distort her report when she receives an uncertain reward by a QSR. Employing Palley's (2013) model of prospect theory with an endogenous reference point, we highlight how loss aversion can account for why an agent may both report $r = 1/2$ for a range of moderate beliefs and bias her reports towards 1/2 for beliefs closer to 0 or 1.

The main contribution of our paper is the introduction of a generalized asymmetric QSR, the L-adjusted rule, which eliminates the incentives for conservative reports and enables the elicitation of true probabilistic beliefs. We use previous experimental work estimating population parameters to derive an off-the-shelf variant of this L-adjusted QSR that requires no prior agent-specific calibration. In an experiment, we demonstrate its effectiveness in recovering truthful and precise probability assessments, and show that it alleviates the shortcomings associated with the classical QSR. In agreement with previous results, we find that in response to the classical QSR, agents tend to report the implicit benchmark probability of 1/2 for a wide range of beliefs near 1/2 in order to ensure a certain payoff. By matching the choice of L to previous empirical estimates of parameters for the overall population, we also obtain a modified QSR that recovers truthful beliefs experimentally. In doing so, we provide a practical and simple off-the-shelf scoring rule that encourages agents to report their beliefs truthfully.

Recently, several related approaches have been suggested to recover true beliefs from conservative reports. Offerman et al. (2009) propose a revealed preference technique that allows the researcher to correct the reported beliefs of agents who are scored according to a standard QSR. In this method, agents initially provide reports for a range of objective probabilities, which then yields an optimal response mapping that can be inverted and applied to infer subjective beliefs from later reports. In an experiment, Offerman et al. demonstrate the effectiveness of this approach in recovering beliefs from reports that do not equal the baseline probability of 1/2. Kothiyal, Spinu, and Wakker (2011) extend this method to overcome the problem of discriminating between moderate beliefs in a range around the baseline probability of 1/2, for which agents give the same optimal report. By adding a fixed constant to one of the QSR payoffs, they both eliminate the excess of uninformative baseline reports and yield an invertible response mapping that makes possible the recovery of true beliefs, while maintaining the properness of the original scoring rule. Kothiyal, Spinu, and Wakker do not provide an experimental test of their method.

The approaches taken in Offerman et al. (2009) and Kothiyal, Spinu, and Wakker (2011) are precise and elegant because they do not need to make structural assumptions on how people make decisions under risk. The downside of these methods is that they are laborious to employ, because a sufficiently dense risk-correction map has to be derived for each agent before any inferences can be made. In both decision analysis and many experimental economics applications, the elicitation of beliefs is a secondary goal, and a simpler and quicker approach may be preferred, as long as it does not sacrifice precision. The method presented in this paper pursues this purpose.

Other elicitation methods that do not make use of scoring rules exist as well. For example, if the utility function is unknown, Allen (1987) presents a randomized payment method that relies on the "linearization" of utility through conditional lottery tickets to incentivize truthful reports. Alternatively, Karni (1999) proposes a procedure with two fixed prizes where the payment function is determined by comparing the agent's report to a random number drawn uniformly from $[0, 1]$, analogous to the Becker, DeGroot, Marschak (1964) mechanism. Under this method, if the agent exhibits probabilistic sophistication, she has a dominant strategy to report her true belief, irrespective of her risk attitudes. However, in experiments, subjects have been found to have a hard time understanding Becker-DeGroot-Marschak-type procedures (Rutström, 1998; Plott and Zeiler, 2005; Cason and Plott, 2012), and empirical comparisons of these methods with scoring rules have yielded mixed results (Hao and Houser, 2010; Hollard, Massoni, and Vergnaud, 2010; Trautmann and van de Kuilen, 2011).

The rest of the paper is organized as follows: Section 2 introduces our L-adjusted QSR and characterizes the corresponding optimal reporting policy under the prospect theory model of risk behavior. We discuss how this predicted behavior provides a parsimonious explanation of previously observed conservative reporting patterns and how the parameter L can be calibrated to allow for the recovery of estimates of any probabilistic belief. Readers who are interested mainly in our experimental results may skim Section 2 and refer to Proposition 1 and Corollary 1. Sections 3 and 4 detail the experiment that we carried out to test the usefulness of this adjusted scoring rule in practice and demonstrate its improvements over the classical QSR. Section 5 concludes and Appendix A characterizes reporting behavior for the general asymmetric L-adjusted QSR and contains proofs of all results. Appendix B provides images and instructions from the experimental interface.

2 The Model

We consider an agent who must report a subjective belief about the chances of an uncertain future event A. Her true belief is that event A will occur (X=1) with probability p and its complement $\bar{A}$ will occur (X=0) with probability $1 - p$. She submits a reported probability $r \in [0, 1]$ that A will occur and receives a payoff according to an L-adjusted QSR, a generalization of the asymmetric QSR introduced by Winkler (1994).

Definition 1 (*L*-adjusted Quadratic Scoring Rule) *The L-adjusted asymmetric QSR is defined by*

$$S_L(X,r) = \begin{cases} \frac{(1-c)^2-(1-r)^2}{c^2 L} & \textit{if } A \textit{ occurs and } r < c, \\ \frac{c^2-r^2}{c^2} & \textit{if } \bar{A} \textit{ occurs and } r < c, \\ \frac{(1-c)^2-(1-r)^2}{(1-c)^2} & \textit{if } A \textit{ occurs and } r \geq c, \\ \frac{c^2-r^2}{(1-c)^2 L} & \textit{if } \bar{A} \textit{ occurs and } r \geq c. \end{cases}$$

In general, the L-adjusted QSR can be centered around any baseline probability c of the event A occurring,[2] but for most of the paper we will focus on the typical case of a symmetric baseline $c = 1/2$. When $L = 1$ this scoring rule reduces to the asymmetric QSR and when $L = 1$ and $c = 1/2$ it reduces to the classical binary QSR.

To understand how an agent will respond to this risky payoff function, we apply a prospect theory model of risk preferences. Prospect theory applies psychological principles to incorporate several important and frequently observed behavioral tendencies into the neoclassical expected utility model of preferences. This more flexible formulation provides a useful descriptive model of choice under risk (Camerer, 2000) and generally includes four main behavioral components:

1. **Reference Dependence:** The agent evaluates outcomes as differences relative to a reference point rather than in absolute levels.

2. **Loss Aversion:** Outcomes that fall below the reference point ("losses") are felt more intensely than equivalent outcomes above the reference point ("gains").

3. **Risk Aversion in Gains, Risk Seeking in Losses, and Diminishing Sensitivity to Both Gains and Losses:** The agent tends to prefer for a sure moderate-sized outcome over an equal chance of a large gain or zero gain, but prefers an equal chance of taking a large loss or avoiding the loss altogether over a sure moderate-sized loss. In addition, the marginal effect of changes in the outcome for the agent diminish as the outcome moves away from the reference point.

4. **Probability Weighting:** The agent overweights probabilities close to 0 and underweights probabilities close to 1.

To simultaneously model each of these behavioral tendencies, we assume that the agent possesses a reference-dependent utility function of the form of Palley (2013), in which the

[2] The assessor may find it useful to select an asymmetric baseline $c \neq 1/2$ if he expects the assessed probability of the event to be particularly low (e.g. the probability of rain on a given day in a desert location) or high (e.g. the probability of rain on a given day in a rainforest location). However, in practice this baseline usually remains equal to $c = 1/2$, as in the classical QSR.

agent develops an expectation E about her outcome S from the scoring rule, and this expected outcome then forms a natural reference point for her to evaluate the outcome that she ultimately receives. When her outcome exceeds this expectation, she feels an additional gain of $(S-E)^\alpha$, where $\alpha \in (0,1]$ specifies the curvature of her risk preferences. When her outcome falls below her expectation, she perceives this as an additional loss equal to $-\lambda(E-S)^\alpha$, where $\lambda \geq 1$ additionally parameterizes the agent's degree of loss aversion. Mathematically, this utility function is specified by

$$v(S,E) = \begin{cases} E - \lambda(E-S)^\alpha & \text{if } S < E \\ E + (S-E)^\alpha & \text{if } S \geq E. \end{cases}$$

If $\alpha = \lambda = 1$, then this formulation coincides with the risk-neutral objective of maximizing expected payoff that the definition of a proper scoring rule implicitly assumes.[3]

In addition, we assume that the agent applies probability weighting functions $w_+(p)$ and $w_-(p)$ for scores that fall above and below E (positive and negative events), respectively. $w_+(\cdot)$ and $w_-(\cdot)$ are assumed to be strictly increasing with $w_+(0) = w_-(0) = 0$, $w_+(1) = w_-(1) = 1$, and $w_+(p) + w_-(1-p) = 1$ for all $p \in [0,1]$.[4] The *ex ante* expected-valuation that an agent receives from responding to a binary scoring rule is then given by a probability-weighted sum over the possible scores; $V(E) = \sum_S w(p_S)v(S,E)$.[5]

As in Palley (2013), the reference point E is determined endogenously according to the consistency equation $V(E) = E$. In other words, the agent's expected-valuation of the prospect when using E as her reference point should be equal to E itself. In this sense, a consistent reference point E is the expectation that perfectly balances the agent's potential gains against her potential losses, weighted according to her beliefs of their respective likelihoods. The consistent reference point E is the natural evaluation of a prospect for

[3]Several existing studies estimate average parameters λ and α for the general population using the classical Cumulative Prospect Theory model of Tversky and Kahneman (1992), who find $\lambda = 2.25$ and $\alpha = 0.88$. In recent work, Tu (2005) finds $\lambda = 3.18$ and $\alpha = 0.68$, Abdellaoui, Bleichrodt, and Paraschiv (2007) find $\lambda = 2.54$ and $\alpha = 0.72$, and Booij, van Praag, and van de Kuilen (2010) estimate $\lambda = 1.58$ and $\alpha = 0.86$. While these estimates are derived from a different model of risk (CPT with a fixed rather than endogenous reference point), their interpretation corresponds directly to our model, so we use an average of these estimates as a rough benchmark for a representative agent of the general population.

[4]These assumptions hold trivially for the unweighted case $w(p) = p$ and approximately for most existing estimates of weighting functions that overweight low probabilities and underweight high probabilities. For example, using Goldstein and Einhorn's (1987) parameterization $w(p) = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma}$, Abdellaoui (2000) finds $\delta_+ = 0.65$, $\gamma_+ = 0.60$, $\delta_- = 0.84$ and $\gamma_- = 0.65$, Abdellaoui, Vossmann, and Weber (2005) find $\delta_+ = 0.98$, $\gamma_+ = 0.83$, $\delta_- = 1.35$ and $\gamma_- = 0.84$ and Booij, van Praag, and van de Kuilen (2010) estimate $\delta_+ = 0.77$, $\gamma_+ = 0.62$, $\delta_- = 1.02$ and $\gamma_- = 0.59$. We use a rough average of these existing estimates as a benchmark for a representative agent for the overall population, but a number of other functional forms for the weighting functions could be used as well (see, e.g., Prelec (1998)).

[5]Further details about the mechanics and intuition of this model can be found in Palley (2013). In this case decision weights reduce simply to weighted probabilities because there are only two possible outcomes.

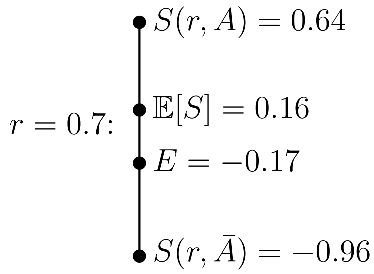


Figure 1: Two examples of an agent's possible report choices and corresponding *ex ante* reference point formation in response to a QSR with baseline $c = 0.5$ when the agent believes the probability of event A is $p = 0.7$, has prospect theory parameters $\lambda = 2.4$ and $\alpha = 1$, and does not apply probability weighting.

