Belief elicitation when more than money matters[*]

Jean-Pierre Benoît
London Business School

Juan Dubra
Universidad de Montevideo

Giorgia Romagnoli
University of Amsterdam

March, 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 skill 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 find that under the ostensibly incentive compatible probability matching mechanism (Grether (1981) and Karni (2009)), subjects report beliefs at least 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.

Keywords: Elicitation, Overconfidence, Control. Experimental Methods.

Journal of Economic Literature Classification Numbers: D3

As economists have come to embrace the experimental paradigm long found in psychology and other disciplines, they have emphasized the benefits of incentivizing subjects. This incentivization is intended both to ensure that subjects participate in a meaningful way and to guide experimenters in their interpretations of subjects' actions. Typical incentive

[*] Contact: jpbenoit@london.edu, dubraj@um.edu.uy, G.Romagnoli@uva.nl. We thank Michael Jetter and Madan Pillutla 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.

protocols in use rely on monetary payments and an (often implicit) assumption that money is the only argument in subjects' utility functions. Thus, an incentive compatible mechanism for eliciting subjects' 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 therefore important to have an idea of the magnitude of possible distortions. Can they safely be 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. Nonetheless, the amount we can measure is not large enough to overwhelm typical findings **in the context where we apply our new elicitation mechanism**. **For example, when the method of Grether (1981) and Karni (2009) is measured to use beliefs, at least 26.8% of what was previously interpreted as overconfidence in fact seems to come from the desire for control.**

Consider the numerous experiments that derive subjects' beliefs about themselves by presenting them with the opportunity to win a prize, either based on their performance on a task or from a random draw. In one format, subjects choose between a bet that yields a 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 a subject who chooses to bet on her performance believes she has a probability of at least x of placing in the top half. In another format, subjects are asked for the probability x that their performance will place them in the top half. The answer x determines, in an incentive compatible manner, the probability the subject will earn a prize based on her performance rather than from a random draw (for example, Hollard Massoni Vergnaud (2010), Andreoni and Sanchez (2014), Benoît, Dubra and Moore. (2015), and Möbius, Niederle, Niehaus and Rosenblat (2014)). The experimenter concludes that the subject believes she has a probability exactly x of placing in the top half.

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

1. 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; this may lead

them to favour payments based on their performance over payments determined by probabilistically equivalent random devices.

2. 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 exagerrate their odds of doing well.

We discuss some of the literature on these non-monetary considerations below.

The presence of non-monetary considerations 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 a subject may choose to bet on her performance even if she thinks the probabilities do not warrant it from a monetary perspective. For instance, a subject who thinks she has only a 60% chance of placing in the top half 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. However, by comparing the choices subjects make under different conditions, we isolate and obtain a measure of the bias for betting on oneself. The bias for positive statements remains unmeasured. **I would delete this last sentence. at this point what we do is not clear, and you only say that we measure the total, and that should be enough for now.**

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 probability matching method of Grether (1981, 1992) and Karni (2009), and replicates previous literature. Under the second mechanism, subjects effectively choose between betting on themselves on one task and betting on themselves on another task. This mechanism involves a novel design that eliminates the control bias: no matter how subjects choose, they can only be paid for a successful performance. Both mechanisms are incentive compatible in monetary terms.

The implicit assumption in most existing literature is that the differences in the designs of the two mechanisms should make no difference in the elicited beliefs, as only the probabilities of winning the money matters. Nevertheless, we find evidence for a significant control effect

– with the elicitation method that duplicates the prior literature, subjects inflate their beliefs by at least 7% in order to shift weight towards bets on themselves (at the cost of reducing the overall chance of winning money). The experiment is run in the context of research on overconfidence and at least 27% of what would otherwise be measured as overconfident beliefs is shown to instead be a willful inflation.

A different approach to eliciting beliefs has been proposed by Hossain and Okui (2013) and Schlag and van der Weele (2013). They 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, this mechanism is subject to self-regard and signalling caveats, as the creators of the rule realise.

Less obviously, the binarized scoring rule is also vulnerable to control issues, as we show in Section 4. Analysis of the 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 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.[1] We find that the control motivation manifests itself only as a desire for betting on doing well; payment for doing badly is perceived as a negative and of the same magnitude as the positive desire for bets on success.

While our study is carried out within the overconfidence paradigm, its applicability is general (see surveys by Schlag, Tremewanz and van der Weele (2015) and Schotter and Trevino (2014) on incentive compatible elicitation). **Given the importance of control, and the difficulty it introduces in interpreting subjects' choices, Owens et al. (2014) have studied issues similar to those we analyze; we discuss their findings in detail in Section 3.3.** The experimental design 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

[1] Subjects are remunerated for each quiz question they answer correctly and are not forewarned of future bets, so their incentive is to do well on the quiz.

be adapted to a variety of mechanisms, including state-of-the-art schemes like the binarized scoring rule.

1 Overstatement

In this section, we discuss some of the literature in economics and psychology on non-monetary concerns which may lead subjects to misreport their beliefs, either implicitly by their actions or explicitly by their words.

1.1 Betting on Yourself: Control

Several studies conclude that people prefer bets on themselves to bets on probabilistically equivalent random devices. Goodie (2003), Goodie and Young (2007), and Heath and Tversky (1991, experiments 1, 2, and 3) use the following methodology. Subjects answer a series of multiple choice questions and, for each question, indicate the likelihood that their answer is correct. The reporting of the likelihoods is unincentivized and subjects do not realize how their 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 according to p_i. In the second group, each subject picks either (a) a bet that pays off with an objective probability p_i or (b) the certainty-equivalent. Subjects in the first group choose the bet (a) over the certain payment (b) 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 themselves and on a random device are not a simple reflection of the probabilities involved. Rather, subjects tend to display a bias towards self-bets, especially when they have done well.[2] For the most part,

[2] It is not entirely clear from the experiments whether subjects always have a preference for betting on themselves or only have this preference when they have done well. To keep our modeling simple – and since our subjects perform well on both our tasks – we will assume they always have the preference. In any case, we test whether, on average, this preference exists, and we find that it does.

this bias is more pronounced when subjects are more confident in their answers and when the questions are drawn from a single category. Heath and Tversky argue that people have a preference for betting on themselves in domains in which they are competent, while Goodie and Young dispute this interpretation and maintain that people have a general preference for control.[3] This distinction 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. As Goodie describes it, for control to be in play the nature of the task should be such that "a particpant could take steps to favorably alter the success rate in subsequent administrations of the task."

While the findings of these papers are revealing, their methods do not permit a measurement of the value of control or of the amount by which control leads people to overstate their beliefs.[4] Our experiment allows us to estimate both.

Owens, Grossman, and Fackler (2014) contrast betting on one's own performance with betting on someone else's. Subjects report their beliefs that they will answer a question correctly and their beliefs that a randomly matched subject will answer a different question correctly. Subjects also 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 their results is clouded by the fact that the mechanism used for eliciting subjects' beliefs is itself prone to control issues. We discuss this experiment in some 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. Gul and Pesendorfer (2015) develops a formal model of preferences that incorporates source dependence.

[3] Interestingly enough, and as the authors note, the results in these and related papers go counter to findings of ambiguity aversion in other contexts. Klein et al. (2010) explores the relation between ambiguity, controllability and competence.

[4] Subjects' in the experiments were overconfident 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. It is impossible to tell to what extent this loss reflected overconfidence and to what extent a sacrifice for non-monetary objectives.

1.2 Saying Nice Things: Self-regard and Signalling

People like to say nice things about themselves, both out of self-regard and because sending 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 lead subjects to overstate their beliefs. Ewers and Zimmerman (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.

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.

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 the experiments that follow. In Section 3.2, we discuss conclusions that are independent of the specific modelling we adopt.

