Belief elicitation when more than money matters*

Jean-Pierre Benoît
London Business School

Juan Dubra
Universidad de Montevideo

Giorgia Romagnoli
University of Amsterdam

August, 2019

Abstract

Incentive compatible mechanisms for eliciting beliefs typically presume that money is the only argument in people's utility functions. However, subjects may also have non-monetary objectives that confound the mechanisms. In particular, psychologists have argued that people favour bets where their ability is involved over equivalent random bets – a so-called preference for control. We propose a new belief elicitation method that mitigates the control preference. With the help of this method, we determine that under the ostensibly incentive compatible matching probabilities method (Ducharme and Donnell (1973)), our subjects report beliefs 7% higher than their true beliefs in order to increase their control. Non-monetary objectives account for at least 27% of what would normally be measured as overconfidence. Our paper also contributes to a refined understanding of control. We find that control manifests itself only as a desire for betting on doing well; betting on doing badly is perceived as a negative.

Keywords: Elicitation, Overconfidence, Control. Experimental Methods.

Journal of Economic Literature Classification Numbers: D3

*Contact: jpbenoit@london.edu, dubraj@um.edu.uy, G.Romagnoli@uva.nl. We thank Michael Jetter, Lewis Kornhauser, Madan Pillutla, Joël van der Weele and Peter Wakker for comments, and Facundo Danza and Juan Pedro Gambetta for outstanding research assistance. Juan Dubra acknowledges the financial support of ANII and Fondo Clemente Estable. The money for the experiments was provided by London Business School and New York University.

As economists have come to embrace the experimental paradigm long found in other disciplines, they have emphasized the benefits of incentivizing subjects. Incentives both encourage subjects to participate in a meaningful way and guide experimenters in their interpretations of subjects' actions. Typical incentive protocols rely on monetary payments and an (often implicit) assumption that money is the only argument in individuals' utility functions. Thus, an incentive compatible mechanism for eliciting beliefs is taken to be a mechanism in which subjects maximize their utility of money by truthfully reporting their beliefs.

However, while money is important, people also have non-monetary concerns. Researchers who ignore these concerns may end up with a distorted understanding of subjects' actions and beliefs. It is important to have an idea of the magnitude of possible distortions. Can they be safely neglected or do they undermine findings? We report on a new experimental design that allows us to obtain a measure of one type of distortion, which we summarize under the designation control, and to obtain a lower bound on the total non-monetary distortion present. We find that the distortions are notable. At the same time, the amount we can measure is not overwhelming. For example, when the matching probabilities method of Ducharme and Donnell (1973)[1] is used to elicit beliefs, at least 27% of what would usually be interpreted as overconfidence comes instead from a desire for control.

Numerous experiments determine subjects' beliefs about themselves by presenting them with the opportunity to win a prize either based on their performance on a task or based on a random draw. In one format, subjects choose between a bet that yields the prize if their performance places them in, say, the top half of subjects and a bet that yields the prize with objective probability x (see, for example, Hoelzl and Rustichini (2005), Grieco and Hogarth (2009), Benoît, Dubra and Moore (2015), and Camerer and Lovallo (1999), which uses a similar format). The experimenter concludes that subjects who choose to bet on their performance believe they have a probability at least x of placing in the top half. In another format, subjects are asked to report the probability that their performance will place them in the top half. A subject's report determines, in an incentive compatible manner, the probability that she will earn a prize based on her performance rather than from a random draw (see, for example, Hollard Massoni Vergnaud (2010), Andreoni and Sanchez (2014),

[1] This method seems to have been invented by Smith (1961) and implemented by Ducharme and Donnell (1973), following on the Becker-Degroot-Marshack mechanism. It has also been adapted by Grether (1981), Holt (2006), and Karni (2009). There does not appear to be a consistent name used for it in the literature.

Benoît, Dubra and Moore (2015), and Möbius, Niederle, Niehaus and Rosenblat (2014)). The experimenter concludes that subjects who report the number y believe they have a probability exactly y of placing in the top half.

Yet, social scientists have identified (at least) two reasons the above conclusions about subjects' beliefs may overstate their actual beliefs.

1. **Control.** People may have a preference for betting on themselves. Indeed, a long tradition in psychology holds that people have a desire for control in their lives, which may lead them to favour payments contingent on their performance over payments determined by probabilistically equivalent random devices.

2. **Positive statements.** People may derive benefits from making positive statements about themselves, either because they savour positive self-regard or to induce favourable treatment from others. This may lead them to exaggerate their odds of doing well on a task.

The presence of such non-monetary concerns is problematic for the experimenter. As Heath and Tversky (1991) write, "If willingness to bet on an uncertain event depends on more than the perceived likelihood of that event and the confidence in that estimate, it is exceedingly difficult – if not impossible – to derive underlying beliefs from preferences between bets." Heath and Tversky have in mind that subjects may choose to bet on their performance even if they believe the probabilities do not warrant it from a monetary perspective. For instance, a subject who thinks she has a 60% chance of placing in the top half of performers on a task may nonetheless choose to bet on this eventuality over a lottery with an objective 65% chance of paying off.

It is indeed difficult to disentangle subjects' beliefs from their disparate motivations by observing discrete choices they make. But, by comparing the choices subjects make under different conditions, we manage to isolate the desire for control and obtain a measure of the bias it introduces.

In our first experiment, beliefs are elicited using two different mechanisms. Under the first mechanism, subjects effectively choose between betting on themselves and betting on an objective random device. This mechanism employs the matching probabilities method, replicating previous literature. Under the second mechanism, subjects effectively choose between betting on themselves on one task and betting on themselves on another task. This novel design mitigates the control bias: no matter how subjects choose, they are betting on themselves. Both mechanisms are incentive compatible in money.

The implicit assumption in most of the existing literature is that the differences in the designs of the two mechanisms should not affect the elicited beliefs. The two mechanisms are incentive compatible, so they should yield the same distribution of reports if only money matters. Nevertheless, we find evidence for a significant control effect. With the matching probabilities mechanism that duplicates prior studies, subjects inflate their beliefs by 7% in order to shift weight towards bets on themselves (at the cost of reducing the overall chance of obtaining money). This experiment is run in the context of research on overconfidence and, as indicated above, at least 27% of what would otherwise be measured as overconfidence turns out to be a willful inflation.

We run a second experiment in order to better understand control, and its effect on different approaches to elicitation, such as that of Hossain and Okui (2013) and Schlag and van der Weele (2013). Building on previous work by Allen (1987) and McKelvey and Page (1990), the two papers develop essentially equivalent ways of adapting proper scoring rules in a way that avoids the problem of risk aversion. We analyze in detail the mechanism of Hossain and Okui, which they term the *binarized scoring rule.*

With this rule, a subject reports, say, the probability that she will place in the top half of quiz takers and is given a bet that sometimes pays off when she places in the top half, sometimes when she places in the bottom half. Clearly, the rule is subject to self-regard and signalling caveats, as the creators of the rule realise. Less obviously, the mechanism is also vulnerable to control issues, as we show in Section 4.

The analysis of the binarized scoring rule requires a refined understanding of the preference for betting on oneself. Is it that people like to bet that they have done *well* on a task or do they like to bet on their performance, regardless of its quality? If the former, are people neutral about betting that they have done (unintentionally) poorly or do they actively dislike it and, if so, to what extent? In our second experiment, we address these questions by running a series of treatments in which subjects sometimes bet on doing well on a quiz and sometimes bet on having failed to do well.[2] We find that the control motivation manifests itself only as a desire for betting on doing well; a payment for doing badly is perceived as a negative.

[2]Subjects are (stochastically) remunerated for each quiz question they answer correctly and they are not forewarned that they might later bet on a poor performance, so their incentive is to do well on the quiz.

While our study is carried out within the overconfidence paradigm, its applicability is general (see surveys by Schlag, Tremewan and van der Weele (2015) and Schotter and Trevino (2014) on incentive compatible elicitation). The elicitation technique we introduce rewards subjects for their performance on one of two tasks, rather than either rewarding them for their performance on a task or the result of a random device. This design idea can be used independently of a desire to measure control and can be adapted to a variety of mechanisms, including state-of-the-art schemes like the binarized scoring rule and the randomized scoring rule of Schlag and van der Weele (2013). The design can be used beyond studies that elicit beliefs about the self. For instance, studies in which subjects answer multiple choice questions, and are asked for their confidence in their responses, are subject to control and could benefit from our new design.

In the economics literature, Owens, Grossman, and Fackler (2014) also investigates the implications of control for the interpretation of choices between bets. We discuss their results in sections 1.1 and 3.3.

1 Overstatement

In this section, we discuss some of the economics and psychology literature on non-monetary concerns that can lead subjects to misrepresent their beliefs.

1.1 Betting on Yourself: Control

Several studies conclude that people prefer bets on themselves to bets on probabilistically equivalent random devices.

In Goodie (2003), Goodie and Young (2007), and Heath and Tversky (1991, experiments 1, 2, and 3) subjects begin by answering a series of multiple choice questions and, for each question, reporting the likelihood that their answer is correct. They do not realize how these reports will be subsequently used.

Consider subjects who declare they have answered question i correctly with probability (about) p_i. In Goodie and in Goodie and Young, these subjects are split into two groups. In the first group, each subject chooses between (a) a bet that pays off if her answer to question i is correct and (b) the certainty-equivalent payment according to p_i. In the second group, each subject chooses between (a) a bet that pays off with an objective probability p_i and (b) the certainty-equivalent payment. Subjects in the first group choose the bet over the certainty-equivalent more often than subjects in the second group. In Heath and Tversky, each subject is given the choice between (a) a bet that pays off if her answer to question i is correct and (b) a bet that pays off with the objective probability p_i. Subjects take the first bet more often than the second bet, in domains in which they are competent.

These papers find that subjects' choices between betting on their answers and betting on a random device are not a simple reflection of the probabilities involved. Rather, subjects tend to display a bias towards betting on themselves–the more so, the more confident they are in their answers. Notice that when subjects choose to bet on themselves, they are choosing an ambiguous bet over an objective one. The interpetation is that the desire for control overcomes ambiguity aversion, at least when subjects have enough confidence in their answers. (Klein et al. (2010) explores the relation between ambiguity, controllability and competence).

Heath and Tversky argue that people have a special preference for betting on their answers in domains in which they are competent, while Goodie and Young dispute this interpretation and maintain that people have a general preference for control. As Goodie describes it, control is in play whenever the nature of the task is such that "a participant could take steps to favorably alter the success rate in subsequent administrations."[3] Goodie's notion translates immediately from bets on answers to bets on any skilled activity, while Heath and Tversky's notion requires some adaptation for activities. The exact reason a person might favour betting on herself – be it control, competence, or something else – is immaterial for our purposes and we, somewhat abusively, refer to any preference for betting on one's performance on a task as a *control* motivation.

While the findings of these papers are revealing, their methodologies do not permit a measurement of the value of control or the amount by which a preference for control would lead people to overstate their beliefs.[4] Moreover, the findings are weakened by the fact that

[3] Goodie talks of future administrations of the task as the subjects cannot better themselves in the current experiment and he wishes to distinguish control from the illusion of control. In the latter category, Li (2011) finds that subjects prefer a lottery in which they choose numbers to one in which the numbers are randomly selected, even though they recognize that the probabilities of winning is the same in the two. Our modelling accommodates both notions.

[4] Subjects' typically displayed overconfidence in that they answered questions correctly less often than the average likelihood they reported. As a result, they lost money by favouring bets on themselves – as much as 15% of earnings in one experiment in Heath and Tversky. It is impossible to tell to what extent these losses reflected overconfidence and to what extent a sacrifice for non-monetary objectives.

subjects' reports of their likelihoods of correct answers are unincentivized. These papers provide some motivation for our study but we do not undertake to match their frameworks.[5] Neither do we investigate the reasons a person might favour self-bets; instead we evaluate the impact of such a preference.

Owens et al. (2014) contrasts betting on one's own performance with betting on someone else's. Subjects are incentivized to report their beliefs that they will answer a question correctly and their beliefs that a randomly matched participant will answer a different question correctly. They are also asked to choose between a bet on their answer and a bet on the matched subject's answer. Based on the reported beliefs, if subjects care only about money they should choose to bet on themselves 56% of the time. Instead, subjects choose to bet on themselves 65% of the time, pointing to a preference for control. However, the interpretation of the results is somewhat clouded by the fact that the mechanism used for eliciting subjects' beliefs is itself prone to control issues. We discuss this experiment in further detail in Section 3.3.

These four papers, and ours, can be viewed as exploring special cases of *source dependence* (Tversky and Wakker (1995)), whereby subjects consider the source of the uncertainty in addition to the probabilities involved. For axiomatizations that allow for source dependence, see Klibanoff et al. (2005), Chew and Sagi (2008) and Gul and Pesendorfer (2015).

1.2 Positive Statements: Self-Regard and Signalling

People like to say nice things about themselves, both out of self-regard and because sending out positive signals may induce favourable treatment from others. As Baumeister (1982) writes "The desire to be one's ideal self gives rise to motivations affecting both the private self and the public self ... It may also cause individuals to want an audience to perceive them as being the way they would like to be... The experimenter constitutes a real and important 'public' to the subject".

Burks et al. (2013) runs an experiment in which subjects take a quiz and are asked to predict the quintile into which they will place. Subjects also answer a personality traits questionnaire, which reveals that people with a high concern for social image tend to place themselves in high quintiles. The authors conclude that social signalling motives may lead

[5] Indeed, there are elements of these papers which we want to avoid. For instance, in Heath and Tversky's second and third experiments, subjects are asked to rate their knowledge of the subject matter in addition to their probability of answering a question correctly, which could have an effect on their subsequent behavior.

subjects to overstate their beliefs. Ewers and Zimmermann (2015) asks subjects whether they believe their performance on a quiz was better or worse than the average performance of another group. Subjects' reports are either (a) only entered privately onto a computer screen or (b) entered onto a computer screen and also given orally in front of other subjects. The latter more public reporting results in significantly higher self-assessments. The authors conclude that subjects inflate their assessments in order to appear skillful to others.[6]

On the other hand, Benoît et. al. (2015) varies the perceived importance of a task that subjects carry out. Although a more important task should give subjects a greater motive to appear competent to others, the variation produces no effect on reported placements. It should be noted there are potential costs to inflating one's self-assessment, including looking or feeling foolish when actual performance falls short of predictions.

2 Formalism

We now incorporate the desire for control and for saying nice things into a model of utility. For ease of exposition, we develop our formalism in the context of the experiments we run, rather than setting out the most general formulation. Our simple model allows us to identify the effect of control in our experiments. In Section 3.2 we discuss conclusions that are independent of the specific modelling we adopt.

