Eliciting Beliefs: Proper Scoring Rules, Incentives, Stakes and Hedging*

Olivier Armantier†    Nicolas Treich‡

August 30, 2012

Abstract

Proper Scoring Rules (PSR) are popular incentivized mechanisms to elicit an agent’s beliefs. This paper combines theory and experiment to characterize how PSR bias reported beliefs when i) the PSR payments are increased, ii) the agent has a financial stake in the event she is predicting, and iii) the agent can hedge her prediction by taking an additional action. In contrast with previous literature, the PSR biases are characterized for all PSR and all risk averse agents. Our results reveal complex distortions of reported beliefs, thereby raising concerns about the ability of PSR to recover truthful beliefs in general decision-making environments.

*We would especially like to thank Amadou Boly and Alexander Wagner for their work on this research project. We also thank the associate editor, three anonymous referees and Roberta Anderson, Monique Chyba, Edi Karni, Alexander Koch, Julie Lamisse, Peter Wakker, and Bob Winkler as well as seminar participants at the NHH of Bergen, the university of Canterbury at Christchurch, the CIRPEE-WZB workshop in Berlin and at the ESA conferences in Tucson (2009) and Melbourne (2010) for helpful discussions. All remaining errors are ours. The views expressed in this paper are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of New York or the Federal Reserve System. Nicolas Treich acknowledges the chair “Marché des risques et création de valeurs” Fondation du risque/SCOR and the ANR project “NanoInnov” for financial support.

†Federal Reserve Bank of New York, U.S. E-mail: olast1@gmail.com.

‡Toulouse School of Economics (LERNA, INRA), France.
1. Introduction

Introduced by statisticians in the 1950s, Proper Scoring Rules (PSR hereafter) are belief elicitation techniques designed to provide an agent the incentives to report her subjective beliefs in a thoughtful and truthful manner (Savage 1971). Although it is well known that the most common PSR are only incentive compatible under expected payoff maximization, the literature does not provide a systematic characterization of the biases — i.e. systematic differences between subjective and reported beliefs — produced by PSR. The object of this paper is to better understand the properties of PSR in general economic environments. More precisely, we characterize the possible PSR biases under three effects: i) a change in the PSR payments, ii) the introduction of a financial stake in the event predicted, and iii) the possibility for the agent to hedge her prediction by taking an additional action. In contrast with the existing literature, our results hold for all PSR and all risk averse agents. The empirical significance of the biases identified is then tested in a between-subject experiment.

Accurate measurements of probabilistic beliefs have become increasingly important both in practice and in academia. In practice, firms are increasingly turning to their employees to forecast (e.g.) sales, completion dates, or industry trends.\footnote{Such firms include Microsoft, Google, Chevron, General Electric, and General Motors.} Likewise, numerous websites now collect and report predictions (about e.g. sporting or political events) and opinions (about e.g. consumer products, movies, or restaurants).\footnote{Such opinion websites include ePinion, Ebay, Zagat, or Amazon. Prediction websites include the Iowa Electronic Market, the Hollywood Stock Exchange, or Intrade.} To be meaningful, the predictions and opinions reported must be informative. Precise belief assessments are also important in academia. In the literature on subjective expectations (Manski 2004), choice data are complemented with elicited beliefs to explain decisions...
related to e.g. health, education, labor or retirement. Likewise, experimental economists are increasingly eliciting their subjects’ beliefs to understand observed behavior better.

Because they are incentive compatible under expected payoff maximization, PSR have long been one of the most popular belief elicitation techniques, with applications to accounting, business, education, psychology, finance, and economics.3 Over the past decade, with the rapid development of prediction markets and opinion websites, there has been renewed interest in PSR. In particular, several mechanisms based on PSR have been recently proposed to promote honest feedback when collecting opinions, reviews or reputation assessments online.4 In addition, Market Scoring Rules have become a popular mechanism to overcome the liquidity problems that have affected early prediction markets.5 In short, Market Scoring Rules may be described as follows. A group of agents is sequentially asked to make a prediction about a particular event. Each agent is paid for her prediction according to a PSR, but she also agrees to pay the previous agent for his prediction according to the same PSR. Because of their attractive properties, Market Scoring Rules have been rapidly adopted by several firms operating prediction markets.6

It is well known however that PSR generate biases under risk aversion (Winkler and Murphy 1970), or under non-expected utility (Offerman et al. 2009). Up to this point, however, the exact nature of the biases induced by

---

4E.g. the “Peer Prediction” mechanism of Miller, Resnick, and Zeckhauser (2005), or the “Collective Revelation” mechanism of Goel, Reeves and Pennock (2009). See Chen and Pennock (2010) for a review of similar mechanisms.
6Companies proposing Market Scoring Rules include Inkling Markets, Consensus Point, Yoopick, Crowdcast, and Predictalot. Companies using Market Scoring Rules internally include Microsoft, General Motors, Boeing, Motorola, and the International Monetary Fund.
PSR has only been characterized for a few specific PSR, typically the quadratic scoring rule. Furthermore, PSR have usually been analyzed in simple decision-making environments in which the agent’s wealth varies only with the PSR payments. This paper contributes to the literature by conducting a more systematic characterization of the biases produced by all PSR in response to three effects relevant to economists.

First, we consider how varying the PSR payments affects reported probabilities. Although experimental economists have long debated how incentives affect choices (Camerer and Hogarth 1999), to the best of our knowledge, the problem has not been explicitly addressed for PSR.\footnote{Some have compared financially versus non-financially incentivized belief elicitation techniques (Rutström and Wilcox 2009). We extend this analysis by considering variations of strictly positive financial incentives.} We show theoretically that the impact of the PSR payments on reported probabilities is ambiguous. In particular, in contrast with a common belief that PSR with low payments induce near risk neutral responses, we show that smaller PSR payments can either reduce or reinforce the PSR biases depending on whether the utility function displays increasing or decreasing relative risk aversion.

Second, we consider an environment in which the agent has a financial stake in the event, a situation common in practice. For instance, an agent may be asked to make a prediction about an economic indicator (e.g. the stock market, the inflation rate) or an event (e.g. a flood, his company future sales). Similarly, an agent facing a Market Scoring Rule always has a stake in the event she predicts, as her payment to the previous predictor depends on the outcome of the event. Finally, subjects in public good experiments are often asked to predict the contributions of others.\footnote{See e.g. Croson (2000), Gächter and Renner (2010), or Fischbacher and Gächter (2010).} In all those cases, independently of the PSR payments, the agent has a stake, as her income depends on the outcome...
of the event she is predicting. We generalize Kadane and Winkler (1988) by showing how violations of the “no stake” condition affect the probabilities reported under any PSR. Furthermore, we provide experimental evidence to support this theoretical result.

Third, the agent has the possibility to hedge her prediction by taking an additional action whose payoff also depends on the event. For instance, in the previous examples, the agent predicting an economic indicator may also choose how to diversify her portfolio, while the agent predicting a catastrophic event may also decide on her insurance coverage. Likewise, subjects in public good experiments have to choose their own contributions.\(^9\) As we shall see, because they are not independent, the prediction and the additional action are in general different from what each decision would be if made separately. More specifically, we show theoretically how hedging can create complex distortions in the probabilities reported with a PSR. Most significantly, reported probabilities may not vary when subjective beliefs change. To the best of our knowledge, this is the first characterization of the PSR biases in a model with hedging opportunities. The empirical relevance of this hedging effect is then illustrated in our experiment.

The remainder of the paper is organized as follows. The theoretical results are derived in Section 2. The experimental treatments are described in Section 3. The results from the experiment are summarized in Section 4. Finally, we discuss in Section 5 some implications of our results for the elicitation of beliefs in general economic environments.

\(^9\)Although experimental economists have been aware of stakes and hedging opportunities (Palfrey and Wang 2009, Andersen et al. 2010, Blanco et al. 2010), these issues have often been ignored (Costa-Gomes and Weizsäcker 2008, Fischbacher and Gächter 2010). Concerns have also been raised that eliciting beliefs of subjects engaged in a game, even absent any stakes and hedging considerations, may lead to more strategic thinking and therefore affect behavior (Croson 2000, Rutström and Wilcox 2009, Gächter and Renner 2010).
2. Theory

We assume that an agent holds a unique subjective probability \( p \in [0,1] \) that an event occurs. We assume that \( p \) cannot be affected by the agent. We consider a standard expected utility framework. We assume that the agent’s von Neumann Morgenstern utility function over income, denoted \( u(.) \), is thrice differentiable, strictly increasing, and state-independent.

Let \( q \in [0,1] \) be the agent’s reported probability. A scoring rule gives the agent a monetary reward \( S_1(q) \) if the event occurs and a monetary reward \( S_0(q) \) if the event does not occur. We assume that the scoring rule is differentiable and real-valued, except possibly that \( S_1(0) = -\infty \) and \( S_0(1) = -\infty \). A scoring rule is said to be proper if and only if a risk neutral agent truthfully reveals her subjective probability.