Consider an experiment where a subject performs a task in which her performance is described by a variable $\theta \in \left\{\theta^L, \theta^H\right\}$, 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 has performed well, $\theta = \theta^H$, and she is asked for a report p of this belief. She can earn an amount of money m, depending on her performance, 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 for money that is obtained from her performance, rather than through a random device. More precisely, beyond the utility of the money itself, she gets an extra utility benefit of 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.[5] If 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 doing well, other times for having done poorly, so that in general the expected utility gain is $c_H q_H + c_L q_L$. Perhaps the most natural reading of the literature is that a subject derives extra utility 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 has belief μ that $\theta = \theta^H$ and reports p, gets an extra utility kick of $n(p, \mu)$ from the report, where the partial derivatives satisfy $n_1 \geq 0$ and $n_{11} \leq 0$. (Additional plausible assumptions could include $n(\mu, \mu) = 0$ and $n_{12} < 0$ – more skilled participants have less reason to inflate their reports.) 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

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

Consider, for a moment, an individual who 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 $\mu(w+m)+(1-\mu)(w)+\mu c_H$. The expected control benefit is μc_H, which is increasing in the probability of a correct answer. 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 conversely.

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

With this modelling, people with larger self-beliefs have a greater control bias towards bets on themselves. This feature is consistent with the experimental findings discussed 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 evidence from our experiment that people with a greater self-belief inflate more for control reasons.[6]

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 actually paid for her performance, while the self-regard/signalling benefit always accrues, by virtue of a declared belief. Our formalism can capture other motivations, or variations on the ones we have given. For instance, according to cognitive evaluation theory, a worker's intrinsic motivation is higher when her salary provides information about her competence level (see Ryan, Mims and Koestner (1983)). As a result, workers respond more productively to rewards that are contingent on their good performance. An extra utility kick c_H for performance is one way of modelling this. In a related vein, many studies have found that people have a preference for 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. As to self-regard, it could be that statements made to an experimenter and statements made as inputs on a computer yield different benefits, so that n 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, in conjunction with another experiment by the same authors. The subject pool consists of 313 undergraduate students from the university. The experiment was pre-registered and the pre-registration materials can be found in Appendix C.[7]

[6] Kruger and Dunning (1999) find that unskilled people are especially prone to having overconfident-looking beliefs. This finding is not directly related to the present study, as it is about actual beliefs, rather than their reports. Furthermore, control is not implicated in their elicitations, which are unincentivized.

[7] The pre-registration is also available at https://aspredicted.org/zu3pc.pdf. The second experiment in the pre-registration is a test of Kruger and Dunning's (1999) "unskilled and unaware" hypothesis. The appendix reports instructions for both experiments.

The experiment comprises two treatments that allow us to isolate and measure the control motivation. The first treatment closely follows the probability-matching method developed by Grether (1981) and elaborated upon by Karni (2009), which has been widely used to elicit beliefs, notably in studies on overconfidence (for example, Möbius et al. (2014) and Benoît et al. (2015)). Under 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 own performance, on one of two tasks.

The main hypothesis is that there are control motives 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. In fact, the distribution of placements in Treatment 1 should first order stochastically dominate the placements in Treatment 2.

We describe the two treatments in terms of the probabilities induced during the experiment. The appendix describes how these probabilities were generated and gives the instructions that were used. The two treatments share the following steps:

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 that are similar to questions they will later answer in a logic quiz. The subjects are subsequently told the median quiz score of participants who took the same quiz on prior occasions.

4. Each subject is asked to report the chance that she will place in the top half of quiz-takers. One of two (monetary) incentive compatible methods, one for each treatment, is used to incentivize the reports.

5. Subjects take a logic quiz in which they answer 12 multiple choice questions. The subjects are ranked according to their scores, with ties broken randomly.

6. Subjects are paid based on their reported beliefs and their performances in the visual task and the quiz, in one of two ways, depending on the treatment to which they are assigned.

Treatment 1

Suppose a subject has indicated a probability p_1 of placing in the top half of quiz-takers (Step 4 above). A number $x \in [0, 1]$ is drawn uniformly. If $x \leq p_1$, the subject wins R lottery tickets if her placement on the quiz is in the top half of subjects. If $x > p_1$, then with probability x she wins R lottery tickets. In all other cases, she wins nothing.

Treatment 2

Suppose a subject 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 placement on the quiz is in the top half of subjects. If $x > p_2$, then with probability x she wins T lottery tickets if she was successful in a randomly drawn instance of the visual task. 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 lottery tickets that can be won in Treatment 2 yield a $\frac{3}{10s_i}$ chance of obtaining €20 (recall that s_i is the fraction of correct answers on the visual task). Subjects are told the numerical value of $\frac{3}{10s_i}$ without being apprised of its dependence on s_i.

With these parameters, in both treatments an expected utility maximizer that cares only about her monetary payoffs will truthfully report her subjective belief that she will end up in the top half of subjects (as shown below).[8] But a subject that is also motivated by self-regard or control concerns may report a higher number.

Let us first undertake an informal analysis, which does not depend on our model. Consider a subject who estimates her chance of placing in the top half to be μ.

- Suppose the subject participates in Treatment 1. Any utility she derives from positive statements about herself, provides her an incentive to exaggerate her reported belief, p_1. Moreover, 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 solely on a random device. Utility she derives from betting on herself, provides a further incentive to inflate her report, in order to shift weight onto earning money for doing well, rather than for being lucky.

[8] Subjects are advised that they maximise their expected payments by accurately reporting their beliefs.

- Suppose the subject participates in Treatment 2. As in Treatment 1, she may exaggerate her report, p_2, 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. Here, a preference for betting on herself does not provide a further incentive to distort.

Because a preference for control gives an incentive to inflate in Treatment 1 but not in Treatment 2, we expect $p_2 < p_1$ if subjects have control motives.

We now reason formally. We adopt the normalisations $u(w) = 0$ and $u(w+20) = U$, where w is the subject's initial wealth, and make the substitutions $N = \frac{10}{3}\frac{n}{U}$ and $C_H = \frac{c_H}{U}$.

In Treatment 1, a subject who believes she has a probability μ of being in the top half and 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. In addition to a potential monetary gain, she derives a control benefit c_H when she is paid for having doing well. 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 placed in the bottom half – is $p_1\mu\frac{3}{10}$. The subject also obtains a self-regard benefit $n(p_1,\mu)$ from her report. The subject has a total expected utility of

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

This is maximized by a report that satisfies

$$p_1^* = \mu(1+C_H) + N_1(p_1^*,\mu), \tag{1}$$

using the aformentioned substitutions and setting $N_1 = \frac{\partial N}{\partial p}$.[9]

If a subject cares only about money, so that $N_1 \equiv 0 = C_H$, then $p_1^* = \mu$. Hence, the mechanism is monetarily incentive compatible. If $N_1 > 0$ and/or $C_H > 0$, a subject will overstate her beliefs. We can interpret μC_H as the subject's overstatement due to control concerns, N_1 as the overstatement due to self-image concerns, and $\mu C_H + N_1$ 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 $p_2\mu\frac{3}{10} + (1-p_2)\frac{(1+p_2)}{2}s_i\frac{3}{10s_i}$ of winning €20. The probability that she earns the money for her performance is also $p_2\mu\frac{3}{10} + (1-p_2)\frac{(1+p_2)}{2}s_i\frac{3}{10s_i}$. Her total expected utility is

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

[9] More precisely, we should write $p_1^* = \min\{\mu(1+C_H) + N_1, 1\}$. About 7% of subjects across the two treatments declare a probability of 1.

This is maximized by a choice of

$$p_2^* = \mu + \frac{N_1\left(p_2^*, \mu\right)}{C_H + 1}. \tag{2}$$

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

3.1 Identification

We adopt a between subject design, with each subject partipating 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 – $E\left(\mu_1\right) = E\left(\mu_2\right) = E\left(\mu\right)$ in both populations. To discuss 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 – $\bar{\mu}_1 = \bar{\mu}_2 = E\left(\mu\right)$ and denote by $\bar{N}_1$ the mean value of N_1.

Let $\bar{p}_i$ be the mean report across subjects in Treatment i and 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, assuming subjects behave as in our model and letting $\bar{p}_i^*$ denote the average of individuals' optimal reports, from (1) we obtain $\bar{p}_1^* = \bar{\mu} + \bar{\mu}C_H + \bar{N}_1$ and an alternative possibility is that $\bar{\mu} = \frac{1}{2}$, but $C_H > 0$ and/or $\bar{N}_1 > 0$. Here, $\bar{\mu}C_H$ is the mean overstatement due to control concerns, $\bar{N}_1$ is the mean overstatement due to self-image concerns, and $\bar{\mu}C_H + \bar{N}_1$ 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.