Consider an experiment where a subject undertakes a task for which her performance is described by a variable $\theta \in \{\theta^L, \theta^H\}$, where θ^L indicates a low, or poor, performance and θ^H indicates a high performance. The subject believes there is a chance μ that she will perform well, $\theta = \theta^H$, and she is asked for a report p of this belief. She might earn an amount of money m, depending on how well she does, the number p she indicates, and random draws. If she has an initial wealth w and earns the amount m with probability $r(p, \mu)$ and the amount 0 with probability $(1 - r)$, her expected monetary utility from the experiment is $ru(w + m) + (1 - r)u(w)$. We add two elements to this standard utility function:

1. **Control**. A subject derives an extra utility kick from money that is obtained for her performance, rather than through a random device: when she is paid m for achieving

[6] More precisely, Ewers and Zimmermann conclude that their findings are consistent with some people making reports that are higher than their actual beliefs and some having overconfident beliefs. Schwardmann and van der Weele (2019) find that people who can earn money by convincing others that they are high performers, may deceive themselves as well.

performance θ^i she derives extra utility c_i beyond the utility of the money itself.[7] More precisely, a subject earns an extra utility c_i when she earns m and $\theta = \theta^i$, but she would have earned 0 if instead $\theta = \theta^{j \neq i}$, ceteris paribus. When the elicitation mechanism is such that this happens with probability $q_i(p, \mu)$, the expected utility gain is $c_i q_i$. A complex bet might involve the possibility of sometimes paying a subject for having done well, other times for having done poorly, so that in general the expected utility gain from control is $c_H q_H + c_L q_L$. Perhaps the most natural reading of the literature is that a subject derives a control benefit only from money obtained for having done well, not from money obtained for having done (unintentionally) poorly, so that $c_H > 0$ but $c_L \leq 0$. Experiment 1 examines the nature of c_H, while Experiment 2 also examines c_L.

2. **Self-regard and signalling**. A subject who believes $\theta = \theta^H$ with probability μ and reports p, gets an extra utility kick of $n(\mu)p$ from the report, where $n(\cdot) \geq 0$. If $n(\mu) \equiv 0$, then people derive no benefit from their reports per se. If $n'(\mu) < 0$, then higher types see less reason to inflate their reports. A more general formulation would give the kick as $x(\mu, p)$, with $x_2 \geq 0$. As self-regard/signalling motives are tangential to our study, we use the formulation $x(\mu, p) = n(\mu)p$ which simplifies the analysis. We briefly discuss the more general formulation in Section 3.2. (This is a reduced form approach to incorporating self-regard and signalling benefits. See Burks et al. (2013) for a derivation of a signalling motive.)

A subject's total expected utility from participating in the experiment is

$$ru(w+m) + (1-r)u(w) + c_H q_H + c_L q_L + np$$

Consider, for a moment, an experiment in which an individual is given a lottery ticket that pays m if she answers a question correctly and 0 otherwise. If her belief in her answer is μ then, factoring in control, the expected utility of the lottery is $\mu u(w+m) + (1-\mu)u(w) + \mu c_H$. The expected control benefit is μc_H, which is increasing in the subjective probability of a correct answer, when $c_H > 0$.

Intuitively, a person who believes she has only a small chance of answering the question correctly, perceives little expected control benefit to being paid for a correct answer, and

[7] We safely omit any dependence of c_i on the amount of money m, as this amount does not vary within any of our experiments.

conversely. This feature of our modelling is consistent with experimental findings, noted in Section 1.1, that subjects are more likely to exhibit a bias towards bets on their answers when they have a greater belief in the answers. In Section 5.1.1, we present suggestive evidence from our experiment that people with greater self-belief inflate their reports more for control reasons. The evidence is only suggestive, as we did not plan to test this feature, which would have required a larger sample size. (Our main purpose in this paper is to identify control and measure its distorting effects). Although the effects are not always significant, the signs are always in the right direction.

2.1 Other Motives

Technically, the difference between the two non-monetary elements, control and self-regard/signalling, as we have modelled them, is that the control benefit is contingent, only accruing when a subject is paid for her performance, while the self-regard/signalling benefit always accrues, by virtue of her report. The formalism itself can capture additional non-monetary motivations and variations on the two we have considered. For instance, according to cognitive evaluation theory, a person's intrinsic motivation is higher when payment provides information about her competence level (see Ryan, Mims and Koestner (1983)). As a result, people respond more productively to rewards that are contingent on their good performance. An extra utility kick c_H for paid performance is one way of modelling this. Several studies have found that people prefer taking decisions based on their own judgements rather than ceding control to an algorithm, even when the algorithm is demonstrably superior, which can also be modelled in this way.[8] As to variations, it could, for example, be that statements made to an experimenter and statements made as inputs on a computer yield different benefits, so that $n(\mu)\,p$ is in fact the result of two different components.

3 First Experiment: Controlling for "Control"

This experiment was run at the CREED Lab at the University of Amsterdam. Our sample consisted of 313 undergraduate students from the university. The experiment was pre-

[8] Logg, Minson, and Moore (2018) argues that, although there is a widespread impression that people prefer human over algorithmic judgements, following the seminal Meehl (1954), actually people have a preference for decisions based on *their own* judgements rather than a general preference for human judgements.

registered and the pre-registration materials can be found in Appendix C.[9]

The experiment comprises two treatments that allow us to isolate and measure the control motive. The first treatment closely follows the matching probabilities method, which has been widely used to elicit beliefs, notably in studies on overconfidence (for example, Möbius et al. (2014) and Benoît et al. (2015)).[10] With this design, beliefs are elicited by having subjects compare bets on their performance on a task with bets on a random device. The second treatment uses a new design in which beliefs are elicited by having subjects compare bets which all depend on their performance, on one of two tasks.

The main hypothesis is that there is a control motive to overstate placement in Treatment 1 but not in Treatment 2, while self-regard/signalling motives are the same in the two treatments. As a result, the average reported placement should be higher in Treatment 1 than in Treatment 2. The difference in average reports can be used to measure the control motive. Moreover, the distribution of reported beliefs in Treatment 1 should first order stochastically dominate the distribution of beliefs in Treatment 2.

Timeline of the experiment

The two treatments share the following timeline.

1. Subjects undertake a visual task in which, on 10 occasions, a blinking string of numbers appears on a computer screen, after which they are asked to reproduce the string. The difficulty of the task varies across repetitions in the length of the string and the duration of the blinks. All the subjects see the same sequence of strings.

2. Call s_i the share, or fraction, of the ten repetitions of the task in which subject i correctly identifies the string. Each subject i is told s_i.

3. Subjects answer three sample questions, similar to questions they will later answer in a logic quiz. Before they answer the sample questions, they are informed of the similarity

[9] The pre-registration is also available at https://aspredicted.org/zu3pc.pdf. The second study in the pre-registration is a test of Kruger and Dunning's (1999) "unskilled and unaware" hypothesis. The Appendix reports only the portion of the instructions which is relevant for the present study. The online Appendix reports the instructions for both studies.

[10] Our implementation of the matching probabilites method differs slightly from the usual format in that subjects win lottery tickets rather than cash directly. This is to make the treatment comparable to Treatment 2, where lottery tickets are needed to preserve risk neutrality.

and of the fact that they will need to form an incentivized assessment of their quiz performance compared to others.

4. Subjects are told the median quiz score of people that took the same quiz on prior occasions. Each subject is asked to report the chance that she will place in the top half of quiz-takers. One of two (monetarily) incentive compatible methods, one for each treatment, is used to incentivize the reports. Details are given below.

5. Subjects take a logic quiz in which they answer twelve multiple choice questions. The subjects are ranked according to their scores, with ties broken randomly.

6. Subjects are paid based on their performances in the visual task and the quiz and the accuracy of their reported beliefs, in a way which depends on the treatment assignment and is further elaborated upon below.

Two mechanisms for belief elicitation

The two treatments differ solely in the way in which the beliefs elicited in step 4 above are incentivized. The incentive mechanisms are summarized below; the instructions given in the experiment are in the Appendix.

Treatment 1. Suppose subject i has indicated a probability p_1 of placing in the top half. A number $x \in [0, 1]$ is drawn uniformly. If $x \leq p_1$ the subject wins R lottery tickets if her score is above the median score of the experimental session, with ties broken randomly. If $x > p_1$, with probability x she wins R lottery tickets. In all other cases, she wins nothing.

Treatment 2. Suppose subject i has indicated a probability p_2 of placing in the top half. A number $x \in [0, 1]$ is drawn uniformly. If $x \leq p_2$ the subject wins R lottery tickets if her score is above the median score of the experimental session, with ties broken randomly. If $x > p_2$, with probability x she wins T_i lottery tickets if she was successful in a randomly drawn instance of the visual task (Note that here the subject's skill is at play even when $x > p_2$.) In all other cases, she wins nothing.

The R lottery tickets that can be won in each treatment yield a $\frac{3}{10}$ chance of obtaining €20. For subject i, the T_i lottery tickets that can be won in Treatment 2 yield a $\frac{3}{10 s_i}$ probability of obtaining €20 when $s_i \geq \frac{3}{10}$ (recall that s_i is the fraction of correct answers on the visual task); when $s_i < \frac{3}{10}$ the probability is capped at 1. Subjects are told the numerical value of $\frac{3}{10s_i}$ without being apprised of its dependence on s_i.

To understand the incentive properties of the mechanisms in the two treatments, consider a subject i who estimates her chance of placing in the top half to be μ.

First suppose that she cares only about money. When $s_i \geq \frac{3}{10}$, she should truthfully report her subjective belief that she will place in the top half of subjects, regardless of the treatment in which she participates. To see this, think of the choice between i) a *placement bet* which yields €20 with probability $\frac{3}{10}$ if the subject places in the top half and ii) a *random bet* which, with probability x, yields €20 with probability $\frac{3}{10}$. The subject prefers the placement bet if $\mu > x$ and the random bet if $\mu < x$. The mechanisms in the two treatments implement this preference by effectively asking her for the threshold probability p that causes her choice to switch from the placement bet to the random bet. Clearly, she optimizes by declaring $p = \mu$. When $s_i < \frac{3}{10}$, the mechanism in Treatment 2 is no longer (monetarily) incentive compatible, since the random bet yields €20 with probability less than $\frac{3}{10}$; the mechanism in Treatment 1 remains incentive compatible.

In the theoretical analysis that follows, we focus on the 91% of subjects for whom $s_i \geq \frac{3}{10}$. In the empirical discussion, we analyze both this restricted group and the entire sample; the results are not markedly different.

Now suppose the subject also has non-monetary concerns. We first carry out an informal analysis not tied to our specific modelling.

Consider a subject who estimates her chance of placing in the top half to be μ.

- In Treatment 1, any utility she derives from making positive statements about herself gives her an incentive to exaggerate her reported belief p_1. On top of this, a declaration p_1 means that with probability p_1 winning the €20 is dependent on her performance on the quiz, while with probability $(1 - p_1)$ winning depends completely on a random device. Utility she derives from betting on herself gives her a further incentive to inflate her report, in order to shift weight onto earning money for doing well rather than for being lucky.

- In Treatment 2, as in Treatment 1, the subject may exaggerate her report in order to say nice things about herself. Now, however, she can only earn money when she has performed well, either on the quiz or on the visual task. Utility derived from betting on herself no longer gives a further incentive to distort.

Because a preference for control provides an incentive to inflate in Treatment 1 but not in Treatment 2, we expect $p_1 > p_2$ if subjects have control motives. The difference in the reports, $p_1 - p_2$, can be used to establish measures of the control effect and the total distortion due to non-monetary objectives.

We now reason formally, adopting the normalizations $u(w) = 0$ and $u(w+20) = U$, where w is a subject's initial wealth.

In Treatment 1, a subject who believes she has a probability μ of being in the top half but reports a probability p_1 has a subjective probability $p_1 \mu \frac{3}{10} + (1-p_1)\frac{(1+p_1)}{2}\frac{3}{10}$ of winning the €20. Note that $(1-p_1)$ is the chance that the random draw x is above p_1 and $\frac{(1+p_1)}{2}$ is then the average value of x. In addition to the potential money gain, the subject derives a control benefit c_H when she is paid for doing well on the quiz. The probability that she is paid for doing well – that is, the probability she earns money when she places in the top half but would not have earned it had she not placed in the top half – is $p_1 \mu \frac{3}{10}$. The subject also obtains a self-regard benefit $n(\mu) p_1$ from her report. She has a total expected utility of

$$\left(p_1 \mu + \frac{1-p_1^2}{2}\right)\frac{3}{10}U + p_1 \mu \frac{3}{10} c_H + n(\mu) p_1. \tag{1}$$

This is maximized by a report

$$p_1^* = \mu(1 + C_H) + N(\mu), \tag{2}$$

making the substitutions $N(\mu) = \frac{10}{3}\frac{n(\mu)}{U}$ and $C_H = \frac{c_H}{U}$.[11]

If a subject cares only about money, so that $N(\mu) \equiv 0 = C_H$, then $p_1^* = \mu$. Hence, the mechanism is monetarily incentive compatible. If $N(\mu) > 0$ and/or $C_H > 0$, the subject overstates her beliefs. We can interpret μC_H as the subject's overstatement due to control concerns, $N(\mu)$ as the overstatement due to self-image concerns, and $\mu C_H + N(\mu)$ as the total distortion.

In Treatment 2, a subject who believes she has a probability μ of being in the top half and reports a probability p_2 has a subjective probability $p_2 \mu \frac{3}{10} + (1-p_2)\frac{(1+p_2)}{2} s_i \frac{3}{10 s_i}$ of winning €20 (for $s_i \geq \frac{3}{10}$). The probability that she earns the money for her performance, either on the quiz or on the visual task, is also $p_2 \mu \frac{3}{10} + (1-p_2)\frac{(1+p_2)}{2} s_i \frac{3}{10 s_i}$. Her total expected utility

[11] More precisely, we should write $p_1^* = \min\{\mu(1+C_H) + N(\mu), 1\}$. About 7% of subjects across the two treatments declare a probability of 1.

is[12]

$$\left(p_2\mu + \frac{1-p_2^2}{2}\right)\frac{3}{10}\left(U + c_H\right) + n\left(\mu\right)p_2. \tag{3}$$

This is maximized by a report

$$p_2^* = \mu + \frac{N\left(\mu\right)}{C_H + 1}, \tag{4}$$

making the same substitutions.

If $N(\mu) \equiv 0$ then $p_2^* = \mu$, so the mechanism is monetarily incentive compatible. If $N(\mu) > 0$ then $p_2^* > \mu$ – a subject with self-regard/signalling objectives overstates. Note that a control motivation, $C_H > 0$, does not give a reason to overstate; on the contrary, it dampens the self-image inflation. The reason for this dampening is that the control incentive reinforces the impetus to report truthfully, since $p_2 = \mu$ maximizes both the probability that the subject earns money and the probability that she earns it for doing well (as doing well is the only way she can earn money).