**Definition. Proper Scoring Rule.** A scoring rule \( S = (S_1(q), S_0(q)) \) is proper if and only if:

\[
p = \arg \max_{q \in [0,1]} q S_1(q) + (1 - q) S_0(q)
\]

(2.1)

In the remainder of this section, we will illustrate some of our results using the popular quadratic scoring rule (QSR), defined by

\[
S_1(q) = 1 - (1 - q)^2
\]  \( (2.2) \)

\[
S_0(q) = 1 - q^2
\]

The QSR is represented in Figure 1a. It is straightforward to show that a QSR satisfies (2.1) and is therefore proper. We now provide a simple characterization of all PSR for binary random variables. This characterization has been
recently proposed in the statistics literature by Gneiting and Raftery (2007)
in the multi-event situation, building on the pioneering work of Savage (1971)
and Schervish (1989).

**Proposition 2.0.** *A scoring rule $S$ is proper if and only if there exists a
function $g(.)$ with $g''(q) > 0$ for all $q \in [0,1]$ such that

\[
S_1(q) = g(q) + (1 - q)g'(q)
\]
\[
S_0(q) = g(q) - qg'(q)
\]

The sufficiency of Proposition 2.0 is easy to prove.\(^{10}\) Indeed, under (2.3),
the agent’s expected payoff $\pi(q) = g(q) + (p - q)g'(q)$ reaches its unique max-
imum at $q = p$ since $p$ is the unique zero of $\pi'(.)$ and $\pi''(p) = -g''(p) < 0$.

Proposition 2.0 indicates that a PSR can be fully characterized by a single
function $g(.)$ and a simple property on this function, namely, its convexity.\(^{11}\)
Observe also that

\[
S'_1(q) = (1 - q)g''(q) > 0
\]
\[
S'_0(q) = -qg''(q) < 0
\]

\(^{10}\)The proof of the necessity of this Proposition, as well as the proofs of all the Propositions
derived in this section may be found in Appendix A.

\(^{11}\)In particular, it is easy to show that $g(q) = 0.5(q^2 + (1 - q)^2 + 1)$ yields the traditional
QSR in (2.2). Likewise, the traditional logarithmic scoring rule $S = (\log q, \log(1 - q))$ and
spherical scoring rule $S = (q^{\eta-1}, (1 - q)^{\eta-1})/(q^\eta + (1 - q)^\eta)^{(1-\eta)/\eta}$ with $\eta > 1$, are obtained
with $g(q) = q \log q + (1 - q) \log(1 - q)$ and $g(q) = (q^\eta + (1 - q)^\eta)^{1/\eta}$. Although the QSR
is the most common, other PSR are also used in practice. In particular, the logarithmic
scoring rule has been used in experimental economics (Ledyard, Hanson and Ishikida 2009,
Palfrey and Wang 2009, Healy et al. 2010) and in the field via the Logarithmic Market
Scoring Rule of Hanson (2003, 2007) and the Peer Prediction mechanism of Miller, Resnick
and Zeckhauser (2005).
Hence, when a scoring rule is proper, the convexity of \( g(.) \) implies the intuitive property that \( S_1(q) \) must be increasing, and that \( S_0(q) \) must be decreasing (see Figure 1a). This implies that \( S_1(q) \) and \( S_0(q) \) cross at most once. Note also that PSR are invariant to any transformation of the form \( aS + b \) where \( a \in \mathbb{R}^{++} \) and \( b \in \mathbb{R}^2 \). To simplify the presentation, we often consider in what follows a subset of all PSR, the “standard” PSR, satisfying

\[
g'(1/2) = 0 \tag{2.5}
\]

so that \( S_1(.) \) and \( S_0(.) \) cross at 1/2. Observe that this symmetry condition is satisfied for the most common PSR. In particular, the traditional quadratic, logarithmic and spherical scoring rules all verify (2.5).

### 2.1. Risk Aversion

From now on, we relax the assumption of risk neutrality and allow for risk aversion. Under expected utility, this is equivalent to assuming a concave utility function. Winkler and Murphy (1970), Kadane and Winkler (1988) and Offerman et al. (2009) have examined the probabilities reported by risk averse agents facing a QSR. They show that risk aversion leads agents to report probabilities skewed toward one half in the case of binary events. The intuition is that reporting more uniform probabilities makes a risk-averter better off since this reduces the difference across terminal payoffs. We first generalize this result to the class of all standard PSR satisfying (2.3) and (2.5), then to the class of all PSR in subsection 2.3.\(^\text{12}\)

\(^{12}\)Kothiyal, Spinu and Wakker (2011) examine theoretically optimal reported beliefs for all bounded binary scoring rules under risk aversion and under non-expected utility.
We define the response function $R(p)$ as follows

$$R(p) = \arg \max_{q \in [0,1]} pu(S_1(q)) + (1 - p)u(S_0(q))$$

(2.6)

where $S$ is a PSR as defined in (2.3) and where $u$ is concave. Our objective in the remainder of this section is to analyze the properties of this response function. In particular, we want to characterize the response function’s “bias”, $|R(p) - p|$.

Using (2.4), the first order condition of the program above can be written as follows

$$f(p, q) = p(1 - q)u'(S_1(q)) - (1 - p)qu'(S_0(q)) = 0.\quad (2.7)$$

It is easy to see that $\frac{\partial f(p, q)}{\partial q} < 0$, so that the program is concave and $R(p)$ is unique. It is also easy to check that it is optimal to report 0 when $p = 0$ and to report 1 when $p = 1$. Observe, moreover, that $\frac{\partial f(p, q)}{\partial p} > 0$, so that the response function $R(p)$ is strictly increasing. That is, for all PSR defined by (2.3), $R'(p) > 0$ together with $R(0) = 0$ and $R(1) = 1$.

Next, we show that truthful revelation of subjective probabilities is in general not optimal under risk aversion. Moreover, we show that the deviation from truth telling is systematic and depends on $p$. To do so, we examine the sign of $f(p, p)$ which captures the marginal benefit of increasing $q$ at $q = p$. It is easy to show that this sign depends on the sign of $S_0(p) - S_1(p)$. That is, under risk aversion the response function $R(p)$ is larger (lower) than $p$ when $S_1(p)$ is lower (larger) than $S_0(p)$. In particular, for a standard PSR defined by (2.3) and (2.5), we have $S_0(p) \geq S_1(p)$ if and only if $p \leq 1/2$. This implies that the agent reports more uniform probabilities in the following sense:
the response function is higher than \( p \) when \( p < 1/2 \) and lower than \( p \) when \( p > 1/2 \), as stated in the following Proposition.

**Proposition 2.1.** For all standard PSR defined by (2.3) and (2.5), and for all \( p \in (0, 1) \), \( R(p) \geq p \) if and only if \( p \leq 1/2 \).

The response function is therefore “regressive” (i.e., it crosses the diagonal from above), with a fixed point equal to one half. Figure 1b displays such a regressive response function for the QSR in (2.2) together with a quadratic utility function \( u(x) = -(2 - x)^2 \) with \( x \leq 2 \).

Note that Proposition 2.1 implies \( R(1/2) = 1/2 \), so that the agent truthfully reveals her subjective probability at \( p = 1/2 \). This result is due to the condition in (2.5). However, if one applies a positive affine transformation to one of the two PSR payoffs, \( S_1(1/2) \) and \( S_0(1/2) \) would differ, while the scoring rule remains proper. In that case, the result that risk aversion leads to reporting more uniform probabilities does not hold anymore. This effect is studied in more details in subsection 2.3.

Proposition 2.1 can easily be generalized to show that more risk averse agents (in the classical sense of Pratt 1964) always report more uniform probabilities (see Armantier and Treich 2011 for a formal proof). An increase in risk aversion therefore leads the response function to increase before the fixed point, and to decrease afterwards. The response function thus moves further away from \( p \), which can naturally be interpreted as an increase in the response function’s bias.

**2.2. Incentives**

The theoretic literature on PSR does not provide clear guidance on whether or not risk averse agents may be induced to provide more truthful responses by
selecting appropriate PSR payments. To address this issue, we analyze how changing the incentives provided by the PSR affects the response function. More precisely, we study the effect of changing $a > 0$ on the response function

$$R(p, a) = \arg \max_{q \in [0, 1]} pu(aS_1(q)) + (1 - p)u(aS_0(q))$$

(2.8)

We show that this effect depends on the relative risk aversion coefficient $\gamma(x) = \frac{-xu''(x)}{u'(x)}$.

**Proposition 2.2.** For all standard PSR defined by (2.3) and (2.5), for all $a > 0$ and all $p \in (0, 1)$, and under $\gamma'(x) \geq (\leq)0$, $\frac{\partial R(p, a)}{\partial a} \geq (\leq)0$ if and only if $p \leq 1/2$.