Nevertheless, the two treatments can be fruitfully combined. From (1) and (2), we obtain

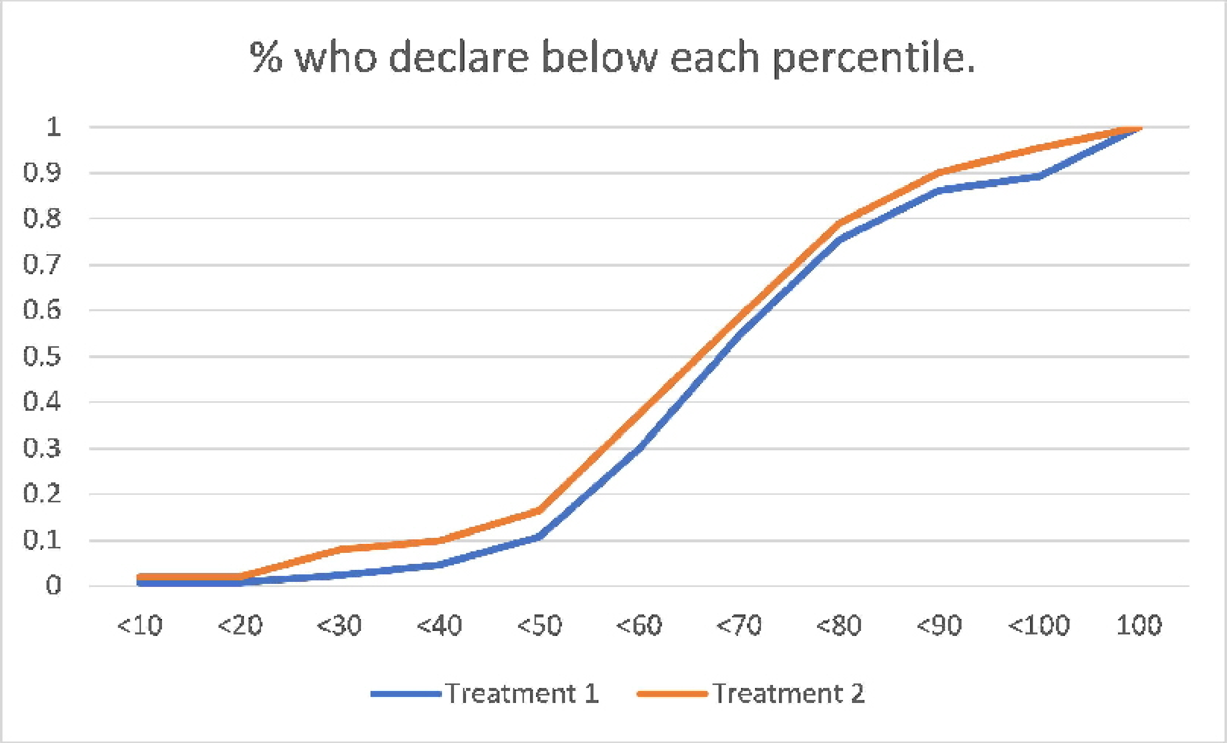
$$\bar{\mu}C_H + \bar{N}_1\left(p_1^*, \bar{\mu}\right) = \bar{p}_1^* - \bar{p}_2^* + \frac{\bar{N}_1\left(p_2^*, \bar{\mu}\right)}{C_H + 1} \geq \bar{p}_1^* - \bar{p}_2^*. \tag{3}$$

As indicated in our informal reasoning, $\bar{p}_1^* - \bar{p}_2^*$ gives a lower bound on the overstatement in Treatment 1, $\bar{\mu}C_H + \bar{N}_1(p_1^*, \mu)$, that is due to non-monetary concerns rather than to overconfidence.

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 distribution of p_1's first order stocahstically dominates the distribution of p_2's, as predicted by the control hypothesis.

as on Screen Picture then

JPG



1.jpg

Treatment 1, which uses a standard-type incentive mechanism, 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. Hence, at least $26.81\% = \frac{4.33}{16.15}$ of the measured overconfidence is due to control or self-regard/signalling distortions.

We can be more specific about the control mark-up. Using (1) and (2), we have

$$C_H = \frac{\bar{p}_1^* - \bar{p}_2^*}{\bar{p}_2^*} + \frac{\bar{N}_1(p_2^*, \bar{\mu}) - \bar{N}_1(p_1^*, \bar{\mu})}{\bar{p}_2}. \tag{4}$$

Since $N_{11} \geq 0$ and the distribution of p_1^*s first order stochastically dominates that of p_2^*s, we obtain $\bar{N}_1(p_2^*, \bar{\mu}) \geq \bar{N}_1(p_1^*, \bar{\mu})$. Therefore,

$$C_H \geq \frac{\bar{p}_1 - \bar{p}_2}{\bar{p}_2} = 7\%.$$

On average, each subject in Treatment 1 inflates her report by a factor of at least 7% to obtain control benefits (μC_H is the control markup).

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

3.2 Modelling

Let us step back for a moment to consider what conclusions can be obtained without adopting our 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. This mitigation leads to the prediction that $\bar{p}_1 > \bar{p}_2$ (and that reports in Treatment 1 should first order stochastically dominate reports in Treatment 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 mechanism design.[10]

When subjects with high beliefs inflate their reports for non-monetary reasons, they may hit the reporting constraint of 100%. If, for illustration, all subjects wanted to inflate their beliefs by a factor of 15%, then all those with beliefs above 87% would make reports of 100%, causing a cluster at this number. Since Treatment 1 has the additional non-monetary control motivation relative to Treatment 2, this suggests that in Treatment 1 we should expect a jump in the number of reports of 100%, as compared to Treatment 2. And, in fact, 10.7% of subjects report 100% in Treament 1 compared to 4.6% of subjects in Treament 2 – a difference which is significant at the 5% level.[11]

Our modelling permits a sharper analysis, at the cost of added assumptions. In particular, our model assumes that money earned for success on the quiz and money earned for success on the visual task yield the same control benefit. This assumption leads to control incentives

[10] Arguably, the mechanism in Treatment 2 is somehwat more complicated than the mechansm in Treatment 1 but it is not clear what impact, if any, this might have on reports. In a different context, Experiment 2 finds no impact from experimental variations.

[11] When we place subjects into reporting bins of size 10% plus a bin at 100%, we ony find a significant jump at 100%.

merging completely with monetary incentives in Treatment 2. 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.[12] It is easy to adapt the model to the case where payments for the two tasks yield different control benefits. If being rewarded for a draw from the performance on the somewhat unconventional visual task yields a smaller control benefit than being rewarded for placement on the logic quiz, then our calculations underestate the lower bound on the effect of control, and conversely.

3.3 Betting on yourself versus 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 subject bets on herself if $\mu_s - \mu_m > c$, for some number c. If $c = 0$, the subject maximizes her expected monetary payoff; if $c < 0$ the subject values control and is willing to sacrifice money in order to bet on herself; if $c > 0$ 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 as positive signal.

The beliefs μ_s, μ_m are not known to the experimenters. Rather, subjects are asked to make reports p_s and p_m of their beliefs. The reports are incentivized using a probability matching 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 $p_s - p_m > \hat{c}$, for some number $\hat{c}$; if $\hat{c} < 0$ the subject is said to exhibit a preference for control. The authors determine that the behaviour of 82% of subjects is consistent with a cutoff strategy.

Let us apply our modeling 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.

[12] The tasks were purposely designed to be dissimilar in their natures, as opposed to their succes rates, as we did not want a subject's performance on one task to yield (much) information about performance on the other.

Under these assumptions, for the elicited beliefs p_s and p_m we have

$$p_s = \mu_s\left(1 + C_H\right) \text{ and } p_m = \mu_m, \tag{5}$$

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

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

$$\mu_s u\left(w + 20\right) + \left(1 - \mu_s\right) u\left(w\right) + \mu_s c_H = \mu_s U + \mu_s C_H U, \tag{6}$$

while the payoff for betting on the match is

$$\mu_m u\left(w + 20\right) + \left(1 - \mu_m\right) u\left(w\right) = \mu_m U. \tag{7}$$

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 $c = -\mu_s C_H$ is negative.