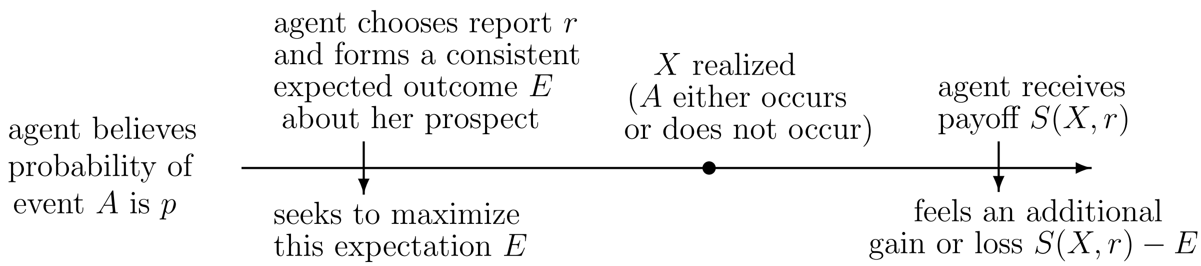


Figure 2: Timeline of the agent's report choice, reference point formation, and ex post evaluation of the event.

an agent who carefully contemplates the possible outcomes and anticipates her possible *ex post* feelings, providing a summary measure of how the agent evaluates the risk in an *ex ante* sense. An agent who initially forms a reference point higher than E will find that her expected losses $-\lambda(E - S)^\alpha$ outweigh her expected gains $(S - E)^\alpha$, causing her to adjust her expectation downwards. Conversely, an agent whose reference point is initially lower than E will find that her expected gains outweigh her expected losses, causing her adjust her reference point upwards. A thoughtful agent will thus converge to a unique consistent expectation E. This notion of expectations as an endogenously determined reference point is introduced and developed in the models of Shalev (2000), Kőszegi and Rabin (2006, 2007), and Kőszegi (2010).

Figure 1 displays an example of this reference point formation process. We see that a loss-averse agent with subjective beliefs of $p = 0.7$ would derive an *ex ante* expectation of -0.17 from truthfully reporting $r = 0.7$ in response to a QSR, while deriving an *ex ante* expectation of -0.05 from reporting $r = 0.6$. Both of these reports would therefore be dominated by reporting the baseline $r = 0.5$, which yields an outcome of 0 with certainty. We assume that the agent seeks to maximize her expected outcome E over all possible reports $r \in [0,1]$, subject to the consistency requirement, which essentially means that the agent will consider her ex post prospects when she chooses her report and forms her ex ante expectation about her outcome. The timeline of events is displayed in Figure 2.

Proposition 1 *The optimal consistent report function when $c = 0.5$ is given by*

$$r_L^*(p) = \begin{cases} \frac{\Lambda(p)^{\frac{1}{\alpha}}}{\Lambda(p)^{\frac{1}{\alpha}}+L}, & p < \min\left\{w_-^{-1}\left(\frac{L^\alpha}{\lambda+L^\alpha}\right), \frac{1}{2}\right\} \\ \frac{1}{2}, & \min\left\{w_-^{-1}\left(\frac{L^\alpha}{\lambda+L^\alpha}\right), \frac{1}{2}\right\} \leq p \leq \max\left\{w_+^{-1}\left(\frac{\lambda}{L^\alpha+\lambda}\right), \frac{1}{2}\right\} \\ \frac{L}{\Lambda(1-p)^{\frac{1}{\alpha}}+L}, & p > \max\left\{w_+^{-1}\left(\frac{\lambda}{L^\alpha+\lambda}\right), \frac{1}{2}\right\}, \end{cases}$$

where $\Lambda(p) = \frac{\lambda w_-(p)}{w_+(1-p)}$ is the agent's loss-weighted odds ratio of event A.

The optimal consistent response function for more general (asymmetric) baseline probabilities c can be found in Appendix A.[6]

Proposition 2 *For any positive linear rescaling of the payoffs $\tilde{S}_L(r) \equiv aS_L(r) + b$, $a > 0, b \in \mathbb{R}$, the optimal consistent report remains $\tilde{r}_L^*(p) = r_L^*(p)$ and the corresponding optimal ex ante expected outcome is rescaled according to $\tilde{E}^*(p) = aE^*(p) + b$.*

In other words, in contrast to the predictions of a classical prospect theory model with a fixed reference point and many classical utility formulations, the agent's behavior will be invariant to positive linear rescaling of the payoffs. This means, for example, that the agent's optimal behavior would not change if the assessor decided to pay her in a different currency with exchange rate $a : 1$ or pay her an additional fixed fee b for providing the report.

Figure 3 displays the shape of optimal reports as a function of the agent's beliefs p in response to the classical QSR. For a large region of moderate beliefs near 1/2, the agent will prefer to simply report 1/2 in order to receive a payoff of 0 with certainty. While the width of this region depends jointly on λ, α, $w_+(\cdot)$, and $w_-(\cdot)$, it is largely driven by the loss aversion parameter λ. The shape of the optimal consistent report function closely mirrors the theoretical results of Offerman et al. (2009). Here the assessor cannot simply provide the agent with the classical QSR and then infer her true beliefs from her report

[6] If $\alpha = 1$, $\lambda = 1$, and $w(p) = p$, then the optimal report is $r^*(p) = p$ and $E^*(p) = \mathbb{E}[S(X,p)]$, the expected score function in the simpler risk-neutral model (see Winkler, 1994). This behavioral model therefore includes the risk-neutral model that proper scoring rules are based on as a special case, yielding a consistent prediction regarding the reports and expected outcome. If the agent is risk-neutral, does not over- or under-weight probabilities, and does not exhibit loss aversion, then the QSR retains its ex ante incentives for truthful reporting.

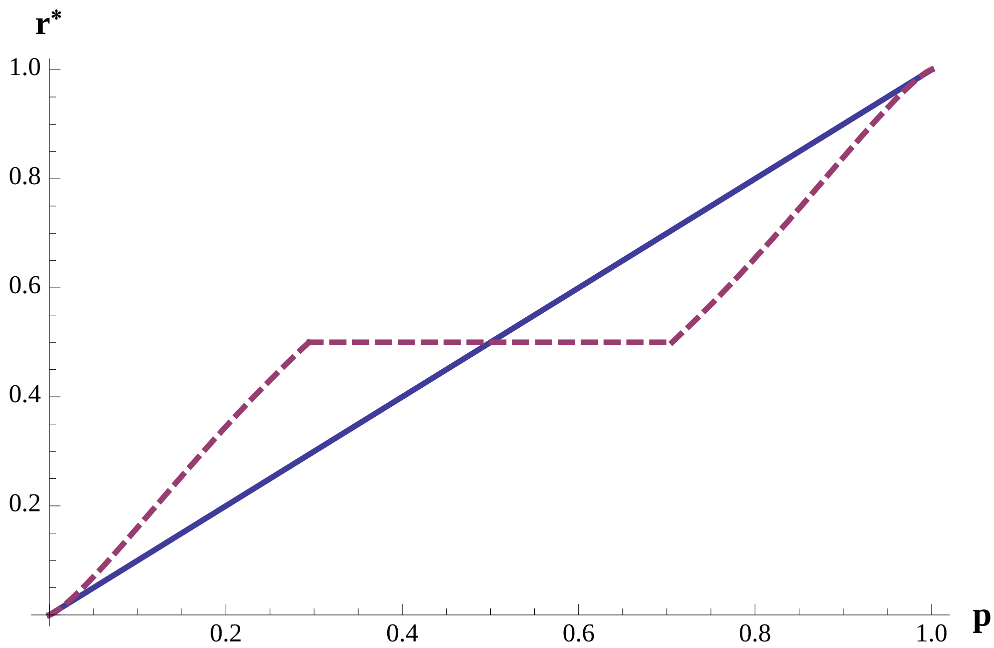


Figure 3: **Optimal consistent report $r^*(p)$ (the dashed line) to the classical QSR ($c = 0.5$, $L = 1$) for $\lambda = 2.4$, $\alpha = 0.8$, and $w_-(p) = w_+(p) = p$ versus truthful reporting (the solid line).**

because the resulting response function $r^*(p)$ is not invertible. All beliefs p in the interval $[w_-^{-1}(\frac{1}{\lambda+1}) \leq p \leq w_+^{-1}(\frac{\lambda}{\lambda+1})]$ are mapped to the conservative risk-free report of 1/2 (this is the flat region of the optimal report function). This means that observing a report of 1/2, which may happen quite frequently if the agent is loss-averse and has moderate beliefs, tells the assessor only that the agent's beliefs lie somewhere within that interval.

2.1 Determining the L-adjustment

To recover true beliefs, the assessor needs to instead adjust the scoring rule to eliminate the "flat region" of conservative reports of 1/2, which will allow him to invert the agent's optimal report function $r^*(p)$ and estimate p according to $r^{*-1}(r)$. Sensitivity analysis suggests that loss aversion accounts for the largest proportion of this conservative behavior. The best way to counteract this phenomenon, then, is to adjust the scoring rule so that negative outcomes are less severe by a factor of $\frac{1}{L}$. By computing the value L^* that solves

$$w_-^{-1}\Big(\frac{L^\alpha}{\lambda + L^\alpha}\Big) = w_+^{-1}\Big(\frac{\lambda}{L^\alpha + \lambda}\Big), \quad (1)$$

the assessor can squeeze the endpoints of the "flat region" of conservative reports of 1/2 together and recover the agent's true beliefs.

Corollary 2 *The optimal adjustment when $c = 0.5$ is given by*

$$L^* = \Lambda(1/2)^{1/\alpha} = \Big(\frac{\lambda w_-(1/2)}{w_+(1/2)}\Big)^{1/\alpha}.$$

This calibration of $L = L^*$ eliminates the agent's incentive to provide these uninformative reports even for very moderate beliefs close to 1/2, and also removes almost all of her distortion in the optimal reporting function. After receiving her report, the assessor can apply the inverse of the optimal report function to the observed report r to recover the agent's exact truthful beliefs $p = r_L^{*-1}(r)$. In the absence of utility curvature and probability weighting ($\alpha = 1$ and $w(p) = p$), the optimal adjustment is simply equal to the loss aversion parameter ($L^* = \lambda$) and the inversion step is unnecessary because the optimal report function is truthful ($r_\lambda^*(p) = p$).

In practice, an agent's report may include a noisy error term ϵ, so that the agent reports $r_L^*(p)+\epsilon$ instead. This means that the inferred beliefs will also contain an error of $r_L^{*-1}(r_L^*(p)+\epsilon) - r_L^{*-1}(r_L^*(p))$. However, since $r_L^{*-1}(\cdot)$ is differentiable and close to the identity function for a broad range of reasonable parameter values, the resulting error in inferred beliefs simply scales roughly equally to the size of the original reporting error. Another concern with the L-adjustment method is that it may become laborious if agents are very heterogeneous. In such a setting, the model parameters α, λ, and $w(p)$ and the corresponding L^* would have to be estimated individually. Our experimental results show, however, that heterogeneity is only of secondary importance and that our method does a remarkable job even without a correction of individual differences.

Figure 4 displays the optimal reports in response to an L-adjusted scoring rule, which is calibrated to average parameter estimates $\lambda = 2.4$, $\alpha = 0.8$, $\delta_+ = 0.8$, $\gamma_+ = 0.7$, $\delta_- = 1.1$ and $\gamma_- = 0.7$ (yielding $L^* = 3.7$) from the studies discussed in footnotes 2 and 3 for the general population, for an agent with various actual loss aversion parameters λ. As might be expected, given that the adjustment is primarily designed to address distortions due to loss aversion, optimal report functions are most sensitive to misestimation of the parameter of loss aversion λ, and are less sensitive to variations in α and the probability weighting functions. This suggests that if the assessor does not want to assess individual parameters, the most important measurement to focus on is λ. We also see that if L^* is miscalibrated due to errors in parameter estimates, he may observe reports both above or below the true beliefs p, depending on whether the estimate is too high or too low.

