WORKING PAPERS

Victor H. Gonzalez-Jimenez

Contracting Probability Distortions

January 2019

Working Paper No: 1901

DEPARTMENT OF ECONOMICS

UNIVERSITY OF VIENNA

Contracting Probability Distortions *

Victor H. Gonzalez-Jimenez †

December 10, 2018

Abstract

I introduce a contract designed to take advantage of the regularity that individuals distort probabilities. With this contract, the principal could activate the probability distortions that are inherent in the agent and use these distortions to incentivize the agent to perform a relevant task. This is because in the contract, the principal could choose the probability that the agent's compensation depends on his own performance on the task. Distortions of such probability generate higher or lower motivation to perform the task. A theoretical framework and an experiment demonstrate that the proposed contract yields higher output than a traditional contract when both contracts offer similar monetary incentives. However, the probability specified by the principal is critical to achieving this result. Small probabilities yield higher levels of performance, whereas medium or high probabilities yield no performance differences between the contracts. The degree to which individuals overweight small probabilities explains these findings.

JEL Classification : C91, C92, J16, J24.

Keywords: Contracts, Probability Weighting Functions, Experiments,Rank-Dependent Utility.

*I am indebted to Wieland Müller, Patricio S. Dalton, Charles Noussair, Cole Williams, Stephanos Vlachos, Geoffrey Castillo, Sandra Rincon, and Stephanie Koornneef for their valuable comments. I am grateful to participants in seminars at Tilburg University, Maastricht University, University of Vienna, the Foundations of Utility and Risk conference at York, the ESA World Meeting 2018, and the Bogota Experimental Economics Conference 2018 for their comments. The experiments reported in this paper were funded by the Center for Economic Research (CentER) at Tilburg University.

†University of Vienna, Department of Economics and Vienna Center for Experimental Economics, Oskar-Morgestern-Platz 1. 1090 AT, Vienna, Austria. E-mail: victor.gonzalez@univie.ac.at.

1 Introduction

The traditional focus of contract theory has been the optimal implementation of monetary incentives (Gibbons and Roberts, 2013). For instance, Holmstrom, in his seminal work, demonstrates how in settings with information asymmetry, the principal could motivate the agent to perform a burdensome task by offering a moderate monetary compensation (Holmstrom, 1999; Holmstrom and Milgrom, 1991; Holmstrom, 1979). Recent literature has also highlighted the favorable influence of non-monetary incentives on the agent's motivation. Tools such as production targets set by the principal (Corgnet et al., 2015; Gómez-Miñambres, 2012), contests for status designed by the principal (Ashraf et al., 2014; Bandiera et al., 2013; Moldovanu et al., 2007), and environments that enhance peer pressure (Falk and Ichino, 2006), among others, have proven to bolster the agent's performance even if monetary incentives are already at work.

In this paper, I introduce a new contract featuring an innovative class of non-monetary incentives. The contract seeks to motivate agents by taking advantage of the regularity that they distort probabilities systematically. Empirical evidence from the literature of decision-making shows that when making decisions under uncertainty, individuals tend to overweight small probabilities and underweight moderate to large probabilities (Prelec, 1998; Wu and Gonzalez, 1996; Tversky and Fox, 1995; Tversky and Kahneman, 1992). The contract that I propose has the capacity to activate these probability distortions and direct them towards the enhancement of the agent's performance on the task that is of relevance to the principal. Note that these incentives are absent in more traditional contracts that rely only on monetary incentives to incentivize the agent; hence, this contract has the potential to improve upon the efficiency of the contracts that are recommended by traditional contract theory.

In the contract, called the "probability contract", the agent is offered a payment scheme that resembles a lottery. The two possible outcomes contained in this scheme are a monetary prize, with a size that increases with the agent's performance on the task, and no payment for performance. Think for instance of a performance bonus that realizes with a probability that cannot be influenced by the agent or a stock option offered by the firm with a dividend that is paid with some likelihood that the agent cannot alter. In addition, the contract allows the principal to set the probability that the monetary prize is paid and makes this decision before the agent performs the task. Therefore, the probability chosen by the principal has the capacity to influence the agent's decision about how much output to supply; not only through the corresponding changes in the monetary incentives that reward performance, but also, and more importantly, through the greater or lower motivation induced by the agent's perception of the probability that his compensation depends on his own performance.

To understand how the probability contract could outperform a traditional contract, consider a setting in which both contracts are cost-equivalent for the principal. The expected monetary reward of producing an additional unit of output under the probability contract is thus equivalent to the monetary reward of producing an additional unit of output under the traditional contract. When facing these incentives, an expected-value maximizer is equally motivated under both contracts and would deliver the same performance. However, if the assumption that the agent perceives probabilities accurately is relaxed and it turns out that he overweights the probability that the prize is paid, he would be motivated to supply more output under the probability contract. This result is robust to the possibility that the agent has a concave utility function, if the probability that the prize is paid is largely overweighted so that the risk-aversion stemming from the utility curvature is outweighed.

I present a theoretical framework to pin down the conditions guaranteeing the greater efficiency of the probability contract with respect to a cost-equivalent traditional contract. The agent not only needs to have risk preferences that admit probability distortions, but also, must overweight the probability that is set by the principal to a degree such that he exhibits a preference for risky contracts. Hence, when the agent has risk preferences characterized by rank-dependent utility (Quiggin, 1982) or by cumulative prospect theory (Tversky and Kahneman, 1992), the probability contract yields higher performance if the principal is able to implement it with a probability that induces risk-seeking attitudes in the agent.

In a controlled laboratory experiment I demonstrate that the probability contract yields higher effort in a real-effort task than a cost-equivalent piece rate if the probability of the contract is set at $p = 0.10$. Moreover, implementing the probability contract with larger probabilities, namely $p = 0.30$ or $p = 0.5$, yields no differences in performance with respect to the piece rate. In addition, the experiment features an elicitation of the utility functions and probability weighting functions of the subjects. These functions are used to show that the risk attitudes generated by probability distortions are causing the observed boost in performance when the contract is implemented with $p = 0.10$.

Although the proposed contracting modality may seem abstract at first glance, it has several applications in the field. One such application is corporate finance. A well-known empirical regularity of this literature is that volatile firms provide option stocks to lower-level employees, who in turn accept them (Spalt, 2013). This regularity is at odds with standard theory analyses, which suggest that risk-averse employees require unusually high premiums to accept such option stocks. In an alternative analysis, Spalt (2013) demonstrates that with this practice firms are capitalizing on the risk attitudes that emerge from the probability weighting function of the employees, who overweight the probability that these options yield large dividends. This paper shows that with option stocks firms are also generating higher motivation and performance on the employees, who might overweight the probability that their performance has an impact on the dividends that the firm distributes. Moreover, my results demonstrate that there is a region of probabilities, the region of maximal overweighting of probabilities, that if anticipated by the firms and implemented in the option stocks package, could maximize the benefits of using this payment modality. I discuss more applications of the proposed contract in the conclusion.

This paper contributes to several strands of literature. First, it adds to the literature of behavioral contract theory (See Koszegi (2014) for a review). The idea that a principal can take advantage of the cognitive biases of consumers and employees is not new and has been previously studied. For instance, DellaVigna and Malmendier (2004) and Heidhues and Koszegi (2010) study how firms can design contracts to take advantage of time-inconsistent individuals; Fehr et al. (2007), Goette et al. (2004), and Herweg and Mueller (2010) study the effects of higher powered incentives as well as the optimal design of incentives when workers form reference points that generate loss aversion; and Sandroni and Squintani (2007) study the effect of contracts when individuals are overconfident. To my knowledge this is the first paper studying the optimal design and implementation of labor contracts that take advantage of the regularity that individuals distort probabilities.

The results of this study also contribute to the literature of decision-making that specializes in the judgment of probabilities (see Wakker (2010) pp. 204 for an extensive list of references). I contribute to this literature in three ways. First, I elicit the probability weighting functions of the subjects using the two-step method developed by Abdellaoui (2000) and find that some of his results do not replicate when smaller monetary incentives and a larger sample are used. Second, I show that in the presence of large individual variation in the data, statistical analyzes that are commonly used in this literature could yield biased conclusions. Instead, I propose that an unbiased method to analyze these data are regressions at the individual level. Third, when regressions at the individual level are used, I find that individuals overweight small probabilities to a larger extent than documented by previous studies. Also, the data suggest that individuals also overweight medium-ranged probabilities but do not underweight large probabilities. These results correspond to a strong inverse-s shaped probability weighting function with unusually high elevation.

Finally, the results of this also study add to the methodological literature that studies the correct incentivization of subjects in the laboratory. I implement the probability contract using the random-lottery incentive system. According to Azrieli et al. (2018) this system is incentive compatible and preferable to offering pay for all decisions/rounds under very general conditions. However, some studies have found that when faced with this system, subjects may exhibit a form of reduction of probabilities that invalidate it (Holt, 1986). I contribute to this literature in two different ways. First, the results of the real-effort task show that subjects facing the random-lottery incentive system on average evaluate each experimental round as an isolated decision. If this was not the case then a) there would not be average performance differences across the treatments, and b) the degree to which individuals distort probabilities would not explain the totality of these performance differences. This finding is in line with the results by Hey and Lee (2005), Cubitt et al. (1998), and Starmer and Sugden (1991). Second, the results of the experiment present a caveat for the implementation of the random-lottery system when the researcher is interested in tasks that require effort from the part of the subjects: this mechanism could leverage higher performance than paying for all tasks/rounds.

2 The model

The theoretical framework considers a principal (she) that incentivizes an agent (he) to work on a task by offering a contract. The agent makes a decision about the output level, $y \in [0, \bar{y}]$, that he produces. His decision depends on the incentives included in the contract that is chosen by the principal. The following subsection presents the two contracts that are studied.

2.1 The contracts

2.1.1 Piece rate contract

As a benchmark of traditional contracts, I use a piece rate. A piece rate contract offers constant marginal monetary incentives, which means that the agent is incentivized to produce an additional unit of output irrespective of how many units he may already have delivered. This is a desirable property inasmuch as the agent delivers output up to some level at which offering an additional unit is too costly for him. Thus, the principal is able to incentivize the agent to to deliver as much performance as he could using the piece rate.[1]

Formally, the agent is offered $W = ay$, where $a > 0$ represents the monetary reward for every additional unit of output that he delivers. The agent derives utility from the monetary rewards and I assume that such utility can be represented by $b(\cdot)$, an increasing and two-times continuously differentiable function.