The proposition below uses the differences in the two treatments to examine the nature of the control benefit. As a preliminary to the proposition, consider a lottery ticket that pays €20 if a subject places in the top half of participants and 0 otherwise, yielding an expected utility of $\mu u(w+20) + (1-\mu)u(w) + \mu c_H = \mu U + \mu C_H U$, where $\mu C_H U$ is the control benefit yielded by the lottery. Previous research suggests that people derive utility from being rewarded for doing well, so that $C_H > 0$. A weaker postulate is that people certainly do not dislike being paid for doing well, so that $C_H \geq 0$. A still weaker assumption is that, in any case, receiving the lottery ticket cannot be a bad thing, so that $\mu U + \mu C_H U \geq 0 \Rightarrow C_H \geq -1$, for $\mu > 0$. We use this weaker assumption in the following proposition.

Proposition 1 *Suppose that $C_H \geq -1$ and consider a subject in our experiment for whom $\mu > 0$. Her optimal reported beliefs p_1^* and p_2^* satisfy i) $p_1^* = p_2^*$ if and only if $C_H = 0$, and ii) $p_1^* > p_2^*$ if and only if $C_H > 0$.*

Proof. For i), note that $p_1^* = p_2^*$ if and only if $\mu(1+C_H) + N(\mu) = \mu + \frac{N(\mu)}{C_H+1}$, which holds if and only if $C_H = 0$ (recall that $N(\mu) \geq 0$).

For ii), first suppose that $-1 \leq C_H \leq 0$. Then $\mu(1+C_H) + N(\mu) \leq \mu + \frac{N(\mu)}{C_H+1}$. Conversely, $C_H > 0$ implies $\mu(1+C_H) + N(\mu) > \mu + \frac{N(\mu)}{C_H+1}$. ■

[12] We assume that control benefits do not depend on the task involved. See Section 3.2 for a discussion of relaxing our assumptions.

3.1 Identification

We adopt a between subject design, with each subject participating in either Treatment 1 or Treatment 2. The two groups are drawn from the same pool, hence we make the standard assumption that the expected values of their beliefs are the same – $E(\mu_1) = E(\mu_2) = E(\mu)$. To achieve identification (the statistical analysis will follow in Section 5.1), we treat our samples as large, so that mean beliefs in the two groups are the same – $\bar{\mu}_1 = \bar{\mu}_2 = E(\mu)$ –, the sample average reported beliefs, $\bar{p_1}$ and $\bar{p_2}$, satisfy $\bar{p_1} = E_\mu(p_1^*) \equiv \bar{p_1}^*$ and $\bar{p_2} = E_\mu(p_2^*) \equiv \bar{p_2}^*$, and the mean value of N is the same in the two groups – $\bar{N}(\mu_1) = \bar{N}(\mu_2) = \bar{N}$.

Consider Treatment 1. The standard interpretation of results in this type of experiment is that a finding of $\bar{p}_1 > \frac{1}{2}$ indicates the population is overconfident, since the mechanism is incentive compatible and the mean belief in a well-calibrated population should be $\frac{1}{2}$ (see Benoît and Dubra (2011)). However, using (2) and averaging, we obtain $\bar{p}_1^* = \bar{\mu} + \bar{\mu}C_H + \bar{N}$ and an alternative possibility is that $\bar{\mu} = \frac{1}{2}$ but $C_H > 0$ and/or $\bar{N} > 0$. Then $\bar{\mu}C_H$ is the mean overstatement due to control concerns, $\bar{N}$ is the mean overstatement due to self-image concerns, and $\bar{\mu}C_H + \bar{N}$ is the mean total distortion. It is impossible to tell on the basis of Treatment 1 alone to what extent, if any, a finding of $\bar{p}_1 > \frac{1}{2}$ reflects non-monetary concerns rather than overconfident self-evaluations.

However, Treatment 1 and 2 can be combined to elucidate the role of non-monetary concerns. First, Proposition 1 yields a test for the sign of C_H. A significant difference in treatment averages, $\bar{p}_1 - \bar{p}_2 > 0$, will imply that $C_H > 0$, so that the desire for control distorts reported beliefs. Our experimental findings, discussed in greater statistical detail in Section 5.1, are that $\bar{p}_1 = 66.15\%$ and $\bar{p}_2 = 61.82\%$. The difference $\bar{p}_1 - \bar{p}_2 = 4.33\%$ is significant at the 5% level, confirming the hypothesis that $\bar{p}_1 > \bar{p}_2$. Moreover, the empirical distribution of p_1's almost first order stochastically dominates the distribution of p_2's, as predicted by the control hypothesis (see Figure 1 in section 5.1.)

We can leverage the model further. Using (2) and (4), and averaging within the groups, we obtain

$$\bar{\mu}C_H + \bar{N} = \bar{p}_1^* - \bar{p}_2^* + \frac{\bar{N}}{C_H + 1} \geq \bar{p}_1^* - \bar{p}_2^* = \bar{p}_1 - \bar{p}_2.$$

Thus, $\bar{p}_1 - \bar{p}_2$ gives a lower bound on the overstatement $\bar{\mu}C_H + \bar{N}$ in Treatment 1 that is due to non-monetary concerns rather than to overconfidence. This treatment uses a standard-type incentive mechanism and finds that, on average, people report an overestimate of their chances of being in the top half of 16.15 percentage points. However, of this, at least 4.33

percentage points come from a willful inflation rather than a miscalibration. Put differently, at least 26.8 1% $= \frac{4.33}{16.15}$ of the measured overconfidence in this experiment comes from control and self-regard/signalling distortions.

We can be more specific about the control mark-up μC_H. Again using (2) and (4), we have

$$C_H = \frac{\bar{p}_1^* - \bar{p}_2^*}{\bar{p}_2^*} = \frac{\bar{p}_1 - \bar{p}_2}{\bar{p}_2} = 7\%. \tag{5}$$

On average, each subject in Treatment 1 inflates her report by a factor of 7% to derive control benefits 0.07μ.

Recall that the marginal benefit of control is $c_H = C_H U$, where $U = u(w+20) - u(w)$, so that $c_H = 0.07(u(20+w) - u(w))$. In words, the marginal utility from inflating for control reasons is 7% of the added utility from a gain of €20.

3.2 Modelling

Let us step back for a moment to consider what conclusions obtain without adopting a specific model.

On the face of it, the mechanism in Treatment 2 mitigates control incentives, since a subject can only earn money for a successful performance on one of two tasks, whereas in Treatment 1 a subject can earn money either for her performance or from a random draw.[13] This mitigation leads to the prediction that $\bar{p}_1 > \bar{p}_2$ without any formal modelling. The confirmation we obtain of this prediction is good evidence for the existence of a control effect and for the effectiveness of the new elicitation design.

When subjects with high beliefs inflate their reports for non-monetary reasons, they may hit the reporting constraint of 100%. If, as an illustration, all subjects wanted to inflate their beliefs by a factor of 10%, then those with beliefs above 91% would all make reports of 100%, causing a cluster at this number. If subjects wanted to inflate their beliefs by a factor of 15%, then those with beliefs above 87% would all make reports of 100%. A greater desire to inflate leads to more bunching at 100%. Since Treatment 1 has an additional control reason to inflate relative to Treatment 2, we should expect a relative jump in the number of reports clustering at 100% in Treatment 1. In line with this prediction, 10.7% of subjects report 100% in Treatment 1 compared to 4.6% of subjects in Treatment 2 – a difference which is

[13] Arguably, the mechanism in Treatment 2 is somewhat more complicated than the mechanism in Treatment 1 but it is unclear what impact, if any, this might have on reports. In a different context, Experiment 2 finds no impact simply from experimental variations.

significant at the 5% level. Moreover, when we place subjects into reporting bins of size 10%, plus a bin at 100%, we only find a significant jump at 100%.

Our model permits sharper conclusions, at the cost of added assumptions. The model assumes that money earned for success on the quiz and money earned for success on the visual task yield the same control benefit. The assumption is plausible, especially given that the success rates on the two tasks are similar – 61% and 63% – but we do not test it. The two tasks were expressly constructed to be dissimilar in their natures, as opposed to their success rates, as we did not want a subject's performance on one task to yield (much) information about performance on the other. Due to design constraints, the timing of the tasks is asymmetric: beliefs are elicited after the visual task but before the quiz. The model assumes that these differences are unimportant.[14] These assumptions lead to control incentives merging completely with monetary incentives in Treatment 2. It is easy to adapt the model to the case where the two tasks yield different control benefits. If being rewarded for success on the somewhat unconventional visual task yields a smaller control benefit than being rewarded for placement on the logic quiz, our calculations understate the effect of control, and conversely.

Our model assumes that self regard motives yield a benefit $n(\mu)\,p$, rather than using a more general formulation $x(\mu,p)$. The more general $x(\mu,p)$ yields similar results if the function is "well-behaved".[15]

3.3 Betting on yourself or someone else

In Owens, Grossman, and Fackler (2014), subjects choose between a bet that will pay $20 if they answer a question correctly and a bet that will pay $20 if a matched subject answers a different question correctly. Let μ_s be a subject's belief that she will answer her question correctly and μ_m be her belief that the matched subject will answer his question correctly. The easiest behaviour to interpret is the use of a *cutoff* strategy. With a cutoff strategy, a

[14] For Goodie (2003), the fact that the visual task has already been completed does not affect the control motive, which depends upon the participant being able to improve her performance in subsequent trials. (See Section 1.1.)

[15] In particular, with the formulation $x(\mu,p)$ suppose that $x_2 \geq 0$, and $x_{22} \leq 0$. Writting $X = \frac{10}{3}\frac{x}{U}$, we now have $p_1^* = \mu(1+C_H) + X_2(\mu, p_1^*)$, $p_2^* = \mu + \frac{X_2(\mu,p_2^*)}{C_H+1}$ and $\bar{\mu}C_H + \bar{X}_2(p_1^*, \bar{\mu}) = \bar{p}_1^* - \bar{p}_2^* + \frac{X_2(\bar{\mu},p_2^*)}{C_H+1} \geq \bar{p}_1^* - \bar{p}_2^*$, which mirrors our previous analysis. If in addition X_{22} is small, then $C_H = \frac{\bar{p}_1 - \bar{p}_2}{\bar{p}_2} + \frac{\bar{X}_2(\mu,p_2^*) - \bar{X}_2(\mu,p_1^*)}{\bar{p}_2} \approx \frac{\bar{p}_1 - \bar{p}_2}{\bar{p}_2}$. Otherwise, since the distribution of p_1s almost first order stochastically dominates that of p_2s, $C_H \geq \frac{\bar{p}_1 - \bar{p}_2}{\bar{p}_2}$ for "reasonable" functions.

subject bets on herself if $\mu_s - \mu_m > k$, for some number k. If $k = 0$, the subject maximizes her expected monetary payoff; if $k < 0$ the subject values control and is willing to sacrifice money in order to bet on herself; if $k > 0$ the subject prefers to bet on someone else. Owens et al. use the word *control* as an "umbrella term" that encompasses any reason a person might favour a bet on herself. This includes choosing to bet on yourself to send a positive signal.

The beliefs μ_s, μ_m are not known to the experimenters. Rather, subjects are incentivized to make reports p_s and p_m of their beliefs, using a matching probabilities method similar to the one we use in Treatment 1. The subjects' behaviour is evaluated with respect to these (observable) reports. That is, a subject is deemed to follow a cutoff strategy if she bets on herself when $p_s - p_m > \overline{k}$, for some number $\overline{k}$. When $\overline{k} < 0$, the subject is said to exhibit a preference for control. The authors determine that the behaviour of 82% of the subjects is consistent with a cutoff strategy.

Let us apply our modelling to this experiment. To begin, we keep things simple and assume that a) subjects have only a pure control motive, so that $c_H > 0$ but $n(\cdot) \equiv 0$, and b) they evaluate money won for someone else's performance purely in monetary terms. Under these assumptions, the elicited beliefs are given by

$$p_s^* = \mu_s (1 + C_H) \text{ and } p_m^* = \mu_m, \tag{6}$$

using the normalizations $u(w) = 0, u(w+20) = U$, and $C_H = c_H/U$.

Now consider a subject's decision whether to bet on herself or bet on her match. Using our modelling, her payoff for betting on herself is

$$\mu_s u(w+20) + (1-\mu_s) u(w) + \mu_s c_H = \mu_s U + \mu_s C_H U, \tag{7}$$

while the payoff for betting on her match is

$$\mu_m u(w+20) + (1-\mu_m) u(w) = \mu_m U. \tag{8}$$

A subject chooses to bet on herself if $\mu_s - \mu_m > -\mu_s C_H$. If $c_H > 0$, as we find on average, then the unobservable cutoff $k = -\mu_s C_H$ is negative.

In terms of observables, from (6) we have that $\mu_s - \mu_m > -\mu_s \frac{c_H}{U}$ if and only if $p_s - q_s > 0$. Although the true cutoff k is negative, the measured cutoff $\hat{k}$ should be zero. Put differently, we have $\hat{k} = 0$ even for a subject with a positive control motivation (or a negative one, for that matter). In line with this reasoning, in one of their analyses, Owens et al. determine that, of the subjects with a cutoff behavior, 65% have a behavior that is consistent with a cutoff of 0. When these subjects are counted as not having a control motivation, our analysis implies that control is under-measured. In their conclusion, Owens et al. also reason that they have found a lower bound on the effect of control incentives.

Although the above reasoning suggests that 0 should be the measured cutoff, 26% of subjects display a strictly negative cutoff (and 9% a strictly positive cut-off). This discrepancy can be reconciled with our modelling in several ways.

1. When given a direct choice between a bet on themselves and a bet on another person, some subjects may feel an extra push to bet on themselves. This push could be because of the positive signal sent by betting on oneself over someone else, because of the inherently larger ambiguity in a bet on someone else, or for some other reason. Such a push is consistent with the discussion in Owens et al. of the various reasons subjects may prefer bets on themselves. In terms of the above analysis, the simplifying assumptions a) and b) may not both hold.

2. The incentive to inflate for control may be especially salient to subjects in this experiment, where subjects are presented with a direct choice between two bets in contrast to the more elaborate matching probabilities mechanism.

3. Procedural details in this experiment and in ours may (inadvertently) play a role in the results.

The distinction between i) self-bets versus bets on someone else and ii) self-bets versus bets on a random device is an interesting one that our experiment and theory does not explore.

4 Second Experiment: The Meaning of Control

Experiment two was also run at the CREED Lab at the University of Amsterdam, this time with one hundred ninety-six undergraduates. There was no overlap in the samples of the two experiments.

This experiment seeks a better understanding of the control motivation. Our first experiment showed that people have a positive bias for bets that pay off when they do well. But how do they feel about bets that pay for an (unintentional) poor performance? Do these bets also yield a control benefit or are they undesirable in this regard? The answers are not only important for a proper understanding of the control motivation but are also crucial for the analysis of some incentive mechanisms.