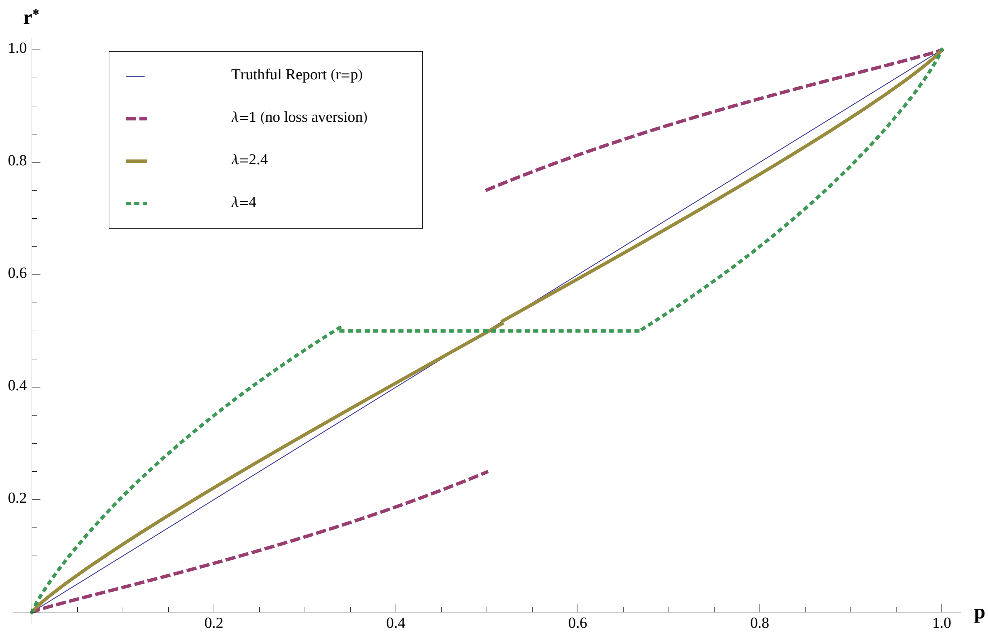


Figure 4: **Optimal consistent reports $r^*_L(p)$ in response to the L-adjusted asymmetric QSR with $L^* = 3.7$ and $c = 0.5$ for varied values of λ when $\alpha = 0.8$, $\delta_+ = 0.8$, $\gamma_+ = 0.7$, $\delta_- = 1.1$ and $\gamma_- = 0.7$ versus truthful report (the solid line).**

Next, note that any remaining difference between the optimal report function in response to the L^*-adjusted rule and truthful reporting, which in theory could be corrected by applying $r^{*-1}(\cdot)$ to the observed report, would be completely swamped by any noise in reports and the distortionary effects of errors in the parameter estimates. As a result, in practice there is very little benefit to attempting to carry out this second inversion step on the reports r. A more practical approach is to simplify the assessment process by eliminating this second inversion step and taking the reported probability as our estimate of the agent's true beliefs. In doing so, it is very important to keep in mind the remaining potential for distortion, which is mainly caused by incorrect estimation of the agent's parameters, and accept that her reports may be somewhat noisy due to this miscalibration.

If the assessor wishes to avoid the laborious process of individually assessing any parameter values for each agent beforehand, a simple approach is to simply present the agent with the L-adjusted QSR with $L^* = 3.7$ and take her resulting report as the estimate of her true beliefs. If the assessor does want to spend some time and effort to estimate the agent's parameter values ahead of time, he should focus on accurately assessing her parameter of loss-aversion λ, taking the other general population average parameter estimates $\alpha = 0.8$, $\delta_+ = 0.8$, $\gamma_+ = 0.7$, $\delta_- = 1.1$ and $\gamma_- = 0.7$ as given, since changes in these values don't have such a large effect on the optimal report function. Once the assessor has an updated estimate of λ, he can solve for L^* as before.

3 Experiment

Offerman et al. (2009) show that, in practice, proper scoring rules fail to elicit truthful reports from human agents, with patterns of reporting behavior that match the theoretical model of this paper. In this experiment we confirm the predictions of the preceding theory and demonstrate the feasibility of the L-adjusted scoring rule in recovering truthful beliefs from human subjects. In doing so, we demonstrate that the L-adjusted rule provides a simple modification of the QSR that can be used for most agents to obtain relatively accurate reports from the general population without having to arduously assess individual parameter and curvature estimates. In addition, we test several values of L and show that the proposed rescaling $L^* = 3.7$ is indeed the most effective at eliciting truthful beliefs from agents.

3.1 Experimental Design and Procedures

The computerized experiment was carried out at the CREED laboratory of the University of Amsterdam. Subjects were recruited from the undergraduate population using the standard procedure, with a total of 133 subjects participating in the experiment. Subjects earned on average 12.85 euros (€) for an experiment that lasted approximately 35 minutes. At the beginning of the experiment, participants were given an explanation of the L-adjusted QSR centered around $c = 0.5$ and a tabular depiction of how their possible payoffs would change depending on what probability they reported. The scoring rules were in units of euros rescaled by a factor of 3 and shifted upwards by 12, so that payments ranged between a minimum of €3 and a maximum of €15, and participants could assure themselves a payoff of €12 by always reporting $r = 0.5$. Subjects read the instructions on their screen at their own pace. After finishing the instructions, they had to correctly answer some control questions that tested their understanding before they could proceed to the experiment. Subjects also received a handout with a summary of the instructions before beginning the experiment (a copy of the instructions for one of the treatments is provided in Appendix B).

We employed a between-subjects design, in which each subject participated in exactly one of the three treatments. The treatments differed only in the size of the loss correction applied to the QSR. In the control treatment we used $L = 1$, which therefore corresponds to the classical QSR that has been previously employed in many experiments. We refer to this treatment as NC (mnemonic for No Correction). In treatment MC (Medium Correction), we applied a moderate-sized correction of $L = 1.5$ and in treatment LC (Large Correction) we applied the large loss correction of $L = 3.7$ derived and predicted to be optimal in Section 2.1. We ran two separate sessions for each treatment. In total, 45 subjects participated in NC, 42 subjects in MC and 46 subjects in LC.

In each treatment, subjects were informed that the experiment would last for 20 rounds and that at the end of the experiment one of the rounds would be randomly selected and used for actual payment. In each round, a subject was asked to give a probability judgment that a randomly drawn number from the set $\{0, 1, \ldots, 99, 100\}$ would be in the range $\{0, 1, \ldots, X\}$. The randomly drawn number was an integer and subjects knew that each number in the set $\{0, 1, \ldots, 99, 100\}$ was equally likely. The range was given at the start of a round and differed across rounds. The lower bound of the range was 0 and the upper bound, which determined the true objective probability, differed across rounds. In the 20 rounds we used the X values $\{5, 10, \ldots, 30, 33, 35, 40, \ldots, 95\}$, in a random order. For example, in the round that used $X = 45$, the subject was asked to give the probability judgment that the randomly drawn integer would fall in the set $\{0, 1, \ldots, 45\}$. While the subject was free to report any probability that he or she wanted, the objective probability of this event is given by $p = \frac{X+1}{101}$, so in the example of $X = 45$ the true probability was $p = \frac{46}{101}$. Each subject was presented with the ranges in a random order to prevent the possibility that order effects might confound the results. Subjects did not receive any feedback between successive rounds, so there was no opportunity to learn from previous rounds.

To provide a visual understanding of this scoring rule, all possible payoffs were listed in a payoff table that was printed on a handout and given to the subjects at the start of the experiment. Appendix B includes the three payoff tables that we used in the experiment. When a subject had tentatively decided which report r he or she wanted to provide in a given round, they were asked to type this probability judgment into a box on the upper part of the screen. Once this response was entered, the lower part of the screen then automatically displayed the relevant part of the payoff table with the current decision highlighted. Using arrows, subjects could scroll through the payoff table and if they desired, increase or decrease their report until they settled upon an ultimate response. Their choice was not finalized until they clicked the button "Satisfied with choice" (Appendix B shows the decision screen). After a subject had provided all 20 responses, the computer randomly selected exactly one round (indexed by the upper bound of its range X), which then determined his or her payment as follows: First, the computer drew a random integer from the set $\{0, 1, \ldots, 99, 100\}$ and determined whether the number was in the range $\{0, 1, \ldots, X\}$ of that round or not. Second, the payoff was determined by inputting both the realization of whether the number was in the range or not and the subject's probability judgment r for that round into to the scoring rule that the subject had faced. At the end of the experiment subjects filled out a questionnaire and were privately paid their earnings.

4 Experimental Results

Figure 5 provides an overview of the results by graphing the average reported probabilities in each treatment as a function of the true objective probability p. The solid black line presents the ideal report function of correct objective probabilities $r = p$. The control treatment NC displays a commonly observed pattern for data collected with uncorrected scoring rules. Subjects overwhelmingly bias their reports in the direction of risk aversion by reporting probabilities that are closer to 50% than the true probabilities. In the treatment with a medium correction MC, subjects' these differences are substantially diminished compared to the control treatment, but a systematic bias in the direction of risk aversion still survives. However, under the treatment with a large loss correction LC, the systematic bias vanishes and the average reported probabilities are almost identical to the true probabilities across the whole range of $p \in [0, 1]$.

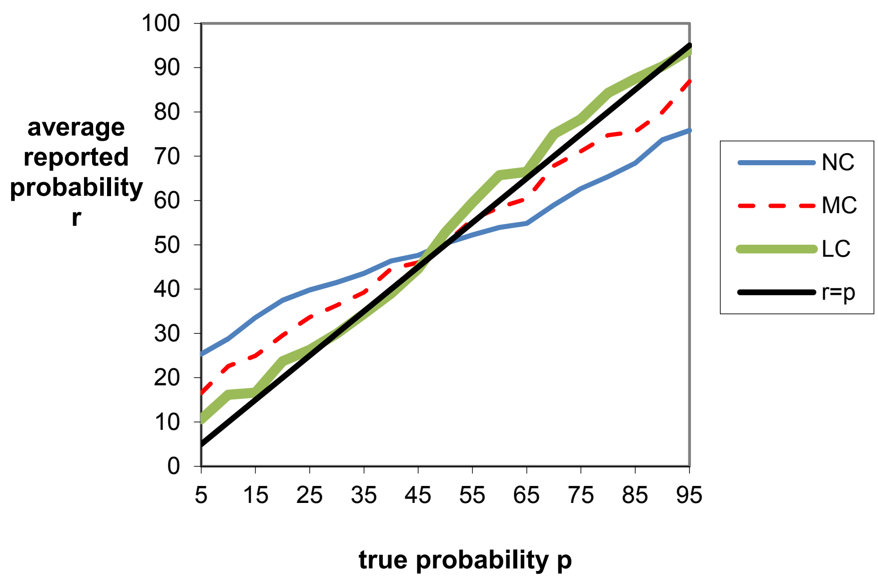


Figure 5: **Average reported probability function $r(p)$ for each treatment versus the true objective probability report $r = p$. Note that probabilities in the graph are written in percentage terms (% from 0 to 100) rather than decimal units (0 to 1).**

A good elicitation method not only avoids systematic biases but also minimizes variance in the reported probabilities, so that reports will be both honest on average and relatively