Assumption 1 (Basic utility from monetary rewards): $b(\cdot)$ is a C^2 function with $b(0) = 0$ and $b_y(\cdot) > 0$ for all y.

Note that I do not make any assumption about the sign of the second derivative of the basic utility. This is because the results of the model are going to be evaluated under the two signs that this derivative can attain.

The agent also experiences disutility from producing output, which captures the notion that working on the task requires attention, persistence, and effort. I model this disutility through

[1]This description of a piece rate is adequate for settings that focus on the intensive margin of labor supply and ignore the influence of incentive schemes on the extensive margin and labor search process. For a more comprehensive discussion about piece rates see Lazear (1986) and Gibbons (1987).

the function $c(y, \theta)$, an increasing, two-times-differentiable and strictly convex function.

Assumption 2 (Cost of producing output): $c(y, \theta)$ is a C^2 function with $c_y(0, \theta) = 0$ $c_y(y, \theta) > 0$, $c_{yy}(y, \theta) > 0$, and $c_{y\theta}(y, \theta) < 0$ for all y,θ.

The parameter $0 < \theta \leq 1$ represents the agent's ability. The last expression in Assumption 2 captures that agents with higher ability have flatter cost functions, which means that for them it is less costly to deliver higher levels of output as compared to agents with lower abilities.

All in all, the agent's utility when they are offered the piece rate can be written as

$$U(y) = b(ay) - c(y, \theta). \tag{1}$$

The agent maximizes this utility delivering a production level y^* that satisfies the following first-order condition[2]

$$ab_y(ay^*) - c_y(y^*, \theta) = 0. \tag{2}$$

Equation (2) shows that the optimal production level y^* increases with the agent's abilities θ. The assumption $c_{y\theta}(y, \theta) < 0$ guarantees this comparative static. Also, the optimal production level increases with higher powered incentives a. The assumption that ensures that y^* maximizes $U(y)$ guarantees this comparative static. Hence, higher abilities make it less costly for an agent to deliver high levels of output and higher monetary incentives increase the marginal utility of an additional unit of output, both of which motivate the agent to deliver higher output levels.

For illustrative purposes, consider the functional forms $c(y, \theta) = \frac{(y/\theta)^2}{2}$ and $b(ay) = \frac{(ay)^{1-\gamma}}{1-\gamma}$, with $-1 < \gamma < 1$. For these forms the optimal output level delivered by the agent has the closed-form solution $y^* = \left(\frac{\theta^2}{a^{\gamma-1}}\right)^{\frac{1}{1+\gamma}}$, which depicts the positive relationship between abilities and output, and monetary incentives and output .

2.1.2 The probability contract

Consider now the situation in which the agent is offered the probability contract. This incentive scheme also offers a monetary compensation that depends on the agent's performance on the task. However, in contrast to the piece rate, the agent's compensation depends on his own performance on the task with some probability. The scheme has the particularity that it allows

[2]A necessary condition for (2) to attain a maximum of $U(y)$ is $a^2 b_{yy} < c_{yy}(y, \theta)$. This implies that the cost function exhibits more curvature than the function capturing the utility from monetary rewards. Since I am particularly interested in situations in which offering one of the contracts yields changes in labor supply at the intensive margin, I impose this assumption wherever necessary.

the principal to choose this probability, giving her the potential to influence the agent's decision about how much output to deliver.

The timing of the contract stipulates that the principal moves first. Her choice consists of selecting $p \in (0, 1]$, the probability that the agent's compensation depends on his performance on the task. Once this choice has been made made, it is communicated to the agent before he performs the task. This implies that the agent's choice about how much output to deliver is influenced by the probability that performance counts toward his earnings. Finally, once the agent has worked on the task, a random device determines the agent's compensation depends on his performance on the task.

Given the aforementioned description, the probability contract can be formally written as the lottery $V = (By, p; 0, 1-p)$, where $B > 0$ represents the monetary compensation offered to the agent for each unit of output that is delivered. When facing this contract, the utility of the agent whose risk preferences are characterized by expected utility is

$$E(U(y)) = pb(By) - c(y, \theta). \tag{3}$$

The agent maximizes (3) choosing the production level y^{**} that satisfies the following first-order condition

$$pBb_y(By^{**}) - c_y(y^{**}, \theta) = 0. \tag{4}$$

Equation (4) shows that a higher probability p generates higher output. The intuition behind this comparative static is that the agent is motivated to deliver more output if it is more likely that output is paid. Moreover, equation (4) also shows that that higher abilities on the task, θ and higher monetary incentives, B yield higher optimal output.

As an illustration, consider the functional forms $c(y, \theta) = \frac{(y/\theta)^2}{2}$ and $b(By) = \frac{(By)^{1-\gamma}}{1-\gamma}$ with $-1 < \gamma < 1$. The optimal output level under these functional forms can be written as $y^{**} = \left(\frac{\theta^2 p}{B^{\gamma-1}}\right)^{\frac{1}{1+\gamma}}$. This closed-form solution depicts the positive effect of the agent's higher abilities, of higher monetary incentives, and of higher probabilities on output.

2.1.3 The probability contract and agents who distort probabilities

So far I have assumed that the agent evaluates probabilities accurately and the results of the previous analysis hinge on this assumption. In this subsection, I let the agent distort probabilities systematically as suggested by empirical evidence from the literature of decision-making (Bleichrodt and Pinto, 2000; Abdellaoui, 2000; Gonzalez and Wu, 1999; Wu and Gonzalez, 1996; Tversky and Fox, 1995; Tversky and Kahneman, 1992). As it will become evident, whether an agent distorts or not the probability that performance is evaluated is key to the effectiveness of the probability contract.

Consider now an agent who weights the probability p using the probability weighting function $w(p)$ which satisfies the following properties

Assumption 3 (Probability weighting function): $w(p) : [0,1] \to [0,1]$ with:

- $w(p)$ is C^2.
- $w_p(p) > 0$ for all $p \in [0,1]$.
- $w(0) = 0$ and $w(1) = 1$.
- $\lim_{p\to 0^+} w_p(p) = \infty$ $\lim_{p\to 1^-} w_p(p) = \infty$
- $\exists \tilde{p} \in (0,1)$ such that $w(\tilde{p}) = \tilde{p}$, $w(p) > p$ if $p \in [0,\tilde{p})$, and $w(p) < p$ if $p \in (\tilde{p},1]$.
- $\exists \hat{p} \in (0,1)$ such that $w_{pp}(p) < 0$ if $p \in [0,\hat{p})$ and $\pi_{pp}(p) > 0$ if $p \in (\hat{p},1]$.
- $\frac{-w_{pp}(p)}{w_p(p)} \in \Re$

According to Assumption 3, the probability weighting function $w(p)$ is a two-times continuously differentiable function that maps the unit interval onto itself and exhibits a positive slope everywhere. This function infinitely-overweights infinitesimal probabilities and infinitely-underweights near-one probabilities.

Moreover, the probability weighting function contains three fixed points: one at $p = 0$, another at $p = 1$, and an interior fixed point $\tilde{p} \in (0,1)$. I also assume that $w(p)$ has an inverse-s shape, which implies concavity up to a point $\hat{p}$ after which the function becomes strictly convex. Note that I do not assume that $\tilde{p} = \hat{p}$, which is a characteristic of early representations of probability weighting functions, and instead I let these two values differ. Finally, I impose the assumption that the curvature of the probability weighting function is bounded. Hence, the risk attitudes generated by the weighting function, alone, do not induce infinite risk-aversion or infinite risk-seeking.

The assumption that the agent evaluates probabilities according to $w(p)$ is not a sufficient characterization of the agent's preferences under risk. This is because a probability weighting function is not a theory of risk itself. Hence, I assume that the agent's preferences are characterized either by rank-dependent utility (Quiggin, 1982), (RDU from here onward) or by cumulative prospect theory (Tversky and Kahneman, 1992), (CPT from here onward). These two theories of risk accommodate probability weighting functions.

Under RDU the utility that the agent derives from monetary outcomes is represented by $b(\cdot)$ with the additional restriction that this function is either concave or linear. Thus, the rank-dependent utility of the agent is similar to that in equation (3), with the differences that p is replaced by $w(p)$ and that $b_{yy} \leq 0$.[3]

The agent maximizes his rank-dependent utility, choosing the production level, y_R^{**} that satisfies the following first-order condition.

$$w(p)Bb_y(By_R^{**}) - c_y(y_R^{**}, \theta) = 0. \tag{5}$$

The influence of the parameters of the model on the optimal production level is similar to those presented in previous analyses.The output level chosen by the agent increases with higher skills θ and with higher monetary incentives B. Additionally, higher probabilities of p generate higher output. However, in contrast to the case in which the agent evaluates probabilities accurately, this increment is non-linear: a probability increment within the interval $p \in [0, \hat{p})$ leads to smaller production increments as compared to an equally large probability increment taking place over the interval $p \in (\hat{p}, 1]$.

For the sake of illustration, consider the functional forms $c(y, \theta) = \frac{(y/\theta)^2}{2}$, and $b(By) = \frac{(By)^{1-\gamma}}{1-\gamma}$ with $0 < \gamma < 1$. The optimal output level under these functional forms has the closed-form solution, $y_R^{**} = \left(\frac{\theta^2 w(p)}{B^{\gamma-1}}\right)^{\frac{1}{1+\gamma}}$. This expression not only shows that higher monetary incentives and abilities raise output levels, but also that higher p leads to higher output.

Under CPT, which is a more descriptive version of RDU, the agent also distorts probabilities systematically, but in contrast to the agent with RDU preferences he evaluates the monetary outcomes offered by the contract relative to a reference point $r \geq 0$. The evaluation of outcomes relative to r is captured by the value function $v(y, r)$ that has the following properties.

Assumption 4 (CPT value function): $v(y, r)$ is the piecewise function,

$$v(y, r) = \begin{cases} b(By - \frac{r}{p}) \text{ , if } y \geq \frac{r}{pB}, \\ -\lambda b(\frac{r}{p} - By), \text{ if } y < \frac{r}{pB}. \end{cases}$$

with $r \geq 0$, $\lambda > 1$, $b(0) = 0$, $b_y(By - \frac{r}{p}) \geq 0$ for all y, $b_{yy}(By - \frac{r}{p}) < 0$ for $y > \frac{r}{pB}$, and $b_{yy}(By - \frac{r}{p}) > 0$ for $y < \frac{r}{pB}$.