In other words, when the relative risk aversion with respect to income is increasing (decreasing), raising the PSR payments leads the agent to report more (less) uniform probabilities. One can present the intuition as follows. There are two effects when the PSR payments increase: i) a wealth effect, as the agent gets a higher reward for any given reported probability, and ii) a risk effect, as the difference between the rewards in the two states becomes more important. The sign of the derivative of the relative risk aversion $\gamma(x)$ ensures that one effect always dominates the other. In particular, when relative risk aversion is increasing, the risk effect dominates the wealth effect so that the agent reports more uniform probabilities to reduce the variability of her payoff.

An implication of this result is that it may not be possible to mitigate the bias of the response function by adjusting the incentives of the PSR. In particular, changing the PSR payments has no effect on the response function when the utility exhibits constant relative risk aversion (CRRA) with respect
Furthermore, a reduction of the PSR payments may in fact exacerbate the PSR bias when the utility function displays decreasing relative risk aversion (DRRA) with respect to income. Hence, the common belief that paying agents smaller amounts necessarily induces more truthful reports is misleading in general.  

The result of Proposition 2.2 is illustrated in Figure 1c. The added response function compared to Figure 1b is calculated for $a = 2$. Observe that both response functions are regressive with a fixed point at 1/2. Yet, the increased incentives lead to reporting more uniform probabilities. This is because the quadratic utility function used for the numerical example displays increasing relative risk aversion.

2.3. Stakes

Kadane and Winkler (1988) define the presence of a stake as a situation in which, absent any PSR payments, the agent’s final wealth varies depending on whether or not the event occurs. Using a QSR, they show that, unless the agent is risk neutral, the presence of a stake leads to biased reported beliefs.

To illustrate, consider a CRRA utility function $u(x) = (1 - \gamma)^{-1}x^{1-\gamma}$ with $\gamma > 0$ and a spherical scoring rule $S = (q, (1 - q))(q^2 + (1 - q)^2)^{-1/2}$. This combination yields a closed-form solution $R(p, a) = p^{1/(1+\gamma)}(p^{1/(1+\gamma)} + (1 - p)^{1/(1+\gamma)})^{-1}$, which is indeed independent of $a$. This is not incompatible with a well-known result in the literature showing that the agent reveals her beliefs truthfully when $a$ tends toward 0, i.e. $\lim_{a \to 0} R(p, a) = p$ (Kadane and Winkler 1988, Jaffray and Karni 1999, Karni 1999). Indeed, these authors consider a utility function of the form $u(x) = U(w + x)$, where $w$ is the agent’s initial wealth, and with $U''(w) < \infty$. This implies that $\gamma'(0) = -U''(w)/U'(w)$ is strictly positive under risk aversion. Therefore, the agent necessarily displays increasing relative risk aversion with respect to income as $x$ tends toward 0. However, the utility function may still be DRRA locally for some $x > 0$. An example of such a utility function is $u(x) = -\exp(1/(w + x))$. Consistent with Proposition 2.2, a reduction of the income through $a$ could then initially move the response function away from $p$ for this utility function, but, once $a$ gets sufficiently close to 0, $R(p, a)$ would start converging toward $p$. In other words, although it is correct that $\lim_{a \to 0} R(p, a) = p$ when $U'(w)$ is finite, a reduction in $a$ does not guarantee more truthful responses.
Below, we generalize this result to the class of all PSR, and we characterize the direction of the PSR biases when the agent has a stake in the event predicted.

Consistent with the definition of Kadane and Winkler (1988), we introduce a stake by assuming that income increases by an exogenous amount $\Delta \in \mathbb{R}$ when the event occurs. Observe, however, that this is formally equivalent to adding a constant $\Delta$ to $S_1(q)$ in (2.3). As a result, it is immediate that any PSR remains proper in the presence of a stake. Note also that if $(S_0(q), S_1(q))$ is a standard PSR, then the PSR defined by $(S_0(q), S_1(q) + \Delta)$ with $\Delta \neq 0$ does not satisfy condition (2.5). Consequently, the results in this section generalize the analysis under risk aversion to the class of all PSR.

The response function is defined by

$$R(p, \Delta) = \arg \max_{q \in [0,1]} pu(\Delta + S_1(q)) + (1 - p)u(S_0(q))$$

in which the added reward $\Delta$ is a finite (positive or negative) “stake”. The first order condition can be written

$$h(\Delta, q) \equiv p(1 - q)u'(\Delta + S_1(q)) - (1 - p)qu'(S_0(q)) = 0$$

As before, $\frac{\partial h(\Delta, q)}{\partial q} < 0$ under risk aversion, so that the program is concave. Also observe that $\frac{\partial h(\Delta, q)}{\partial \Delta} < 0$ so that the response function is decreasing in $\Delta$ under risk aversion. That is, for all PSR defined by (2.3), we have $\frac{\partial R(p, \Delta)}{\partial \Delta} \leq 0$.

The intuition for this result is straightforward. Under risk aversion, an increase in $\Delta$ reduces the marginal utility when the event occurs. Therefore, to compensate for the difference in marginal utility across states, the agent wants to increase the reward of the PSR when the event does not occur. This can be done by reducing the reported probability that the event occurs. We
can now state a Proposition that characterizes the response function when the agent has a stake.

**Proposition 2.3.** For all PSR defined by (2.3) and for all \( p \in (0,1) \), the response function \( R(p, \Delta) \) is characterized as follows:

i) if there exists a \( \hat{p} \) such that \( \Delta + S_1(\hat{p}) = S_0(\hat{p}) \), then \( R(p, \Delta) \geq p \) if and only if \( p \leq \hat{p} \),

ii) if \( \Delta + S_1(p) \geq (\leq)S_0(p) \) for all \( p \), then \( R(p, \Delta) \leq (\geq)p \)

Proposition 2.3 is illustrated in Figure 1d. Compared to Figure 1b, there are two additional response functions in Figure 1d, one for \( \Delta = 1/2 \) and the other for \( \Delta = 1 \). Both response functions are regressive, but the first has a fixed point at 1/4 and the second is below the diagonal everywhere. Note also that the presence of a stake does not necessarily increase the PSR bias. In particular, for any \( p \leq 1/4 \), \( |R(p) - p| \) is smaller when \( \Delta = 1/2 \) than when \( \Delta = 0 \).

### 2.4. Hedging

We now assume that the agent, in addition to her prediction, can make another decision whose payoff depends on the outcome of the event. This hedging opportunity can create a stake in the event, and it may therefore be interpreted as an endogenous stake. Below, we provide what we believe to be the first theoretical analysis of the biases produced by PSR in a model in which the agent has an hedging opportunity. We show that the presence of hedging can produce complex, yet predictable, biases.

Suppose the agent receives an endowment \( \bar{\omega} \) and can invest an amount \( \alpha \) in \([0, \bar{\omega}]\) in a risky asset. Per unit of money invested, this asset returns \( k + 1 \) if the event occurs, and 0 (i.e., the investment is lost) if the event does not occur. We assume that \( k \) and \( \bar{\omega} \) are strictly positive and finite. The problem
becomes

\[
\max_{q \in [0,1], \alpha \in [0,\bar{\alpha}]} pu(S_1(q) + k\alpha + \bar{\alpha}) + (1 - p)u(S_0(q) - \alpha + \bar{\alpha})
\]  

(2.11)

The Proposition below presents properties of the response function when this particular form of hedging is available. It shows that there are two critical threshold values,

\[
p(k) = \frac{u'(S_0(\frac{1}{1+k}) + \bar{\alpha})}{ku'(S_1(\frac{1}{1+k}) + \bar{\alpha}) + u'(S_0(\frac{1}{1+k}) + \bar{\alpha})}
\]

and

\[
\bar{p}(k) = \frac{u'(S_0(\frac{1}{1+k}))}{ku'(S_1(\frac{1}{1+k}) + (k + 1)\bar{\alpha}) + u'(S_0(\frac{1}{1+k}))}
\]

that shape the optimal investment rule in the sense that the agent does not invest at all for low enough probabilities \( p \leq p(k) \), and invests the maximum amount \( \bar{\alpha} \) for high enough probabilities \( p \geq \bar{p}(k) \).

**Proposition 2.4.** For all standard PSR defined by (2.3) and (2.5), the solutions \( R(p) \) and \( \alpha(p) \) to program (2.11) satisfy the following properties:

i) for \( p \leq p(k) \), we have \( \alpha(p) = 0 \) and \( R(p) \in [0, \frac{1}{1+k}] \) with \( R'(p) > 0 \),

ii) for \( p \) in \([p(k), \bar{p}(k)]\), we have \( \alpha(p) \in [0, \bar{\alpha}] \) with \( \alpha'(p) > 0 \) and \( R(p) = \frac{1}{1+k} \),

iii) for \( p \geq \bar{p}(k) \), we have \( \alpha(p) = \bar{\alpha} \) and \( R(p) \in [\frac{1}{1+k}, 1] \) with \( R'(p) > 0 \).