Consider the binarized scoring rule. While the rule does not explicitly present subjects with a trade-off between winning based on their performance and winning based on a random draw, it does so implicitly. To see exactly in what way, we describe the rule for a subject who is asked for the probability that her performance on a task is high ($p\left(\theta = \theta^H\right)$). In order to focus on control issues, we ignore self-regard concerns in what follows, setting $n \equiv 0$.

The binarized scoring rule works as follows. After a subject reports a probability p of being in the top half, a random number z is drawn uniformly from $[0, 1]$. The subject wins an amount m if and only if (a) $\theta = \theta^H$ and $z \geq (1-p)^2$ or (b) $\theta = \theta^L$ and $z \geq p^2$.

Suppose that $p \geq \frac{1}{2}$. If $z \geq p^2$, she wins m regardless of her performance; if $z < (1-p)^2$ she wins nothing regardless of her performance. In both cases, control plays no role. Control is at play when $(1-p)^2 \leq z < p^2$, as the subject then wins m if and only if she performs well. Setting $u(w+m) = 1$ and $u(w) = 0$, the expected utility from a report $p \geq \frac{1}{2}$ is

$$1 - p^2 + \left(p^2 - (1-p)^2\right)\mu + \left(p^2 - (1-p)^2\right)c_H\mu,$$

which is maximized at $p^* = \mu + c_H\mu$.

Similar reasoning shows that if $p < \frac{1}{2}$, control is at play when $p^2 \leq z < (1-p)^2$, as she then wins m if and only if $\theta = \theta^L$. Note that she now earns money for a poor performance. Her expected utility is

$$1 - (1-p)^2 + (1-\mu)\left((1-p)^2 - p^2\right) + c_L(1-\mu)\left((1-p)^2 - p^2\right),$$

which is maximized at $p^* = \mu - c_L(1-\mu)$, when this is less than $\frac{1}{2}$.

Experiment 1 established that $c_H > 0$, so that subjects with belief $\mu \geq \frac{1}{2}$ inflate their reports for control reasons. Whether control pushes subjects with belief $\mu < \frac{1}{2}$ to inflate, deflate, or neither depends on the sign of c_L. That is, the impact of control depends upon a subject's feelings about being rewarded for failure.

4.1 Three Treatments

Experiment 2 involves three treatments which share the following steps (the appendix provides the instructions that were used):

1. Subjects take a quiz in which they answer twenty multiple-choice questions. With a 50% chance subjects will be paid €0.50 for each correct answer; with a 50% chance they will be paid according to the incentive compatible mechanism in the second part of the experiment. (At this point, subjects are not aware of the precise nature of the second part so that, presumably, their incentive is to do well on the quiz).

2. Subjects are then asked to report on their placement odds in one of three (monetary) incentive compatible manners.[16]

Treatment 1

Each subject is asked for the probability p_1 that she will place in the top half. A number $x \in [0, 1]$ is drawn uniformly. If $x \leq p_1$ she wins €10 if she lands in the top half, with ties broken randomly. If $x > p_1$ she wins €10 with probability x. In all other cases, she wins nothing.

Treatment 2

Subjects are given the opportunity to bet on a low placement, rather than a high one.

Each subject is asked for the probability q_2 that she will place in the *bottom* half. A number $x \in [0, 1]$ is drawn uniformly. If $x \leq q_2$ she wins €10 if she lands in the bottom half, with ties broken randomly. If $x > q_2$ she wins €10 with probability x. In all other cases, she wins nothing.

Treatment 3

This treatment is a mixture of the first two.

Each subject is asked to report her belief p_3 that she will place in the top half of quiz takers. A coin is flipped and a number $x \in [0, 1]$ is drawn. Suppose the coin comes up heads. Then if $x \leq p_3$, she wins €10 if she lands in the top half, while if $x > p_3$ she wins €10 with probability x. Suppose the coin comes up tails. Then if $x \leq 1 - p_3$, she wins €10 if she lands in the bottom half, while if $x > 1 - p_3$ she wins €10 with probability x.[17]

[16] In contrast to Experiment 1, subjects make their predictions after having taken the test rather than after only having seen sample questions, since they will sometimes bet on doing poorly. Because of this and other differences, the beliefs elicited in Experiments 1 and 2 are not directly comparable. This has no consequences for our analysis.

[17] In actuality, for half of the subjects in this treatment, the question was framed as a bet on placing in the bottom half, rather than in the upper half. To both groups it was explained that, depending on the results of the toss of the coin flip, they would end up betting either on their placement in the upper half or in the lower half. We found no difference between the two frames of choice (p-value = 0.677)

On a conceptual level, Treatment 1 here mimics Treatment 1 in the first experiment. Subjects have an incentive to inflate their reports, both for self-regard/signalling reasons and in order to bet on themselves doing well.

Treatment 2 has no parallel in Experiment 1. While self-regard/signalling concerns operate exactly as in Treatment 1 – subjects have an incentive to underreport the probability of placing in the bottom half, which is equivalent to overreporting the chance they end up in the top half –, control considerations are different. Here, subjects can be rewarded for doing poorly but not for doing well. In terms of our formalism, the parameter c_L, rather than c_H, now plays a role.

4.2 Reporting Incentives

We first analyze reporting incentives, adopting the normalizations $u(w+0) = 0$ and $u(w+10) = 1$, where w is a subject's initial wealth, and making the substitution $q_2 = 1 - p_2$.

Consider a subject who estimates her chance of placing in the top half to be μ and reports this chance as: p_1 if in Treatment 1; effectively reports it as $p_2 = 1 - q_2$ if in Treatment 2; and reports it as p_3 if in Treatment 3.

In Treatment 1, she has an expected utility of

$$p_1\mu + \frac{1-p_1^2}{2} + c_H p_1 \mu + n(\mu)\, p_1,$$

which is maximized at

$$p_1^* = \mu(1+c_H) + n(\mu)\,. \tag{9}$$

In Treatment 2, she has an expected utility of

$$(1-p_2)(1-\mu) + \frac{2p_2 - p_2^2}{2} + c_L(1-p_2)(1-\mu) + n(\mu)\, p_2,$$

which is maximized at

$$p_2^* = \mu - c_L(1-\mu) + n(\mu)\,. \tag{10}$$

In Treatment 3, she has an expected utility of

$$\frac{1}{2}\left(p_3\mu + \frac{1-p_3^2}{2} + c_H p_3 \mu\right) + \frac{1}{2}\left((1-p_3)(1-\mu) + \frac{2p_3 - p_3^2}{2} + c_L(1-p_3)(1-\mu)\right) + n(\mu)\, p_3,$$

which is maximized at

$$p_3^* = \mu + \frac{1}{2}c_H\mu - \frac{1}{2}c_L(1-\mu) + n(\mu)\,. \tag{11}$$

We exploit these expressions in the next section.

4.3 Identification

We again analyze mean behaviour across treatments. From (9), (10), and (11), the theory demands that the optimal choices satisfy $\overline{p}_3^* = \frac{1}{2}\overline{p}_1^* + \frac{1}{2}\overline{p}_2^*$. Thus, Treatment 3 does not add anything to the estimation of the parameters but serves as a consistency check of the theory. The theory receives confirmation – or, at least, is not rejected – as we find that $\overline{p}_1 = 66.2\%$, $\overline{p}_2 = 67.9\%$ and $\overline{p}_3 = 66.7\%$ and, as we show later, we cannot reject $p_1^* = p_2^* = p_3^*$.

Given $\overline{p}_1^* = \overline{p}_2^*$, (9) and (10) together imply that

$$c_L = -c_H \frac{\mu}{1-\mu}. \tag{12}$$

Experiment 1 established a strictly positive, and statistically significant desire for betting on one's success. The results of this experiment show that winning money for doing poorly provides negative utility. This finding is consistent with Heath and Tversky's (1991) finding that subjects favour an ostensibly fair random bet over a bet that pays when they have answered a question *incorrectly*. Our result goes further, indicating that the utility loss from a payment for doing poorly, $c_L(1-\mu)$, is the exact negative of the utility gain from a payment for doing well, $c_H\mu$.

Returning to the binarized scoring rule, control objectives will lead a subject with belief μ to report $p^* = \mu + c_H\mu = \mu - c_L(1-\mu)$, whether her belief μ is above or below $\frac{1}{2}$ (when $n \equiv 0$). Thus, the binarized scoring rule is subject to control distortions similar to those in the matching probabilities method.[18] The mechanism we introduced earlier to eliminate control distortions under the matching probabilities method can be adapted to eliminate control distortions with this rule.

We note that the result $p_1^* = p_2^* = p_3^*$ indicates that subjects do not change their behaviour simply in reaction to different experimental protocols.

5 Experiments: Timelines and Statistical Analysis.

In this section, we give a detailed description of the experiments and provide a statistical analysis of the results.

[18] The quadratic scoring rule is also subject to control issues if, analogously to these findings, subjects value a dollar gained for doing well differently than a dollar gained for doing poorly.

5.1 Experiment 1 - Regression analysis

The experiment was run in the CREED Lab at the University of Amsterdam in Spring 2018, using the software oTree (Chen, Schonger and Wickens, 2016) and in conjunction with another experiment by the same authors. Three hundred and thirteen undergraduates participated and were assigned to either Treatment 1 (N=154) or Treatment 2 (N=159). The experiment's duration was approximately 50 minutes and subjects earned €18.50 on average.

The randomization was successful in ensuring a good gender balance, with 56.49% and 56.33% of female participants in Treatment 1 and 2, respectively. The randomization was also balanced in terms of performance in the sample questions, a predictor of both placement and actual performance in the subsequent test (the mean number of correct sample questions was 2.04, out of 3, in both treatments). Two subjects opted for not reporting their gender. They are excluded from the analysis in accordance with the pre-registration plan (including them in the difference of means test that does not need gender does not change the results).

A second group of subjects deserve a special statistical treatment and is selectively removed in some of the analysis that follows. These are subjects with a success rate below 30% in the visual task. These subjects are awarded enough lottery tickets to ensure a capped probability of 1 of receiving the prize, conditional on a successful round of the visual task being drawn for them. However, a probability of 1 is insufficient to achieve incentive compatibility (in monetary terms) for them in Treatment 2. As a result, these subjects have an incentive to over-report in Treatment 2, giving a potential distortion to our measurements. Hence, in most of the analysis that follows we will present results with and without this group of subjects. When we exclude low performing subjects, we do so for both treatments to avoid introducing a selection effect.[19]

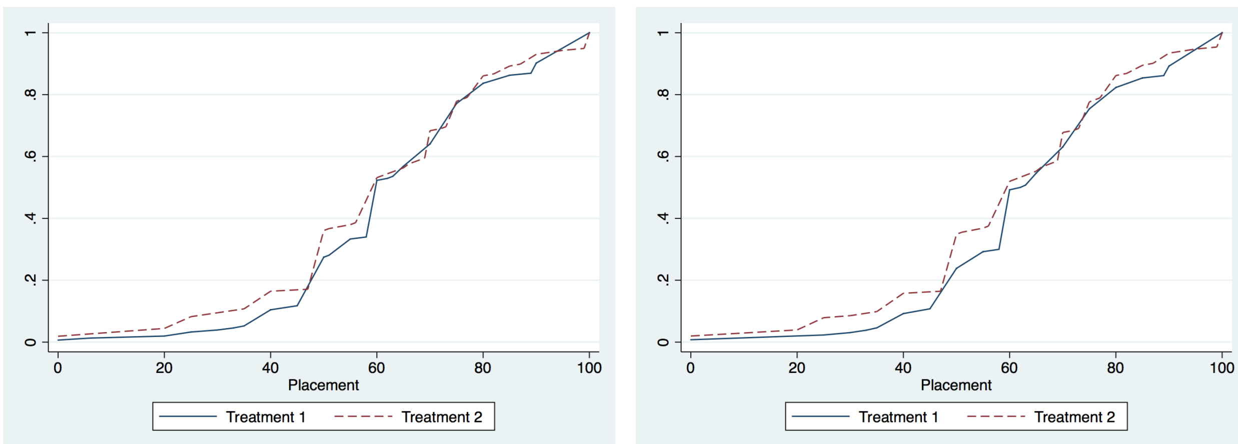
The main hypothesis is the existence of control motives to overstate beliefs in Treatment 1 but not in Treatment 2, while self-regard/signalling motives are the same in the two treatments. Formally, as pre-registered, we test if the average placement $\bar{p}_1$ in Treatment 1 is statistically larger than the placement $\bar{p}_2$ in Treatment 2, by performing an independent two-sample one-sided t-test. The test supports the hypothesis (p-value = 0.0728, N= 311. The $t-$test for the 282 subjects with $s_i \geq \frac{3}{10}$, whose reports were properly incentivized, appears as Model A in Table 1, and Treatment 1 is significant at the 5% level, with a

[19] We note that we did not anticipate the failure of incentive compatibility for some subjects and consequently we did not pre-register their exclusion from the analysis.

$p-$value of 0.03).

The model predicts not only that $\bar{p}_1 > \bar{p}_2$, but also that the distribution of reported beliefs in treatment 1 first order stochastically dominates the distribution of beliefs in treatment 2. We explore this hypothesis in Figure 1, where we plot the cumulative distribution of placement by treatment. Panel 1.a includes all 311 subjects. Panel 1.b only includes subjects for whom incentive compatibility holds.

Figure 1: Cumulative distributions



(a) All subjects (N=311)

(b) Only incentive compatible (N=282)

The cumulative distribution of p_1 lies below the one for p_2 for the bulk of the domain. This indicates that the result $p_1 > p_2$ is not due to only a handful of participants. Because we had not anticipated testing for first order stochastic dominance, we calculated our sample size only to test whether $p_1 > p_2$. Perhaps as a result, the Kolmogorov-Smirnov test of difference between distributions does not reject the hypothesis that the two distributions are the same ($p-value = 0.61$ when the sample includes all subjects and $p-value = 0.295$ when including only subjects for whom incentive-compatibility holds). It is well known that the Kolmogorov-Smirnov test has low power (see, for example, Razali and Wah (2011) and Kim and Whitt 2015), especially with small samples. Thus, in simulations using normal distributions with a common standard deviation of 0.2 and a difference in means of 4.33, as in our data, we reject equality in only 39% of cases, with sample sizes of 150 from each distribution. More data would be needed to reach conclusive results on distributional differences; we only envisioned and pre-registered running the much more powerful difference-in-means test.

Next, we report the results of the regressions including controls in Table 1. The first three columns — models A-C — report results only for subjects for whom the mechanism is incentive compatible. The table also presents, in model D, the analysis for all subjects;[20] the analysis is fundamentally unchanged. The dependent variable is *Placement*, the reported probability of being in the upper half of the scores distribution. The main variable of interest is *Treatment-1*, a dummy taking value 1 if the observation belongs to Treatment 1. In accordance with the pre-registration plan, the table reports the p-values of the one-sided test for the hypothesis $p_1^* > p_2^*$, though we also report the $p-$values for the two-sided test (in parenthesis). All our one-sided tests including controls are significant at 5%.