The reference point $r \geq 0$ represents a monetary amount that the agent expects to receive

[3]According to RDU preferences an individual facing a lottery $(x_1, p_1; x_2, p_2; ...; x_n, p_n)$ ranks the prizes of the lottery in an increasing arrangement $x_1 < x_2 < ... < x_n$ and assigns decision weights to each of these outcomes according to their rank in the following way $\pi_n = w(p_n), \pi_{n-1} = w(p_{n-1} + p_n) - w(p_n)), ..., \pi_1 = 1 - \sum_j^n w(p_n)$. The ordered outcomes of the lottery enter the utility function through the function $b(\cdot)$ which is assumed to be concave or linear. In the context of our model the decision-weights are equivalent to the probability weighting functions, because in our setting an agent faces the lottery $(p, By; 1-p, 0)$, so that the decision weights are $\pi = w(p)$ and $\pi = 1 - w(p)$.

(Koszegi and Rabin, 2006; Pokorny, 2008; Abeler et al., 2011), a monetary amount that he received in the past, or an amount of money that he owns (Kahneman et al., 1991). Under the probability contract, the reference point is formulated as $\frac{r}{p}$. This is done to let the agent adjust his reference with respect to the different probabilities embedded in the contract. For instance, when the principal chooses lower probabilities, the agent adjusts his reference upward. This means that the agent accounts for the small likelihood that performance is paid and internalizes that he needs to deliver larger quantities of output to achieve the monetary target r.

Unlike the utility function in EUT and in RDU, the value function in CPT allows the agent to have different risk attitudes for gains, all monetary outcomes above the reference point, and losses, all monetary outcomes below the reference point. Above the reference point, the value function is concave and below the reference point this function exhibits convexity. This implies that the agent exhibits risk-averse attitudes in gains and risk-seeking attitudes in losses. Additionally, the agent is loss-averse, which means that for him losses loom larger than gains. This is represented by the parameter $\lambda > 1$ which enters the value function only for the domain of losses. Hence, to be compensated for a loss amounting $u(q)$, the agent is required to be paid an indemnity accruing $\lambda u(-q)$.

Finally, an agent with CPT preferences evaluates the probabilities that gains realize using the weighting function $w(p)$. In addition, this agent evaluates the probabilities that losses realize using a different weighting function, which I call $z(p)$. The functions $w(p)$ and $z(p)$ relate through the duality $z(p) = 1 - w(1-p)$. Hence, the probability weights that result from ordering the outcomes according to a rank from most desirable to least desirable is equivalent to the probability weights that result from ordering the outcomes according to a rank from least desirable to most desirable. [4]

All in all, the utility of the agent with CPT preferences is equal to

$$U(y,r) = \begin{cases} w(p)v(y,r) + (1 - w(p)v(0,r) - c(y,\theta), \text{ if } y \geq \frac{r}{pB} = 0, \\ w(p)v(y,r) + z(p)v(0,r) - c(y,\theta), \text{ if } y \geq \frac{r}{pB} > 0, \\ z(p)v(y,r) + w(1-p))v(0,r) - c(y,\theta), \text{ if } \frac{r}{pB} > y > 0. \end{cases} \tag{6}$$

Note that there is a key assumption made throughout the three considered theories of risk: the monetary outcomes, whether evaluated according to final positions, as in EUT or RDU, or relative to a reference point, as in CPT, are represented by the same function $b(\cdot)$. This assumption is introduced to simplify the comparison between the two contracts.

[4]According to CPT an individual facing a lottery $(x_1, p_1; x_2, p_2; ...; x_n, p_n)$ ranks the outcomes using an increasing arrangement $x_1 < x_2 < ...x_{r-1} < r < x_{r+1}... < x_n$ and evaluates the outcomes of the lottery relative to r through the function $v(y,r)$. The lottery outcomes $x_{r+1}, ...x_n$ are gains and the outcomes $x_1, ...x_{r-1}$ are losses. The individual assigns decision weights to gains in the following way $\pi_n = w(p_n), \pi_{n-1} = w(p_{n-1} + p_n) - w(p_n)), ..., \pi_{r+1} = 1 - \sum_{j=r+1}^{n} w(p_j)$ and assigns decision weights to losses in the following way $\pi_1 = z(p_1), \pi_2 = z(p_1 + p_2) - z(p_1)), ..., \pi_{r-1} = 1 - \sum_{j=r-1}^{n} z(p_j)$.

The agent with CPT preferences delivers a level of output y_C^{**}, which satisfies the following system of equations

$$Bw(p)b_y\Big(By_C^{**} - \frac{r}{p}\Big) - c_y(y_C^{**}, \theta) = 0, \text{ if } y \geq \frac{r}{pB}, \tag{7}$$

$$Bz(p)\lambda b_y\Big(\frac{r}{p} - By_C^{**}\Big) - c_y(y_C^{**}, \theta) = 0, \text{ if } y < \frac{r}{pB}. \tag{8}$$

Let us first consider the case in which the agent is in the domain of gains. According to equation (7), output increases with the monetary incentives offered by the contract B and with the abilities of the agent θ. Also, higher probabilities increase output in a non-linear way, with probability increments within the region $p \in (0, \hat{p})$ yielding smaller increases in output than equivalent probability increments in the region $p \in (\hat{p}, 1)$.

When the agent is in the domain of losses, the parameters of the model have a similar influence on the agent's decision. Specifically, higher abilities, higher monetary incentives, and higher probabilities increase output. Additionally, higher values of the loss aversion parameter, λ lead to higher output. This comparative static captures that the agent is willing to deliver higher levels of output to avoid experiencing losses.

Finally, the effect of a higher reference point r on output is ambiguous for both domains. Higher reference points shift to the right the marginal value function in both domains, $Bw(p)b_y(By - \frac{r}{p})$ and $Bz(p)\lambda b_y(\frac{r}{p} - By)$. This shift might suggest that higher reference points yield higher output. However, the solution to equations (7) and (8) features multiple equilibria, with equilibria at low levels of output becoming lower as r increases.[5] The intuition of this result is that higher reference points yield higher output up to a point after which they become unattainable and demotivate the agent. This is a well-known regularity of endogenous reference points (Dalton et al., 2016; Corgnet et al., 2015; Wu et al., 2008; Heath et al., 1999).

2.2 Contract comparisons

We are now in the position to compare the piece rate and the probability weighting contracts with respect to the output that they deliver. To simplify the analysis, I let the two contracts deliver similar monetary incentives for performance. This equivalence allows me to focus on

[5] To see how, consider a reference point of zero. In such a case only the curve capturing the marginal utility of gains, $Bw(p)b_y(By - \frac{r}{p})$, is relevant since only this curve attains positive solutions. The unique crossing point of this concave curve with the marginal cost of output $c_y(y, \theta)$ determines the optimal output level for this situation. Now, consider a small increment of r. Since higher values of the reference point shift the curve $Bw(p)b_y(By - \frac{r}{p})$ to the right, the crossing point between $c_y(y, \theta)$ and $Bw(p)b_y(By - \frac{r}{p})$ also shifts to the right. However, as r increases, the convex curve $Bz(p)\lambda b_y(\frac{r}{p} - By)$ becomes relevant. Among the multiple crossings between $Bz(p)\lambda b_y(\frac{r}{p} - By)$ and $c_y(y, \theta)$, there are equilibria at low levels of output which become lower as r increases.

the motivational effect of choosing different probabilities p. Formally, let $B = a/p$, so that $E(V) = ay$, which implies $E(V) = W$.

Note that under the assumed equivalence, the probability contract nests the piece rate contract. When the principal chooses to compensate output constantly, this is as $p \to 1$, then $B \approx a$. Moreover, in a setting in which the principal decides to evaluate output with very little frequency, this is as $p \to 0^+$, the monetary incentives for delivering one additional unit of output become very large to compensate for the low frequency at which performance in the task is paid.

I rewrite the first-order conditions describing the optimal output levels delivered by the agent under the different contracts in terms of the parameters a and p. Equation (4) can be written as

$$ab_y\left(\frac{a}{p}y^{**}\right) - c_y(y^{**}, \theta) = 0, \tag{9}$$

equation (5) becomes

$$\frac{w(p)}{p}ab_y\left(\frac{a}{p}y_R^{**}\right) - c_y(y_R^{**}, \theta) = 0, \tag{10}$$

and equations (7) and (8) become

$$a\frac{w(p)}{p}b_y\Big(\frac{a}{p}y_C^{**} - \frac{r}{p}\Big) - c_y(y_C^{**}, \theta) = 0, \text{ if } y \geq \frac{r}{a}, \tag{11}$$

$$a\frac{z(p)}{p}\lambda b_y\Big(\frac{r}{p} - \frac{a}{p}y_C^{**}\Big) - c_y(y_C^{**}, \theta) = 0, \text{ if } y < \frac{r}{a}. \tag{12}$$

Let us start comparing y^* and y^{**} from equations (2) and (9). This analysis compares the output levels delivered by an agent that has risk preferences characterized by expected utility when he faces each of the contracts. To build intuition about how these two production levels compare, consider the case in which the functional forms $c(y, \theta) = \frac{(y/\theta)^2}{2}$ and $b(By) = \frac{(By)^{1-\gamma}}{1-\gamma}$ are assumed. Under these forms, a necessary condition for $y^{**} > y^*$ is $\gamma \in (-1, 0]$.This condition implies that to supply higher production under the probability contract, the agent needs to have risk-seeking attitudes. Hence, the effectiveness of the probability contract depends, in this case, on the agent's curvature of the basic utility $b(\cdot)$. Proposition 1 generalizes this result using general $b(\cdot)$ and $c(y, \theta)$. All proofs are relegated to Appendix A.

Proposition 1: For an agent with ability $\tilde{\theta} \in (0, 1)$ and who evaluates probabilities using $w(p) = p$, then $y^{**} > y^*$ if $b_{yy}(\cdot) \geq 0$.