According with intuition, the investment opportunity is not attractive when \( p \) is low and becomes more attractive as \( p \) increases. Note that when \( p \) is high enough so that the agent invests the maximum amount \( \bar{\alpha} \), then the investment opportunity has an effect similar to a stake equal to \( (k + 1)\bar{\alpha} \). In the intermediate range where \( p \) belongs to \([p(k), \bar{p}(k)]\), the agent invests some strictly positive amount in \((0, \bar{\alpha})\). Perhaps most interestingly, the agent reports
probabilities that are constant in this interval, and are therefore independent of $p$. The intuition is that the PSR is used as a transfer scheme across states, while the investment opportunity is used to adjust risk exposure to changes in $p$. Note finally, that when $k > 1$, we have $p(k) < 1/(1 + k)$, so that it can be optimal to invest in the risky asset even when its expected return is negative. This shows that the presence of the PSR can also alter the investment decision. We plot in Figures 1e and 1f the optimal investment share $\alpha(p)/\bar{\alpha}$ and the response function under $k = 1$ and $\bar{\alpha} = 0.5$.

3. The Experiment

The subjects are presented with three different series of 10 events. In this paper, we focus on the first series of 10 events. Each event describes the possible outcome resulting from the roll of two 10-sided dice (one black, the other red). The red die determines the first digit and the black die determines the second digit of a number between 1 and 100. For instance, we described the event with an objective probability of 25% as “the number drawn is between 1 (included) and 25 (included)”. The first 10 events have objective probabilities 3%, 5%, 15%, 25%, 35%, 45%, 61%, 70%, 80%, and 90%. A complete description of the events and the order in which they were presented to subjects may be found in Appendix B.

For each of the 30 events, a subject is asked to make a choice consisting of selecting 1 out of 149 possible options called “choice numbers”. To each choice number corresponds two payments generated with a QSR. As further explained below, the first is the payment to the subject when the event occurs, while the second is the payment to the subject when the event does not occur. A sub-

\cite{Armantier:2011} for an analysis of the other two series of 10 events.

Although PSR are commonly implemented by directly asking subjects for a prediction
ject’s set of possible choice numbers, as well as their corresponding payments, was presented in the form of a “Choice Table” (see Appendix B). Observe that the Choice Table is ordered such that, as the choice number increases, the payment when the event occurs increases, while the payment when the event does not occur decreases. Note also that the choice number 75 guarantees the same payment to a subject regardless of the roll of the dice.

3.1. The Control Treatment

The subjects’ payments in the control treatment ($T_0$) are generated with a QSR of the form $S_1(q) = a \cdot [1 - (1 - q)^2]$ and $S_0(q) = a \cdot [1 - q^2]$, where $a = 4,000$ FCFA in our experiment. Each entry in the choice table, and in particular the link between choices and payments, was explained in details and illustrated through several examples (see Appendix B). After reading the instructions, the subjects’ understanding of the table was submitted to a test, which was then solved by the experimenter. The subjects were then presented with the list of 30 events. No time limit was imposed, and the subjects could modify any of their previous choices at any time.

expressed as a probability or a frequency, we were concerned that such framed instructions may influence behavior. Instead, consistent with recent PSR analyses (Offerman et al. 2009, Andersen et al. 2010, Hollard, Massoni and Vergnaud 2010, Hao and Houser 2012), we decided to present the PSR as a choice between lotteries. This feature is often considered a positive attribute of PSR. For instance, Camerer (1995 p 592) states: “Besides being incentive compatible, scoring rules enable judgement of probability to be elicited without mentioning the word probability or defining it.”

How best to present PSR to subjects remains an open question. Tables, although not ideal, have been often adopted in part because they are simple to implement (see, e.g., McKelvey and Page 1990, Sonnemans and Offerman 2004, Rutström and Wilcox 2009, Blanco et al. 2010). A possible issue with this approach, however, is that subjects may be attracted to salient choices (e.g. in the middle of the table). If such an effect exists, then it could affect the interpretation of the results in specific treatments, but it would not explain differences across treatments since they are conducted with the same table.

The Franc CFA is the currency in Burkina Faso where the experiment was conducted (see Section 3.6 for details). The conversion rate at the time was roughly $1$ for 455 FCFA.
Once all subjects had completed their task, the experimenter randomly selected one of the 30 events and rolled the two dice once to determine whether this event occurred or not. All subjects in a session were then immediately paid according to their choice number for the event randomly drawn. For instance, if a subject selects the choice number 30 for the event randomly selected, she receives either 1,440 FCFA if the event obtains or 3,840 FCFA if the event does not obtain (see Appendix B). This amount constitutes the entirety of a subject’s payments, as no show-up fee was provided in the control treatment.

Based on the theoretical analysis conducted in the previous section, we can frame the experimental hypotheses for each treatment in terms of the properties of the response function $R(p)$. In particular, assuming subjects in our experiment are risk averse, we can use Proposition 2.1 to formulate our first hypothesis.

$H_0$: The response function in $T_0$ is i) regressive (i.e. it crosses the diagonal from above) and ii) has a fixed point at $1/2$.

3.2. The “High Incentives” and “Hypothetical Incentives” Treatments

Two treatments were conducted to study the effect of incentives. As indicated in Table 1, where the differences between treatments are summarized, the “High Incentives” treatment ($T_1$) is identical to the control treatment except that every payment in the choice table is now multiplied by 10. The “Hypothetical Incentives” treatment ($T_2$) is identical to the “High Incentives” treatment except that payments are now hypothetical. More specifically, subjects in $T_2$ were asked to make their choices as if they would be paid the amounts in the choice table. Yet, they knew they would receive only a flat fee
of 3,000 FCFA for completing the task, regardless of their choices.

Proposition 2.2 shows that higher incentives affect the response function only when relative risk aversion is non-constant. To derive hypothesis $H_1$, we chose to follow most of the experimental economics literature (including recent papers on scoring rules such as Offerman et al. (2009) or Andersen et al. 2010) and assume constant relative risk aversion with respect to income.

$H_1$: The response function in $T_1$ is identical to the response function in $T_0$.

Since subjects’ choices are not incentivized in $T_2$, no theoretical prediction can be derived. However, Holt and Laury (2002) suggest that treatments with high hypothetical payments and treatments with low real payoffs yield similar results. This leads to our next hypothesis.

$H_2$: The response function in $T_2$ is identical to the response function in $T_0$.

### 3.3. The “Low Stake” and “High Stake” Treatments

Two treatments were conducted to study the effect of stakes. They are identical to $T_0$ except that subjects receive a bonus when the event occurs. The bonus is 2,000 FCFA in the “Low Stake” treatment ($T_3$) and 8,000 FCFA in the “High Stake” treatment ($T_4$).

We have demonstrated in the previous section that adding a positive stake when the event occurs lowers the response function of a risk averse agent. Therefore, the response functions in $T_3$ and $T_4$ are predicted to be less elevated everywhere than the response function in $T_0$. Moreover, Proposition 2.3 identifies two special cases, one in which the response function is regressive with an interior fixed point and one in which the response function is below

---

19 Read (2005) argues that the comparison between real and hypothetical treatments in Holt and Laury (2002) may have problems. Note also that Noussair, Trautman and van de Kuilen (2011) find no differences between real and hypothetical conditions.
the diagonal for all \( p \) in \((0,1)\) (and therefore has no interior fixed point). It is easy to show, given the size of the stakes and the specific QSR we used, that \( T_3 \) corresponds to the first case with a fixed point at \( 1/4 \), while \( T_4 \) corresponds to the second case. This leads to the hypotheses:

\( H_3: \) The response function in \( T_3 \) i) is lower than in \( T_0 \) and ii) has a fixed point at \( 1/4 \).

\( H_4: \) The response function in \( T_4 \) is i) lower than in \( T_3 \) and ii) lower than \( p \).

### 3.4. The “Low Hedging” and “High Hedging” Treatments

Two treatments were conducted to study the effect of hedging. They are identical to \( T_0 \) except that subjects are asked to make an additional decision for each event. Namely, subjects were given 2,000 FCFA and offered the opportunity to bet a share of this endowment. If the event does not occur, the bet is lost. If the event occurs, the bet multiplied by 2 (respectively, 4) is paid to the subjects in the “Low Hedging” treatment \( T_5 \) (respectively, the “High Hedging” treatment \( T_6 \)). Finally, in both states of the world, the subjects retain the part of the 2,000 FCFA they did not bet. In other words, in addition to selecting a choice number, subjects in the hedging treatments have to make a simple portfolio decision between a riskless and a risky asset.

Proposition 2.4 shows that, with the possibility for hedging, the response function of a risk averse agent is regressive, yet constant when the share of the endowment invested is in \((0,1)\). Proposition 2.4 also shows that a subject invests all her endowment when \( p \) is high enough. In this case, the hedging opportunity operates as a stake of 4,000 FCFA (respectively, 8,000 FCFA) in treatment \( T_5 \) (respectively, \( T_6 \)), and the response functions are reduced accordingly compared to \( T_0 \). Observe finally that \( T_5 \) corresponds to the case
$k = 1$ and $T_6$ to the case $k = 3$ in section 2.4. It is then easy to show that the fixed points in $T_5$ and $T_6$ are respectively $1/2$ and $1/4$. This leads to the following hypotheses:

$H_5$: The response function in $T_5$ is i) regressive with a fixed point at $1/2$, ii) equal to $1/2$ when the share invested is in $(0,1)$, and iii) lower than in $T_0$ when $p$ is close to $1$.