In terms of observables, from (5) we have that $\mu_s - \mu_m > -\mu_s \frac{c_H}{U}$ if and only if $p_s - q_s > 0$. This means that, although the true cutoff is negative, the measured cutoff $\hat{c}$ should be zero. Put differently, we have $\hat{c} = 0$ 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 the findings in the paper under-measure control. In their conclusion, Owens et al. also suggest that they have found a lower bound on the effect of control incentives.

Although the above reasoning suggests that the measured cutoff should be 0, the behaviour of 26% of subjects displays a strictly negative cutoff (and for 9%, a strictly positive cut-off). This discrepancy can be accounted for 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 choose the self-bet. This push could be because of the positive signal a self-bet sends, because of the inherent ambiguity in a bet on someone else (in contrast to the objectively random bets used in the elicitation of probabilities), or for some other reason. In terms of our analysis, the simplifying assumptions a) and b) may not jointly hold. An extra push is consistent with the discussion in Owens et al. of the various reasons subjects may favour bets on themselves which their use of the word *control* encompasses.

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

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

We note that the difference between i) self-bets versus bets on someone else and ii) self-bets versus bets on a random device is an interesting wrinkle that our experiment and theory does not explore, as we only consider mechanisms that offer a choice between a self-bet and a bet on a random device.

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 subject pools. **As I recall, thie experiment was not preregistered, right?** **correct**

This experiment seeks a better understanding of the control motivation. Our first experiment shows that people have a positive bias for bets that pay off when they do well. But how do people 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 it does not explicitly compare a bet on performance with a random bet, there is an implicit comparison, so that control concerns are important. 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 $\left(p\left(\theta = \theta^H\right)\right)$. It is immediate that self-image concerns may lead her to overstate this probability. In order to focus on control issues, we ignore these self-image concerns in what follows. That is, we set $n \equiv 0$.

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 $z \geq p^2$, the subject wins m regardless of her performance; if $z < (1-p)^2$ she wins nothing regardless of her performance. In both cases, there is no control issue. A potential control benefit arises when $(1-p)^2 \leq z < p^2$, as she then wins m if and only if $\theta = \theta^H$. Setting $u(w+m) = 1$ and $u(w) = 0$, the expected utility from a report p is

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

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

Similar reasoning establishes that if $p < \frac{1}{2}$, a potential control issue arises 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 establishes that $c_H > 0$, so that subjects want to inflate their reports for control reasons when $\mu \geq \frac{1}{2}$. Whether control pushes subjects to inflate, deflate, or neither when $\mu < \frac{1}{2}$ 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. They are paid €0.50 for each correct answer. (At this point, subjects are not aware of the bets to follow 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.[13]

Treatment 1

[13] In contrast to Experiment 1, subjects make their preditictions after having taken the test rather than after only having seen sample questions. This difference is due to unrelated purposes of the experiment, which we report in another paper about the unskilled and unaware effect (also mentioned in the pre-registration). 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.

Each subject is asked for the probability p_1 that she will place in the top half. With probability p_1 she wins €10 if she lands in the top half; with probability $(1-p_1)\frac{1+p_1}{2}$, she wins €10 based on a random draw. 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. Specifically, each subject is asked for the probability q_2 that she will place in the *bottom* half. Then, with probability q_2 she wins €10 if she lands in the bottom half; with probability $(1-q_2)\frac{1+q_2}{2}$, she wins €10 based on a random draw. In all other cases, she wins nothing.

Treatment 3

The third 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. If it comes up heads, then with probability p_3 the subject wins €10 if her placement is in the *top* half of subjects and with probability $(1-p_3)\frac{1+p_3}{2}$ she wins €10 based on a random draw. If the coin comes up tails, then with probability $(1-p_3)$ she wins €10 if her placement on the quiz is in the *bottom* half of subjects and with probability $p_3\frac{2-p_3}{2}$ she wins €10 based on a random draw.[14]

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 now plays a role.

4.2 Individual Incentives

We analyze individual 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

[14] 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)

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

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

which is maximized at

$$p_1^* = \mu\left(1 + c_H\right) + n_1. \tag{8}$$

In Treatment 2, she has an expected utility of

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

which is maximized at

$$p_2^* = \mu - c_L\left(1 - \mu\right) + n_1. \tag{9}$$

In Treatment 3, she has an expected utility of

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

which is maximized at

$$p_3^* = \mu + \frac{1}{2}c_H\mu - \frac{1}{2}c_L\left(1 - \mu\right) + n_1. \tag{10}$$

4.3 Identification

Again, we analyze mean behaviour across treatments.

From (8), (9), and (10), 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 $\overline{p}_1^* = \overline{p}_2^* = \overline{p}_3^*$. Note that the result $\overline{p}_1^* = \overline{p}_2^* = \overline{p}_3^*$ indicates that subjects do not change their behaviour simply in reaction to different experimental protocols.

Given $\overline{p}_1^* = \overline{p}_2^*$, (8) and (9) together imply that

$$c_L = -c_H \frac{\mu}{1 - \mu}. \tag{11}$$

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 than Heath and Tversky, 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, $c_H\mu$. Plugging (11) into (9) yields

$$1-q_2^*=p_2^*=\mu-c_L(1-\mu)+n_1=\mu(1+c_H)+n_1,$$

Hence, when a subject is paid for doing poorly, she distorts away from her poor performance to the same extent that she distorts towards a good performance when paid for doing well.

Returning to the binarized scoring rule, control objectives will lead a subject with belief μ to report $p^*=\mu-c_L(1-\mu)=\mu+c_H\mu$, whether 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 probabilty matching method.[15] The mechanism we introduced in Treatment 2 can be adapted to this rule to eliminate control distortions.

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.

5.1 Experiment 1 - Regression analysis

The experiment was run in the CREED Lab at the University of Amsterdam in Spring 2018. Three hundred and thirteen undegraduates participated and were assigned to either Treatment 1 (N=154) or Treatment 2 (N=159). 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).

The main hypothesis is the existence of control motives to overstate placement 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. We also report the results of

[15] 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.

the regressions including controls in Table 1.[16] 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-2*, a dummy taking value 1 if the observation belongs to Treatment 2. 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. All our tests including controls are significant at 5%.

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 for every subject to ensure incentive compatibility. Specifically, subjects with a success rate of s_i on the visual task stand to win lottery tickets that yield a probability $\frac{3}{10s_i}$ of winning the the €20. However, twenty-nine subjects performed particularly poorly in the visual task – achieving a success rate of 20% or lower – making it impossible to achieve full incentive compatibility for them in Treatment 2.[17] Since incentive compatibility was not achieved for these subjects, we perform the analysis excluding them (from both treatments to avoid introducing a selection problem) in the first three columns. The table also presents, in column 4, the analysis for all subjects with complete data; the analysis is fundamentally unchanged.

The last control variable is the score 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).[18] Not surprisingly, 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).

gender should be male in the table. I had changed it before. R^2 in caps.

[16] The table excludes two subjects who did not complete the gender question. This exclusion was pre-registered.

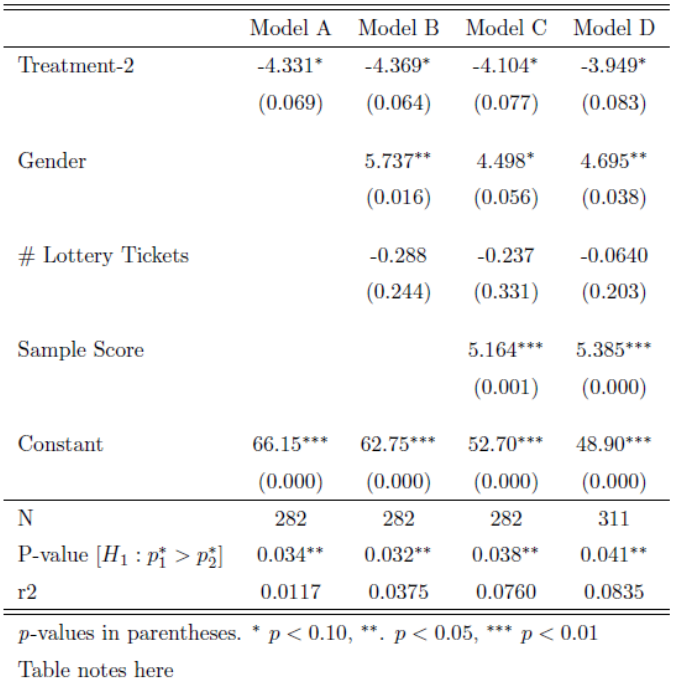
[17] For these subjects, the chance of a successful round of the visual task being selected for payment was so low that even awarding them lottery tickets with a 100% probability of winning the lottery did not ensure incentive compatibility.