Proposition 1 shows that when the agent has risk attitudes characterized by expected utility theory, his risk attitudes, which stem from the curvature of $b(\cdot)$, determine the effectiveness of the probability contract. An agent with a curvature $b_{yy}(\cdot) < 0$ experiences disutility from the risk that is introduced by this contract and because of this disutility, he exhibits lower performance as compared to the case in which he was offered the piece rate. Alternatively, an agent with curvature $b_{yy}(\cdot) > 0$ derives utility from the risk introduced by the probability contract, which motivates him supply more performance than in the case in which he was offered the piece rate. These results hold for any $p \in (0, 1]$ that is chosen by the principal. The first prediction of the model is based on this result.

Prediction 1: Agents with a concave basic utility deliver lower output under the probability contract. This result is independent of the principal's choice $p \in (0, 1]$.

Let us now compare y^* and y_R^{**} from equations (2) and (10). This analysis compares the performance levels of an agent whose risk preferences are characterized by RDU when facing each of the contracts. To build intuition about how these two output levels compare, I first analyze the case in which the functional forms $c(y, \theta) = \frac{(y/\theta)^2}{2}$ and $b(By) = \frac{(By)^{1-\gamma}}{1-\gamma}$, are assumed. In contrast to the result from Proposition 1, an agent exhibiting concave utility for money, this is an agent with $0 < \gamma < 1$, might perform better under the probability weighting contract. This is possible since the inequality $y_R^{**} \geq y^*$ holds if $\frac{w(p)}{p^{1-\gamma}} > 1$, which does not only depend on the curvature of the basic utility γ, but also on p, which is chosen by the principal.

To understand the influence of the principal's choice p on the agent's performance, consider an agent with $\gamma = 0$. For him, $y_R^{**} \geq y^*$ is ensured if $\frac{w(p)}{p} > 1$, which holds when the principal chooses $p \in (0, \tilde{p})$. Hence, if the agent overweights the probability that is chosen by the principal, he is motivated to deliver higher output under the probability contract. Let us now relax the assumption that $\gamma = 0$. As the agent's utility for money becomes more concave, the condition ensuring $y_R^{**} \geq y^*$ becomes more stringent. Thus, to ensure $\frac{w(p)}{p^{1-\gamma}} > 1$ the principal needs to choose probabilities that induce larger degrees of overweighting of probabilities, which occurs at subsets of smaller probabilities of $p \in (0, \tilde{p})$.[6]

Proposition 2 generalizes the result that agents who distort probabilities according to $w(p)$, could deliver higher performance under the probability contract, even when they have risk averse attitudes stemming from the curvature of the basic utility $b(\cdot)$.

Proposition 2: For an agent with ability $\tilde{\theta} \in (0, 1)$, basic utility with $b_{yy}(\cdot) \leq 0$, $b_{yp}(\cdot) > 0$

[6]For $\gamma > 0$, the inequality $\frac{w(p)}{p^{1-\gamma}} > 1$ can be expressed as $\frac{ln(w(p))}{ln(p)} < (1 - \gamma)$, note that for this inequality to hold, the expression $\frac{ln(w(p))}{ln(p)}$ needs to become smaller as $\gamma \to 1$, which holds for smaller probabilities which ensure that $ln(p)$ attains larger negative values than $ln(w(p))$.

and $\lim_{p \to 0+} b_{yp}(\frac{ay}{p}) = 0$, and who evaluates probabilities according to $w(p)$, then $\exists p^* \in (0, \tilde{p})$ such that $y_R^{**} > y^*$ if $p < p^*$.

Proposition 2 shows that the principal could attain higher levels of performance if he offers the probability contract. However, to achieve this result, she is required to choose a probability p that induces a sufficiently large overweighting of probabilities in the agent. This result is robust to the agent having a concave function $b(\cdot)$, which induces risk averse attitudes.

The intuition behind this result is that the overweighting of small probabilities acts as a risk-seeking mechanism: by perceiving small probabilities to be larger, individuals more often choose lotteries that are unlikely to happen. Hence, by choosing a probability that is overweighted by the agent, the principal induces risk-seeking attitudes, which, when sufficiently strong, lead the agent to derive more utility under contracts that feature risk. Moreover, since the agent's decision depends on the marginal utility of output, the risk seeking agent supplies more output under the probability contract than under the piece rate.

There are two additional assumptions regarding the basic utility $b(\cdot)$ that guarantee the result presented in Proposition 2. The condition $b_{yp}(\frac{ay_R^{**}}{p}) > 0$ is a reasonable and intuitive restriction, it captures that lower probabilities induce higher basic disutility in the agent, which generates to lower output. Moreover, $\lim_{p \to 0+} b_{yp}(\frac{ay_R^{**}}{p}) = 0$ captures that close to zero probabilities yield a very high disutility, that motivate the agent not to deliver any output. A number of commonly used utility functions in the literature satisfy this requirement, for example $\left(\frac{ay}{p}\right)^r$ with $0 < r < 1$.

The result that the agent supplies more output under the probability contract depending on the probability chosen by the principal constitutes Prediction 2.

Prediction 2: Agents who distort probabilities systematically deliver higher output under the probability contract when the choice of p yields a sufficiently large degree of overweighting of probabilities.

Finally, l analyze the case in which the agent's risk preferences are characterized by CPT. It turns out that the conditions presented in Proposition 2 suffice to motivate the agent to supply more output under the probability contract. Thus, setting $p < p^*$ induces risk-seeking attitudes in the agent in both domains, making him more productive in the probability contract than in the piece rate. The complete analysis is presented in Appendix B

Prediction 2a: Agents that evaluate outcomes around a reference point and who distort probabilities systematically deliver higher output under the probability contract when the principal's choice, p, yields a sufficiently large overweighting of probabilities.

With these predictions in mind, I run a controlled laboratory experiment in which subjects exert effort on a task and are randomly assigned to different incentive schemes representing the different contracts studied in this model. The experiment investigates the effectiveness of the probability contract when it features either a high, medium, or a low probability. According to Prediction 1, irrespective of the probability at which it is implemented, the probability contract is outperformed by the piece rate if subjects have a concave $b(\cdot)$. However, if subjects distort probabilities systematically, according to an inverse s-shape probability weighting function, we can expect that the probability at which the contract is implemented matters. Specifically, subjects assigned to the treatment with low probability should exhibit higher performance if this probability is overweighted to a large degree. Moreover, the probability contract with high probability should yield lower performance if the subjects underweights the high probability to a large degree.

To conclude this section, let me emphasize that the focus of this theoretical framework was the agent's incentive compatibility constraint in the absence of any information asymmetries. In Appendix C, I provide a more general setup in which the principal implementing the probability contract also faces a participation constraint. Moreover, introducing information asymmetries about the risk preferences of the agent and how this shapes the contracts that are offered by the principal is a topic that is studied in a companion paper.

3 Experimental Method

The experiment was conducted at Tilburg University's CentERLAB in April 2017. The participants were all students at the university and were recruited using an electronic system. The data set consists of 15 sessions with a total of 172 subjects. On average, a session lasted approximately 80 minutes. Between eight and eighteen subjects took part in a session. The currency used in the experiment was Euros. I used Z-Tree (Fischbacher, 2007) to implement and run the experiment. Subjects earned on average 15.83 Euros. The instructions of the experiment are presented in Appendix D.

The experiment consisted of two parts. Upon arrival, participants were informed that their earnings from either part one or those from part two would become their final earnings and that this would be decided by chance at the end of the experiment. In the first part of the experiment, subjects performed a task that required their effort and attention. The task consisted of summing five two-digit numbers.[7] Each summation featured randomly drawn numbers by the computer, ensuring similar levels of difficulty among participants. When a

[7] This task has been widely used by other researchers (See for instance Niederle and Vesterlund (2007), and Buser et al. (2014))

participant knew the answer to a summation, he/she submitted it using the computer interface. Immediately after submission, a new summation appeared on the participant's screen and the participant was invited to solve it. In total, subjects had 10 rounds of four minutes to complete as many summations as they could.

There were four treatments that differed with respect to how monetary incentives were given to the subjects. Participants were randomly assigned to one of these four treatments. The baseline treatment is *Piecerate*. Subjects assigned to this treatment were paid 0.25 euros for every correctly solved summation. The other three treatments also offered monetary incentives contingent on the subject's performance on the task. However, in these treatments performance in some of the rounds, chosen at random at the end of the experiment, counted toward the subject's earnings.

The treatments *LowPr*, *MePr* and *HiPr* featured a low, a medium, and a high probability, respectively, that performance in a round counted toward earnings. These treatments represent the probability contract implemented with different probabilities. Specifically, in LowPr one round was randomly chosen at the end of the experiment and only performance in that round was paid. Similarly, in MePr and HiPr, three and five rounds, respectively, were randomly chosen at the end of the experiment and performance in those rounds was paid. [8]

As in the theoretical framework, the monetary compensation offered in Piecerate, LowPr, MePr and HiPr was calibrated such that subjects received similar monetary incentives across these treatments. For instance, a subject assigned to LowPr received 2.50 Euros for each correct summation in the round that was chosen for compensation. This payment was tenfold of what a subject assigned to Piecerate earned for each correct summation. Such difference in monetary payments exactly accounts for the likelihood difference across treatments that performance in a round is paid. Similarly, subjects assigned the MePr and HiPr treatments received a compensation of 0.85 and 0.50 Euros for a correctly solved task for the rounds that were chosen for compensation.[9]

The probabilities governing the treatments LowPr, MePr and HiPr were chosen according to the most common findings in the literature of decision-making: subjects distort probabilities according to an inverse-s shape probability weighting function with a fixed point at approximately $p = 1/3$ (see Wakker (2010) pp.204 for a complete list of references). If subjects in this experiment are not the exception, then they should overweight the probability that a round

[8] Note that the experiment was designed under the assumption of isolation or narrow bracketing, which is strongly supported by the literature of experimental economics (See for instance Hey and Lee (2005) and Cubitt et al. (1998)). Specifically, if a subject was to choose the amount of effort to exert in a task as if each round was isolated, then indeed these treatments generate uncertainty about whether the supplied performance in a round counts towards performance. However, if a subject treats the whole real-effort task as one decision, these treatments do not generate uncertainty.

[9] These compensations correspond to three times and two times what a subject received in Piecerate, respectively.

is chosen with 10% chance, underweight the probability that a round is chosen with 50% chance, and approximately evaluate accurately the probability that a round is chosen with 30% chance. Therefore, the experiment was designed to observe performance differences across the treatments if the incentives from probability distortions generate strong incentives.