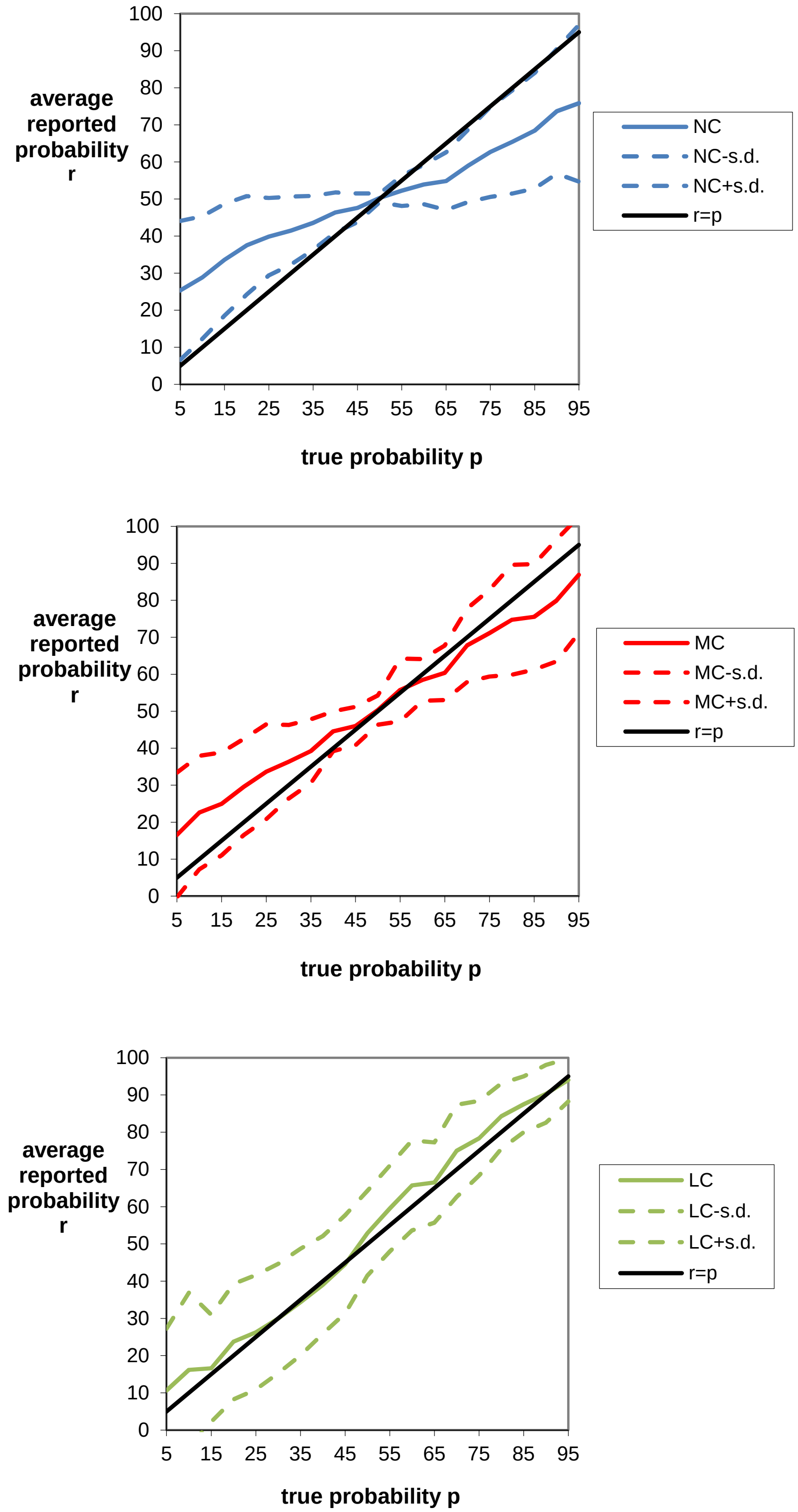


Figure 6: **Average reported probability function $r(p)$ with $+/-$ one standard deviation for each treatment versus the true objective probability report $r = p$. Note that probabilities in the graphs are written in percentage terms (% from 0 to 100) rather than decimal units (0 to 1).**

precise, meaning that a typical deviation from the honest report will not be too large. The lower part of Figure 6 provides a more detailed view of reports in each treatment by adding standard deviations above and below the average reported probabilities. For treatments NC and MC, the standard deviation is smallest for the true probability of 50% and increases proportionally with the distance between the true probability and 50%. The picture is somewhat different for treatment LC, in which the standard deviation gradually diminishes as the probability increases. Figure 7 displays the median reports in each treatment, which provides another perspective of the 'typical' behavior under each treatment. We can see that median reports in the control treatment NC display a wide flat region of uninformative reports near 0.5 that is predicted by the preceding theory. This characteristic flat region, which is highlighted more readily by the computation of the proportion of 50% reports in Table 1 below, is masked in the graphs in Figure 6 because the underlying flat region is averaged against more extreme reports.

Table 1 compares the performance of the treatments with respect to five measures. First, for each subject we computed the average absolute difference between the reported and true probabilities. Both treatments MC and LC that apply a loss correction perform substantially better than the control treatment without such a correction, with absolute errors roughly halved. Mann-Whitney tests reveal that the differences between both MC and NC and LC and NC are significant. Treatment LC performs on average somewhat better than MC, but this difference is not significant. A similar picture emerges for our second error measure, which is based on subjects' average squared differences between reported and true probabilities. Again, the MC and LC treatments substantially and significantly outperform the control treatment NC, and while LC additionally seems to do a somewhat better job than MC, the latter difference is not significant.

As a third measure, we computed the Spearman rank correlation between reported and true probabilities for each subject. Ideally, a belief elicitation measure would elicit beliefs that perfectly correlate with true probabilities. In studies that employ uncorrected scoring rules, it is well known that a few subjects are very much attracted by the sure payoff corresponding to a report of 50%, which results in a poor correlation between reported and true probabilities. Table 1 shows that indeed MC and in LC produce substantially and significantly higher Spearman rank correlation coefficients than NC does. Interestingly, with respect to this measure, LC also performs significantly better than MC, even though the difference remains moderate. As a fourth measure, we compare the treatments to the extent that they induce uninformative 50% reports. If subjects were to always report true probabilities, reports of 50% should occur in only 1/20th of the cases. NC and MC substantially overshoot this ideal benchmark, with frequencies of 50% reports equaling 39.8% and 23.5%, respectively.

Table 1: Comparison between treatments.

Treatment	$\|p-r\|$	$(p-r)^2$	Spearman Rank ρ	Frequency of 50% Reports	Risk Bias
NC	12.9 (11.8)	305.9 (524.5)	0.75	39.8%	12.0 (13.3)
MC	8.5 (10.1)	173.7 (358.1)	0.87	23.5%	5.7 (12.3)
LC	7.9 (10.9)	180.9 (541.2)	0.92	10.1%	-1.0 (13.5)
		Mann-Whitney probability			
NC versus MC	0.00	0.01	0.00	0.02	0.00
NC versus LC	0.00	0.00	0.00	0.00	0.00
MC versus LC	0.43	0.46	0.01	0.01	0.00

Note that probabilities in the table are written in percentage terms (% from 0 to 100) rather than decimal units (0 to 1). Each cell lists the average for the relevant statistics, with the standard deviations in parentheses. p denotes the true probability (determined by the range X according to $p = \frac{X+1}{101}$) and r denotes the reported probability. Risk bias equals $p - r$ if $p > 0.5$, and equals $r - p$ if $p < 0.5$ (cases where $p = 0.5$ are excluded). Mann-Whitney tests use average statistics per subject as data points (45 subjects in NC, 42 subjects in MC, and 46 subjects in LC).

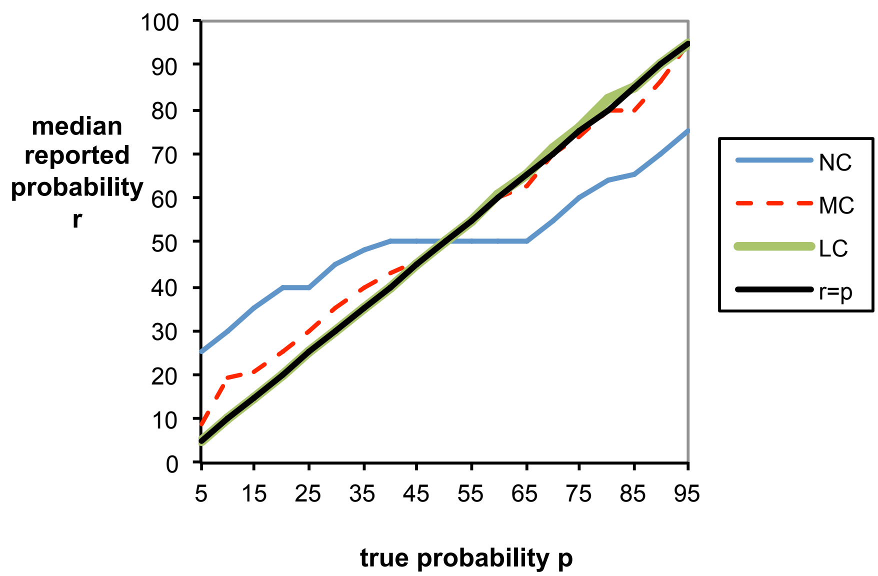


Figure 7: **Median reported probability function $r(p)$ for each treatment versus the true objective probability report $r = p$. Note that probabilities in the graph are written in percentage terms (% from 0 to 100) rather than decimal units (0 to 1).**

In comparison, LC performs very well, producing such reports only 10.1% of the time. All pairwise differences between the treatments are significant with respect to this frequency of 50% reports. Finally, we make precise the extent to which the three treatments suffer from systematic risk biases. For each subject, we computed how much on average a subject biased the report in the direction of 50%. If the average risk bias is positive (negative) then this provides evidence that subject are risk averse (risk seeking). In line with Figure 6, the final column of Table 1 shows that subjects are very biased in the direction of risk aversion in treatment NC. In treatment LC, there is virtually no bias, and the bias in treatment MC falls roughly in the middle of the two other treatments. All risk bias differences between the treatments are highly significant.

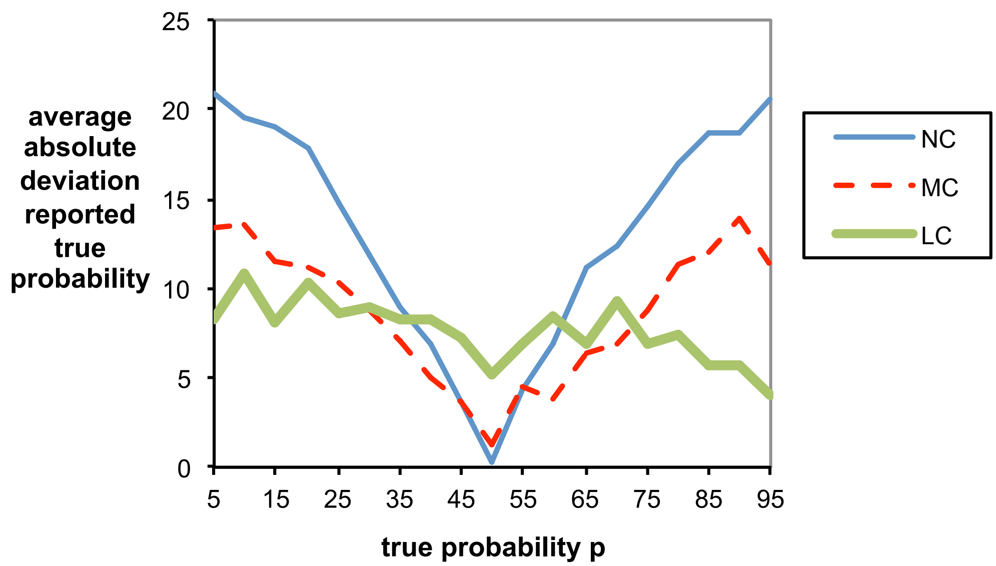


Figure 8: **Average absolute error $|r-p|$ in the reported probability function $r(p)$ for each treatment. Note that probabilities in the graph are written in percentage terms (% from 0 to 100) rather than decimal units (0 to 1).**

Figure 8 shows how average absolute errors $|r - p|$ in the report vary with the objective probability p in each of the treatments. The uncorrected scoring rule performs well precisely where we would expect it to—the incentives to make a conservative baseline report of 50% impel almost unanimously honest reporting when the objective probability is in fact very close to 50%. However, the uncorrected scoring rule performs far worse than the loss-corrected scoring rules when the true probabilities are larger than approximately 65% or smaller than approximately 35%. In other words, errors in the uncorrected scoring rule occur exactly in cases where the effects of loss and risk aversion kick in most heavily. Overall, the uncorrected scoring rule thus proves to be unreliable for eliciting subjective beliefs, since the assessor does not know which of these regions the true probability belongs to.