$H_6$: The response function in $T_6$ is i) regressive with a fixed point at $1/4$, ii) equal to $1/4$ when the share invested is in $(0,1)$, and iii) lower than in $T_5$ when $p$ is close to $1$.

### 3.5. Implementation of the Experiment

The experiment took place in Ouagadougou, the capital of Burkina Faso, in June 2009. The choice of location was motivated by two factors. First, we wanted to take advantage of a favorable exchange rate i) to create salient financial differences between treatments (e.g. between the control and the “Hypothetical Incentives” treatments), and ii) to provide subjects with substantial incentives so that risk aversion had a fair chance to play a role.\(^{20}\) Second, one of the authors had conducted several experiments in Ouagadougou over the past three years (see e.g. Armantier and Boly 2011, 2012).

Building on our experience, we asked a local recruiting firm (Opty-RH) to place fliers around Ouagadougou. To be eligible, subjects had to be at least 18 years old and be current or former university students. After their credentials were validated, subjects were randomly assigned to a treatment and a session. The sessions were conducted in a high school we rented for the occasion. Upon arrival, the subjects were gathered in a large room. Instructions were read

\(^{20}\)The maximum payment of 40,000 FCFA in the “High Incentives” treatment slightly exceeds the monthly average entry salary for a university graduate.
aloud, followed by questions. Each subject was then assigned to a workstation where he could make his choices privately. Two sessions were conducted for each treatment, with each session taking on average 90 minutes to complete.

As indicated in Table C1 provided in Appendix C, a total of 301 subjects participated in the experiment, with a minimum (maximum) of 41 (48) subjects per treatment. The subjects were composed mostly of men (74%) and students currently enrolled at the university (68%), ranging in age between 19 and 38 (with a median age of 25). In a post-experiment survey, slightly more than half the subjects reported having taken a probability or a statistics class at the university. Finally, most of the subjects (86%) reported not having participated in a similar economic or psychology experiment. Excluding the “Hypothetical Incentives” treatment (where earnings were fixed at 3,000 FCFA), the average earnings of a subject were 8,861 FCFA. As indicated in Table C1, however, earnings varied greatly across subjects and treatments (the smallest amount paid was 100 FCFA and the maximum was 40,000 FCFA).

4. Experimental Results

4.1. The Control Treatment ($T_0$)

Figure 2.1 shows the subjects’ average responses to the first 10 events in the control treatment. According with hypothesis $H_0$, the response function in the control treatment is regressive. Moreover, it exhibits the traditional inverse S-shape with a fixed point around 1/2. This observation is consistent with Offerman et al. (2009) who identified similar shapes in a comparable experiment in which beliefs about objective probabilities are elicited with a QSR.

To test our hypotheses more formally, we compare statistically the subjects’
choices across treatments with the following reduced form model:

\[ \hat{P}_{it} = \varphi_i(P_t) + u_{it} \]  

(4.1)

where \( \hat{P}_{it} \) is the probability corresponding to the choice number \( N_{it} \) selected by subject \( i \) for event \( t = 1, ..., 10 \) (i.e. \( \hat{P}_{it} = 2/3 \cdot N_{it} \)), \( P_t \) is the objective probability of occurrence of event \( t \), \( u_{it} \) follows a normal distribution truncated such that \( \hat{P}_{it} \in [0, 1] \), and \( \varphi_i(.) \) is a continuous function satisfying \( \varphi_i(0) = 0 \), \( \varphi_i(1) = 1 \) and \( \varphi_i'(.) > 0 \).

Consistent with previous literature, we consider a function that can exhibit an inverse S-shape:

\[ \varphi_i(P_t) = \exp \left( \left[ \ln P_t \right]^{b_i} \cdot \left[ \ln(a_i) \right]^{1-b_i} \right) \]  

(4.2)

where \( a_i \in (0, 1] \) and \( b_i > 0 \) are parameters that may vary across subjects. Observe that under the reparametrization \( \left\{ b_i = \alpha_i; a_i = \exp \left( -\beta_i^{1-\alpha_i} \right) \right\} \), (4.2) is in fact the probability weighting function \( w_i(P_t) = \exp \left( -\beta_i \left[ - \ln P_t \right]^{\alpha_i} \right) \) proposed in a different context by Prelec (1998). The specification in (4.2) was preferred to Prelec’s because the parameters are easier to interpret with our experimental data. Indeed, observe that \( \varphi_i(a_i) = a_i \) and \( \varphi_i'(a_i) = b_i \). In other words, \( a_i \) captures where the function \( \varphi_i \) crosses the diagonal, while \( b_i \) captures the slope of \( \varphi_i \) at this fixed point.

To control for possible treatment effects we define:

\[
\begin{align*}
a &= a_0 + a_1 \cdot T_0 \\
b &= b_0 + b_1 \cdot T_0
\end{align*}
\]

where \( T_0 \) is a dummy variable equal to 1 when the observation was collected in the control treatment. To complete the econometric specification we assume
that $a_i = a$, so that unobserved heterogeneity is entirely captured through the random parameter $b_i = b \cdot \eta_i$ where $\ln(\eta_i)$ is a zero-mean normally distributed random effect.\textsuperscript{21}

The parameters, estimated by Maximum Simulated Likelihood, are reported in Table 2.\textsuperscript{22} Observe that in the control treatment $a_0$ is not significantly different from $1/2$, while $b_0$ is significantly smaller than 1. In other words, we find statistical evidence suggesting that the response function exhibits an inverse S-shape and crosses the diagonal near $1/2$. This result is therefore consistent with hypothesis $H_0$.

\textbf{4.2. The Incentives Treatments ($T_1$ and $T_2$)}

Figure 2.1 indicates that the response function in the “High Incentives” treatment is flatter than in the control treatment, although it still cuts the diagonal around $1/2$. This observation finds statistical support in Table 2. Indeed, $a_1$ is not significantly different from 0 in $T_1$, thereby suggesting that the fixed points of the response functions cannot be distinguished statistically across the two treatments. In contrast, we find the parameter $b_1$ to be positive and significant in $T_1$ which suggest that, compared to the control treatment, the

\textsuperscript{21}This specification of the parameters is somewhat consistent with heterogeneity in risk aversion. Indeed, as explained at the end of section 2.1, an increase in risk aversion leads the response function to increase before the fixed point and to decrease afterwards. Consistent with theory, our reduced form model can capture the fact that the fixed point of the response function is not affected by heterogeneity in risk attitude, while the slope at the fixed point is closer to zero for a more risk averse subject.

\textsuperscript{22}The robustness of our results to the parametric assumptions finds support in Appendix C where we report the results obtained under two alternative specifications. In Table C2, unobserved heterogeneity is entirely captured through the random parameter $a_i$. Following Wilcox (2010), the response function in Table C3 is specified as the cumulative function of a Beta distribution. Observe that unlike (4.2), this second specification of the response function satisfies the symmetry condition $R(p) = R(1-p)$, which must hold theoretically when a QSR (2.2) is used to elicit beliefs (Offerman et al. 2009).
response function is more biased in the “High Incentives” treatment. Observe that in addition to being statistically significant, the increase in the PSR bias can be substantial. In particular, the average probability reported for the event with objective probability 3% is nearly three times higher in $T_1$ (32.1%) than in $T_0$ (11.5%).

These results do not comply with hypothesis $H_1$ derived under the assumption of constant relative risk aversion. Instead, as explained in Section 2.4, choices in $T_1$ are consistent with subjects exhibiting increasing relative risk aversion. In a similar belief elicitation experiment, Andersen et al. (2010) also conclude that their subjects’ behavior may be best described under increasing relative risk aversion. It is also interesting to note that our results imply that paying more does not necessarily yield “better” answers. Instead, we find that, because of our subjects’ specific form of relative risk aversion, using a PSR that provides higher incentives generates more biases.

As for the “Hypothetical Incentives” treatment $T_2$, Figure 2.1 reveals that subjects’ average responses, although still exhibiting the inverse S-shape, are slightly closer to the diagonal than in the control treatment. This observation is confirmed statistically in Table 2 as $b_1$ is found to be negative and significant in $T_2$. Note also that $\sigma_u$, the standard deviation of the error term $u_{it}$ in (4.1), is significantly larger for $T_2$ than for $T_0$. These results therefore do not support hypothesis $H_2$. Instead, we find that the beliefs elicited with a QSR differ when subjects face real or hypothetical incentives.

Our conclusions are only partially consistent with the literature. Like Gächter and Renner (2010), we find that financial incentives reduce the noise.