Table 1: The effect of control on placement

	Placement	Placement	Placement	Placement
	Model A	Model B	Model C	Model D
Treatment 1	4.331*	4.369*	4.104*	3.960*
	(0.069)	(0.064)	(0.077)	(0.079)
Male		5.737**	4.498*	4.609**
		(0.016)	(0.056)	(0.042)
# of Lottery Tickets		-0.288	-0.237	-0.287
		(0.244)	(0.331)	(0.101)
Sample Score			5.164***	5.344***
			(0.001)	(0.000)
Constant	61.82***	58.38***	48.60***	48.90***
	(0.000)	(0.000)	(0.000)	(0.000)
H_1: Treatment $1 > 0$ (p-values)	0.0346	0.0322	0.0385	0.0469
N	282	282	282	311
R^2	0.0117	0.0375	0.0760	0.0867

NOTES: The dependent variable is placement (the reported belief that own performance in the quiz is above the median). Models A-C only include subjects for whom the elicitation is incentive compatible. Model D reports results for the full dataset (excluding two subjects who did not report their gender). *P*-values in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The variable *# of Lottery Tickets* is tied to the performance in the visual task and measures how many lottery tickets are awarded to subjects who end up betting on the visual task, conditional on having been successful in the selected round. The amount is calibrated

[20] In model D, we exclude the two subjects who did not report their gender, as the regression requires that data.

for every subject to ensure incentive compatibility and, as expected, the coefficients associated to this variable are insignificant. The last control variable is *Sample Score*, i.e., the number of correct answers given in the three sample questions. This variable is a signal that subjects can use to infer how well they will perform in the quiz (which, they are told, is based on questions similar to the sample questions).[21] As expected, a better performance on the sample quiz significantly increases the reported placement probability. Gender also correlates with placement reports; males tend to assign a significantly higher probability to the event that they will place in the top half of test takers than females. This is in line with previous findings that men are more confident-looking than women (see Barber and Odean (2001), and Niederle and Vesterlund (2007), and the references therein).

When comparing model D with the other models, we note that when the analysis exclusively focuses on subjects for whom the mechanism is incentive compatible, the point estimates of the difference between the two treatments is larger and the statistical significance improves. This is not surprising since the failure of incentive compatibility gives subjects an extra (monetary) incentive to over-report in Treatment 2, so as to reduce the likelihood that payment depends on the visual task. It stands to reason that the removal of these subjects strengthens the results.

5.1.1 Heterogeneous effects

We now examine possible heterogeneous effects of control motivations. In particular, we explore whether control motives affect reported beliefs to different degrees depending on the confidence level μ. As mentioned in Section 1.1, Goodie (2003), Goodie and Young (2007), and Heath and Tversky (1991) provide some evidence that people with greater confidence in their performance exhibit a greater bias towards betting on themselves.[22]

We first consider the theoretical predictions our model makes. The model is built on the premise that a subject reaps control benefits upon being paid for a successful performance. Hence, there is an intuitive sense that distortions brought about by a manipulation of control motives, such as in our experiment, should be increasing in a subject's perceived likelihood of success.

[21] Subjects are not told their scores on the sample questions, but it is likely that they formed some beliefs about their performance in the sample. They are told the median score in previous sessions, and that the sample and quiz questions are similar. This enables them to transform their absolute inference into a relative belief.

[22] This is true of Heath and Tversky's first three experiments but not their fourth.

Recall that optimal reported beliefs are $p_1^* = \mu(1 + C_H) + N(\mu)$ and $p_2^* = \mu + \frac{N(\mu)}{C_H+1}$. Thus,

$$\begin{aligned} \frac{dp_1^*}{d\mu} &= (1 + C_H) + N'(\mu) \\ \frac{dp_2^*}{d\mu} &= 1 + \frac{N'(\mu)}{C_H + 1} \end{aligned}$$

With the minimal assumptions we have made on N so far, namely $N(\mu) \geq 0$, it is possible that $\frac{dp_i^*}{d\mu} < 0$. That is, people with lower beliefs could make higher reports. Clearly, making an inflated report for self-regard/signalling reasons does not make much sense if high reports indicate low beliefs. Indeed, a fully articulated signalling model would start with assumptions to ensure that higher types make higher reports. Accordingly, let us now assume that $(1 + C_H) > -N'(\mu)$. The model then predicts that $\frac{dp_1^*}{d\mu}, \frac{dp_2^*}{d\mu} > 0$.

This prediction is of limited interest, obtaining even if people have no control or signalling objectives ($C_H = 0 \equiv N$). A more interesting prediction comes from examining $p_1^* - p_2^*$. The difference $p_1^* - p_2^*$ comes from the control motivation in Treatment 1. We have that

$$\frac{d(p_1^* - p_2^*)}{d\mu} = C_H(1 + \frac{N'(\mu)}{C_H + 1}) > 0$$

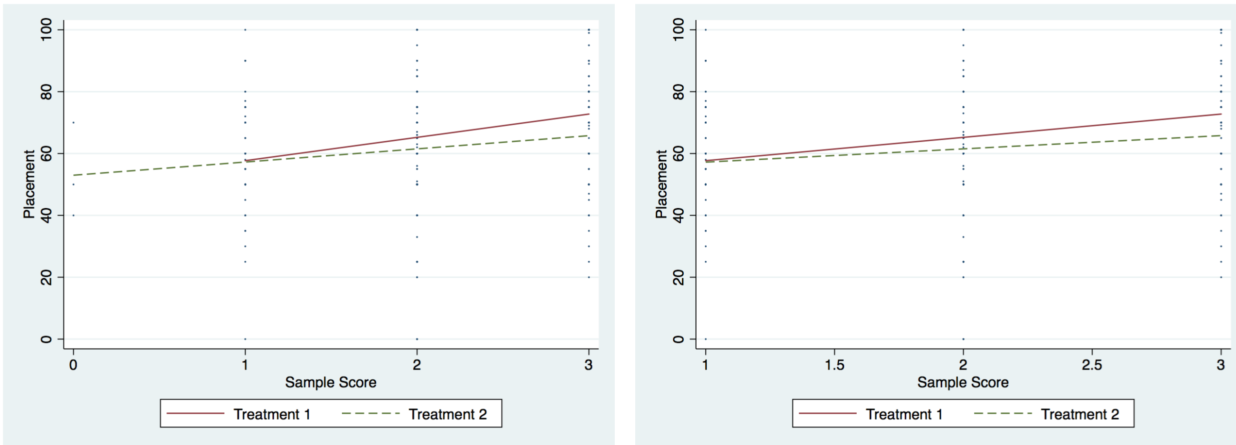
Thus, the model predicts that the distortion due to control is increasing in beliefs.

We now turn to our data. Using subjects' placement as an estimator for confidence levels would lead to obvious endogeneity problems when testing whether $p_1 - p_2$ is increasing in μ. We instead use the score obtained on the sample questions as a proxy for confidence. This score is a strong predictor of both performance in the actual quiz and reported placement probability (p-values ~ 0.000 in OLS regressions, as reported in the Appendix), suggesting that performance in the sample questions is both a valuable signal for future performance in the quiz and an information seemingly used by subjects to determine their confidence level.

Equipped with this proxy for confidence, we set to explore its impact on the distance between reported probabilities in the two treatments. Figure 2 plots Placement against Sample Score and, in addition, presents two separate regression lines for the two treatments. Panel 2.a includes only subjects for whom the elicitation is incentive compatible. Panel 2.b further excludes 3 outliers who answered zero sample questions correctly.

The scatter plots inclusive of all subjects are not reported but are similar to the ones in the figure. The plots suggest that the degree of the distortion is slightly increasing in

Figure 2: Heterogeneous effects



(a) Only incentive compatible (N=282)

(b) Exclusion of outliers (N=279)

confidence. Interestingly, the point estimate of the effect of control (as captured by our treatment effect) for subjects who answered one sample question correctly is zero. However, the slope coefficients are not statistically different from one another (p-value = 0.2673).

We further explore the relationship between confidence μ and treatment differences with a second set of specifications in which we additionally include the interaction variable *Treatment-1* $\times$ *Sample Score*.

If treatment differences are indeed increasing in confidence, we should observe a larger treatment effect as the sample score (and therefore μ) increases. Hence our estimation for the coefficient of the interaction term should be positive and significant. The results are presented in Table 2. Models E-G are OLS regressions, showing that the interaction term is positive and only insignificant when both $Treatment - 1$ and $Sample\ Score$ are included as regressors. Model H is a semi-structural estimation imposing a model restriction that derives from assuming the presence of control motives. More specifically, as highlighted by Proposition 1, $C_H > 0$ implies $p_1^* > p_2^*$ for all μ. In accordance with this restriction, we run a constrained regression imposing that the intercept estimated for treatment 1 must lie weakly above the intercept for treatment 2. The estimation is reported under Model H, which is essentially equivalent to model F given the negative sign of the estimated *Treatment-1* coefficient in Model G, which is the unrestricted version of model H. The effect of the interaction term becomes significant at the 5% level. The regression analysis with the exclusion of the 3

Table 2: Heterogeneous effect - The role of confidence

	Placement Model E	Placement Model F	Placement Model G	Placement Model H
Treatment 1	-10.61**		-2.811	0
	(0.048)		(0.677)	$(\cdot)$
Sample Score		4.332***	3.813*	4.332***
		(0.007)	(0.061)	(0.007)
Treat.1 X Sample Score	7.005***	2.064**	3.245	2.064**
	(0.002)	(0.046)	(0.283)	(0.046)
Male	5.031**	4.526*	4.524*	4.526*
	(0.032)	(0.054)	(0.054)	(0.054)
Constant	54.57***	46.17***	47.40***	46.17***
	(0.000)	(0.000)	(0.000)	(0.000)
N	282	282	282	282
R^2	0.0648	0.0761	0.0767	0.0761

NOTES: The dependent variable is placement (the reported belief that own performance in the quiz is above the median). Models E-G are OLS regressions. Model H is a semi-structural estimation imposing the model's restriction that Treatment 1 > 0. P-values in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

outliers (who scored zero in the sample questions) is fundamentally unchanged. All things considered, we find suggestive but not particularly robust evidence that control motives are increasing in confidence levels. One must still keep in mind that this is an "out of sample" prediction of our model (it was not built to yield this prediction), that the data is in line with the prediction (and, importantly, does not reject it), and that the sample size of the experiment was intended for the more basic test of $\bar{p}_1 > \bar{p}_2$.

Note that our prediction that more confident individuals inflate more for control reasons is unrelated to the Kruger and Dunning (1999) *unskilled and unaware* effect, which maintains that unskilled people are especially overconfident in their beliefs, as this effect is about people's actual beliefs, not their reports of these beliefs. Moreover, control is not implicated in the Kruger and Dunning experiments, which elicit beliefs in an unincentivized manner.

5.2 Experiment 2

Experiment 2 was run in the same laboratory in Fall 2016. One hundred ninety-six undergraduates participated, drawn again from the University of Amsterdam. No subject took part in both experiments.

The three treatments exhibit basically the same average estimate of p_i. In Treatment 1, with 68 subjects, $\bar{p}_1 = 66.2\%$; in Treatment 2, with 61 subjects, $\bar{p}_2 = 67.9\%$; in Treatment 3, with 67 subjects, $\bar{p}_3 = 66.7\%$. There are large standard deviations of comparable magnitude across treatments (16.9, 18.5 and 19.8 for Treatments $1-3$ respectively).

We perform two tests. With the Wilcoxon rank sum (Whitney-Newey) test, the p value for equality of distributions is 0.61 for Treatments 1 and 2, 0.76 for treatments 2 and 3, and 0.78 for Treatments 1 and 3. We also run the corresponding t test for difference of means and we do not reject equality (p value $= 0.58$ for Treatments 1 and 2, 0.72 for Treatments 2 and 3, and 0.88 for Treatments 1 and 3).

These findings establish that in the experiment the positive control kick for a favourable bet on self is the same as the negative kick for a bet that the subject will place in the bottom half. They also constitute evidence in favour of the theory, which requires $p_3 = \frac{1}{2}p_1 + \frac{1}{2}p_2$, and is not rejected by the data.

6 Conclusion

Social scientists are interested in people's beliefs about themselves. One way to elicit these beliefs is simply to ask for them. However, with little at stake, people may provide ready answers that have little connection to their actual beliefs. To counter this possibility, researchers have designed payment schemes that reward people for accurately reporting their beliefs. In particular, a variety of payment schemes have been designed so that people maximize their utility of money by reporting their actual beliefs.

However, these schemes remain vulnerable to distortions, as subjects may care about more than money. Our study joins work by Heath and Tversky (1991), Goodie and Young (2007), Burks et al. (2013), Owens, Grossman, and Fackler (2014), and Ewers and Zimmermann (2015), among others, in determining that non-monetary considerations may lead subjects to overstate their beliefs about themselves under ostensibly incentive compatible mechanisms. In one experiment, using the matching probabilities method, subjects inflate their reported beliefs about themselves by 7% for control reasons; non-monetary considerations account for at least 27% of what would otherwise be estimated to be overconfidence.

Our study differs from earlier ones in that we introduce a new design that eliminates the control bias for self-bets. This design can be used in a variety of contexts where control is a factor. These contexts go well beyond elicitations of beliefs about the self.

7 Appendix A. Using Sample Score as a proxy for confidence

In this section, we discuss the merits of utilizing the variable Sample Score (i.e. the number of correct answers in the sample questions) as a proxy for the unobservable true beliefs, μ. While we do not observe true beliefs, the intuition suggests (and our model predicts) that they correlate with reported beliefs, as captured by the variable Placement. Table 3 shows that Sample Score is indeed a strong predictor of both reported beliefs (used as dependent variable in models J and K) and actual performance (measured by a dummy variable indicating whether a subject placed in the upper half of the score distribution — models L and M). We conclude that Sample Score is a strong signal for future performance and that, conceivably, subjects use it to predict how well they will do in the quiz. Each additional correct answer in the sample questions is associated to an increase of 5 percentage points in subjective placement and of 24 percentage points in the actual chances of belonging to the top half of quiz takers. Interestingly, subjects seem to underestimate the predictive power of the signal they are provided. Yet, the strong correlations observed, allow us to confidently use Sample Score as a proxy for true beliefs in the analysis on heterogenous effects.

8 Appendix B. Instructions for Experiment 1

We present a short version of the instructions for Experiment 1. In particular, we have edited out the portion of the instructions pertaining to another experiment about the Kruger and Dunning (1999) “unskilled and unaware” effect, which was run in conjunction with Experiment 1. We refer to the paper’s online Appendix for the full version of the instructions.