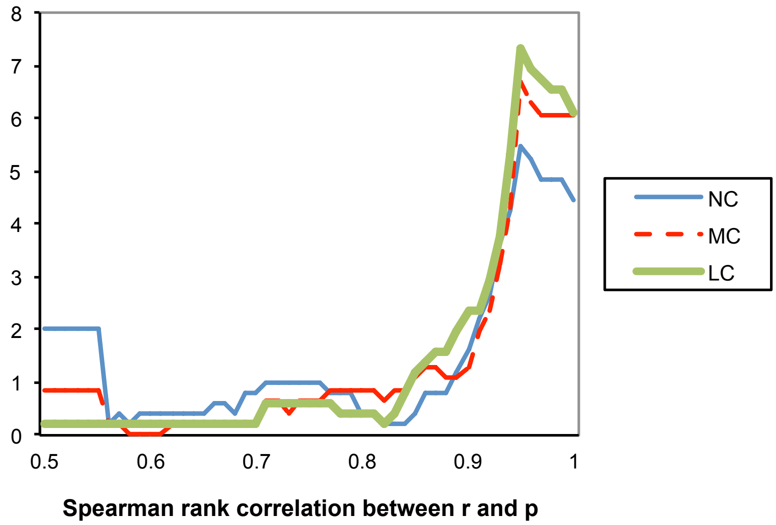


Figure 9: **Histogram of the Spearman-rank correlation (SRC) between the true probabilities p and the subject's reported probabilities r. The figure displays for each SRC the percentage of subjects that fall in the interval [SRC - 0.05, SRC + 0.05]. The few observations where SRC< 0.5 are added to SRC$= 0.5$.**

Figure 9 displays the empirical density of the Spearman-rank correlation coefficients in the three treatments. In all treatments most subjects have a fairly high Spearman-rank correlation coefficients larger than 0.9, while a few subjects have very low coefficients smaller than or equal to 0.5. The treatments differ primarily in the relative frequency of these two categories of correlation coefficient (high or low). The proportion of overly cautious or haphazard reporters with a low coefficient of less than or equal to 0.5 equals 20.0% in NC, 7.1% in MC and only 2.2% in LC. [7]

5 Discussion

In practice, quadratic and other proper scoring rules can fail to recover the true probabilistic beliefs that they are designed to elicit. Distortions in agents' reports generally take one of two forms: First, a risk-averse agent may bias her report away from categorical beliefs

[7] In total, 6 subjects in the NC treatment and 2 subjects in the MC treatment reported 50% in every round, while no subject in the LC treatment reported 50% in every round. In contrast, the proportion of consistent reporters with a coefficient of at least 0.9 equaled 60.0% in NC, 73.8% in MC and 80.4% in LC.

of 0 and 1, as predicted by, for example, the theory of Winkler and Murphy (1970). Second, a risk-averse agent with moderate beliefs close to the baseline probability of 1/2 may revert to simply reporting 1/2 in order to receive a risk-free payoff. In other words, under proper scoring rules such as the classical QSR, we should expect to see a large proportion of uninformative reports of $r = 0.5$, and even strong beliefs near 0 or 1 will be skewed towards this focal point of $c = 0.5$. This pattern of conservative behavior, which has been observed experimentally by, for example, Offerman et al. (2009) and in the experiment of this paper, is explained by the prospect theory model in Section 2 of this paper.[8]

The predictions of this theory reinforce the existing result that agents may not reveal their true beliefs even when assessed by a proper scoring rule, and provides an explanation for when and why we might expect to see these two forms of distortions. As demonstrated in Section 2, both effects appear to be largely driven by loss aversion, which motivates the agents to seek a certain payoff when they have moderate beliefs and to lower their risk by generally shading their reports closer towards 1/2 for stronger beliefs. The intuition here is that reporting something other than 1/2 introduces uncertainty into the payoffs, so that some outcomes will be felt as gains and some outcomes will be felt as losses. As a result, a loss-averse agent with beliefs close to 1/2 (who doesn't have much better information than the default baseline prediction) will not find it worthwhile to expose herself to the possibility of these losses. The L^*-adjusted QSR, which generalizes the classical QSR, can be calibrated to correct for both forms of distortions predicted by this prospect theory model of optimal reports. The L-adjusted QSR provides a simple scoring rule that can be used in a straightforward manner to elicit an agent's true subjective probabilistic beliefs. The main challenge in successfully implementing this rule is that the optimal choice of L^* requires an accurate estimate of the agent's parameters α, λ, and $w(p)$. In particular, when applying this adjustment the assessor needs to be careful not to use an unsuitable value of L. For example, an agent who is really risk neutral will respond to an L-adjusted scoring rule by biasing her reports away from 1/2 for any choice of $L > 1$.

Our experimental results demonstrate that people do behave as predicted by the theory. Our data suggests that the optimal calibration of $L^* = 3.7$ for the average population does indeed perform better than the other treatments, but even the moderate-sized correction of $L = 1.5$ provides a vast improvement over the classical unadjusted QSR. The major potential benefits of this L-adjustment include eliminating the flat region of reports $r = 1/2$ for moderate beliefs, which are uninformative and prevent the optimal report function from

[8] In situations where agents receive rewards from reporting beliefs and from making additional decisions, other distortions may emerge. In such cases, agents may hedge their beliefs, for example in order to guarantee a minimal payoff. The extent to which hedging biases reports when beliefs are incentivized is discussed in Blanco et al. (2010) and Armantier and Treich (2013).

being inverted, and de-biasing reports, so that they provide truthful subjective beliefs on average.

In theory, when processing reports, an assessor would need to implement an additional second step of computing $r_L^{*-1}(\cdot)$ and inferring true beliefs according to $r_L^{*-1}(r)$ rather than simply using the raw report r as the estimate. In practice, however, the impact of this additional step will be very small and likely dwarfed by noise in the reports and errors in the calibration of L to the agent. Our experiment confirms that the second step is indeed unnecessary, and that reports can be simply recorded as provided in a straightforward manner.

For the general population, $L = 3.7$ does seem to be the best adjustment to use, as predicted by applying existing empirical estimation of population parameters to our theoretical results and as evidenced in our experiment. However, if any additional information such as individual-specific parameters is available, an assessor should adjust the size of L^* according to Corollary 1 in order to achieve more accurate results, with likely lower noise in the reports.

Finally, we would like to emphasize that while we only formally examined L-adjustments to a QSR, an exactly analogous adjustment could be applied to any other proper scoring rule with bounded payoffs. Applying the same analysis of behavior under risk will yield similar results; we would expect loss aversion to induce both a region of uninformative baseline reports for moderate beliefs and reports that are biased away from the agent's true belief for stronger beliefs. The same L-adjustment should be equally effective at recovering informative responses by pulling the endpoints of the interval of baseline reports together until this "flat region" in the response function is eliminated. While there exists a closed-form solution for these results under the QSR, these optimal response functions and L-adjustments would have to be solved numerically for more general scoring rules.

References

Abdellaoui, M. (2000). Parameter-free elicitation of utilities and probability weighting functions. *Management Science* **46(11),** 1497–1512.

Abdellaoui, M., Bleichrodt, H., Paraschiv, C. (2007). Loss Aversion Under Prospect Theory: A Parameter-Free Measurement. *Management Science* **53(10),** 1659–1674.

Abdellaoui, M., Vossmann, F., Weber, M. (2005). Choice-based elicitation and decomposition of decision weights for gains and losses under uncertainty. *Management Science* **51,** 1384–1399.

Allen, F. (1987). Discovering Personal Probabilities When Utility Functions Are Unknown. *Management Science* **33(4),** 542–544.

Andersen, S., Fountain, J., Harrison, G.W., Rutström, E. (2010). Estimating Subjective Probabilities. *Discussion Paper, Georgia State University.*

Armantier, O., Treich, N. (2013). Eliciting beliefs: Proper scoring rules, incentives, stakes and hedging. *European Economic Review.* **62,** 17–40.

Becker, G.M., DeGroot, M.H., Marschak, J. (1964). Measuring Utility By A Single-Response Sequential Method. *Behavioral Science* **9,** 226-232.

Blanco, M., Engelmann, D., Koch, A.K., Normann, H. T. (2010). Belief Elicitation in Experiments: Is There a Hedging Problem? *Experimental Economics* **13,** 412–438.

Booij, A.S., van Praag, B.M.S., van de Kuilen, G. (2010). A parametric analysis of prospect theory's functionals for the general population. *Theory and Decision* **68,** 115–148.

Brier, G.W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review* **78,** 1–3.

Camerer, C.F. (2000). Prospect theory in the wild: Evidence from the field. in Camerer, C.F., Loewenstein, G. and Rabin, M., eds., *Advances in Behavioral Economics,* Princeton University Press, Princeton, NJ.

Cason, T.N., Plott, C.R. (2012). Misconceptions and Game Form Recognition of the BDM Method: Challenges to Theories of Revealed Preference and Framing. *Social Science Working Paper 1364, California Institute of Technology.*

Costa-Gomes, M.A., Weizsäcker, G. (2008). Stated Beliefs and Play in Normal-Form Games. *Review of Economic Studies* **75,** 729–762.

Ghysels, E. (1993). On Scoring Asymmetric Periodic Probability Models of Turning-point Forecasts. *Journal of Forecasting* **12,** 227–238.

Goldstein, W.M., Einhorn, H.J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review* **94,** 236–254.

Diebold, F.X., Rudebusch, G.D. (1989). Scoring the Leading Indicators. *The Journal of Business* **62(3),** 369–391.

Hao, L., Houser, D. (2010). Getting It Right the First Time: Belief Elicitation With Novice Participants. *George Mason University Department of Economics Paper No. 10-12.*

Hollard, G., Massoni, S., Vergnaud, J. (2010). Subjective beliefs formation and elicitation rules: experimental evidence. *Université Paris CES Working Paper 2010.88.*

Huck, S., Weizsäcker, G. (2002). Do Players Correctly Estimate What Others Do? Evidence of Conservatism in Beliefs. *Journal of Economic Behavior and Organization* **47,** 71–85.

Kahneman, D., Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica* **47(2),** 263–292.

Karni, E. (2009). A Mechanism for Eliciting Probabilities. *Econometrica* **77(2),** 603–606.

Keeney, R.L. (1982). Decision Analysis: An Overview. *Operations Research* **30(5),** 803–838.

Koszegi, B. (2010). Utility from anticipation and personal equilibrium. *Economic Theory* **44,** 415–444.

Koszegi, B., Rabin, M. (2006). A Model of Reference-Dependent Preferences. *The Quarterly Journal of Economics* **121(4),** 1133–1165.

Koszegi, B., Rabin, M. (2007). Reference-Dependent Risk Attitudes. *The American Economic Review* **97(4),** 1047–1073.

Kothiyal, A., Spinu, V., Wakker, P. P. (2011). Comonotonic Proper Scoring Rules to Measure Ambiguity and Subjective Beliefs. *Journal of Multi-Criteria Decision Analysis* **17,** 101–113.

McKelvey, R.D., Page, T. (1990). Public and Private Information: An Experimental Study of Information Pooling. *Econometrica* **58(6),** 1321–1339.

Murphy, A.H., Winkler, R.L. (1984). Probability Forecasting in Meteorology. *Journal of the American Statistical Association* **79(387),** 489–500.

Nyarko, Y., Schotter, A. (2002). An Experimental Study of Belief Learning Using Elicited Beliefs. *Econometrica* **70(3),** 971–1005.

Offerman, T., Sonnemans, J., Schram, A. (1996). Value Orientations, Expectations and Voluntary Contributions in Public Goods. *Economic Journal* **106,** 817–845.

Offerman, T., Sonnemans, J., van de Kuilen, G., Wakker, P.P. (2009). A Truth Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes. *The Review of Economic Studies* **76,** 1461–1489.

Palley, A.B. (2013). Great Expectations: Prospect Theory with a Consistent Reference Point. *Working Paper,* Retrieved September 4, 2013 from http://ssrn.com/abstract=2313851.

Plott, C.R., Zeiler, K. (2005). The Willingness to Pay-Willingness to Accept Gap, the Endowment Effect, Subject Misconceptions, and Experimental Procedures for Eliciting Valuations. *American Economic Review* **95(3),** 530–545.

Prelec, D. (1998). The probability weighting function. *Econometrica* **66(3),** 497–527.

Rutström, E.E. (1998). Home-Grown Values and the Design of Incentive Compatible Auctions. *International Journal of Game Theory* **27(3),** 427–441.

Savage, L.J. (1971). Elicitation of Personal Probabilities and Expectations. *Journal of the American Statistical Association* **66(336),** 783–801.

Schervish, M.J. (1989). A General Method for Comparing Probability Assessors. *The Annals of Statistics* **17(4),** 1856–1879.