\footnote{Similar conclusions are reached nonparametrically by using Mann-Whitney tests for each objective probability. Indeed, Table C4 reported in Appendix C shows that, except for some objective probabilities around $1/2$, the distributions of responses are closer to the diagonal in $T_0$ than in $T_1$.}
in the beliefs elicited. In contrast with our experiment, however, Sonnemans and Offerman (2004) and Trautmann and van de Kuilen (2011) find that predictors perform with equal accuracy when rewarded with a QSR or a flat fee.

4.3. The Treatments with Stakes \((T_3 \text{ and } T_4)\)

For the smallest objective probabilities, no obvious difference is visible in Figure 2.2 between the response functions in the control treatment and those in the low and high stakes treatments. In contrast, Figure 2.2 shows that, compared to \(T_0\), responses for the highest objective probabilities are lower in \(T_3\) and lowest in \(T_4\). The estimation results in Table 2 confirm these observations. Indeed, \(a_1\) and \(b_1\) are positive and significant for both treatments \(T_3\) and \(T_4\) but significantly larger for treatment \(T_4\). This implies that, compared to \(T_0\), the response function becomes lower and flatter in the “Low Stake” treatment and that the magnitude of this effect is stronger in the “High Stake” treatment. In other words, part i) of predictions \(H_3\) and \(H_4\) (i.e. the response function is less elevated in \(T_3\) than in \(T_0\), and in \(T_4\) than in \(T_3\)) is verified. Behavior in the experiment, however, is not fully consistent with our predictions. Indeed, observe in Table 2 that the parameter \(a_0\) is significantly different from 1/4 in \(T_3\), and from 0 in \(T_4\), thereby contradicting part ii) of \(H_3\) and \(H_4\).\(^{24}\)

To sum up, we find that when they have a stake in the event, subjects in our experiment tend to smooth their payoffs across the two states, especially when the event is likely to occur. The direction of this treatment effect is only partially in agreement with the theory: the direction is correct, but the magnitude is insufficient.

\(^{24}\)The non parametric tests in Table C4 of Appendix C confirm that for most objective probabilities above 25% the samples of responses are stochastically higher in \(T_0\) than in \(T_3\) and higher in \(T_3\) than in \(T_4\).
4.4. The Treatments with Hedging ($T_5$ and $T_6$)

The response functions for the two hedging treatments plotted in Figure 2.3 reveal several differences with the control treatment. First, for the highest objective probabilities, the response function becomes substantially lower in $T_5$ and lowest in $T_6$. In particular, compared to $T_0$, the average probability reported for the event with objective probability 90% is 14% lower in $T_5$ and 30% lower in $T_6$. Second, although the fixed point of the response function is also around 1/2 in $T_5$, it is slightly above 40% in $T_6$. Third, the response functions appear flatter (but not perfectly flat) around the diagonal in both hedging treatments.

Most of these observations are confirmed statistically in Table 2. In particular, observe that the estimate of $a_1$ is insignificant in $T_5$, while it is positive and significant in $T_6$. The former is consistent with part i) of hypothesis $H_5$, as we cannot exclude that, as in the control treatment, the response functions in the “Low Hedging” treatment cut the diagonal at 1/2. In the “High Hedging” treatment, however, the parameter $a_0$ is found to be significantly greater than 1/4, which contradicts part i) of hypothesis $H_6$. Observe also that $b_1$ is significant and positive in both $T_5$ and $T_6$ thereby indicating flatter responses around the diagonal in the two hedging treatments compared to $T_0$.\(^{25}\) This result therefore supports part iii) of hypotheses $H_5$ and $H_6$.

Turning now to the subjects’ betting behavior in the two hedging treatments, we can see in Figure 2.4 that subjects in the “High Hedging” treatment invest more in the risky asset for any objective probability than in the “Low Hedging” treatment.\(^{26}\) The subjects’ betting behavior, however, is not fully

\(^{25}\)The nonparametric tests in Table C4 (Appendix C) confirm that compared to $T_0$ responses for most objective probabilities above 60% are significantly lower in $T_5$ and lowest in $T_6$.

\(^{26}\)This observation is confirmed statistically by the nonparametric tests reported in the
consistent with the theory. In particular, we can see in Figure 2.4 that, on average, subjects invest strictly positive amounts even for low probabilities, while they do not invest all of their endowments even for high probabilities.

To summarize, although not fully consistent with the theory, subjects in our experiment take advantage of their hedging opportunities. In particular, we find that subjects tend to bet high on the most likely events, while simultaneously making lower predictions than in the control treatment. In other words, it seems that subjects are willing to take some risk on the bet, while using the scoring rule as an insurance in case the event does not occur. Our results are therefore consistent with Blanco et al. (2010) who find that transparent and strong hedging opportunities affect behavior and belief elicitation.

5. Conclusion

Proper Scoring Rules (PSR) have been one of the most popular incentivized mechanisms to elicit an agent’s beliefs. We have shown theoretically how incentives, stakes, and hedging opportunities may lead to substantial distortions in the probabilities reported with a PSR. As indicated in Table 3 where our experimental results are summarized, most of these effects are found to be empirically relevant.

We believe our results have implications for the elicitation of subjective probabilities with PSR in general field settings. As argued in the introduction, eliciting beliefs in the field typically involves some form of a stake or the possibility to hedge one’s prediction. In particular, agents who participate in Prediction Markets based on PSR (the so-called Market Scoring Rule) always have a stake in the event they predict. In addition, as in our experiment, the last column of Table C4 in Appendix C.
stakes in field environments are likely to be far larger than the prediction’s reward. Consider for instance the stakes the agent may have when predicting the stock market, a natural catastrophe, or the future of her company. Our results therefore suggest that in general field settings stakes and hedging may distort substantially the beliefs reported under a PSR.

Furthermore, the stakes and hedging opportunities an agent may have in the field are typically unobserved by the analyst. In such cases, theory cannot be used to predict and correct the distortions generated by PSR. Suppose, nevertheless, that an agent has no stake in the event she is predicting. Then, two additional issues arise. First, one may be concerned that the beliefs elicited are not informative since the agent had no incentives ex-ante to acquire information about the event. Second, one cannot rule out the possibility that the agent will look for hedging opportunities ex-post. The presence of such opportunities may lead the agent to bias her reported beliefs even though she has no stake when making her prediction. In other words, stakes and hedging may lead to unpredictable distortions in reported beliefs, thereby rendering PSR ineffective in recovering subjective probabilities in general field environments.

Our results may also be relevant for traditional lab experiments. As argued earlier, incentivized belief elicitation in many lab experiments involves a stake or a hedging opportunity. Consistent with our experimental outcomes, Blanco et al. (2010) find that, when faced with transparent and strong hedging opportunities, many subjects in their lab experiment indeed distort both their beliefs and their actions. Because the analyst has more control in the lab than in the field, possible remedial measures may nevertheless be designed to try to correct the PSR biases.

Using a PSR that pays smaller amounts has been the measure most frequently implemented in the lab (e.g. Nyarko and Schotter 2002, Rutström
and Wilcox 2009). We have shown theoretically, however, that this approach can in fact exacerbate the PSR biases if the agent exhibits decreasing relative risk aversion. Recently, various “truth serums” have been proposed to correct PSR biases (e.g. Offerman et al. 2009, Andersen et al. 2010), but these approaches do not address stakes and hedging issues. Conversely, the methods proposed to immunize PSR from stakes and hedging (Armantier and Treich 2009, Blanco et al. 2010) do not correct the PSR biases due to risk aversion.

To the best of our knowledge, making all payments (i.e. for the prediction, stakes and hedging) in lottery tickets that give the subject a chance to win a prize (as suggested by Roth and Malouf 1979 and Allen 1987) is the only approach that can be shown to be incentive compatible under expected utility when stakes or hedging opportunities are present. In practice, however, doubts have been expressed about the ability of this approach to control for risk attitude (Davis and Holt 1993, Selten, Sadrieh and Abbink 1999).

Note also that the adverse effects of stakes and hedging are not specific to PSR and also apply to other incentivized belief elicitation techniques such as the standard lottery mechanism (Kadane and Winkler 1988), the direct revelation mechanism recently proposed by Karni (2009), or the various “truth serums” documented in Trautmann and van de Kuilen (2011). In fact, Karni and Safra (1995) show that unbiased belief elicitation based on marginal rates of substitution is impossible when stakes are not observed by the experimenter. This impossibility result holds even if the utility function is observable and even if several experiments can be implemented.

In this context, one may want to consider the merits of eliciting beliefs without offering any financial reward for accuracy. This approach is simple, transparent, and commonly used outside experimental economics and decision theory. In his review of the subjective expectations literature, Manski
(2004) concludes that the beliefs elicited in such a way are informative. Consistent with Trautmann and van de Kuilen (2011), our results support this view. However, economists and more specifically experimental economists, are often skeptical about non-incentivized methods. In particular, it may be argued that paying subjects for their predictions promotes at least thoughtful, if not truthful responses. Given the increasing role played by subjective beliefs in economics, more evidence should therefore be gathered the settle the methodological debate about the best way to elicit beliefs.