[18] Subjects are not told their scores on the sample questions, but it is likely that they form some beliefs about their performance in the sample.

add as model D the test of difference in means with 311 subjects (that is what we pre-registered)

add table notes. e.g. models A-C for subjects who score 3 or more in the Visual Task. Modeld D-E for full sample.

Table 1: Effect of control



	Model A	Model B	Model C	Model D
Treatment-2	-4.331*	-4.369*	-4.104*	-3.949*
	(0.069)	(0.064)	(0.077)	(0.083)
Gender		5.737**	4.498*	4.695**
		(0.016)	(0.056)	(0.038)
# Lottery Tickets		-0.288	-0.237	-0.0640
		(0.244)	(0.331)	(0.203)
Sample Score			5.164***	5.385***
			(0.001)	(0.000)
Constant	66.15***	62.75***	52.70***	48.90***
	(0.000)	(0.000)	(0.000)	(0.000)
N	282	282	282	311
P-value $[H_1 : p_1^* > p_2^*]$	0.034**	0.032**	0.038**	0.041**
r2	0.0117	0.0375	0.0760	0.0835

p-values in parentheses. * $p < 0.10$, **. $p < 0.05$, *** $p < 0.01$

Table notes here

When the analysis uniquely focuses on subjects for which the mechanism is incentive compatible, the point estimates of the difference between the two treatments is larger and the statistical significance improves.[19]

5.1.1 Further tests

In this section, we examine some further implications of our modelling.

Recall that $p_1^* = \mu(1 + C_H) + N_1(p_1^*, \mu)$ and $p_2^* = \mu + \frac{N_1(p_2^*,\mu)}{C_H+1}$. The role played by control in determing the p_i's is straighforward, having been derived as an expected benefit. The role of self-regard and signalling is more complicated, involving interactions between reports and beliefs in $N_1(p_i^*, \mu)$. With the minimal assumptions we have made on $N_1(p_1^*, \mu)$ so far, the incentive to inflate can behave bizarrely enough that people with lower self-beliefs make higher reports. Clearly, though, inflating beliefs for self-regard and signalling reasons does

[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.

not make sense if high reports correspond to low beliefs. Let us minimize interaction effects and assume from now on that all the second derivatives N_{ij}, $i, j = 1, 2$ are small. The model then predicts that $\frac{dp_1^*}{d\mu}, \frac{dp_2^*}{d\mu} > 0$.

We cannot observe subjects' beliefs but a reasonable proxy for these beliefs is sample scores, which correlate well with performance. As we have already noted, our regressions in Table 1 show that p_1^* and p_2^* are indeed both increasing in sample scores. While the prediction $\frac{dp_1^*}{d\mu}, \frac{dp_2^*}{d\mu} > 0$ is confirmed, it is not very restrictive; many models would predict that reports are increasing in beliefs. Indeed, this feature might be considered the starting point of a signalling model. A more interesting prediction comes from examing $p_1^* - p_2^*$.

Treatment 2 elicits smaller reports than Treatment 1 because it eliminates the control motivation for inflating. Since people derive a larger (expected) control benefit when they have a greater self-belief, one would expect the reduction in reports, $p_1^* - p_2^*$, to be increasing in μ. A simple calculation shows that the model predicts that $p_1^* - p_2^*$ is increasing in μ, when the N_{ij}'s are small. We put this hypothesis to the test with a second set of specifications in which we additionally include the interaction variable *Treatment-2* × *Sample Score*. Hence, we leverage the observed correlations, and use *Sample Score* as a proxy for expected and actual performance, that is, as a proxy for the unobservable skill level μ. The model predicts that we should observe a larger treatment effect as the sample score (and therefore μ) increases. In other words, the estimated coefficient associated to *Treatment-2* × *Sample Score* should be negative and significant. The results are presented in Table 2. The hypothesis is confirmed. **Gender should be male, and add table notes.**

Table 2: Control- Heterogeneous effects

	Model E	Model F	Model G	Model H
Treatment-2 × Sample Score	-1.408*	-2.866	-1.477**	-4.292
	(0.052)	(0.348)	(0.039)	(0.132)
Gender	4.494*	4.494*	4.694**	4.696**
	(0.056)	(0.056)	(0.038)	(0.038)
# Lottery Tickets	-0.221	-0.201	-0.0630	-0.0531
	(0.364)	(0.413)	(0.205)	(0.294)
Sample Score	5.947***	6.769***	6.171***	7.727***
	(0.000)	(0.003)	(0.000)	(0.000)
Treatment		4.796		9.213
		(0.623)		(0.307)
Constant	50.92***	48.70***	47.55***	43.79***
	(0.000)	(0.000)	(0.000)	(0.000)
N	282	282	311	311
P-value [H_1: Treatment-2 × Sample Score<0]	0.026**	0.174	0.018**	0.066*
R2	0.0781	0.0789	0.0872	0.0903

p-values in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table notes here

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 61% for Treatments 1 and 2, 76% for treatments 2 and 3, and 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 $= 58\%$ for Treatments 1 and 2, 72% for Treatments 2 and 3, and 88% for Treatments 1 and 3).

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 Zimmerman (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 probability matching method, subjects inflate their reported beliefs about themselves by at least 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.

7 Appendix A. 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. 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.

8 Appendix B. 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].

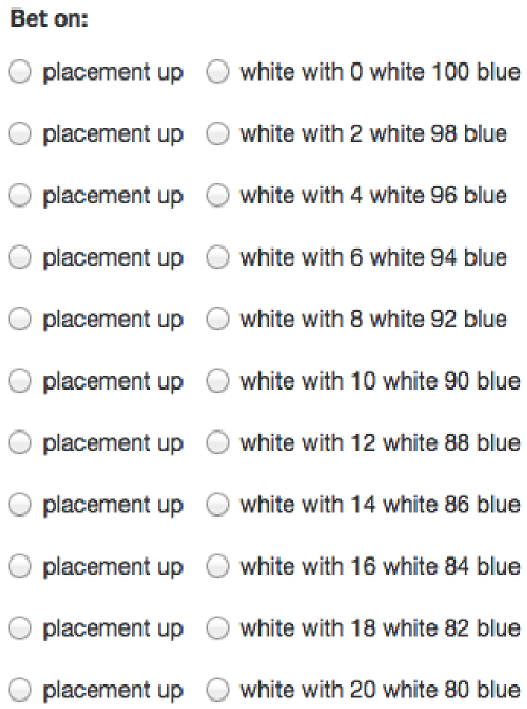
Second part (Treatment 1 - Betting up)

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.



Bet on:

○ placement up ○ white with 0 white 100 blue

○ placement up ○ white with 2 white 98 blue

○ placement up ○ white with 4 white 96 blue

○ placement up ○ white with 6 white 94 blue

○ placement up ○ white with 8 white 92 blue

○ placement up ○ white with 10 white 90 blue

○ placement up ○ white with 12 white 88 blue

○ placement up ○ white with 14 white 86 blue

○ placement up ○ white with 16 white 84 blue

○ placement up ○ white with 18 white 82 blue

○ placement up ○ white with 20 white 80 blue

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.