Once the last round of the real-effort task was over, participants were asked to state their beliefs about how well they did in the task. This belief elicitation is used to investigate whether subjects anticipated the effect of the treatments on performance. A subject received a bonus of one euro if his answer was exactly equal to the number of correct summations that he performed over the ten rounds. This elicitation was unanticipated and the monetary compensation given to the subject„ when he provided a correct answer, was small as compared to the other sources of earnings in the experiment. These two characteristics ensure incentive compatibility since a subject was rewarded for reporting accurate beliefs and he could not use the bonus to hedge against his own performance in the real-effort task (Blanco et al., 2010).

In the second part of the experiment, the subjects' task was to choose between two binary lotteries in multiple occasions. This part of the experiment was designed to elicit the utility and the probability weighting functions of every subject. To elicit these two functions, I used the two-step method developed by Abdellaoui (2000). This method has the advantage of not making assumptions about the way in which subjects evaluate probabilities nor about the way in which subjects evaluate monetary outcomes. [10]

This part of the experiment consisted of 11 decision sets. Each decision set elicited an outcome or a probability that made a subject indifferent between two binary lotteries. Indifference was found with a sequence of choices that reduced the space of outcomes/probabilities through bisection. Specifically, a subject needed to express his preference between two initial lotteries. After having made a choice, either an outcome or the probability, depending on the decision set, of one lottery changed in a way that depended on the subject's choice. When facing the new situation, the subject was invited to choose between the two available lotteries again. This process was repeated four times in every decision set. Table 1 presents an example illustrating the bisection procedure. The left panel shows an example of the bisection procedure used to elicit outcomes and the right panel shows an example of the bisection procedure used to elicit probabilities. Participants knew that one of their choices would be chosen at random at the end of the experiment, and that that random chosen lottery would be played to determine their earnings of the second part of the experiment.

Decision sets 1 to 6 constitute the first step of Abdellaoui (2000)'s methodology. These decision sets elicit a sequence of outcomes $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ that made the subject indifferent

[10] A drawback of this method is error propagation. Since the choices of a subject are chained, a mistake in a choice could lead to mistakes in subsequent choices. I overcome this disadvantage by adding questions in between that are not used in the analysis.

between a lottery $L = (x_{j-1}, 2/3; 0.5, 1/3)$ and a lottery $R = (x_j, 2/3; 0, 1/3)$ for any $j = \{1, ..., 6\}$. These lotteries were designed so that the elicited sequence of outcomes yields utility levels that are equally spaced, i.e. $u(x_j) - u(x_{j-1}) = u(x_{j-1}) - u(x_{j-2})$.[11] The starting point of the program, x_0, was set such that the monetary outcomes used in the lotteries always reflected the earnings of the subject in the first part of the experiment. Specifically, x_0 was set at 2/5 of what a subject earned in the first part of the experiment. The advantage of using monetary outcomes of similar magnitude as the incentives offered in the real-effort task, is that I can correlate the behavior of the subjects in the task with the elicited preferences. Subjects were not informed about this calibration and were not informed about their earnings in the first part of the experiment until the end of the session.

Decision sets 7 to 11 constitute the second step and were designed to elicit a sequence of probabilities $w^{-1}(p_1), w^{-1}(p_2), w^{-1}(p_3), w^{-1}(p_4)$ and $w^{-1}(p_5)$ with $p_{j-1} = j - 1/6$ and $j = 2, ..., 6$. These probabilities made the subjects indifferent between the lottery $L = (x_6, p_{j-1}; x_0, 1 - p_{j-1})$ and the degenerate lottery x_{j-1}. The elicited probabilities yield equally spaced probability weights, i.e. $w(p_j) - w(p_{j-1}) = w(p_{j-1}) - w(p_{j-2})$.

Once the second part of the experiment was over, subjects were presented with feedback about their performance in the real-effort task, the round(s) that counted toward payment if assigned to LowPr, MePr or HiPr and whether their belief was correct. Also, subjects were informed about the lottery that was chosen for compensation for the second part of the experiment, its realization, and their final earnings. Finally, participants completed a questionnaire that asked them about their willingness to take risks, such as willingness to take health-related risks, willingness to take job-related risks, willingness to take risks while driving, and general willingness to take risks. These questions where taken from Dohmen et al. (2012).The questionnaire also featured measures of self-efficacy and a self-reported measure of mathematical abilities. Appendix D presents the questionnaire.

4 Treatment effects

4.1 Performance

I compare the average performance produced by each treatment. Performance is defined as the total number of correctly solved summations by a participant in all rounds. Table 2 shows the descriptive statistics of performance by treatment. This table suggests that the probability contract implemented with 10% probability delivers higher performance than the piece rate contract. Specifically, a subject assigned to the LowPr treatment solves on average 20.56 %

[11] Note that indifference between L and R implies $w(1/3)u(x_{j-1}) + (1 - w(1/3))u(0.5) = w(1/3)u(x_j) + (1 - w(1/3))u(0)$ which is equivalent to $u(0.5) - u(0) = u(x_j) - u(x_{j-1})$ for any $j = \{1, ..., 6\}$

Table 1: Example of the Abdellaoui's (2000) algorithm

#	Alternatives	Interval	Choice	Alternatives	Probabilities	Choice
1	L=$(1, .66; .50, .33)$ R=$(3.7, .66; 0, .33)$	[1, 6.4]	L	L=$(x_1, 1)$ R=$(x_6, .50; 1, 0.5)$	$[0, 1]$	L
2	L=$(1, .66; .50, .33)$ R=$(5.05, .66; 0, .33)$	[3.7,6.4]	R	L=$(x_1, 1)$ R=$(x_6, .75; 1, 0.25)$	$[.5, 1]$	L
3	L=$(1, .66; .50, .33)$ R=$(4.38, .66; 0, .33)$	[3.7,5.05]	R	L=$(x_1, 1)$ R=$(x_6, .87; 1, 0.13)$	$[.75, 1]$	R
4	L=$(1, .66; .50, .33)$ R=$(4.04, .66; 0, .33)$	[3.7,4.38]	L	L=$(x_1, 1)$ R=$(x_6, .81; 1, 0.19)$	$[.75, .87]$	L
5	L=$(1, .66; .50, .33)$ R=$(4.21, .66; 0, .33)$	[4.04,4.38]	L	L=$(x_1, 1)$ R=$(x_6, .85; 1, 0.15)$	$[.81, .87]$	L
End		$x_1 \in [4.21, 4.38]$			$p_1 \in [85, 87]$	

Note: This table illustrates the bisection method used to elicit utility and probability functions. The lotteries in this table are expressed in the form $(A, p; B, 1-p)$ where A and B are prizes, and p is a probability. The left panel presents the bisection method to elicit utility and the right panel presents the bisection method to elicit probability functions.

more summations than a subject assigned the Piecerate ($t(84.454) = 2.361, p = 0.010$). The effect size of this difference is 0.5 standard deviations which is significantly different from zero.[12]

By contrast, the probability contract implemented with higher probabilities produces similar average performance as compared to the piece rate. A subject assigned the MePr treatment solved on average 87.9 correct summations, and a subject assigned the HIPR treatment solved on average 83.7 correct summations, neither of which are statistically different from the average correct summations under Piecerate, 81.37 summations.[13]

Among the three treatments representing the probability contract, the LowPr produces higher average performance. This contract yields an increase in average performance of 17% as compared to the HiPr ($t(75.215) = 2.232, p = 0.014$), and an increase in average performance of 11% when compared to MePr ($t(79.575) = 1.478, p = 0.0716$). All in all, the analysis of the descriptive statistics of performance demonstrates that LowPr yields higher average performance than the Piecerate and HiPr treatments.[14]

To control for factors that may influence these results other than the treatment assignment,

[12] The significance of this effect size was evaluated with a bootstrapped 95% confidence interval with 10000 repetitions.

[13] The t-tests of these comparisons are ($t(83) = 1.005, p = .159$) and ($t(82.44) = -0.386, p = .692$), respectively.

[14] An alternative analysis is to compare average performance by round and by treatment rather than analyzing the sum of correct summations by treatment. Overall, I find the same qualitative results. I find evidence to reject the null hypothesis that LowPr and Piecerate produce on average the same performance in a round ($p = 0.010$). Moreover, I find evidence to reject the null hypothesis that LowPr and HIPR yield on average the same performance by round ($p = 0.013$), but I find no evidence to reject the null that LowPr and MePr produce on average the same performance by round once the Bonferroni correction is performed, ($p = 0.07$). Finally, I find no evidence to reject the null hypothesis that MePr and HiPr yield the same average performance by round ($p = 0.47$) and that Piecerate produces the same average performance than MePr and HiPr, ($p = 0.69$) and ($p = 0.32$), respectively.

Table 2: Descriptive statistics of performance by treatments

Treatment	LowPr	MePr	HiPr	Piecerate	Total
Mean	98.116	87.9	83.75	81.377	87.686
Median	91	87	82.5	77	85
St.dev.	34.659	28.134	24.358	31.684	30.412
N	43	40	44	45	172

Note: This table presents the average, median and standard deviations of performance in the experiment by experimental treatment. Performance is defined as the total number of summations produced by a subject in the real-effort task.

I regressed the subject's performance on treatment dummies, the subject's gender, self-reported measures of risk attitudes, a self-reported measure of self-efficacy on the task, a self-reported measure of mathematical skills, and performance beliefs. Table 3 presents the estimates of the regression. All in all, the results of the regression analysis confirm the aforementioned findings. First, the coefficient associated with the LowPr treatment is significant and positive, which supports the result that a subject assigned to this treatment produces higher average performance than a subject assigned the Piecerate treatment. Second, the estimate of LowPr is significantly higher than the estimate associated with HiPr ($F(1, 159) = 6.58$). Also, the estimate of LowPr is significantly higher that of MePr ($F(1, 159) = 6.02$). Therefore, among the probability contracts, the LowPr yields the highest performance.