Schlag, K., Tremewan, J., van der Weele, J. (2013). A Penny for Your Thoughts: A Survey of Incentives and Belief Elicitation. *Discussion Paper, University Vienna.*

Shuford, E.H., Albert, A., Massengill, H.E. (1966). Admissible Probability Measurement Procedures. *Psychometrika* **31(2),** 125–145.

Shalev, J. (2000). Loss aversion equilibrium. *International Journal of Game Theory* **29,** 269–287.

Trautmann, S.T., van de Kuilen, G. (2011). Belief Elicitation: A Horse Race Among Truth Serums. *Tilburg University Center for Economic Research Discussion Paper 2011-117.*

Tu, Q. (2005). Empirical analysis of time preferences and risk aversion. *CentER PhD Thesis* **142,** Tilburg University.

Tversky, A., Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty* **5,** 297–323.

Winkler, R.L. (1969). Scoring Rules and the Evaluation of Probability Assessors. *Journal of the American Statistical Association* **64(327),** 1073–1078.

Winkler, R.L. (1994). Evaluating Probabilities: Asymmetric Scoring Rules. *Management Science* **40(11),** 1395–1405.

Winkler, R.L. (1996). Scoring Rules and the Evaluation of Probabilities. *Test* **5(1),** 1–60.

Winkler, R.L., Murphy, A.H. (1970). Nonlinear Utility and the Probability Score. *Journal of Applied Meteorology* **9(1),** 143–148.

6 Appendix A

6.1 Reporting Under a General Asymmetric L-adjusted QSR

To derive a closed-form solution for the optimal reporting strategy in response to a general asymmetric L-adjusted QSR, we need to make an additional assumption on reporting behavior:

Definition 3 (Directional Reporting) *We say that the agent's reporting preferences are* **directional** *if* $p \leq c \Rightarrow r \leq c$ *and* $p \geq c \Rightarrow r \geq c$ *for any beliefs* p.

Directional reporting holds automatically in the optimal reporting strategy when the baseline $c = 0.5$ (this follows through a symmetry argument, see the proof of Proposition 1 for details) as in the classical QSR, and approximately for a broad range of realistic parameter values when $c \neq 0.5$. Assuming directional reports is natural in the context of real-world agents who are asked to provide a report of their beliefs relative to some baseline, asserting that reporting behavior will be restricted to shading reports either towards or away from the baseline probability of c. While this assumption is not required to compute an agent's optimal consistent report function, we will assume that it holds for the analysis that follows because it allows for closed-form solutions.

Proposition 3 *If the agent reports directionally, the optimal consistent report function to an asymmetric L-adjusted QSR is*

$$r_L^*(p) = \begin{cases} \frac{\Lambda(p)^{\frac{1}{\alpha}}}{\Lambda(p)^{\frac{1}{\alpha}}+L}, & p < \min\left\{c, w_-^{-1}\left(\frac{(\frac{cL}{1-c})^\alpha}{\lambda+(\frac{cL}{1-c})^\alpha}\right)\right\} \\ c, & \min\left\{c, w_-^{-1}\left(\frac{(\frac{cL}{1-c})^\alpha}{\lambda+(\frac{cL}{1-c})^\alpha}\right)\right\} \le p \le \max\left\{c, w_+^{-1}\left(\frac{\lambda}{(\frac{(1-c)L}{c})^\alpha+\lambda}\right)\right\} \\ \frac{L}{\Lambda(1-p)^{\frac{1}{\alpha}}+L}, & p > \max\left\{c, w_+^{-1}\left(\frac{\lambda}{(\frac{(1-c)L}{c})^\alpha+\lambda}\right)\right\}, \end{cases}$$

where $\Lambda(p) = \frac{\lambda w_-(p)}{w_+(1-p)}$ *is the agent's loss-weighted odds ratio of event A.*

As before, the results are preserved under arbitrary positive linear rescaling of the payoffs. In other words, for any positive linear rescaling of the payoffs $\tilde{S}_L(r) \equiv aS_L(r)+b$, $a > 0, b \in \mathbb{R}$, the optimal consistent report remains $\tilde{r}_L^*(p) = r_L^*(p)$ and the corresponding optimal expected outcome is simply rescaled according to $\tilde{E}^*(p) = aE^*(p) + b$.

The assessor should then calibrate the L-adjusted QSR by selecting the value of L^* that solves

$$w_-^{-1}\left(\frac{(\frac{cL^*}{1-c})^\alpha}{\lambda + (\frac{cL^*}{1-c})^\alpha}\right) = w_+^{-1}\left(\frac{\lambda}{(\frac{(1-c)L^*}{c})^\alpha + \lambda}\right)$$

in order to eliminate the flat region of uninformative reports of c.

6.2 Proofs of Propositions 1 - 3 and Corollary 1

Proof of Proposition 1: This result is a special case of Proposition 3, with $L = 1$, $a = 1$, and $b = 0$. Note that directionality holds automatically if $c = 0.5$. To prove this, it is sufficient to show that $\check{E}^*(p) \ge \hat{E}^*(p)$ for $p \le 0.5$ and $\hat{E}^*(p) \ge \check{E}^*(p)$ for $p \ge 0.5$ whenever the expectations $\check{E}^*(p)$ and $\hat{E}^*(p)$ are both consistent. Observe that when $c = 0.5$, $\check{E}(r \mid p) = \hat{E}(1 - r \mid 1 - p)$, where $E(r \mid p)$ denotes a consistent expectation from reporting r when the probability beliefs is p. Since $\check{r}^*(p) = 1 - \hat{r}^*(1 - p)$, we have that $\check{E}(\check{r}^*(p) \mid p) = \check{E}(1 - \hat{r}^*(1 - p) \mid p) = \hat{E}(\hat{r}^*(1 - p) \mid 1 - p)$, or $\check{E}^*(p) = \hat{E}^*(1 - p)$. In particular, this means that $\check{E}^*(1/2) = \hat{E}^*(1/2)$ and $\check{E}^*(p)$ and $\hat{E}^*(p)$ are symmetric around $p = 0.5$. Then to prove directionality it suffices to show that $\check{E}^*(p)$ is decreasing in p. As shown in Proposition 3, $\check{E}'(r) < 0$. $\frac{d}{dp}\check{r}^*(p) = \frac{L\left(\frac{d}{dp}\Lambda(p)^{\frac{1}{\alpha}}\right)}{\left(\Lambda(p)^{\frac{1}{\alpha}}+L\right)^2}$, where $\Lambda(p)^{\frac{1}{\alpha}} \ge 0$ and $\frac{d}{dp}\Lambda(p)^{\frac{1}{\alpha}} = \frac{1}{\alpha}\Lambda(p)^{(\frac{1}{\alpha}-1)}\frac{w_+(1-p)\lambda w_-'(p) - \lambda w_-(p)w_+'(1-p)}{(w_+(1-p))^2} \ge 0$, so $\frac{d}{dp}\check{r}^*(p) \ge 0$. Then by the chain rule $\frac{d}{dp}\check{E}^*(p) \le 0$, which implies that $\check{E}^*(p)$ is decreasing in p, $\hat{E}^*(p)$ is increasing in p, and the agent will prefer to report $r \le 0.5$ for $p \le 0.5$ and $r \ge 0.5$ for $p \ge 0.5$.

Proof of Proposition 2: This result also follows from Proposition 3, by setting $L = 1$ and comparing the case where $a = 1$, and $b = 0$ to the case of general a and b. The optimal consistent report function is the same in both cases, and after simplifying the expressions for the corresponding consistent expected outcome, we also have that $\tilde{E}^*(p) = aE^*(p) + b$ for all p.

Proof of Proposition 3: By Lemma 1 of Palley (2013), for any risky prospect that yields a payoff of y with probability p and z with probability $1-p$, where $y \leq z$, there exists a unique consistent expected outcome $E \in [y, z]$ such that $V(E) = E$, which means there will be a unique E associated with any report r. Consider an $L-$adjusted asymmetric QSR whose payoffs have been rescaled by an arbitrary positive linear transformation $aS_L(X, r)+b$, $a > 0, b \in \mathbb{R}$. The agent must consider three separate cases when selecting the value of r to report:

Case 1 ($r < c$): $S_L(r) = \begin{cases} a\frac{(1-c)^2-(1-r)^2}{c^2L} + b < b & \text{if } A \text{ occurs} \\ a\frac{c^2-r^2}{c^2} + b > b & \text{if } \bar{A} \text{ occurs,} \end{cases}$ meaning that $a\frac{(1-c)^2-(1-r)^2}{c^2L} + b < E < a\frac{c^2-r^2}{c^2} + b$ and

$$v(S_L(r), E) = \begin{cases} E - \lambda\big(E - (a\frac{(1-c)^2-(1-r)^2}{c^2L} + b)\big)^\alpha & \text{if } A \text{ occurs} \\ E + (a\frac{c^2-r^2}{c^2} + b - E)^\alpha & \text{if } \bar{A} \text{ occurs.} \end{cases}$$

Consistency requires that $w_-(p)\left(E-\lambda(E-a\frac{(1-c)^2-(1-r)^2}{c^2L}-b)^\alpha\right)+w_+(1-p)\left(E+(a\frac{c^2-r^2}{c^2}+b-E)^\alpha\right) = E$, so the consistent expectation for $r < c$ is $\check{E}(r) = a\frac{\frac{c^2-r^2}{c^2}+\Lambda(p)^{\frac{1}{\alpha}}\left(\frac{(1-c)^2-(1-r)^2}{c^2L}\right)}{1+\Lambda(p)^{\frac{1}{\alpha}}} + b$, where $\Lambda(p) \equiv \frac{\lambda w_-(p)}{w_+(1-p)}$. $\check{E}$ is a concave quadratic function of r, so its maximum occurs where $\check{E}'(r) = \frac{2a}{c^2\left(1+\Lambda(p)^{\frac{1}{\alpha}}\right)}\left(\frac{1}{L}\Lambda(p)^{\frac{1}{\alpha}}(1-r) - r\right) = 0$, meaning that $\check{r}^* = \frac{\Lambda(p)^{\frac{1}{\alpha}}}{\Lambda(p)^{\frac{1}{\alpha}}+L}$. However, this is only consistent for $r < c$, so this reporting strategy is optimal only if $\frac{\Lambda(p)^{\frac{1}{\alpha}}}{\Lambda(p)^{\frac{1}{\alpha}}+L} < c$, or equivalently, only if $p < w_-^{-1}\big(\frac{(\frac{cL}{1-c})^\alpha}{\lambda+(\frac{cL}{1-c})^\alpha}\big)$. Then for all $p < w_-^{-1}\big(\frac{(\frac{cL}{1-c})^\alpha}{\lambda+(\frac{cL}{1-c})^\alpha}\big)$, $\check{E}^*(p) = \check{E}(\frac{\Lambda(p)^{\frac{1}{\alpha}}}{\Lambda(p)^{\frac{1}{\alpha}}+L}) > \check{E}(c) = b$, and for all $p \geq w_-^{-1}\big(\frac{(\frac{cL}{1-c})^\alpha}{\lambda+(\frac{cL}{1-c})^\alpha}\big)$, reporting $r < c$ is not optimal.