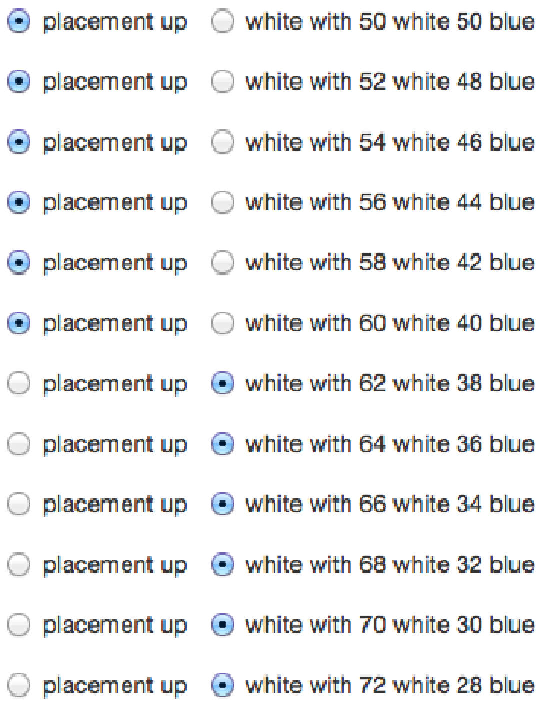
Thresholds: The number of white balls represents your chances of winning when you bet on the virtual bag. The number of white balls increases as you scroll down the list, so the virtual bag becomes more attractive the more down you go on the list. Hence we expect that, if you choose the virtual bag in 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:



● placement up ○ white with 50 white 50 blue

● placement up ○ white with 52 white 48 blue

● placement up ○ white with 54 white 46 blue

● placement up ○ white with 56 white 44 blue

● placement up ○ white with 58 white 42 blue

● placement up ○ white with 60 white 40 blue

○ placement up ● white with 62 white 38 blue

○ placement up ● white with 64 white 36 blue

○ placement up ● white with 66 white 34 blue

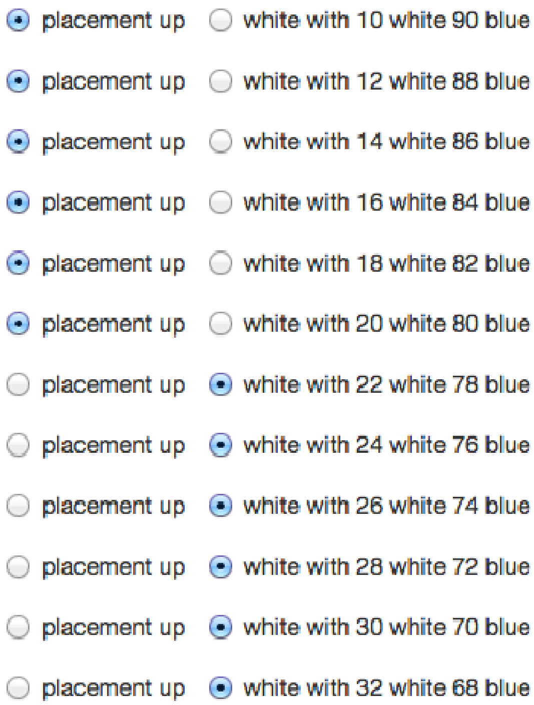
○ placement up ● white with 68 white 32 blue

○ placement up ● white with 70 white 30 blue

○ placement up ● white with 72 white 28 blue

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.



● placement up	○ white with 10 white 90 blue
● placement up	○ white with 12 white 88 blue
● placement up	○ white with 14 white 86 blue
● placement up	○ white with 16 white 84 blue
● placement up	○ white with 18 white 82 blue
● placement up	○ white with 20 white 80 blue
○ placement up	● white with 22 white 78 blue
○ placement up	● white with 24 white 76 blue
○ placement up	● white with 26 white 74 blue
○ placement up	● white with 28 white 72 blue
○ placement up	● white with 30 white 70 blue
○ placement up	● white with 32 white 68 blue

Figure 2. John's Choices

Second part (Treatment 2 - Betting down)

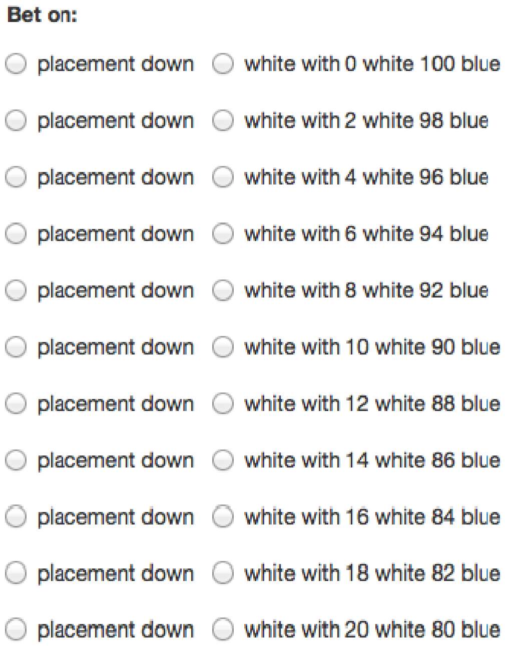
In this second part of the experiment, we ask you to estimate how well you did in the quiz relative to the other subjects. Of course, you cannot know your relative performance for sure so we will ask you for a probability estimate. Specifically, we will ask you with which probability you think you placed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



Bet on:

○ placement down	○ white with 0 white 100 blue
○ placement down	○ white with 2 white 98 blue
○ placement down	○ white with 4 white 96 blue
○ placement down	○ white with 6 white 94 blue
○ placement down	○ white with 8 white 92 blue
○ placement down	○ white with 10 white 90 blue
○ placement down	○ white with 12 white 88 blue
○ placement down	○ white with 14 white 86 blue
○ placement down	○ white with 16 white 84 blue
○ placement down	○ white with 18 white 82 blue
○ placement down	○ white with 20 white 80 blue

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.

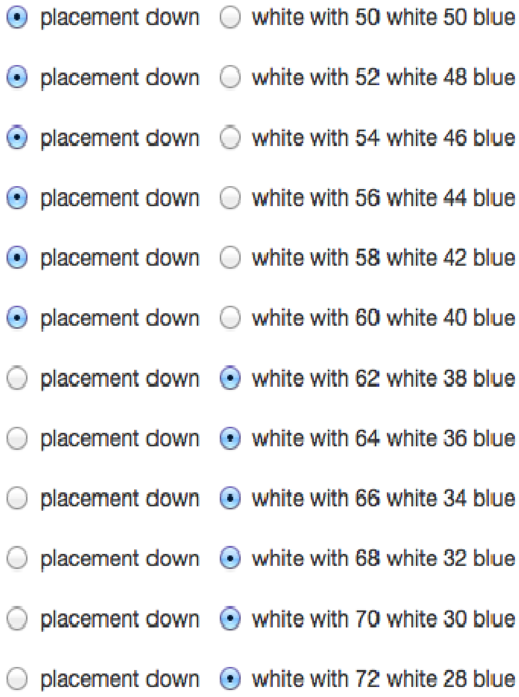
Thresholds: The number of white balls represents your chances of winning when you bet on the virtual bag. The number of white balls increases as you scroll down the list, so the virtual bag becomes more attractive the more down you go on the list. Hence we expect that, if you choose the virtual bag in 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:



● placement down ○ white with 50 white 50 blue

● placement down ○ white with 52 white 48 blue

● placement down ○ white with 54 white 46 blue

● placement down ○ white with 56 white 44 blue

● placement down ○ white with 58 white 42 blue

● placement down ○ white with 60 white 40 blue

○ placement down ● white with 62 white 38 blue

○ placement down ● white with 64 white 36 blue

○ placement down ● white with 66 white 34 blue

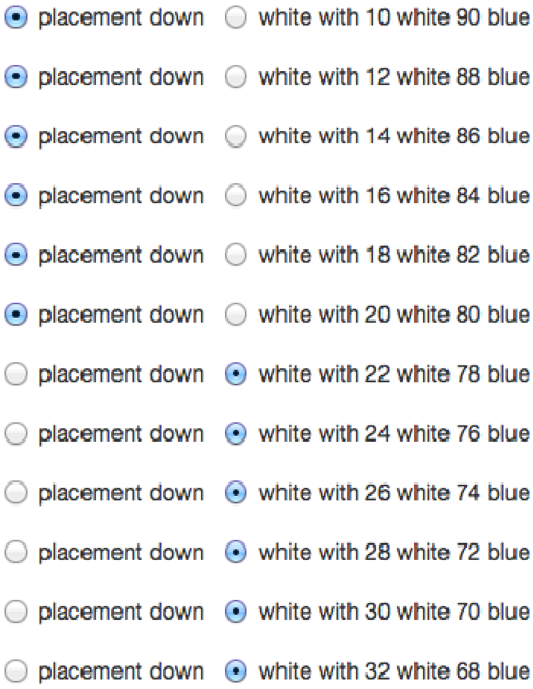
○ placement down ● white with 68 white 32 blue

○ placement down ● white with 70 white 30 blue

○ placement down ● white with 72 white 28 blue

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.