An explanation for these results is that the probability contract with 10% yields higher performance because, as opposed to the piece rate, it circumvents wealth effects. This could explain the higher average performance of subjects assigned to LowPr. To dispel this possibility, I show that this difference in performance appears as of the first round, which demonstrates that in the absence of wealth effects the LowPr outperforms the piece rate. Subjects assigned to LowPr achieve on average 7.44 tasks in the first round, which is significantly higher than the 6.46 average tasks solved by the subjects assigned to Piecerate ($t(85.54) = 1.443, p = 0.07$). Also, I find no average differences in performance in the first round between the Piecerate and MePr and the Piecerate and HiPr.

The results of the real-effort task support Prediction 2 and Prediction 2a. These predictions state that the efficiency of the probability contract depends on the probability at which this contract is implemented. From the performance data it seems that subjects overweight the probability $p = 0.10$ to a large degree, given that they exhibit higher performance under the probability contract with $p = 0.10$ than under Piecerate. However,from this data it also seems that subjects do not underweight the probability $p = 0.5$ since HiPr and Piecerate yield on average similar performance levels. The latter result is at odds with the majority of the results from the literature that find that individuals underweight probabilities after the range $p \in [0.3, 0.4]$. Section 5 demonstrates that subjects in the experiment overweight small probabilities and do not underweight large and moderate probabilities.

4.2 Performance Beliefs

The previous subsection demonstrated that the probability contract yields higher performance than the piece rate. In this subsection, I investigate whether the subjects understand and anticipate the non-monetary incentives included in the probability contract. To that end, I analyze the data on the subject's beliefs about their own performance in the task. If the subjects understood these incentives, their beliefs about their own performance should reflect the performance differences across the treatments.

Table 4 presents the descriptive statistics of the performance beliefs by treatment. Overall, I find no belief differences between the treatments, which suggests that subjects do not anticipate the non-monetary incentives of the contract. Particularly, the average beliefs of the subjects in LowPr, MePr, and HiPr are not statistically different from those of subjects assigned to Piecerate.[15] Also, I find no significance difference in average performance beliefs between the LowPr and the MePr($t(80.749) = 0.206, p = 0.837$), the HiPr and MePr ($t(78.819) = 0.956, p = 0.3418$) or the LowPr and HiPr treatments($t(83.241) = 1.1885, p = .1190$).

To account for factors that may be driving these results other than the treatment assignment, I regress the performance belief of a subject on treatment dummies, the subject's gender, self-reported measures of risk attitudes over different domains, a self-reported measure of self-efficacy on the task, and a self-reported measure of mathematical skills. Table 5 presents the OLS estimates. The regression estimates corroborate the aforementioned results. The coefficients associated with the MePr, HiPr and LowPr treatments are not significant, suggesting no statistical differences between the average beliefs of subjects assigned to these treatments and those of subjects assigned to Piecerate. Furthermore, there is no evidence to reject the null hypothesis that the coefficients associated with LowPr and MePr are equal ($F(1, 160) = 0.29$), as well as no evidence to reject the null hypothesis that the coefficients of LowPr and HiPr are equal ($F(1, 160) = 0.09$).

All in all, the belief data suggest that subjects do not internalize the psychological incentives included in the probability contract. Understanding the reasons behind the gap between performance and performance beliefs is beyond the scope of this paper and requires methodologies that allow the researcher to study in more detail the cognitive processes underlying probability judgments. I conjecture that this gap can be explained in light of the findings of Berns et al. (2008), who show that the perception of probabilities primarily involves the perceptual stage of

[15] The statistics of these t-tests are ($t(85.98) = 1.190, p = 0.1186$), ($t(82.843) = 0.976, p = 0.331$), and ($t(84.91) = -0.10, p = 0.920$), respectively.

Table 3: Regression of performance on treatments

	(1) Performance	(2) Performance	(3) Performance
LowPr	16.739**	16.649**	16.567***
	(7.090)	(6.448)	(6.193)
MePr	6.522	1.839	0.016
	(6.487)	(6.130)	(6.236)
HiPr	2.372	1.844	-0.388
	(5.985)	(5.474)	(5.669)
Gender		-2.326	0.689
		(4.413)	(4.567)
Task difficulty		-6.412***	-6.231***
		(1.920)	(1.920)
Math skills		3.071**	3.864***
		(1.245)	(1.271)
Self-efficacy		0.516	-0.084
		(1.168)	(1.123)
Risk general			-2.431**
			(0.973)
Risk occupation			-0.180
			(0.843)
Risk health			0.012
			(0.947)
Risk drive			2.030**
			(0.960)
Constant	81.378***	74.735***	79.340***
	(4.726)	(9.120)	(9.383)
R^2	0.045	0.25	0.291
Observations	172	172	172

Note: This table presents the estimates of the Ordinary Least Squares regression of the model $Performance_i = \beta_0 + \beta_1 MePr + \beta_2 LowPr + \beta_3 HiPr + Controls'\Gamma + \epsilon_i$, with $E(\epsilon|MepR, LowPr, HiPr, Controls) = 0$. "Performance" is the number of correctly solved sums in the first part of the experiment, "LowPr", "MePr" and "HiPr" are dummy variables that capture whether the subject was assigned to the treatment with low, medium or high probability, respectively. The controls considered in this model are "Gender" a variable which indicates the gender of the participant, "Math Skills " which captures the self-reported mathematical skills of the subject, "Task Difficulty " which captures the self-reported difficulty to perform the task."Risk general", "Risk occupation", "Risk health", and "Risk drive, capture the self-reported willingness to take risks in general, at their studies, with their health and while driving. Robust standard errors in parenthesis. *** denotes significance at the 0.01 level, ** denotes significance at the 0.05 level, * denotes significance at the 0.1 level.

Table 4: Descriptive statistics of performance beliefs by treatments

Treatment	LowPr	MePr	HiPr	Piecerate	Total
Mean	83.86	82.025	74.022	73.177	78.123
Median	75	80	75	64	74.5
St. Dev.	40.864	40.156	36.139	43.318	40.147
N	43	40	44	45	172

Note: This table presents the average, median and standard deviations of performance beliefs by treatment. A performance belief is the estimate of a subject about the number of correct summations solved in the real-effort task.

the cognitive process rather than stages of consciousness that allow individuals to internalize these distortions. Without being able to internalize the perception of probabilities, subjects are unlikely to understand the non-monetary incentives of the probability contract as it is suggested by the beliefs data.

5 The probability weighting functions

A crucial assumption underlying Prediction 2 and Prediction 2a is that the agent distorts probabilities in a systematic way, overweighting small probabilities and underweighting large to moderate probabilities. In this section I investigate whether subjects exhibit such a systematic pattern of probability weighting. To that end I analyze the data from the second part of the experiment, which features the subjects' preferences over lotteries that were designed to elicit their utility and probability weighting functions. An unbiased analysis of the data suggests that subjects severely overweighted small probabilities. Therefore, the risk attitudes that stem from the probability weighting functions of the subjects could explain the aforementioned treatment differences.

As explained in section 3, the second part of the experiment consisted of 11 decision sets. Decision sets 1 to 6 elicited the sequence of outcomes $\{x_1, x_2, x_3, x_4, x_5, x_6\}$. With these data I can analyze the properties of the utility functions of the subjects. I find that the majority of subjects exhibit linear utility functions over the monetary outcomes that were considered with the lotteries. This finding is in line with the notion that individuals exhibit linear utility over small monetary outcomes (Rabin, 2000) and that utility elicitations using Wakker and Deneffe (1996)'s trade-off method require large monetary outcomes to capture the curvature of the utility function of a subject. Appendix E presents the complete analysis of these data.

Decision sets 7 to 11 were designed to elicit the sequence of probabilities

$$\{w^{-1}(p_1), w^{-1}(p_2), w^{-1}(p_3), w^{-1}(p_4), w^{-1}(p_5)\}$$

Table 5: Regression of performance beliefs on treatments

	(1)	(2)	(3)
	Beliefs	Beliefs	Beliefs
LowPr	10.683	7.644	6.861
	(8.977)	(8.409)	(8.507)
MePr	8.847	3.601	2.041
	(9.055)	(8.818)	(9.053)
HiPr	0.845	0.755	-0.571
	(8.452)	(7.983)	(8.183)
Gender		13.912**	14.770**
		(5.816)	(5.857)
Task difficulty		-5.812**	-5.897**
		(2.645)	(2.664)
Math skills		3.907**	4.255**
		(1.766)	(1.856)
Self-efficacy		-0.488	-0.720
		(1.527)	(1.546)
Risk general			-0.792
			(1.301)
Risk occupation			-0.322
			(1.381)
Risk health			1.518
			(1.563)
Risk drive			0.360
			(1.408)
Constant	73.178***	59.412***	59.362***
	(6.461)	(13.487)	(13.877)
R^2	0.014	0.165	0.174
Observations	172	172	172

Note: This table presents the estimates of the Ordinary Least Squares regression of the model $Belief_i = \beta_0 + \beta_1 MePr + \beta_2 LowPr + \beta_3 HiPr + Controls'\Gamma + \epsilon_i$, with $E(\epsilon | MepR, LowPr, HiPr, Controls) = 0$. "Beliefs" is the subject's predicted number of correctly solved sums in the first part of the experiment , "LowPr", "MePr" and "HiPr" are dummy variables that capture whether the subject was assigned to the treatment with low, medium or high probability, respectively. The controls considered in this model are "Gender" which that indicates the gender of the participant, "Math Skills " which captures the self-reported mathematical skills of the subject, "Task Difficulty " which captures the self-reported difficulty to perform the task."Risk general", "Risk occupation", "Risk health", and "Risk drive, capture the self-reported willingness to take risks in general, at their studies, with their health and while driving. Robust standard errors presented in parentheses. *** denotes significance at the 0.01 level, ** denotes significance at the 0.05 level, * denotes significance at the 0.1 level.

. These data is analyzed to understand the properties of the probability weighting function of the subjects. I first analyze these data performing the analyses presented in Abdellaoui (2000) and Bleichrodt and Pinto (2000). These analyses feature individual classifications as well as analyses using averages and medians of each elicited probability. I also analyze the data using a variety of regressions at the individual level. I show that in the presence of large individual variation, the latter analysis yields more reliable conclusions.