Table 3: Sample Score as a predictor of performance and beliefs

	Placement Model J	Placement Model K	Performance Model L	Performance Model M
Sample Score	5.804***	5.257***	0.240***	0.238***
	(0.000)	(0.001)	(0.000)	(0.000)
Male		4.535*		0.0249
		(0.054)		(0.661)
Treatment 1		4.006*		-0.00248
		(0.084)		(0.965)
Constant	51.66***	44.39***	0.0137	-0.0160
	(0.000)	(0.000)	(0.865)	(0.884)
N	282	282	282	282
R^2	0.0501	0.0728	0.137	0.137

NOTES: The dependent variable is Placement in models J and K, and Performance (a dummy variable taking value 1 if the subject placed in the top half of quiz takers) in models L and M. P-values in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Explanatory comments from the authors are, at times, interspersed among the instructions. They are indicated by use of the italic font and enclosed in square brackets.

Instructions

Welcome! This is an experiment in decision-making. If you follow the instructions and make good decisions you will earn a substantial amount of money. The money you earn will be paid to you in CASH at the end of the experiment. The experiment has three parts, and there is a show-up fee of 5 euro that you will earn regardless of your choices. The entire experiment will take place on computer terminals. Please do not talk or communicate with each other in any way and turn off your phones now.

Preamble: Measuring your beliefs about the likelihood of events

[The preamble explains the betting mechanism for eliciting beliefs in general terms. Part 3 adapts the mechanism to the specific setup of Experiment 1].

In this experiment, you will be taking various trivia and logic quizzes. About half of your earnings will depend on how well you did in these quizzes, while the other half will depend on how accurately you evaluate your own performance. In particular, you will be asked the likelihood of certain events, with questions such as "What are the chances that you gave the correct answer in the question you just answered?" or "What are the chances that you performed better than the median subject?".

Here we explain the procedure that will be used throughout the experiment to reward you for the accuracy of your self-assessment.

As an illustration, suppose that you are asked the following question: *Who is the current Prime Minister of the United Kingdom?* to which you answer *Theresa May.* You are then asked: *What are the chances that the answer you just gave was correct?*.

Your answer to this question will be measured in **chances**, which go from 0 (standing for: I am absolutely sure that I gave the wrong answer) to 100 (standing for: I am absolutely sure that I gave the right answer). So for example:

- 50 means that there are exactly equal chances that you were right or wrong;
- 33.3 (that is, one-third of 100) means that you think you have a 1-over-3 chance to be correct, or, in other words, that you have the same chances to be correct as are the chances to cast a 6-face die and draw a number smaller or equal to 2.
- 75 means that you have the same chances to be correct as are the chances that a white ball is drawn from a bag with 75 white balls and 25 blue balls; and so on.

Review questions:

- What are the chances that you toss a fair coin and you get Tails?
- In a multiple-choice question with 4 options, if you blindly pick one at random, what are the chances that it will be correct?

Incentives: How you are rewarded for reporting your chances accurately

We follow a special procedure to reward you for your self-assessment. This procedure is a bit complicated but the important thing to remember is that it is designed so that it is in your best interest to report your most accurate guess about your real chances. The procedure is as follows.

On the screen, you can visualize a virtual bag. The bag is currently empty and will be filled with 100 blue and white balls. The exact composition of the virtual bag will be determined at the end of the experiment by a random device that will pick one of the following possibilities with equal likelihood: (0 white, 100 blue), (1 white, 99 blue), (2 white, 98 blue) ... (99 white, 1 blue), (100 white, 0 blue). There is a prize that you have the chance to win by either betting on your answer being correct or by betting on a white draw from the virtual bag. Whether you prefer to bet on your answer being correct or on the white draw from the virtual bag depends on how many white balls are in the bag. When there are 0 white balls, you probably prefer to bet on your answer being correct as in most situations you have at least some chances to be correct, no matter how small, while you will never draw a white ball from a bag that contains exclusively blue balls. On the other extreme, when there are 100 white balls, you will probably prefer to bet on the virtual bag rather than on your answer, because it is guaranteed that you will win the prize from a bag with such a composition, whereas a grain of doubt may remain about the correctness of your answer. Somewhere in between 0 and 100 there is a number of white balls that makes you indifferent between betting on the correctness of your answer and betting on the virtual bag. We interpret this number as the chances that your answer is correct. In other words, if you are indifferent between betting on the bag with x white balls and betting on the correctness of your answer, it means you think you have exactly x/100 of having answered correctly.

So, to incentivize you to be truthful, after you report your chances p to be correct in a question, your payment will be determined as follows:

- If, at the end of the experiment, in the virtual bag there are more than p white balls, you will bet on the virtual bag. That is, we will draw a random ball from the bag, and, if the ball is white you win 4 euro, if not you earn 0 euro.

- If instead in the virtual bag there are p white balls or less, you will bet on the correctness of your answer. That is, if your answer is correct you win 4 euro and if it is incorrect you win 0 euro.

Take some time to verify that it is indeed in your best interest to state your chances truthfully. Suppose that you believe you have 70/100 chances that your answer was correct. Then it means that you prefer to bet on your answer being correct, rather than to bet on a white draw from the bag, if in the bag there are fewer than 70 white balls. Viceversa, if in the bag there are more than 70 white balls, you prefer to draw from the bag and hope in a white draw, which has more than 70/100 chance to happen. Being truthful ensures that you always get the better deal between the two options, given your beliefs. We will use this procedure several times throughout the experiment so make sure you understand it, and please feel free to ask any questions.

Part 1

[Edited out because not relevant for Experiment 1].

Part 2: Visual Task

In this part of the experiment, you will perform 12 repetitions of the following exercise. You will see a string of numbers blinking on the screen and will then have to type the numbers into the box appearing on the screen.
The duration of the blinks and the number of elements in the string will vary across periods, hence remembering the string will be easier in some periods and more difficult in others.
You will face two practice rounds and then repeat this exercise 10 times for payment.

Payment:
At the end of the experiment, one round will be selected at random. If in that round you reported the string of numbers correctly, you earn 2 euro, otherwise you earn 0 euro.
Click on the Next button to proceed to the two sample rounds.

Part 3: Logic quiz

In this section, you are asked to answer a logic Quiz. The Quiz consists of 12 multiple-choice questions and you have 6 minutes to answer all the questions.

Self-assessment

Before you take the Quiz, we ask you to estimate how well you will do relative to the other subjects. Specifically, we ask you how likely you think it is that you will do better than half of the participants. Here is how. After the quiz is complete, you will be assigned a ranking according to how many questions you answer correctly. The best performer among you will be assigned to rank 1, the second to rank 2 and so on.

We will then list the participants from the highest rank to the lowest rank and divide the subject pool into two equally sized-groups, an upper half and a lower half. For example, with 30 subjects the top 15 will be ranked in the upper half and the other 15 will be ranked in the lower half. If two people are tied for 15th in terms of performance, then one of them will be randomly placed in the top half and one of them in the lower half.

We want you to tell us your best estimate of the probability that you are in the upper half. Your answer to this question will be measured in **chances**, which go from 0 (standing for: I am absolutely sure that my score will not be in the upper half of the distribution) to 100 (standing for: I am absolutely sure that my score will be in the upper half of the distribution). So for example, 50 means that there are exactly equal chances that you score in the upper or the lower half, and so on.

[The following portion of the instructions is different in the two treatments. Instructions for the two treatments are reported one after the other]

Treatment 1: Payment based on lottery tickets and BDM

Your payment for reporting your chances follows a procedure similar to the one outlined in the preamble, that is, you will either bet on your placement in the upper half or on a white draw from the virtual bag. The only difference is that the prize now is 10 lottery tickets (each one worth a 3% chance of winning 20 euro). The procedure will go as follows. You will report your chance p of being in the upper half and then the computer will randomly determine the number of white balls in the virtual bag. Your payment will be determined as follows:

- If the number of white balls is equal to or smaller than p, then you will be betting on your placement in the upper half. That is, you will receive the 10 lottery tickets (worth in total a 30% chance of winning 20 euro) if your score indeed placed in the upper half of the distribution of scores, and otherwise you will get nothing.

- If instead the number of white balls in the virtual bag is larger than p, then you will bet on the virtual bag. That is, a ball will be drawn from the virtual bag and if it is white you will receive the 10 lottery tickets (again worth in total a 30% chance of winning 20 euro), otherwise you will get nothing.

Treatment 2: Payment based on lottery tickets and VisualTask-BDM

Your payment for reporting your chances follows a procedure similar to the one outlined in the preamble with two differences: (1) the prize for winning is now given by a number of lottery tickets (each one worth a 3% chance of winning 20 euro); and (2) your choice will not be between betting on your placement or betting on the virtual bag, but rather between betting on your placement in the Quiz or betting on your performance in the Visual Task.

The procedure is as follows: You will report your chances **p** of being in the upper half and then the computer will randomly determine the number of white balls in the virtual bag. Then:

- If the number of white balls is equal to or smaller than p, then you will be betting on your placement in the upper half. That is, you will receive the 10 lottery tickets (worth in total a 30% chance of winning 20 euro) if your score indeed placed in the upper half of the distribution, and otherwise you will get nothing.

- If instead the number of white balls in the virtual bag is larger than p, then you will bet on the visual task. That is, a ball will be drawn from the virtual bag and one of the 10 rounds that you completed in the visual task will be extracted at random (with each round having the exact same probability of being selected). If the ball is white and you were successful at the visual task in the extracted round, you will receive **N** lottery tickets (worth in total a **M** percent chance of winning 20 euro) otherwise you receive zero euro. The number of lottery tickets that you can win is calibrated on your performance in the visual task to ensure that it is indeed in your best interest to report the chances of being in the upper half accurately. *[Note: in the experimental screen, **N** and **M** were replaced by personalized values, calibrated for each subject depending on their success rate in the visual task].*

[The remaining instructions are common to both treatments]

Before you state your chances of being in the upper half, you will answer 3 sample questions which are comparable in difficulty to the questions that you will find in the Quiz. There is no payment for the sample questions. You are now ready to start the sample questions. Please click on the Next button now.

[The following is the message visualized on the screen after the subjects complete the sample questions]
What are your chances to be in the upper half of the scores' distribution? Type a number between **0** (meaning: I have zero chance to be in the upper half) to **100** (meaning: I am absolutely sure I will be in the upper half of score distribution).

Note:

- The sample questions you just saw are of comparable difficulty to the actual questions you will encounter in the Quiz.
- In past experiments, the better performing half of the subjects answered 7 or more questions correctly, out of a total of 12 questions.

9 Appendix C. Instructions for Experiment 2

We present instructions for Experiment 2. Explanatory comments from the authors are, at times, interspersed among the instructions. They are indicated by use of the italic font and enclosed in square brackets.

Instructions

This is an experiment in decision making. Funds have been provided to run this experiment. If you follow the instructions and make good decisions, you will earn a substantial amount of money. The money you earn will be paid to you in CASH at the end of the experiment. The session will take place through computer terminals. There is a show-up fee of 10 euro that you will earn regardless of your choices. The experiment will consist of two parts. At the end of the experiment, a random device will determine whether you are going to be paid according to your answers in the first part or in the second part of the experiment, with a 50% chance that each part is used for payment.

Please turn off your phones now and do not talk or communicate to each other in any way.

First part

In the first part of the experiment, you are asked to answer a logic quiz. The quiz consists of 20 multiple-choice questions and you have 13 minutes to answer the questions. You will earn 50 cents for each correct answer and zero cents for each incorrect answer. Hence, if this first part of the experiment is randomly drawn and used for payment, you can earn from a minimum of 10 euro to a maximum of 20 euro including the show-up fee.

[The second part is presented separately for each of the 3 treatments].

Second part (Treatment 1 - Betting up)

In this second part of the experiment, we ask you to estimate how well you did in the quiz relative to the other subjects. Of course, you cannot know your relative performance for sure so we will ask you for a probability estimate. Specifically, we will ask you with which probability you think you placed in the upper half of subjects.

You will be assigned a ranking based on how many questions you answered correctly in the quiz you just took. The best performer among you will be assigned to rank 1, the second best performer to rank 2 and so on. We will then list the participants in the experiment from the highest rank to the lowest rank and divide the subject pool into two equally sized-groups, an upper half and a lower half. For example, with 14 subjects the top 7 will be ranked in the upper half and the other seven will be ranked in the lower half. If, say, two people are tied for 7^{th} in terms of performance, then one of them will be randomly placed in the upper half and one of them in the lower half.

We want you to tell us your best estimate of the probability that you are in the upper half. For this purpose, we will use a special payment procedure that rewards you for giving us your best estimate. The procedure is a bit complicated but the most important thing to understand about it is simply that you maximize your expected payment by reporting your best estimate. We now explain this procedure.

At the end of the experiment, the computer will create a virtual bag. The bag will be filled with 100 blue and white balls. The exact composition of the virtual bag will be determined at the end of the experiment by a random device that will pick one of the following possibilities with equal likelihood: (0 white, 100 blue), (2 white, 98 blue), (4 white, 96 blue) ... (98 white, 2 blue), (100 white, 0 blue) - so the virtual bag will have one among all possible combinations of white and blue balls with increments of two.

There is a prize of 10 euro that you have a chance to win by either betting on your placement or by betting on the virtual bag. For each of the possible combinations, we want to know if you prefer to bet on your placement or to bet on a white draw from the virtual bag. Choices will be presented to you in a list of pairwise comparisons, as shown in Figure 1.

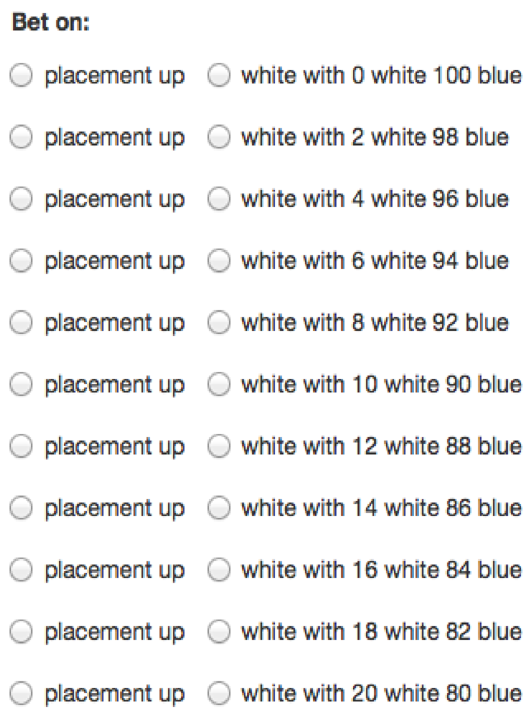


Figure 1. Choices

In each comparison you choose between betting on your placement-up or on the virtual bag:

- If you bet on your **placement-up,** you win 10 euro if you are in the upper half of the ranking and 0 euro otherwise.
- If you bet on the **virtual bag**, you win 10 euro if a white ball is drawn from the virtual bag and 0 euro otherwise.