6. References


Allen, F., 1987, Discovering personal probabilities when utility functions are unknown, Management Science, 33, 542-44.


Armantier, O. and A. Boly, 2011, A controlled field experiment on corruption, European Economic Review, 55, 8, 1072-1082.

Armantier, O. and A. Boly, 2012, Can corruption be studied in the lab?, working paper, université de Montréal.


Hanson, R., 2003, Combinatorial information market design, Information Systems Frontiers, 5, 107-119.


Pratt, J. W., 1964, Risk aversion in the small and in the large, Econometrica, 32, 122-36.

Read, D., 2005, Monetary incentives, what are they good for?, Journal of Economic Methodology, 12, 265-76.


Sonnemans J. and T. Offerman, 2004, Is the quadratic scoring rule really incentive compatible?, mimeo, University of Amsterdam.


Figure 1a represents a quadratic scoring rule: $S_1(q) = 1 - (1-q)^2$ and $S_0(q) = 1 - q^2$. Figures 1b, 1c and 1d represent the response function under a quadratic scoring rule and a quadratic utility function: $u(x) = -(2-x)^2$ with $x \leq 2$. Figure 1c plots the response function for $a=1$ and $a=2$. Figure 1d plots the response function for $\Delta=0$, $\Delta=\frac{1}{2}$ and $\Delta=1$. Figures 1e and 1f plot respectively the optimal investment and the response function under a quadratic scoring rule, a quadratic utility function, a double-or-nothing investment opportunity ($k=1$) and a maximal amount invested of 0.75.
Figure 2.1: The Effect of Incentives

Figure 2.2: The Effect of Stakes

Figure 2.3: The Effect of Hedging on the Response Function

Figure 2.4: The Effect of Hedging on the Bet
Table 1: Financial Differences between Treatments (in FCFA)

<table>
<thead>
<tr>
<th></th>
<th>T0 Control</th>
<th>T1 High Incentives</th>
<th>T2 Hypothetical Incentives</th>
<th>T3 Low Stakes</th>
<th>T4 High Stakes</th>
<th>T5 Low Hedging</th>
<th>T6 High Hedging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show-up-fee</td>
<td>0</td>
<td>0</td>
<td>3,000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum Scoring Rule Payment</td>
<td>4,000</td>
<td>40,000</td>
<td>0</td>
<td>4,000</td>
<td>4,000</td>
<td>4,000</td>
<td>4,000</td>
</tr>
<tr>
<td>Maximum Return on Investment</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2,000</td>
<td>8,000</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2: Estimation of the Response Function

<table>
<thead>
<tr>
<th></th>
<th>T0 Control</th>
<th>T1 High Incentives</th>
<th>T2 Hypothetical Incentives</th>
<th>T3 Low Stakes</th>
<th>T4 High Stakes</th>
<th>T5 Low Hedging</th>
<th>T6 High Hedging</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_0)</td>
<td>0.489***</td>
<td>0.509***</td>
<td>0.463***</td>
<td>0.418***</td>
<td>0.358***</td>
<td>0.498***</td>
<td>0.392***</td>
</tr>
<tr>
<td>((T_0))</td>
<td>(0.017)</td>
<td>(0.005)</td>
<td>(0.021)</td>
<td>(0.019)</td>
<td>(0.038)</td>
<td>(0.009)</td>
<td>(0.038)</td>
</tr>
<tr>
<td>(a_1)</td>
<td>—</td>
<td>-0.048</td>
<td>-0.047</td>
<td>0.045**</td>
<td>0.135***</td>
<td>-0.036</td>
<td>0.127***</td>
</tr>
<tr>
<td>((T_0))</td>
<td>(0.032)</td>
<td>(0.028)</td>
<td>(0.020)</td>
<td>(0.031)</td>
<td>(0.031)</td>
<td>(0.022)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>(b_0)</td>
<td>0.677***</td>
<td>0.235***</td>
<td>0.843***</td>
<td>0.585***</td>
<td>0.489***</td>
<td>0.502***</td>
<td>0.534***</td>
</tr>
<tr>
<td>((T_0))</td>
<td>(0.065)</td>
<td>(0.038)</td>
<td>(0.051)</td>
<td>(0.035)</td>
<td>(0.035)</td>
<td>(0.063)</td>
<td>(0.056)</td>
</tr>
<tr>
<td>(b_1)</td>
<td>—</td>
<td>0.405***</td>
<td>-0.134**</td>
<td>0.108**</td>
<td>0.246***</td>
<td>0.184***</td>
<td>0.163***</td>
</tr>
<tr>
<td>((T_0))</td>
<td>(0.081)</td>
<td>(0.051)</td>
<td>(0.047)</td>
<td>(0.047)</td>
<td>(0.060)</td>
<td>(0.077)</td>
<td>(0.051)</td>
</tr>
<tr>
<td>(\sigma_u)</td>
<td>0.518***</td>
<td>0.775***</td>
<td>0.473***</td>
<td>0.448***</td>
<td>0.500***</td>
<td>0.653***</td>
<td>0.564***</td>
</tr>
<tr>
<td>((T_0))</td>
<td>(0.106)</td>
<td>(0.105)</td>
<td>(0.057)</td>
<td>(0.056)</td>
<td>(0.073)</td>
<td>(0.106)</td>
<td>(0.076)</td>
</tr>
<tr>
<td>(\sigma_u)</td>
<td>0.068***</td>
<td>0.059***</td>
<td>0.109***</td>
<td>0.073***</td>
<td>0.089***</td>
<td>0.080***</td>
<td>0.089***</td>
</tr>
<tr>
<td>((T_0))</td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>(\ln(\mathbb{E}))</td>
<td>-1,206.53</td>
<td>-2,488.76</td>
<td>-2,443.75</td>
<td>-2,305.26</td>
<td>-2,243.38</td>
<td>-2,214.82</td>
<td>-2,012.20</td>
</tr>
</tbody>
</table>

In each cell, the first number corresponds to the point estimate, while the number in parenthesis is the estimated standard deviation of the parameter. ***, **, and * respectively indicate parameters significant at the 1%, 5% and 10% levels.

Table 3: Summary of Results

<table>
<thead>
<tr>
<th>Treatments</th>
<th>Hypotheses</th>
<th>Experimental results</th>
</tr>
</thead>
<tbody>
<tr>
<td>T0 Control</td>
<td>(H_0): The response function in (T_0) is (i) regressive (i.e. it crosses the diagonal from above) (ii) has a fixed point at 1/2</td>
<td>Verified Verified</td>
</tr>
<tr>
<td>T1 High Incentives</td>
<td>(H_1): The response function in (T_1) is identical to the response function in (T_0)</td>
<td>Not Verified</td>
</tr>
<tr>
<td>T2 Hypothetical Incentives</td>
<td>(H_2): The response function in (T_2) is identical to the response function in (T_0)</td>
<td>Not Verified</td>
</tr>
<tr>
<td>T3 Low Stakes</td>
<td>(H_3): The response function in (T_3) is (i) lower than in (T_0) (ii) has a fixed point at 1/4</td>
<td>Verified Not Verified</td>
</tr>
<tr>
<td>T4 High Stakes</td>
<td>(H_4): The response function in (T_4) is (i) lower than in (T_3) (ii) lower than (p)</td>
<td>Verified Not Verified</td>
</tr>
<tr>
<td>T5 Low Hedging</td>
<td>(H_5): The response function in (T_5) is (i) regressive with a fixed point at 1/2 (ii) equal to 1/2 when the share invested is in (0,1) (iii) lower than in (T_0) when (p) is close to 1</td>
<td>Verified Partly Verified Verified</td>
</tr>
<tr>
<td>T6 High Hedging</td>
<td>(H_6): The response function in (T_6) is (i) regressive with a fixed point at 1/4 (ii) equal to 1/4 when the share invested is in (0,1) (iii) lower than in (T_5) when (p) is close to 1</td>
<td>Not Verified Partly Verified Verified</td>
</tr>
</tbody>
</table>
Appendix A: Proofs of the Propositions in Section 2

**Proposition 2.0.** A scoring rule $S$ is proper if and only if there exists a function $g(.)$ with $g''(q) > 0$ for all $q \in [0, 1]$ such that
\[
S_1(q) = g(q) + (1 - q)g'(q) \\
S_0(q) = g(q) - qg'(q)
\]

**Proof:** We proved the sufficiency in the text. We now prove the necessity. Define
\[
g(p) \equiv \max_q pS_1(q) + (1 - p)S_0(q) = pS_1(p) + (1 - p)S_0(p) \tag{6.1}
\]
by the definition of a PSR. Note that $g(p)$ is convex since it is the maximum of linear functions of $p$. By the envelope theorem, we have $g'(p) = S_1(p) - S_0(p)$. Replacing $S_1(p)$ by $g'(p) + S_0(p)$ and $S_0(p)$ by $S_1(p) - g'(p)$ in (6.1) directly gives the result. ■

**Proposition 2.1** For all standard PSR defined by (2.3) and (2.5), and for all $p \in (0, 1)$, $R(p) \geq p$ if and only if $p \leq 1/2$.