● placement down ○ white with 10 white 90 blue

● placement down ○ white with 12 white 88 blue

● placement down ○ white with 14 white 86 blue

● placement down ○ white with 16 white 84 blue

● placement down ○ white with 18 white 82 blue

● placement down ○ white with 20 white 80 blue

○ placement down ● white with 22 white 78 blue

○ placement down ● white with 24 white 76 blue

○ placement down ● white with 26 white 74 blue

○ placement down ● white with 28 white 72 blue

○ placement down ● white with 30 white 70 blue

○ placement down ● white with 32 white 68 blue

Figure 2. 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.[20]

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.

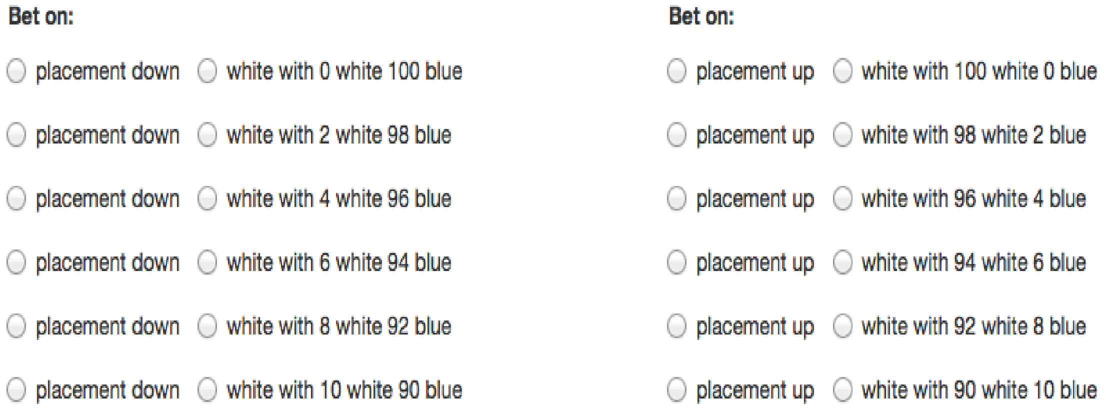
We want you to tell us your best estimate of the probability that you are in the lower half. For this purpose, we will use a special payment procedure that rewards you for giving us your best estimate. The procedure is a bit complicated but the most important thing to understand about it is simply that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

[20] *[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].*

bag. Choices will be presented to you in two groups of pairwise comparisons, as shown in Figure 1.



Bet on:		Bet on:	
○ placement down	○ white with 0 white 100 blue	○ placement up	○ white with 100 white 0 blue
○ placement down	○ white with 2 white 98 blue	○ placement up	○ white with 98 white 2 blue
○ placement down	○ white with 4 white 96 blue	○ placement up	○ white with 96 white 4 blue
○ placement down	○ white with 6 white 94 blue	○ placement up	○ white with 94 white 6 blue
○ placement down	○ white with 8 white 92 blue	○ placement up	○ white with 92 white 8 blue
○ placement down	○ white with 10 white 90 blue	○ placement up	○ white with 90 white 10 blue

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.

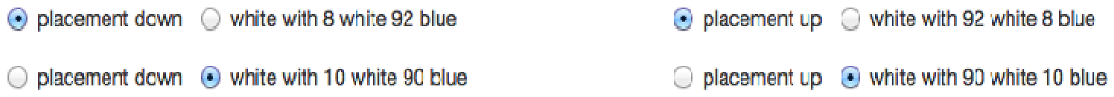
[Note: In 2 (out of 4) sessions of treatment 3, the order of the columns was reversed and the instructions were adjusted accordingly. As a result, subjects would bet on their placement-up in the column on the left, and on their placement-down in the column on the right.]

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 4, 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.

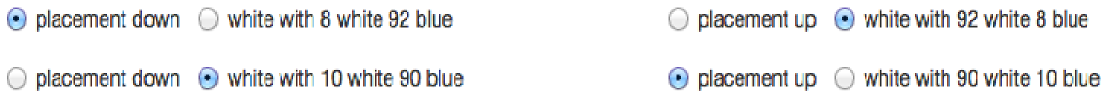
A way to ensure you are meeting this constraint is to verify that, taking two questions placed on the same line, you are betting on the virtual bag in one and only one of them. Figure 2 shows two examples of non-admissible choices. Figure 3 shows two examples of admissible choices. If you make a mistake,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



● placement down ○ white with 8 white 92 blue
○ placement down ● white with 10 white 90 blue

● placement up ○ white with 92 white 8 blue
○ placement up ● white with 90 white 10 blue

Figure 2. Non-admissible choices



● placement down ○ white with 8 white 92 blue
○ placement down ● white with 10 white 90 blue

○ placement up ● white with 92 white 8 blue
● placement up ○ white with 90 white 10 blue

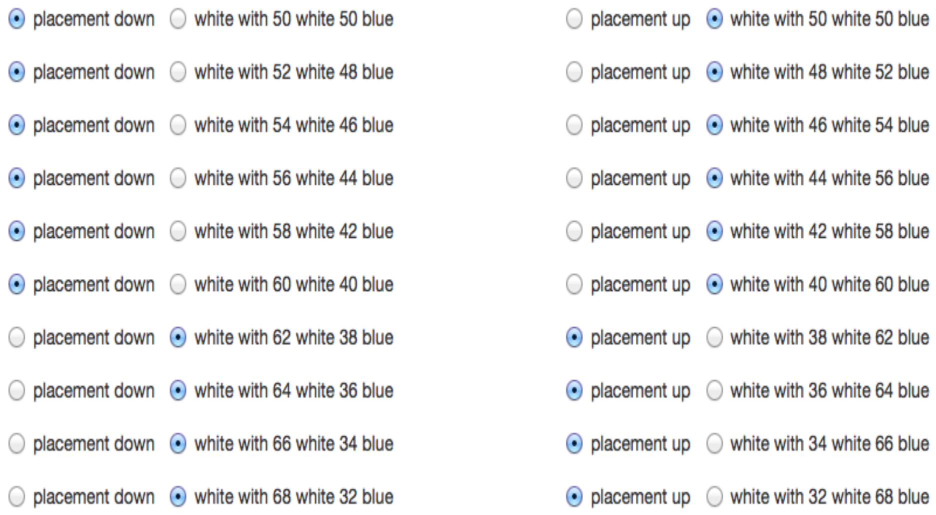
Figure 3. Admissible choices

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.