Analysis at the individual level

The first analysis of the data is done at the individual level and is based on Bleichrodt and Pinto (2000). The aim of this analysis is to perform a classification of subjects according to the shape of their probability weighting function. To perform this classification, I construct the variable $\partial^j_{j-1} \equiv \frac{w(p_j)-w(p_{j-1})}{w^{-1}(p_j)-w^{-1}(p_{j-1})}$, which denotes the average slope of the probability weighting function of a subject between the successive probabilities j and $j-1$, and the variable $\bigtriangledown^j_{j-1} \equiv \partial^j_{j-1} - \partial^{j-1}_{j-2}$, which denotes the change of the average slope of the weighting function between successive probabilities.

Given that the treatments seek to study the motivational effect of probability distortions at small and large probabilities, I am especially interested in the shape of the probability weighting function at the smallest and largest considered probabilities. Thus, the sign of $\bigtriangledown^{0.33}_{0.16}$ and that of $\bigtriangledown^1_{0.83}$ are computed for each subject. If a subject exhibits $\bigtriangledown^{0.33}_{0.16} < 0$, then his probability weighting function has the property of lower subadditivity (LS), implying that near-zero probability intervals have larger weights than mid-range probability intervals. In other words, a subject with LS overweights small probabilities. Moreover, if a subject has $\bigtriangledown^1_{0.83} > 0$, then his probability weighting function exhibits the property of upper subadditivity (US), which means that near-one probability intervals have larger weights than mid-range probability intervals. This implies that this subject underweights large probabilities.

Table 6 shows that 38% of subjects exhibit LS, so the majority of subjects did not overweight small probabilities. Moreover, 75% of the subjects exhibit US and thus underweight large probabilities. Furthermore, only 31% of the subjects present probability weighting functions with both LS and US.

Additionally, the signs of $\bigtriangledown^j_{j-1}$ as j increases allows me to examine the shape of the weighting function of a subject throughout all the considered probabilities. A subject was classified as having a concave probability weighting function if at least three (out of five) $\bigtriangledown^j_{j-1}$ had a negative sign and he did not exhibit US. Alternatively, a subject had a convex probability weighting function if at least three (out of five) $\bigtriangledown^j_{j-1}$ were positive and he did not exhibit LS. Note that these classifications allow for the possibility of response error.

Table 6 shows that 51% of the subjects exhibit concave weighting functions and that 13%

of the subjects exhibit convex probability weighting functions. Thus, under the framework of rank-dependence, more subjects in the experiment were pessimistic than optimistic. This conclusion is consistent with the aforementioned finding that most of the subjects exhibited LS and did not exhibit US. Furthermore, the proportion of subjects in the experiment with either concave or convex probability weighting functions is higher than that reported by Bleichrodt and Pinto (2000), who finds that only 15% of the subjects have probability weighting functions with either of these shapes.

Table 7 shows that the majority of subjects in the experiment have probability weighting functions with the certainty effect (CE) exceeding the possibility effect (PE). For these subjects the probability weighting function is steeper at the highest considered probability than at the lowest considered probability, i.e.$w(\frac{1}{6}) < 1 - w(\frac{5}{6})$. However, the proportion of subjects for which PE exceeds CE is not negligible as it constitutes close to 42% of the subjects.

To account for the possibility that subjects have CPT preferences with reference points other than zero, I perform the previous analyses with the additional assumption that the monetary equivalent of a subject's performance belief in the real-effort task is his reference point.[16] This assumption presumes that a subject evaluates the outcomes of the lotteries relative to his expectations about his current earnings in the experiment (Koszegi and Rabin, 2006). This non-zero reference point is addressed as "Beliefs" from here onward.

It is important to emphasize that the nature and intuition of the classification under the assumption that Beliefs is the reference point differs from the original classification. The reason for this difference is that the data does not admit enough $\bigtriangledown_{j-1}^{j}$s to analyze the shape of the probability weighting function of a subject for the domain of gains as well as for the domain of losses. Instead, I analyze the shape of a subject's probability weighting function for the domain wherein the majority of his $\bigtriangledown_{j-1}^{j}$s lie. Thus, this analysis could shed light on whether subjects who have most of their choices in the domain of losses exhibit weighting functions of different shape than subjects who have most of their choices in the domain of gains.

The results are also presented in Table 6 and Table 7. I find that the main conclusions of the original analysis are robust to the assumption that Beliefs is the reference point. Specifically, I find that there is a larger proportion of subjects with US than those with LS in the domain of gains and in the domain of losses. Moreover, I find that the proportion of subjects exhibiting concave and/or convex weighting functions is very large and comparable across the two domains. Finally, the data suggest that in both domains, more subjects have probability weighting functions with the CE exceeding the PE.

In conclusion, the analysis of the data at the individual level suggest that the majority

[16]Given that the lotteries in the second part of the experiment only feature positive prizes, subjects with CPT preferences with a reference point equal to zero cannot be distinguished from subjects with RDU preferences. However, subjects with CPT preferences with reference points larger than zero can be distinguished from subjects with RDU preferences, since their preferences in the two domains were elicited and can therefore be analyzed.

Table 6: Classification of subjects according to the shape of their probability weighting function

Reference Point	Domain	Convex	Concave	LS	US	LS & US
No/Zero	No/Gains	89	23	66	129	53
Beliefs	Gains	45	9	28	63	22
Beliefs	Losses	44	14	38	66	31

Note: This table presents the classification of subjects according to the shape of their probability weighting function. Subjects are classified as having a probability weighting function with upper subadditivity, lower subadditivity or both. Also, subjects are classified as having a convex or concave probability weighting function if they do not exhibit lower subadditivity and upper subadditivity respectively. This classification depends on the sign of $\bigtriangledown_{j-1}^{j}$. The first row presents the classification when the analysis is performed with all the data. The second and third columns feature the analysis assuming that Beliefs is a reference point. The second row presents the analysis when the monetary outcomes of the lotteries are above the reference point, whereas the third row presents the analysis when the monetary outcomes of the lotteries are below the reference point.

of subjects underweights probabilities. Under the framework of rank-dependence, this finding implies that most of the subjects had a pessimistic attitude about the realization of the outcomes of the lotteries. This conclusion is enhanced by the result that for more subjects the CE exceeds PE. A convex probability weighting function or a function that features underweighting of some of the considered probabilities can capture such pessimism.

Analysis at the aggregated level

The second analysis features the sequence $\{w^{-1}(p_1), w^{-1}(p_2), w^{-1}(p_3), w^{-1}(p_4), w^{-1}(p_5)\}$ aggregated across all the subjects. Table 8 presents the medians, means and standard deviations of the aggregated sequence. Additionally, Table 9 presents sign tests that investigate the number of positive and negative counts of the difference $w^{-1}(p_j) - j/6$ for each $j = \{1, ..., 5\}$. These two tables suggest that subjects on average underweighted small probabilities. Specifically, the weight $w(p) = 1/6$ is on average mapped by the probability $p = .306$, which is significantly larger than 1/6, and the weight $w(p) = 1/3$ is on average mapped by the probability $p = .448$, which is significantly larger than 1/3. Additionally, I do not find evidence that subjects underweighted or overweighted large probabilities. The weight $w(p) = 2/3$ is on average mapped by the probability $p = 0.638$ and the weight $w(p) = 5/6$ is on average mapped by the probability $p = 0.78$, both of which are not statistically different from 2/3 and 5/6, respectively.

I reach similar conclusions when it is assumed that subjects have CPT preferences with Beliefs as the reference point. The probability that maps $w(p) = 1/6$ is significantly larger than 1/6 for the domain of gains and for the domain of losses, and the probability that maps $w(p) = 1/3$ is larger than 1/3 but only for the domain of gains. One difference with respect

Table 7: Classification of subjects according to the steepness near certainty and near possibility

Reference Point	Domain	CE	PE	CE =PE
No/Zero	No/Gains	73	83	16
Beliefs	Gains	31	35	13
Beliefs	Losses	42	48	3

Note: This table presents the classification of subjects according to the slope of their probability weighting function at the lowest and largest probabilities considered. Subjects are classified as having probability weighting function where the certainty effect (CE) exceeds the possibility effect (PE) if the slope of the function at $w^{-1}(5/6)$ exceeds the slope at $w^{-1}(1/6)$. Subjects were classified with a probability weighting where PE exceeds CE if the slope of the function at $w^{-1}(1/6)$ exceeds the slope at $w^{-1}(5/6)$.

to the original analysis which is that I find weak evidence that subjects in the domain of gains underweight large probabilities[17]. I do not find evidence of such underweighting of large probabilities for the domain of losses.

The analysis of the aggregated data also suggests that subjects underweighted small probabilities. This conclusion contradicts most of the regularities from the literature of decision-making and probability judgments. For instance, it contradicts the findings by Abdellaoui (2000), from who the elicitation of probability weights was borrowed, and those by Bleichrodt and Pinto (2000).

Individual variation and regression analyses

The previous analyses of the data could yield erroneous conclusions since they do not account for all sources of individual variation in the data. This is not a minor problem, given that individual variation is sizable in the data as it is evidenced by the standard deviation of the elicited probabilities presented in Table 8. On the one hand, the individual analysis cannot inform us about the relevance of the subjects' deviation from perceiving probabilities accurately. This is, a small deviation from the accurate perception of probabilities is treated in the same way as a large deviation. On the other hand, extrapolating the average probability weighting function from the averages of each of the elicited probabilities fails to take into account that with their choices, individuals could deviate in various and different ways from each average probability. Omitting this variation yields downward biased estimates when a linear relationship between weights and probabilities is assumed. [18]

[17] The difference $w^{-1}(5/6) - 5/6 > 0$ is only significant at the 10% significance level

[18] To see how denote with $\bar{p_j}$ the average of some probability j. Assume that the relationship between weights and probabilities is described by $w(\bar{p_j}) = c_1 + s_1\bar{p_j} + \epsilon_j$. However, the researcher chooses to represent this

Table 8: Means, Medians and Standard Deviation of $w^{-1}(p)$

Probability	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
$w^{-1}(1/6)$	.306	.234	.240	.412	.390	.298	.249	.234	.273
$w^{-1}(2/6)$	.448	.406	.252	.553	.578	.298	.383	. 359	.289
$w^{-1}(3/6)$	.524	.484	.253	.582	.609	.308	.492	.484	.298
$w^{-1}(4/6)$	.638	.671	.244	.677	.765	.299	.628	.640	.282
$w^{-1}(5/6)$	.781	.859	.217	.807	.890	.233	.763	.828	.224
Ref.Point	No/Zero			Beliefs			Beliefs		
Domain	No/Gains			Gains			Losses		
N	173			79			93		

Note: This table presents the average, medians and standard deviation of the sequence $w^{-1}(1/6), w^{-1}(2/6), w^{-1}(3/6), w^{-1}(4/6), w^{-1}(5/6))$ elicited in the second part of the experiment using decision sets 7 until 11. Columns 2,3, and 4, present the mean, median, and standard deviation of the probabilities respectively. Columns 5,6, and 7 present the mean, median, and standard deviation of the sequence of probabilities when the reference point is assumed to be the monetary equivalent of a subject's beliefs about performance and the outcomes implied for the lotteries are below the reference point. Columns 8,9, and 10 present the mean, median, and standard deviation of the sequence of the sequence of probabilities when the lottery prizes are above the reference point.