Thresholds: The number of white balls represents your chances of winning when you bet on the virtual bag. The number of white balls increases as you scroll down the list, so the virtual bag becomes more attractive the more down you go on the list. Hence we expect that, if you choose the virtual bag in one comparison, you will choose the virtual bag in all comparisons that follow below it. In other words, we expect that you will have a **threshold**, that is, a certain amount of white balls such that you bet on your placement-up until that threshold and then switch to bet on the virtual bag if it contains more white balls than the threshold. We will interpret this threshold as the probability that you believe your score falls in the upper half of the distribution.

You can try out different thresholds and your choice will be final only when you click on the Next button. Remember, once again, that you maximize your chances of winning if your threshold is the probability that you assign to having a quiz score in the upper half of the distribution.
At the end of the experiment, a random device will select one of the questions, that is, one of the possible bag compositions. Then one ball will be extracted from the virtual bag. Your payment will depend on the color of the ball and your choice in the selected question. To recap, if, in the selected question:

- You bet on the virtual bag, then you win 10 euro if a white ball is randomly extracted from the bag;
- You chose to bet on you placement-up, then you win 10 euro if you placed in the upper half.

Examples: Lisa thinks there is a 60% chance she placed in the upper half. Hence, she chooses to bet on her placement-up if in the bag there are 60 white balls or fewer and on the virtual bag if it contains more than 60 white balls. She, therefore, clicks all the buttons according to this rule and her choices will look as in Figure 2:

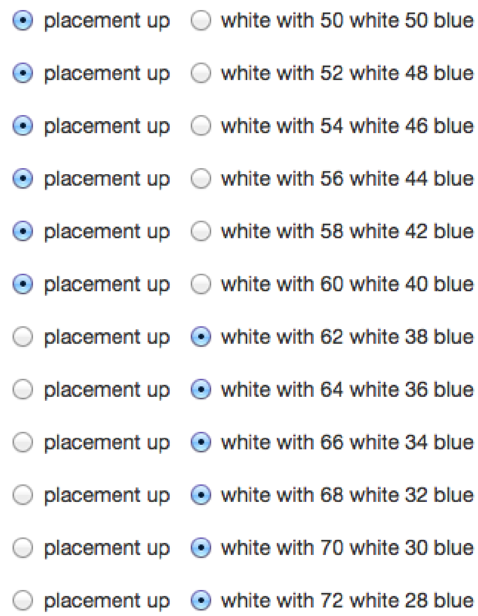


Figure 2. Lisa's Choices

John thinks there is a 20% chance he placed in the upper half. Hence, he chooses to bet on his placement-up if there are 20 white balls or fewer in the virtual bag, otherwise he prefers to bet on the virtual bag. He clicks the buttons according to this threshold and his choices will look as in Figure 3.

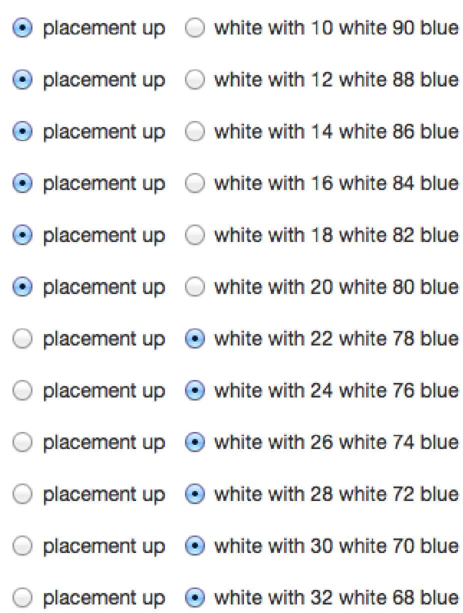


Figure 3. John's Choices

Second part (Treatment 2 - Betting down)

In this second part of the experiment, we ask you to estimate how well you did in the quiz relative to the other subjects. Of course, you cannot know your relative performance for sure so we will ask you for a probability estimate. Specifically, we will ask you with which probability you think you placed in the lower half of subjects.

You will be assigned a ranking based on how many questions you answered correctly in the quiz you just took. The best performer among you will be assigned to rank 1, the second-best performer to rank 2 and so on. We will then list the participants in the experiment from the highest rank to the lowest rank and divide the subject pool into two equally sized-groups, an upper half and a lower half. For example, with 14 subjects, the top 7 will be ranked in the upper half and the other seven will be ranked in the lower half. If, say, two people are tied for 7^{th} in terms of performance, then one of them will be randomly placed in the upper half and one of them in the lower half.

We want you to tell us your best estimate of the probability that you are in the lower half. For this purpose, we will use a special payment procedure that rewards you for giving us your best estimate. The procedure is a bit complicated but the most important thing to understand about it is simply that you maximize your expected payment by reporting your best estimate. We now explain this procedure.

At the end of the experiment the computer will create a virtual bag. The bag will be filled with 100 blue and white balls. The exact composition of the virtual bag will be determined at the end of the experiment by a random device that will pick one of the following possibilities with equal likelihood: (0 white, 100 blue), (2 white, 98 blue), (4 white, 96 blue) ... (98 white, 2 blue), (100 white, 0 blue) - so the virtual bag will have one among all possible combinations of white and blue balls with increments of two.

There is a prize of 10 euro that you have a chance to win by either betting on your placement or by betting on the virtual bag. For each of the possible combinations, we want to know if you prefer to bet on your placement or to bet on a white draw from the virtual bag. Choices will be presented to you in a list of pairwise comparisons, as shown in Figure 1.

Figure 1. Choices

In each comparison you choose between betting on your placement-down or on the virtual bag:

- If you bet on your **placement-down,** you win 10 euro if you are in the lower half of the ranking and 0 euro otherwise.
- If you bet on the **virtual bag**, you win 10 euro if a white ball is drawn from the virtual bag and 0 euro otherwise.

Thresholds: The number of white balls represents your chances of winning when you bet on the virtual bag. The number of white balls increases as you scroll down the list, so the virtual bag becomes more attractive the more down you go on the list. Hence we expect that, if you choose the virtual bag in one comparison, you will choose the virtual bag in all comparisons that follow below it. In other words, we expect that you will have a **threshold**, that is, a certain amount of white balls such that you bet on your placement-down until that threshold and then switch to bet on the virtual bag if it contains more white balls than the threshold. We will interpret this threshold as the probability that you believe your score falls in the lower half of the distribution.

You can try out different thresholds and your choice will be final only when you click on the Next button. Remember, once again, that you maximize your chances of winning if your threshold is the probability that you assign to having a quiz score in the lower half of the distribution.

At the end of the experiment, a random device will select one of the questions, i.e. one of the possible bag compositions. Then one ball will be extracted from the virtual bag. Your payment will depend on the color of the ball and your choice in the selected question. To recap, if, in the selected question:

- You bet on the virtual bag, you win 10 euro if a white ball is randomly extracted from the bag;
- You chose to bet on you placement-down, you win 10 euro if you placed in the lower half.

Examples: Lisa thinks there is a 60% chance she placed in the lower half. Hence, she chooses to bet on her placement-down if in the bag there are 60 white balls or fewer and on the virtual bag if it contains more than 60 white balls. She, therefore, clicks all the buttons according to this rule and her choices will look as in Figure 2:

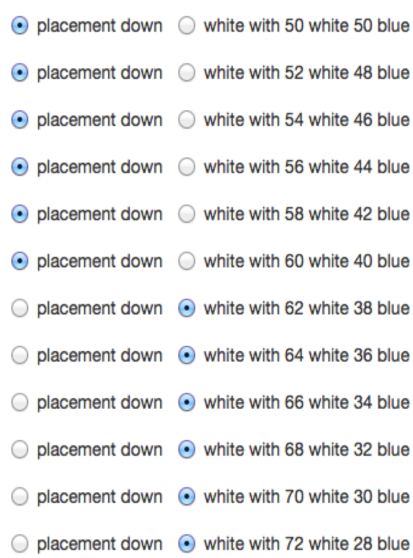


Figure 2. Lisa's Choices

John thinks there is a 20% chance he placed in the lower half. Hence, he chooses to bet on his placement-down if there are fewer than 20 white balls in the virtual bag, otherwise he prefers to bet on the virtual bag. He clicks the buttons according to this threshold and his choices will look as in Figure 3.

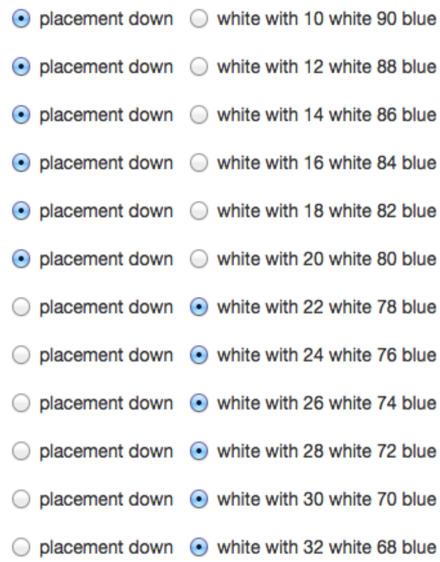


Figure 3. John's Choices

Second part (Treatment 3 - Betting up and down)

In this second part of the experiment, we ask you to estimate how well you did on the quiz relative to the other subjects. Of course, you cannot know your relative performance for sure so we will ask you for a probability estimate. Specifically, we will ask you with which probability you think you placed in the lower half of subjects.[23]

You will be assigned a ranking based on how many questions you answered correctly on the quiz you just took. The best performer among you will be assigned to rank 1, the second best performer to rank 2 and so on. If there are ties, these ties will be broken randomly, so that everyone is assigned a unique rank.

We will then list the participants in the experiment from the highest rank to the lowest rank and divide the subject pool into two equally sized-groups, an upper half and a lower half. For example, with 14 subjects the top 7 will be ranked in the upper half and the other seven will be ranked in the lower half. If, say, two people are tied for 7^{th} in terms of performance, then one of them will be randomly placed in the upper half and one of them in the lower half.

We want you to tell us your best estimate of the probability that you are in the lower

[23] *[Note: In this treatment, subjects bet on both their performance being in the upper part and in the lower part of the distribution. In 2 (out of 4) sessions, the framing of the instructions starts off with betting-down and later introduces betting-up, in the other two treatments the order in which the two types of bets are presented is reversed].*

half. For this purpose, we will use a special payment procedure that rewards you for giving us your best estimate. The procedure is a bit complicated but the most important thing to understand about it is simply that you maximize your expected payment by reporting your best estimate. We now explain this procedure.

At the end of the experiment, the computer will create a virtual bag. The bag will be filled with 100 blue and white balls. The exact composition of the virtual bag will be determined at the end of the experiment by a random device that will pick one of the following possibilities with equal likelihood: (0 white, 100 blue), (2 white, 98 blue), (4 white, 96 blue) ... (98 white, 2 blue), (100 white, 0 blue) - so the virtual bag will have one among all possible combinations of white and blue balls with increments of two.

There is a prize of 10 euro that you have a chance to win by either betting on your placement or by betting on the virtual bag. For each of the possible combinations, we want to know if you prefer to bet on your placement or to bet on a white draw from the virtual bag. Choices will be presented to you in two groups of pairwise comparisons, as shown in Figure 1.

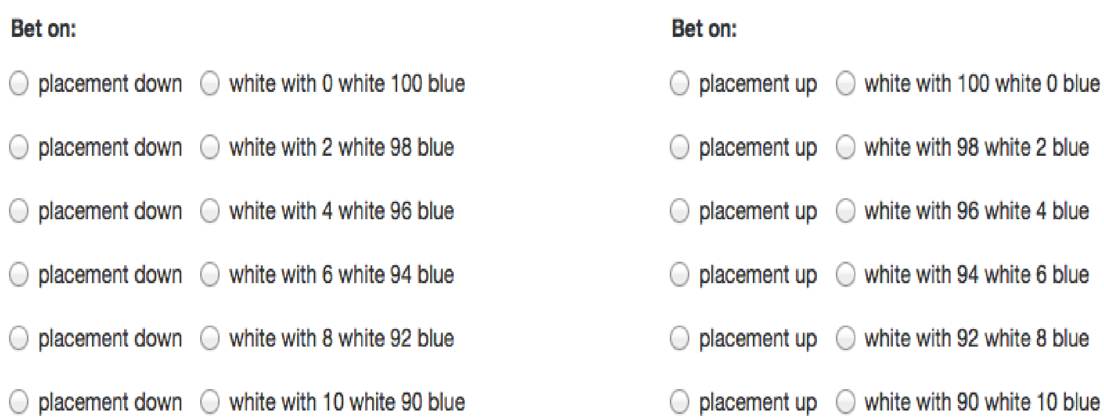


Figure 1. Choices

In the column on the left, you choose between betting on your placement-down or on the virtual bag:

- If you bet on your **placement-down,** you win 10 euro if you are in the lower half of the ranking and 0 euro otherwise.
- If you bet on the **virtual bag**, you win 10 euro if a white ball is drawn from the virtual bag and 0 euro otherwise.

In the column on the right, you choose between betting on your placement-up or on the virtual bag:

- If you bet on your **placement-up,** you win 10 euro if you are in the upper half of the ranking and 0 euro otherwise.
- If you bet on the **virtual bag,** you win 10 euro if a white ball is drawn from the virtual bag and 0 euro otherwise.

[Note: In 2 (out of 4) sessions of treatment 3, the order of the columns was reversed and the instructions were adjusted accordingly. As a result, subjects would bet on their placement-up in the column on the left, and on their placement-down in the column on the right.]

Thresholds: The number of white balls represents your chances of winning when you bet on the virtual bag. In the left column, the number of white balls increases as you scroll down the list, so the virtual bag becomes more attractive the more down you go on the list. Hence we expect that, if you choose the virtual bag in one comparison, you will choose the virtual bag in all comparisons that follow below it. In other words, we expect that you will have a **threshold**, that is, a certain amount of white balls such that you bet on your placement-down until that threshold and then switch to bet on the virtual bag if it contains more white balls than the threshold. We will interpret this threshold as the probability that you believe your score falls in the lower half of the distribution.

Your choices from the right column. In the right column, you are choosing between betting on your placement-up or the virtual bag. Here the order of the virtual bags is reversed: The number of white balls starts at 100 and decreases as you scroll down the list. Here again, you will have a threshold: You will start betting on the virtual bag and then switch at some point to betting on your placement-up. This threshold will tell us the probability with which you believe your score belongs to the upper half of the distribution.

Admissible choices: The choices from the two columns are tied together, that is, **the two thresholds will have to be placed on the same line**. The reason is that if you told us that there is an $x\%$ chance that your rank is in the lower half, we will presume you think there is a $100 - x\%$ chance that your score is in the upper half. In Figure 5, you can see a preview of what it means for the two thresholds to be placed on the same line. We'll go back to it at the end.