Case 2 ($r > c$): $S_L(r) = \begin{cases} a\frac{(1-c)^2-(1-r)^2}{(1-c)^2} + b < b & \text{if } A \text{ occurs} \\ a\frac{c^2-r^2}{(1-c)^2L} + b > b & \text{if } \bar{A} \text{ occurs,} \end{cases}$ meaning that $a\frac{c^2-r^2}{(1-c)^2L} + b < E < a\frac{(1-c)^2-(1-r)^2}{(1-c)^2} + b$ and

$$v(S_L(r), E) = \begin{cases} E + (a\frac{(1-c)^2-(1-r)^2}{(1-c)^2} + b - E)^\alpha & \text{if } A \text{ occurs} \\ E - \lambda(E - \big(a\frac{c^2-r^2}{(1-c)^2L} + b)\big)^\alpha & \text{if } \bar{A} \text{ occurs.} \end{cases}$$

Consistency requires that $w_+(p)\Bigg(E + (a\frac{(1-c)^2-(1-r)^2}{(1-c)^2} + b - E)^\alpha\Bigg) + w_-(1-p)\Bigg(E - \lambda(E - a\frac{c^2-r^2}{(1-c)^2 L} - b)^\alpha\Bigg) = E$, so the consistent expectation for $r > c$ is $\hat{E}(r) = a\frac{\frac{(1-c)^2-(1-r)^2}{(1-c)^2}+\Lambda(1-p)^{\frac{1}{\alpha}}\left(\frac{c^2-r^2}{(1-c)^2L}\right)}{1+\Lambda(1-p)^{\frac{1}{\alpha}}} + b$. $\hat{E}$ is a concave quadratic function of r, so its maximum occurs where $\hat{E}'(r) = \frac{2a}{\left(1+\Lambda(1-p)^{\frac{1}{\alpha}}\right)(1-c)^2}\Bigg((1-r) - \Lambda(1-p)^{\frac{1}{\alpha}}\frac{r}{L}\Bigg) = 0$, meaning that $\hat{r}^* = \frac{L}{\Lambda(1-p)^{\frac{1}{\alpha}}+L}$. However, this is only consistent for $r > c$, so this reporting strategy is optimal only if $\frac{L}{\Lambda(1-p)^{\frac{1}{\alpha}}+L} > c$, or equivalently, only if $p > w_+^{-1}\big(\frac{\lambda}{(\frac{(1-c)L}{c})^\alpha+\lambda}\big)$. Then for all $p > w_+^{-1}\big(\frac{\lambda}{(\frac{(1-c)L}{c})^\alpha+\lambda}\big)$, $\hat{E}^*(p) = \hat{E}(\frac{L}{\Lambda(1-p)^{\frac{1}{\alpha}}+L}) > \hat{E}(c) = b$, and for all $p \leq w_+^{-1}\big(\frac{\lambda}{(\frac{(1-c)L}{c})^\alpha+\lambda}\big)$, reporting $r > c$ is not optimal.

<u>Case 3 ($r = c$):</u> $S(r) = b$, meaning that $E = b$ and $v(S(r), E) = b$. Then consistency is satisfied since $V(E) = b = E$.

Then for any belief p, the agent has three reporting choices:

1. $r < c$, in which case she will receive $\check{E}(\check{r}^*(p))$
2. $r > c$, in which case she will receive $\hat{E}(\hat{r}^*(p))$, or
3. $r = c$, in which case she will receive $E = b$.

If $w_-^{-1}\big(\frac{(\frac{cL}{1-c})^\alpha}{\lambda+(\frac{cL}{1-c})^\alpha}\big) < w_+^{-1}\big(\frac{\lambda}{(\frac{(1-c)L}{c})^\alpha+\lambda}\big)$, then the only consistent report for $p \in [w_-^{-1}\big(\frac{(\frac{cL}{1-c})^\alpha}{\lambda+(\frac{cL}{1-c})^\alpha}\big), w_+^{-1}\big(\frac{\lambda}{(\frac{(1-c)L}{c})^\alpha+\lambda}\big)]$ is $r = c$, $r^* = \frac{\Lambda(p)^{\frac{1}{\alpha}}}{\Lambda(p)^{\frac{1}{\alpha}}+L}$ for $p < w_-^{-1}\big(\frac{(\frac{cL}{1-c})^\alpha}{\lambda+(\frac{cL}{1-c})^\alpha}\big)$, and $r^* = \frac{L}{\Lambda(1-p)^{\frac{1}{\alpha}}+L}$ for $p > w_+^{-1}\big(\frac{\lambda}{(\frac{(1-c)L}{c})^\alpha+\lambda}\big)$.

If $w_-^{-1}\big(\frac{(\frac{cL}{1-c})^\alpha}{\lambda+(\frac{cL}{1-c})^\alpha}\big) \geq w_+^{-1}\big(\frac{\lambda}{(\frac{(1-c)L}{c})^\alpha+\lambda}\big)$, then for $p < w_+^{-1}\big(\frac{\lambda}{(\frac{(1-c)L}{c})^\alpha+\lambda}\big)$ we have $r^* = \frac{\Lambda(p)^{\frac{1}{\alpha}}}{\Lambda(p)^{\frac{1}{\alpha}}+L}$, for $p > w_-^{-1}\big(\frac{(\frac{cL}{1-c})^\alpha}{\lambda+(\frac{cL}{1-c})^\alpha}\big)$ we have $r^* = \frac{L}{\Lambda(1-p)^{\frac{1}{\alpha}}+L}$, and for $p \in [w_+^{-1}\big(\frac{\lambda}{(\frac{(1-c)L}{c})^\alpha+\lambda}\big), w_-^{-1}\big(\frac{(\frac{cL}{1-c})^\alpha}{\lambda+(\frac{cL}{1-c})^\alpha}\big)]$ we have $r^* = \arg\max\big\{\check{E}(\frac{\Lambda(p)^{\frac{1}{\alpha}}}{\Lambda(p)^{\frac{1}{\alpha}}+L}), \hat{E}(\frac{L}{\Lambda(1-p)^{\frac{1}{\alpha}}+L})\big\}$. Under the assumption of directional reporting, this simply reduces to $r^* = \frac{L}{\Lambda(1-p)^{\frac{1}{\alpha}}+L}$ for $p \in [w_-^{-1}\big(\frac{(\frac{cL}{1-c})^\alpha}{\lambda+(\frac{cL}{1-c})^\alpha}\big), c]$ and $r^* = \frac{\Lambda(p)^{\frac{1}{\alpha}}}{\Lambda(p)^{\frac{1}{\alpha}}+L}$ for $p \in [c, w_+^{-1}\big(\frac{\lambda}{(\frac{(1-c)L}{c})^\alpha+\lambda}\big)]$.

The consistent *ex ante* expected outcome corresponding to the optimal consistent report

$r^*(p)$ is

$$E^*(p) = \begin{cases} a\frac{c^2 - \left(\frac{\Lambda(p)^{\frac{1}{\alpha}}}{\Lambda(p)^{\frac{1}{\alpha}}+L}\right)^2 + \frac{\Lambda(p)^{\frac{1}{\alpha}}}{L}\left((1-c)^2 - \left(\frac{L}{\Lambda(p)^{\frac{1}{\alpha}}+L}\right)^2\right)}{c^2\left(1+\Lambda(p)^{\frac{1}{\alpha}}\right)} + b, & p < \min\left\{c, w_-^{-1}\left(\frac{(\frac{cL}{1-c})^\alpha}{\lambda+(\frac{cL}{1-c})^\alpha}\right)\right\}, \\ b, \quad \min\left\{c, w_-^{-1}\left(\frac{(\frac{cL}{1-c})^\alpha}{\lambda+(\frac{cL}{1-c})^\alpha}\right),\right\} \leq p \leq \max\left\{c, w_+^{-1}\left(\frac{\lambda}{(\frac{(1-c)L}{c})^\alpha+\lambda}\right)\right\} \\ a\frac{(1-c)^2 - \left(\frac{\Lambda(1-p)^{\frac{1}{\alpha}}}{\Lambda(1-p)^{\frac{1}{\alpha}}+L}\right)^2 + \frac{\Lambda(1-p)^{\frac{1}{\alpha}}}{L}\left(c^2 - \left(\frac{L}{\Lambda(1-p)^{\frac{1}{\alpha}}+L}\right)^2\right)}{(1-c)^2\left(1+\Lambda(1-p)^{\frac{1}{\alpha}}\right)} + b, & p > \max\left\{c, w_+^{-1}\left(\frac{\lambda}{(\frac{(1-c)L}{c})^\alpha+\lambda}\right)\right\}. \end{cases}$$

Proof of Corollary 1: If $c = 1/2$, then we need L^* to satisfy $w_-^{-1}\left(\frac{(L^*)^\alpha}{\lambda+(L^*)^\alpha}\right) = w_+^{-1}\left(\frac{\lambda}{(L^*)^\alpha+\lambda}\right)$, or $w_-^{-1}\left(\frac{(L^*)^\alpha}{\lambda+(L^*)^\alpha}\right) = 1 - w_-^{-1}\left(\frac{(L^*)^\alpha}{(L^*)^\alpha+\lambda}\right)$. This means that $1/2 = w_-^{-1}\left(\frac{(L^*)^\alpha}{(L^*)^\alpha+\lambda}\right)$, or $(L^*)^\alpha w_-(1/2) + \lambda w_-(1/2) = (L^*)^\alpha$. Rearranging terms, $(L^*)^\alpha = \frac{\lambda w_-(1/2)}{1-w_-(1/2)}$, which yields the desired result.

Appendix B: Experimental Instructions and Interface

Welcome to this experiment on decision-making. Please read the following instructions carefully. As soon as everyone has finished reading the instructions you will receive a handout with a summary. During the experiment you will be asked to make a number of decisions. Your decisions will determine your earnings. (In this experiment, your earnings are not affected by the decisions of other participants). The experiment consists of 20 rounds. At the end of the experiment, one of the 20 rounds is selected at random. Your earnings for the experiment will equal the earnings that you made in this round. Your earnings will be privately paid to you in cash.

THE TASK

In this experiment, you will be asked to give probability judgments. In each round, you will be asked to give your probability judgment that a randomly drawn number will be in a particular range. The randomly drawn number will always be an integer number between 0 and 100, and each of the possible numbers between 0 and 100 will be equally likely. The range will differ across rounds. In the round that is selected for payment, your earnings will be determined as follows.
1. The computer draws a number between 0 and 100 (each number is equally likely).
2. Then it will be determined whether the number is in the range for the particular round or not.
3. You will receive a payoff that depends on your probability judgment for that round and on whether the number was in the range or not.

PAYOFF

On your table you find the list of payoffs that will result for each possible probability judgment, both for the case that the number is in the range and the case that it is outside of the range.

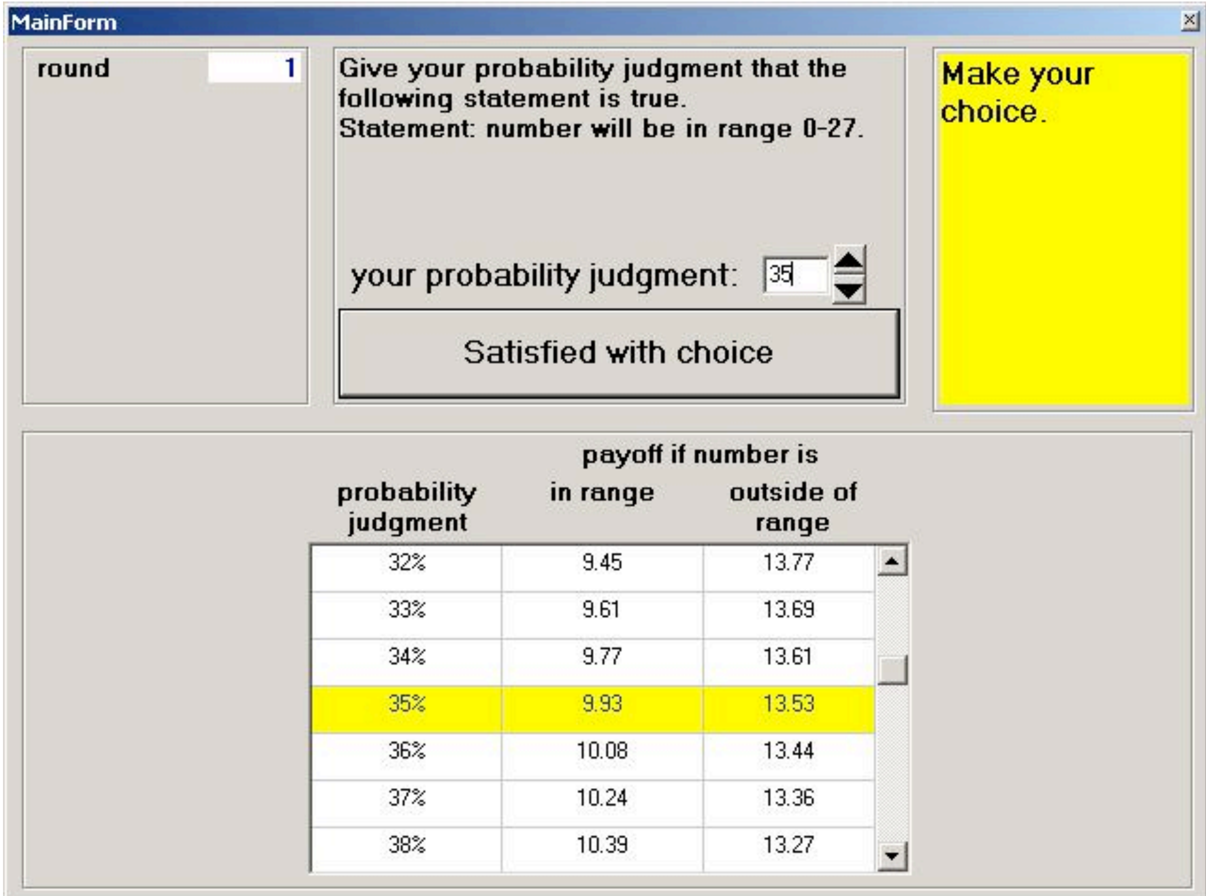
EXAMPLE

(The numbers in this example are arbitrarily chosen and do not indicate how you should make your choices.)
Assume that in the round that is selected for payment you were asked to give your probability judgment that the number will be in the range of 0-27. Say that you chose a probability judgment of 35%. If in that case the randomly drawn number was in the range you will receive a payoff of 9.93 euro and if it was outside of the range you will receive a payoff of 13.53 euro. The borders of the range belong to the range. For instance, if for the round of the example a number of 0 or 27 is drawn, then the number is in the range.