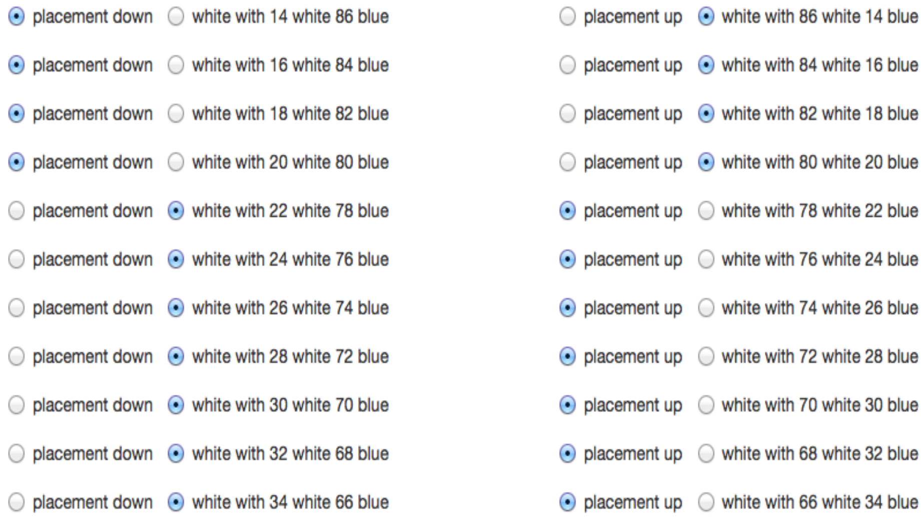
Examples: Lisa thinks there is a 60% chance she placed in the lower half and a 40% chance she placed in the upper half. Hence, she chooses to bet on her placement-down if in the bag there are 60 white balls or fewer. Moreover, she chooses to bet on her placement-up if there are fewer than 40 white balls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



● placement down ○ white with 50 white 50 blue ○ placement up ● white with 50 white 50 blue
● placement down ○ white with 52 white 48 blue ○ placement up ● white with 48 white 52 blue
● placement down ○ white with 54 white 46 blue ○ placement up ● white with 46 white 54 blue
● placement down ○ white with 56 white 44 blue ○ placement up ● white with 44 white 56 blue
● placement down ○ white with 58 white 42 blue ○ placement up ● white with 42 white 58 blue
● placement down ○ white with 60 white 40 blue ○ placement up ● white with 40 white 60 blue
○ placement down ● white with 62 white 38 blue ● placement up ○ white with 38 white 62 blue
○ placement down ● white with 64 white 36 blue ● placement up ○ white with 36 white 64 blue
○ placement down ● white with 66 white 34 blue ● placement up ○ white with 34 white 66 blue
○ placement down ● white with 68 white 32 blue ● placement up ○ white with 32 white 68 blue

Figure 4. Lisa's Choices

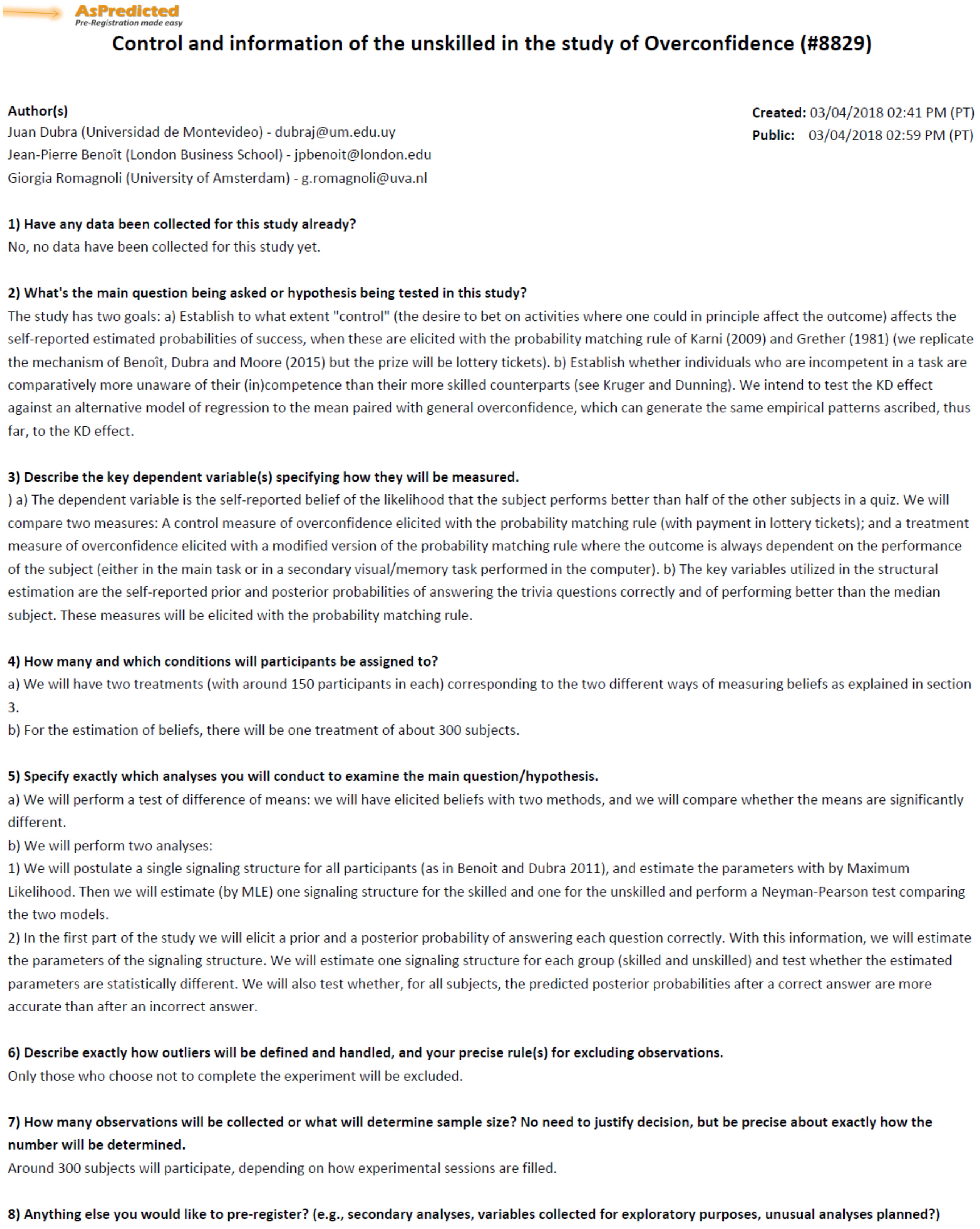
John thinks there is a 20% chance he placed in the lower half. Hence, he chooses to bet on his placement-down rather than on the virtual bag if there are fewer than 20 white balls in the virtual bag, otherwise he prefers to bet on the bag. He clicks the buttons according to this threshold and his choices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



● placement down ○ white with 14 white 86 blue ○ placement up ● white with 86 white 14 blue
● placement down ○ white with 16 white 84 blue ○ placement up ● white with 84 white 16 blue
● placement down ○ white with 18 white 82 blue ○ placement up ● white with 82 white 18 blue
● placement down ○ white with 20 white 80 blue ○ placement up ● white with 80 white 20 blue
○ placement down ● white with 22 white 78 blue ● placement up ○ white with 78 white 22 blue
○ placement down ● white with 24 white 76 blue ● placement up ○ white with 76 white 24 blue
○ placement down ● white with 26 white 74 blue ● placement up ○ white with 74 white 26 blue
○ placement down ● white with 28 white 72 blue ● placement up ○ white with 72 white 28 blue
○ placement down ● white with 30 white 70 blue ● placement up ○ white with 70 white 30 blue
○ placement down ● white with 32 white 68 blue ● placement up ○ white with 68 white 32 blue
○ placement down ● white with 34 white 66 blue ● placement up ○ white with 66 white 34 blue

Figure 5. John's Choices

9 Appendix C. Pre-Registered experiments.



AsPredicted
Pre-Registration made easy

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?)

References

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

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

[3] 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.

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

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

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

[7] Ehrlinger, J., K. Johnson, M. Banner, D. Dunning and J. Kruger (2008), “Why the Unskilled Are Unaware: Further Explorations of (Absent) Self-Insight Among the Incompetent,” *Organizational Behavior and Human Decision Processes* 105(1): 98–121.

[8] Eil, David and Justin Rao (2011), "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself," American Economic J: Microeconomics, 3, 114-38.

[9] 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.

[10] 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.

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

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

[13] Grieco, D. and R. Hogarth (2009) “Overconfidence in absolute and relative performance: The regression hypothesis and Bayesian updating,” *Journal of Economic Psychology* 3, 756-71.

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

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

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

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

[18] 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.

[19] Kunda, Z. (1990), “The case for motivated reasoning,” Psychological Bulletin, 108(3), 480-498.

[20] 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.

[21] 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.

[22] Moore, Don and Paul. J. Healy (2008), “The trouble with overconfidence," Psychological Review, 115(2), 502-517.

[23] 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.

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

[25] 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).

[26] Schotter, A. and I. Trevino (2014), “Belief elicitation in the laboratory,” Annual Review of Economics.