A way to ensure you are meeting this constraint is to verify that, taking two questions placed on the same line, you are betting on the virtual bag in one and only one of them. Figure 2 shows two examples of non-admissible choices. Figure 3 shows two examples of admissible choices. If you make a mistake, an error message will prompt you to correct your entries until only admissible choices are present.

Figure 2. Non-admissible choices

Figure 3. Admissible choices

You can try out different thresholds and your choice will be final only when you click on the Next button. Remember, once again, that you maximize your chances of winning if, in the left column, your threshold is the probability that you assign to having a quiz score in the lower half of the distribution, and, in the right column, you pick as threshold the probability that your score is in the upper half.
At the end of the experiment, a random device will select one of the two groups of questions and one of the possible bag compositions. Then one ball will be extracted from the virtual bag. Your payment will depend on the color of the ball and your choice in the selected question. To recap, if in the selected question:

- You bet on the virtual bag, you win 10 euro if a white ball is randomly extracted from the bag;
- You chose to bet on your placement-down, you win 10 euro if you placed in the lower half;
- You chose to bet on your placement-up, you win 10 euro if you placed in the upper half.

Examples: Lisa thinks there is a 60% chance she placed in the lower half and a 40% chance she placed in the upper half. Hence, she chooses to bet on her placement-down if in the bag there are 60 white balls or fewer. Moreover, she chooses to bet on her placement-up if there are fewer than 40 white balls in the bag. She therefore clicks all the buttons according to this rule and her choices will look as in Figure 4:

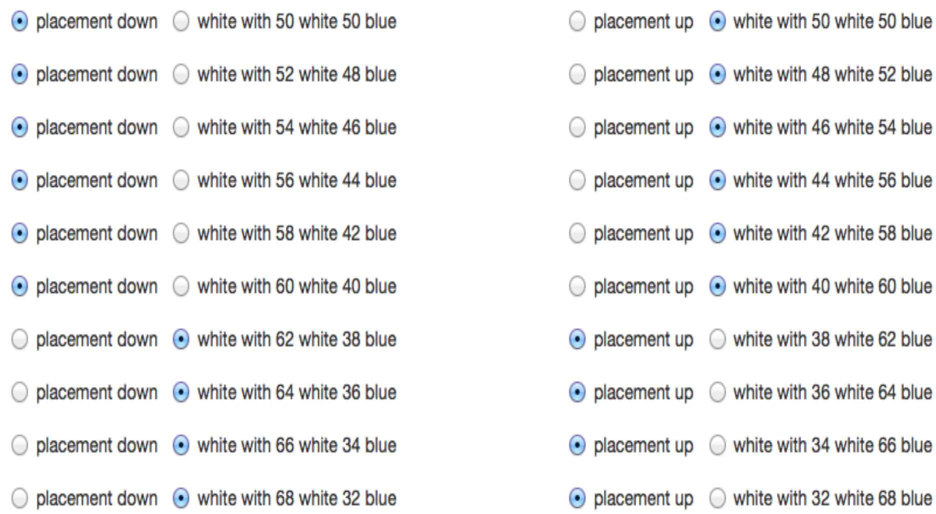


Figure 4. Lisa's Choices

John thinks there is a 20% chance he placed in the lower half. Hence, he chooses to bet on his placement-down rather than on the virtual bag if there are fewer than 20 white balls in the virtual bag, otherwise he prefers to bet on the bag. He clicks the buttons according to this threshold and his choices will look as in Figure 5. This should be consistent with his belief that there is an 80% probability that he scored in the upper half.

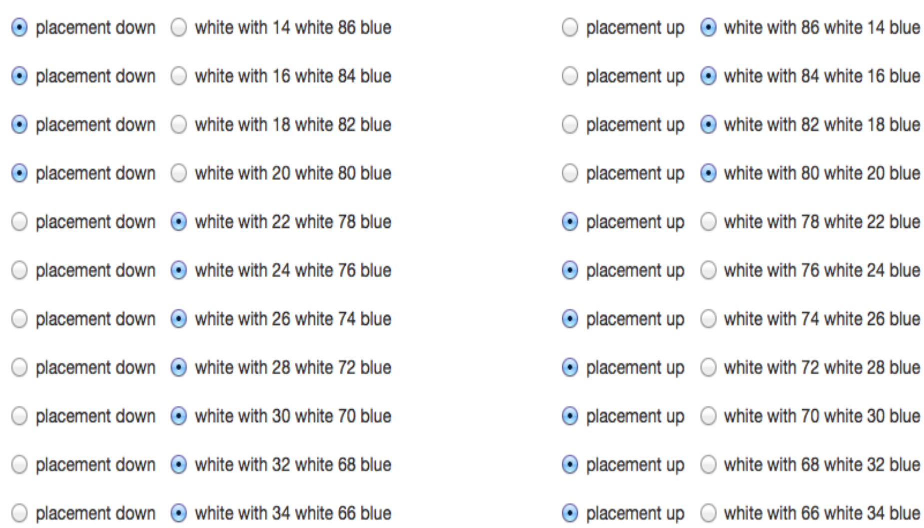


Figure 5. John's Choices

References

[1] Allen, F, (1987), “Discovering Personal Probability when Utility Functions Are Unknown,” *Management Science*, 33(4), pp. 542-544.

[2] Andreoni, J. and A. Sanchez (2014), “Do Beliefs Justify Actions or Do Actions Justify Beliefs? An Experiment on Stated Beliefs, Revealed Beliefs, and Social-Image Motivation” mimeo.

[3] Barber, B. and T. Odean (2001), “Boys will be Boys: Gender, Overconfidence, and Common Stock Investment,” Quarterly Journal of Economics, 116(1), 261-92.

[4] Baumeister, R.F. (1982), “A self-presentational view of social phenomena,” *Psychological Bulletin* 91:3–26.

[5] Benoît, J-P. and J. Dubra (2011), “Apparent Overconfidence” *Econometrica* **79(5)**, 1591-1625.

[6] Benoît, J.-P., Dubra, J. and Moore, D. A. (2015), Does the Better-than-average Effect show that People are Overconfident?: Two Experiments. Journal of the European Economic Association, 13: 293â€“329.

[7] Burks, Stephen, Jeffrey Carpenter, Lorenz Goette and Aldo Rustichini (2013), “Overconfidence and Social Signalling,” Review of Economic Studies, 80(3), 949-83.

[8] Camerer, C., and Lovallo, D. (1999), “Overconfidence and Excess Entry: An Experimental Approach,” The American Economic Review 89.1 : 306â€“318.

[9] Chen, D. L., Schonger, M. and Wickens, C. (2016), “oTree - An open-source platform for laboratory, online, and field experiments,” J. Behav. Exp. Finance 9, 88-97.

[10] Chew S. H., and J. Sagi, (2008), “Small worlds: Modeling attitudes toward sources of uncertainty,” *Journal of Economic Theory*, 139(1), pp. 1-24.

[11] Clark, Jeremy and Lana Friesen (2009), “Overconfidence in Forecasts of Own Performance: An Experimental Study,” Economic Journal, 119(1), 229-51.

[12] Ducharme, W. and Donnell, M. (1973), “Intrasubject Comparison of Four Response Modes for Subjective Probability Assessment,” *Organizational Behavior and Human Performance* 10, 108-117.

[13] Ewers, M. and F. Zimmermann (2015), "Image and Misreporting," Journal of the European Economic Association, 13(2), 363–80.

[14] Goodie, A. S. (2003) The effects of control on betting: paradoxical betting on items of high confidence with low value. J Exp Psychol Learn Mem Cogn: 29, 598-610.

[15] Goodie, Adam and Diana Young (2007), "The skill element in decision making under uncertainty: Control or competence?," Judgment and Decision Making, 2(3), pp. 189-203.

[16] Grether, D. M. (1981) "Financial Incentive Effects and Individual Decision Making," Social Sciences working paper 401, Cal. Tech.

[17] Grether, D. M. (1992) "Testing Bayes Rule and the Representativeness Heuristic: Some Experimental Evidence." *Journal of Economic Behavior and Organization* **17**, 31–57.

[18] Gul, R and W. Pesendorfer (2015) "Hurwicz Expected Utility and Multiple Sources," Journal of Economic Theory.

[19] Heath, Chip and Amos Tversky, (1991) "Preference and Belief: Ambiguity and Competence in Choice under Uncertainty," Journal of Risk and Uncertainty, 4, 5-28.

[20] Hoelzl, E. and A. Rustichini, (2005), "Overconfident: do you put your money on it?" the *Economic Journal*, **115**, pp. 305-18.

[21] Hollard, G., Massoni, S., & Vergnaud, J.-C. (2016), "In Search of Good Probability Assessors: An Experimental Comparison of Elicitation Rules for Confidence Judgments," Theory and Decision. 80(3): 363-387.

[22] Hossain, T. and R. Okui (2013), "The Binarized Scoring Rule," Review of Economic Studies 80, 984-1001.

[23] Karni, E. (2009), "A Mechanism for Eliciting Probabilities," *Econometrica*, **77(2)**, 603–6.

[24] Kim, S-H. and W. Whitt (2015), "The power of alternative Kolmogorov-Smirnov tests based on transformations of the data," *ACM Trans. Model. Comput. Simul.* 25(4), article 24.

[25] Klein, WMP, J.L. Cerully and M.M. Monin (2010), "Ability, chance, and ambiguity aversion: Revisiting the competence hypothesis," *Judgment and Decision Making*, 5(3), 192–9.

[26] Klibanoff, P., M. Marinacci and S. Mukerji (2005), "A smooth model of decision making under ambiguity," *Econometrica* 73(6), pp. 1849-92.

[27] Kruger, Justin and David Dunning (1999), "Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments," *Journal of Personality and Social Psychology*, **77**, 1121–34.

[28] Li, K.K. (2011), "Preference towards control in risk taking: Control, no control, or randomize?" *Journal of Risk and Uncertainty* (43), pp. 39-63.

[29] McKelvey, R. and T. Page (1990), "Public and Private Information: An Experimental Study of Information Pooling," *Econometrica* 58(6), pp. 1321-39.

[30] Merkle, Christoph and Martin Weber (2011), "True Overconfidence: The Inability of Rational Information Processing to Account for Apparent Overconfidence," Organizational Behavior and Human Decision Processes, 116(2), 262-71.

[31] Möbius, M., M. Niederle, P. Niehaus, and T. Rosenblat (2014), "Managing self-confidence: Theory and experimental evidence," Working Papers, No. 11-14, Federal Reserve Bank of Boston.

[32] Niederle, M. and L. Vesterlund (2007), "Do Women Shy Away from Competition? Do Men Compete Too Much?" Qarterly Journal of Economics, 122(3), 1067-1101.

[33] Owens, D., Z. Grossman and R. Fackler (2014), "The Control Premium: A Preference for Payoff Autonomy," American Economic Journal: Microeconomics, 6(4), 138-61.

[34] Razali, N.M. and Y.B. Wah (2011), "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests," *Journal of Statistical Modeling and Analytics*, 2(1), 21-33.

[35] Ryan, R. M., V. Mims and R. Koestner (1983), "Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory," *Journal of Personality and Social Psychology*, 45(4), 736.

[36] Schlag, K., J. Tremewanz, J. van der Weele (2015), "A penny for your thoughts: A survey of methods for eliciting beliefs," *Experimental Economics*, 18:3, 457-490.

[37] Schlag, K. H. and J. van der Weele (2009), "Eliciting Probabilities, Means, Medians, Variances and Covariances without Assuming risk-neutrality", (Working Paper, Universitat Pompeu Fabra, Barcelona).

[38] Schotter, A. and I. Trevino (2014), "Belief elicitation in the laboratory," *Annual Review of Economics* (6), 103-128.

[39] Schwardmann, P. and J. van der Weele (2019), "Deception and Self-Deception," Nature Human Behaviour, July, pp 1-7.

[40] Smith, C, (1961), "Consistency in Statistical Inference and Decision," *Journal of the Royal Statistical Society Series B.* 23(1), pp. 1-37.

[41] Tversky, A. and P. Wakker (1995), "Risk Attitudes and Decision Weights, " *Econometrica* 63(6), 1255-80.

Control and information of the unskilled in the study of Overconfidence (#8829)

Author(s)
Juan Dubra (Universidad de Montevideo) - dubraj@um.edu.uy
Jean-Pierre Benoît (London Business School) - jpbenoit@london.edu
Giorgia Romagnoli (University of Amsterdam) - g.romagnoli@uva.nl

Created: 03/04/2018 02:41 PM (PT)
Public: 03/04/2018 02:59 PM (PT)

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
The study has two goals: a) Establish to what extent "control" (the desire to bet on activities where one could in principle affect the outcome) affects the self-reported estimated probabilities of success, when these are elicited with the probability matching rule of Karni (2009) and Grether (1981) (we replicate the mechanism of Benoît, Dubra and Moore (2015) but the prize will be lottery tickets). b) Establish whether individuals who are incompetent in a task are comparatively more unaware of their (in)competence than their more skilled counterparts (see Kruger and Dunning). We intend to test the KD effect against an alternative model of regression to the mean paired with general overconfidence, which can generate the same empirical patterns ascribed, thus far, to the KD effect.

3) Describe the key dependent variable(s) specifying how they will be measured.
) a) The dependent variable is the self-reported belief of the likelihood that the subject performs better than half of the other subjects in a quiz. We will compare two measures: A control measure of overconfidence elicited with the probability matching rule (with payment in lottery tickets); and a treatment measure of overconfidence elicited with a modified version of the probability matching rule where the outcome is always dependent on the performance of the subject (either in the main task or in a secondary visual/memory task performed in the computer). b) The key variables utilized in the structural estimation are the self-reported prior and posterior probabilities of answering the trivia questions correctly and of performing better than the median subject. These measures will be elicited with the probability matching rule.

4) How many and which conditions will participants be assigned to?
a) We will have two treatments (with around 150 participants in each) corresponding to the two different ways of measuring beliefs as explained in section 3.
b) For the estimation of beliefs, there will be one treatment of about 300 subjects.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
a) We will perform a test of difference of means: we will have elicited beliefs with two methods, and we will compare whether the means are significantly different.
b) We will perform two analyses:
1) We will postulate a single signaling structure for all participants (as in Benoit and Dubra 2011), and estimate the parameters with by Maximum Likelihood. Then we will estimate (by MLE) one signaling structure for the skilled and one for the unskilled and perform a Neyman-Pearson test comparing the two models.
2) In the first part of the study we will elicit a prior and a posterior probability of answering each question correctly. With this information, we will estimate the parameters of the signaling structure. We will estimate one signaling structure for each group (skilled and unskilled) and test whether the estimated parameters are statistically different. We will also test whether, for all subjects, the predicted posterior probabilities after a correct answer are more accurate than after an incorrect answer.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
Only those who choose not to complete the experiment will be excluded.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
Around 300 subjects will participate, depending on how experimental sessions are filled.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)