MAKING YOUR DECISIONS

The input of your probability judgment takes place in two phases: first you type in an integer number between 0 and 100, next you will be shown a menu in which your choice is replicated with the corresponding scores from the table. At that moment you can still alter your choice and choose any other integer number between 0 and 100.

You can do this by selecting the up or down arrow, or by clicking the mouse in the menu and scroll to another probability judgment. Next, when you click on “Satisfied with choice” your choice is final and you continue with the next round. Below you see an example of the decision-making screen that will be used in the experiment. At this moment, you cannot yet make a decision.



In this experiment there are no right or wrong answers; you choose what you want best. On the next screen you will be asked to answer some control questions. Please answer these questions now.

Please answer the following questions:

(1) Is the following statement correct? In this experiment, your earnings will be the sum of what you earn in all rounds. ○ yes ○ no

(2) Is the following statement correct? In the round that is selected for payment, the computer will draw an integer number between 0 and 100, and each of these numbers is equally likely. ○ yes ○ no

(3) THE FOLLOWING DECISIONS ARE IMAGINARY AND DO NOT INDICATE WHAT YOU SHOULD DO IN THE EXPERIMENT. Consider that a round is selected for actual payment in which the range of numbers equals [0-70]. You chose a probability judgment of 55%

(A) How much will you earn (in eurocents) if the number randomly drawn by the computer is 64? [ ]

(B) How much will you earn (in eurocents) if the number randomly drawn by the computer is 92? [ ]

(4) THE FOLLOWING DECISIONS ARE IMAGINARY AND DO NOT INDICATE WHAT YOU SHOULD DO IN THE EXPERIMENT. Consider that a round is selected for actual payment in which the range of numbers equals [0-70]. You chose a probability judgment of 82%

(A) How much will you earn (in eurocents) if the number randomly drawn by the computer is 15? [ ]

(B) How much will you earn (in eurocents) if the number randomly drawn by the computer is 77? [ ]

Payoff Table: Treatment NC

	Payoff (in euros) if number is	
Probability judgment	In range	Outside of range
0%	3.00	15.00
1%	3.24	15.00
2%	3.48	15.00
3%	3.71	14.99
4%	3.94	14.98
5%	4.17	14.97
6%	4.40	14.96
7%	4.62	14.94
8%	4.84	14.92
9%	5.06	14.90
10%	5.28	14.88
11%	5.49	14.85
12%	5.71	14.83
13%	5.92	14.80
14%	6.12	14.76
15%	6.33	14.73
16%	6.53	14.69
17%	6.73	14.65
18%	6.93	14.61
19%	7.13	14.57
20%	7.32	14.52
21%	7.51	14.47
22%	7.70	14.42
23%	7.89	14.37
24%	8.07	14.31
25%	8.25	14.25
26%	8.43	14.19
27%	8.61	14.13
28%	8.78	14.06
29%	8.95	13.99
30%	9.12	13.92
31%	9.29	13.85
32%	9.45	13.77
33%	9.61	13.69
34%	9.77	13.61
35%	9.93	13.53
36%	10.08	13.44
37%	10.24	13.36
38%	10.39	13.27
39%	10.53	13.17
40%	10.68	13.08
41%	10.82	12.98
42%	10.96	12.88
43%	11.10	12.78
44%	11.24	12.68
45%	11.37	12.57
46%	11.50	12.46
47%	11.63	12.35
48%	11.76	12.24
49%	11.88	12.12
50%	12.00	12.00

	Payoff (in euros) if number	
Probability judgment	In range	Outside of range
51%	12.12	11.88
52%	12.24	11.76
53%	12.35	11.63
54%	12.46	11.50
55%	12.57	11.37
56%	12.68	11.24
57%	12.78	11.10
58%	12.88	10.96
59%	12.98	10.82
60%	13.08	10.68
61%	13.17	10.53
62%	13.27	10.39
63%	13.36	10.24
64%	13.44	10.08
65%	13.53	9.93
66%	13.61	9.77
67%	13.69	9.61
68%	13.77	9.45
69%	13.85	9.29
70%	13.92	9.12
71%	13.99	8.95
72%	14.06	8.78
73%	14.13	8.61
74%	14.19	8.43
75%	14.25	8.25
76%	14.31	8.07
77%	14.37	7.89
78%	14.42	7.70
79%	14.47	7.51
80%	14.52	7.32
81%	14.57	7.13
82%	14.61	6.93
83%	14.65	6.73
84%	14.69	6.53
85%	14.73	6.33
86%	14.76	6.12
87%	14.80	5.92
88%	14.83	5.71
89%	14.85	5.49
90%	14.88	5.28
91%	14.90	5.06
92%	14.92	4.84
93%	14.94	4.62
94%	14.96	4.40
95%	14.97	4.17
96%	14.98	3.94
97%	14.99	3.71
98%	15.00	3.48
99%	15.00	3.24
100%	15.00	3.00

Payoff Table: Treatment MC

	Payoff (in euros) if number is	
Probability judgment	In range	Outside of range
0%	6.00	15.00
1%	6.16	15.00
2%	6.32	15.00
3%	6.47	14.99
4%	6.63	14.98
5%	6.78	14.97
6%	6.93	14.96
7%	7.08	14.94
8%	7.23	14.92
9%	7.38	14.90
10%	7.52	14.88
11%	7.66	14.85
12%	7.80	14.83
13%	7.94	14.80
14%	8.08	14.76
15%	8.22	14.73
16%	8.36	14.69
17%	8.49	14.65
18%	8.62	14.61
19%	8.75	14.57
20%	8.88	14.52
21%	9.01	14.47
22%	9.13	14.42
23%	9.26	14.37
24%	9.38	14.31
25%	9.50	14.25
26%	9.62	14.19
27%	9.74	14.13
28%	9.85	14.06
29%	9.97	13.99
30%	10.08	13.92
31%	10.19	13.85
32%	10.30	13.77
33%	10.41	13.69
34%	10.52	13.61
35%	10.62	13.53
36%	10.72	13.44
37%	10.82	13.36
38%	10.92	13.27
39%	11.02	13.17
40%	11.12	13.08
41%	11.22	12.98
42%	11.31	12.88
43%	11.40	12.78
44%	11.49	12.68
45%	11.58	12.57
46%	11.67	12.46
47%	11.75	12.35
48%	11.84	12.24
49%	11.92	12.12
50%	12.00	12.00

	Payoff (in euros) if number	
Probability judgment	In range	Outside of range
51%	12.12	11.92
52%	12.24	11.84
53%	12.35	11.75
54%	12.46	11.67
55%	12.57	11.58
56%	12.68	11.49
57%	12.78	11.40
58%	12.88	11.31
59%	12.98	11.22
60%	13.08	11.12
61%	13.17	11.02
62%	13.27	10.92
63%	13.36	10.82
64%	13.44	10.72
65%	13.53	10.62
66%	13.61	10.52
67%	13.69	10.41
68%	13.77	10.30
69%	13.85	10.19
70%	13.92	10.08
71%	13.99	9.97
72%	14.06	9.85
73%	14.13	9.74
74%	14.19	9.62
75%	14.25	9.50
76%	14.31	9.38
77%	14.37	9.26
78%	14.42	9.13
79%	14.47	9.01
80%	14.52	8.88
81%	14.57	8.75
82%	14.61	8.62
83%	14.65	8.49
84%	14.69	8.36
85%	14.73	8.22
86%	14.76	8.08
87%	14.80	7.94
88%	14.83	7.80
89%	14.85	7.66
90%	14.88	7.52
91%	14.90	7.38
92%	14.92	7.23
93%	14.94	7.08
94%	14.96	6.93
95%	14.97	6.78
96%	14.98	6.63
97%	14.99	6.47
98%	15.00	6.32
99%	15.00	6.16
100%	15.00	6.00

Payoff Table: Treatment LC

	Payoff (in euros) if number is	
Probability judgment	**In range**	**Outside of range**
0%	9.57	15.00
1%	9.63	15.00
2%	9.70	15.00
3%	9.76	14.99
4%	9.82	14.98
5%	9.88	14.97
6%	9.95	14.96
7%	10.01	14.94
8%	10.07	14.92
9%	10.13	14.90
10%	10.18	14.88
11%	10.24	14.85
12%	10.30	14.83
13%	10.36	14.80
14%	10.41	14.76
15%	10.47	14.73
16%	10.52	14.69
17%	10.58	14.65
18%	10.63	14.61
19%	10.68	14.57
20%	10.74	14.52
21%	10.79	14.47
22%	10.84	14.42
23%	10.89	14.37
24%	10.94	14.31
25%	10.99	14.25
26%	11.03	14.19
27%	11.08	14.13
28%	11.13	14.06
29%	11.18	13.99
30%	11.22	13.92
31%	11.27	13.85
32%	11.31	13.77
33%	11.35	13.69
34%	11.40	13.61
35%	11.44	13.53
36%	11.48	13.44
37%	11.52	13.36
38%	11.56	13.27
39%	11.60	13.17
40%	11.64	13.08
41%	11.68	12.98
42%	11.72	12.88
43%	11.76	12.78
44%	11.79	12.68
45%	11.83	12.57
46%	11.87	12.46
47%	11.90	12.35
48%	11.93	12.24
49%	11.97	12.12
50%	12.00	12.00

	Payoff (in euros) if number	
Probability judgment	**In range**	**Outside of range**
51%	12.12	11.97
52%	12.24	11.93
53%	12.35	11.90
54%	12.46	11.87
55%	12.57	11.83
56%	12.68	11.79
57%	12.78	11.76
58%	12.88	11.72
59%	12.98	11.68
60%	13.08	11.64
61%	13.17	11.60
62%	13.27	11.56
63%	13.36	11.52
64%	13.44	11.48
65%	13.53	11.44
66%	13.61	11.40
67%	13.69	11.35
68%	13.77	11.31
69%	13.85	11.27
70%	13.92	11.22
71%	13.99	11.18
72%	14.06	11.13
73%	14.13	11.08
74%	14.19	11.03
75%	14.25	10.99
76%	14.31	10.94
77%	14.37	10.89
78%	14.42	10.84
79%	14.47	10.79
80%	14.52	10.74
81%	14.57	10.68
82%	14.61	10.63
83%	14.65	10.58
84%	14.69	10.52
85%	14.73	10.47
86%	14.76	10.41
87%	14.80	10.36
88%	14.83	10.30
89%	14.85	10.24
90%	14.88	10.18
91%	14.90	10.13
92%	14.92	10.07
93%	14.94	10.01
94%	14.96	9.95
95%	14.97	9.88
96%	14.98	9.82
97%	14.99	9.76
98%	15.00	9.70
99%	15.00	9.63
100%	15.00	9.57