**Proof:** As we argued in the text, the comparative statics depends on the sign of $f(p, p)$. Using $f(p, q)$ in (2.7), we directly obtain
\[
f(p, p) = p(1 - p)(u'(S_1(p)) - u'(S_0(p))
\]
Thus under risk aversion $f(p, p)$ has the sign of $S_0(p) - S_1(p) = -g'(p)$ using (2.3). Since $g'$ is increasing by the definition of a PSR, and under (2.5) i.e. $g'(1/2) = 0$, we directly obtain $S_0(p) \geq S_1(p)$ if and only if $p \leq 1/2$. ■

**Proposition 2.2** For all standard PSR defined by (2.3) and (2.5), for all $a > 0$ and all $p \in (0, 1)$, and under $\gamma'(x) \geq (\leq)0$, $\frac{\partial R(p, a)}{\partial a} \geq (\leq)0$ if and only if $p \leq 1/2$.

**Proof:** The response function $R(p, a) \equiv R$ is defined by the first order condition
\[
M(a) \equiv p(1 - R)au'(aS_1(R)) - (1 - p)Rau'(aS_0(R)) = 0 \tag{6.2}
\]
Since the objective function is concave, the sign of $\frac{\partial R(p, a)}{\partial a}$ is the same as that
of $M'(a)$. We obtain

$$
M'(a) = p(1 - R) a S_1(R) u''(a S_1(R)) - (1 - p) R a S_0(R) u''(a S_0(R))
$$

$$
= p(1 - R) u'(a S_1(R)) \frac{a S_1(R) u''(a S_1(R))}{u'(a S_1(R))} -

(1 - p) R u'(a S_0(R)) \frac{a S_0(R) u''(a S_0(R))}{u'(a S_0(R))}
$$

$$
= p(1 - R) u'(a S_1(R)) [\gamma(a S_0(R)) - \gamma(a S_1(R))]
$$

where the last equality uses (6.2) and the definition of $\gamma(x)$. Notice that $M'(a)$ has the sign of the term in brackets. Using the properties of $S_0$ and $S_1$ in (2.3) and (2.5), and those of $R$, we conclude that, under $\gamma(x)$ increasing (respectively decreasing), $M'(a)$ is positive if and only if $p$ is lower (respectively larger) than $1/2$.

\textbf{Proposition 2.3} For all PSR defined by (2.3) and for all $p \in (0, 1)$, the response function $R(p, \Delta)$ is characterized as follows:

i) If there exists a $\hat{p}$ such that $\Delta + S_1(\hat{p}) = S_0(\hat{p})$, then $R(p, \Delta) \geq p$ if and only if $p \leq \hat{p}$,

ii) If $\Delta + S_1(p) \geq (\leq) S_0(p)$ for all $p$, then $R(p, \Delta) \leq (\geq) p$.

\textbf{Proof:} The response function $R(p, a) \equiv R$ is characterized by the following first order condition

$$
g(\Delta, R) \equiv p(1 - R) u'(\Delta + S_1(R)) - (1 - p) R u'(S_0(R)) = 0
$$

We compute the marginal benefit of increasing $R$ at $p$:

$$
g(\Delta, p) = p(1 - p)[u'(\Delta + S_1(p)) - u'(S_0(p))]
$$

which is positive if and only if

$$
N(p) \equiv \Delta + S_1(p) - S_0(p) \leq 0
$$

We thus have $R \geq p$ if and only if $N(p) \leq 0$. Since, by our assumptions on the PSR, $N(p)$ is strictly increasing, there is at most one $\hat{p}$ satisfying $N(\hat{p}) = 0$. Therefore either $\hat{p}$ exists and we are in case i), or $\hat{p}$ does not exist and we are in case ii). It is then direct to conclude the proof for each case i) and ii).

\textbf{Proposition 2.4} For all standard PSR defined by (2.3) and (2.5), the solu-
tions \( R(p) \) and \( \alpha(p) \) to program (2.11) satisfy the following properties:

i) for \( p \leq p(k) \), we have \( \alpha(p) = 0 \) and \( R(p) \in [0, \frac{1}{1+k}] \) with \( R'(p) > 0 \),

ii) for \( p \) in \( [p(k), \bar{p}(k)] \), we have \( \alpha(p) \in [0, \bar{p}] \) with \( \alpha'(p) > 0 \) and \( R(p) = \frac{1}{1+k} \),

iii) for \( p \geq \bar{p}(k) \), we have \( \alpha(p) = \bar{p} \) and \( R(p) \in [\frac{1}{1+k}, 1] \) with \( R'(p) > 0 \).

**Proof:** The conditions which characterize the interior solutions \((\alpha^*, q^*)\) of the program (2.11) are

\[
pk u'(S_1(q^*) + k\alpha^* + \bar{p}) - (1 - p)u'(S_0(q^*) - \alpha^* + \bar{p}) = 0 \tag{6.3}
\]

\[
p(1 - q^*)u'(S_1(q^*) + k\alpha^* + \bar{p}) - (1 - p)q^*u'(S_0(q^*) - \alpha^* + \bar{p}) = 0 \tag{6.4}
\]

which imply \( q^* = \frac{1}{1+k} \).

Therefore the condition (6.3) writes

\[
pk u'(S_1(\frac{1}{1+k}) + k\alpha^* + \bar{p}) - (1 - p)u'(S_0(\frac{1}{1+k}) - \alpha^* + \bar{p}) = 0 \tag{6.5}
\]

Differentiating with respect to \( p \) and rearranging yields

\[
\frac{\partial \alpha^*}{\partial p} = -\frac{ku'(S_1(\frac{1}{1+k}) + k\alpha^* + \bar{p}) + u'(S_0(\frac{1}{1+k}) - \alpha^* + \bar{p})}{pk^2u''(S_1(\frac{1}{1+k}) + k\alpha^* + \bar{p}) + (1 - p)u''(S_0(\frac{1}{1+k}) - \alpha^* + \bar{p})} > 0
\]

Therefore \( \alpha^* \) can only increase in \( p \); moreover, the condition (6.3) cannot be satisfied at \( p = 0 \) or at \( p = 1 \). Indeed it is strictly negative at \( p = 0 \) and strictly positive at \( p = 1 \). Consequently, \( \alpha(p) \) is first equal to zero, then equal to \( \alpha^* > 0 \) and strictly increasing in \( p \), and finally constant and equal to \( \bar{p} \).

There are thus two critical values of subjective probability denoted \( p(k) \) and \( \bar{p}(k) \) with \( 0 < p(k) < \bar{p}(k) < 1 \) such that the optimal \( \alpha \) is equal to zero for \( p \leq p(k) \) and is equal to \( \bar{p} \) for \( p \geq \bar{p}(k) \). Moreover, the response function \( R(p) = \frac{1}{1+k} \) when \( p \) is in \( [p(k), \bar{p}(k)] \).

We now study more specifically the response function when \( p \leq p(k) \). Since \( \alpha^* = 0 \), the response function \( R(p) \) is equal to the reported probability without hedging effects, as characterized by \( q \) solving

\[
p(1 - q)u'(S_1(q) + \bar{p}) - (1 - p)qu'(S_0(q) + \bar{p}) = 0
\]

The threshold probability \( p(k) \) is defined by the \( p \) solving

\[
pk u'(S_1(\frac{1}{1+k}) + \bar{p}) - (1 - p)u'(S_0(\frac{1}{1+k}) + \bar{p}) = 0
\]
We thus have
\[ p(k) = \frac{u'(S_0(\frac{1}{1+k}) + \bar{\alpha})}{ku'(S_1(\frac{1}{1+k}) + \bar{\alpha}) + u'(S_0(\frac{1}{1+k}) + \bar{\alpha})} \]
as stated in the Proposition. Moreover, the properties of the standard PSR (2.3) and (2.5) imply \( p(k) < \frac{1}{1+k} \) if and only if \( k > 1 \). Notice also that an increase in \( k \) decreases \( p(k) \). Finally, it is easy to check that \( R(p(k)) = \frac{1}{1+k} \).

We finally study the response function when \( p \geq \bar{p}(k) \). Since \( \alpha^* = \bar{\alpha} \), the optimal reported probability \( q \) is defined by
\[ p(1-q)u'(S_1(q) + (k+1)\bar{\alpha}) - (1-p)qu'(S_0(q)) = 0 \]
The threshold probability \( \bar{p}(k) \) is defined by the \( p \) solving
\[ pk u'(S_1(\frac{1}{1+k}) + (k+1)\bar{\alpha}) - (1-p)u'(S_0(\frac{1}{1+k})) = 0, \]
that is by
\[ \bar{p}(k) = \frac{u'(S_0(\frac{1}{1+k}))}{ku'(S_1(\frac{1}{1+k}) + (k+1)\bar{\alpha}) + u'(S_0(\frac{1}{1+k}))}. \]