Table 9: Counts of $w^{-1}(p) - p > 0$ and $w^{-1}(p) - p < 0$

$w^{-1}(p) - p)$	>0	<0	>0	<0	>0	<0
$p = 1/6$	117***	55	61***	18	56**	37
$p = 2/6$	105***	67	55***	24	50	43
$p = 3/6$	79	93	40	39	39	54
$p = 4/6$	87	85	44	35	43	50
$p = 5/6$	89	83	49 *	30	40	53
Ref.Point	No/Zero		Beliefs		Beliefs	
Domain	No/Gains		Gains		Losses	
N	173		79		93	

Note: This table presents the number of positive and negative events of the difference $w^{-1}(p) - p$ for $p = 1/6, 2/6, 1/2, 4/6, 5/6$. The significance of the count is tested with one tailed sign tests. Columns 2 and 3 present the number of events using all the data. Columns 4, 6, 7 and 8 present the number of events when the monetary equivalent of a subject's beliefs about his own performance is assumed to be his reference point. Columns 4 and 5 present the number of negative and positive events when the lottery prizes are above the assumed reference point. Columns 6 and 7 present the number of negative and positive events when the lottery prizes are below this reference point. *** denotes significance at the 0.01 level, ** denotes significance at the 0.05 level, * denotes significance at the 0.1 level.

In this section, I analyze the data on probabilities performing regressions of the elicited probabilities on probability weights at the individual level. Such analysis fully accounts for individual variation in the data, since individual variation between and within probabilities are included in the variance-covariance matrix of the different regressions.[19] Furthermore, throughout this analysis, I compare the estimates of the regressions with those reported in previous studies. These comparatives are useful since they could inform us about the degree to which subjects distorted probabilities in the experiment. However, it is important to emphasize that any resulting difference cannot only be attributed to differences in preferences but also, for example, to the different methods used to elicit preferences, the type of uncertainty used and the monetary stakes used to elicit the subject's preferences.

Table 10 presents the regression estimates when different functional forms are assumed. I use the most relevant and well-known functions proposed in the literature. Let us first study the case in which it is assumed that the probability weighting function follows $w(p) = c + sp$. This linear function is referred as the neo-additive weighting function (Chateauneuf et al., 2007; Wakker, 2010). To allow for $c + s > 1$, which is what the previous analyses of the data suggest, I perform the regressions allowing for truncation at $w(0)$ and $w(1)$. The maximum likelihood estimation yields estimates that comply with $\hat{c} > 0$ and $\hat{c}+\hat{s} < 1$. Therefore, subjects on average overweighted small and medium-ranged probabilities and underweighted large probabilities.[20] Furthermore, the estimation yields a larger constant and a smaller slope than in Abdellaoui et al. (2011), which implies that subjects in my experiment overweighted small probabilities more severely and also overweighted medium-ranged probabilities.

The function proposed by Tversky and Kahneman (1992) describes a relationship between probabilities and weights captured by $w(p) = \frac{p^{\psi}}{(p^{\psi}+(1-p)^{(1-\psi)})^{\frac{1}{\psi}}}$. Table 10 presents the estimates of this parametric form when the non-linear least squares method is used. I find a lower estimate than those reported in previous studies, which are in the range $\hat{\psi} \in [0.60, 0.75]$ (See Bleichrodt and Pinto (2000), Abdellaoui (2000), Wu and Gonzalez (1996), Camerer and Ho (1994), and Tversky and Kahneman (1992)). Thus, subjects in the experiment had an inverse-s average probability weighting function inducing more severe overweighting of small probabilities than that documented in previous studies.

Finally, Prelec (1998) proposed a two-parameter function i.e. $w(p) = exp(-\beta(-ln(p)))^{\alpha}$.

---

relationship using $w(p_{ij}) = c_2 + s_2 p_{ij} + \epsilon_{ij}$. The OLS estimates of each process are $\hat{s_1} = \frac{Cov(\bar{p_j}, w(\bar{p_j}))}{Var(\bar{p_j})}$ and $\hat{s_2} = \frac{Cov(p_{ij}, w(p_{ij}))}{Var(p_{ij})}$. By construction we have that $Var(p_{ij}) \geq Var(\bar{p_j})$. This is because $Var(p_{ij})$ captures individual variation within probabilities (in each p_j) and between probabilities (across $p_j's$), whereas $Var(\bar{p_j})$ only captures individual variation between probabilities. Moreover, given that Abdellaoui's (2000) method imposes $w(\bar{p_{ij}}) = w(p_j)$, then $Cov(p_{ij}, w(p_{ij})) = Cov(p_{ij}, w(\bar{p_j})) \leq Cov(\bar{p_j}, w(\bar{p_j}))$. Therefore, $\hat{s_1} \geq \hat{s_2}$.

[19]However, a disadvantage of such an analysis is that I assume some functional form of the probability weighting function, which imposes structure on the relationship between weights and probabilities.

[20]When estimating the regression without truncation, these conclusions also hold.

When this function is assumed, the resulting estimate $\hat{\alpha}$, which is statistically lower than one, suggests that the average probability function has an inverse-s shape. Moreover, the estimate $\hat{\beta}$, which is also statistically lower than one, suggests high elevation of the weighting function. Previous estimations of this probability weighting function report larger values of α and β (Murphy and Ten Brincke, 2018; Haridon et al., 2018; Fehr-duda, 2012; Abdellaoui et al., 2011; Bleichrodt and Pinto, 2000). Hence, subjects in this experiment had an average probability weighting function with more curvature, this means a function with a stronger inverse-s shape, and higher elevation. These two characteristics imply that on average small probabilities were overweighted to a larger degree than in previous studies and that medium-ranged probabilities were also overweighted.

All in all, the estimates of the regressions are suggestive of subjects exhibiting an average probability weighting function with a strong inverse-s shape and high elevation. These two characteristics entail that subjects overweighted severely small probabilities, even to a larger degree than documented by previous studies, and that subjects also overweighted medium-ranged probabilities.[21] Furthermore, these two characteristics also imply that large probabilities are moderately underweighted or not underweighted at all. This pattern of probability distortion is able to explain the result that LowPr generates more output than Piecerate and the finding that there is no difference in average performance between HiPr and Piecerate. The next subsection intends to reconcile the findings of the first part of the experiment with the findings of the second part of the experiment.

When it is assumed that Beliefs is the reference point, the conclusion that subjects had probability weighting functions with a strong inverse-s shape and with high elevation is robust and holds for both domains. Moreover, this alternative analysis also yields that the probability weighting function in the domain of gains has a higher elevation, which contradicts most empirical evidence on probability weighting functions for gains and losses.[22] Despite this striking result, I can conclude that even when the reference point is shifted from zero to Beliefs, the conclusions that subjects in the experiment severely overweighted small probabilities and also overweighted medium-ranged probabilities hold.

[21] A possible explanation for the finding that elevation is higher than in previous studies is that the probability weighting function is sensitive to the magnitude of the outcomes of the lotteries. According to Bruhin et al. (2010) and Etchart-Vincent (2004) lotteries with low stakes yield probability weighting functions with higher elevation as compared to probability weighting functions elicited using lotteries with higher stakes. I used monetary outcomes that reflect the subjects' monetary gains in the real-effort task, which are usually not higher than 20 euros. Thus, it is possible that I find probability weighting functions with an elevation as in previous studies, had I used similar stakes.

[22] This conclusion must be taken with a grain of salt since this result may stem from the restriction that the prizes of the lotteries considered in this part of the experiment are strictly positive. This implies that the estimations in the domain of losses may be subject to a bias arising from censoring the choice data around zero. The inclusion of lotteries with negative prizes may yield probability weighting functions with higher elevation in the domain of losses.

Table 10: Parametric estimates of the weighting function

Neo-additive (truncated)			
$w(p) = c + s * p$			
$\hat{c}$	.194 ***(.021)	.228*** (.024)	.155*** (.024)
$\hat{s}$	.566*** (.035)	.463*** (.037)	.686 *** (.044)
Log-Likelihood	220.288	75.200	166.842
Tversky & Kahneman (1992)			
$w(p) = \frac{p^{\psi}}{(p^{\psi}+(1-p)^{(1-\psi)})^{\frac{1}{\psi}}}$			
$\hat{\psi}$	.598*** (.016)	.597*** (.012)	.785*** (.037)
Adj. R^2	0.838	0.827	0.866
Prelec (1998)			
$w(p) = exp(-\beta(-ln(p)))^{\alpha}$			
$\hat{\alpha}$	.284***(.025)	.143*** (.025)	.357*** (.033)
$\hat{\beta}$	.841 ***(.0148)	.596 ***(.024)	. 944***(.0195)
Adj. R^2	0.864	0.907	0.851
N	860	304	550
Ref.Point	No/Zero	Beliefs	Beliefs
Domain	No/Gains	Gains	Losses

Note: This table presents the estimates of the probability weighting function when parametric estimates are assumed. The upper panel presents the maximum likelihood estimates of the equation $w(p) = c + s(p)$ when truncation at $p = 0$ and at $p = 1$ is assumed. The second panel from top to bottom presents the non-linear least squares estimation of the function $w(p) = \frac{p^{\psi}}{(p^{\psi}+(1-p)^{(1-\psi)})^{\frac{1}{\psi}}}$. The last panel presents the non-linear least squares estimates of the function $w(p) = exp(-\beta(-ln(p)))^{\alpha}$. The first column in all the panels presents the estimates when all the data is used. The second and third columns present the estimations when it is assumed that Beliefs is the reference point and only data for the domain of gains and the domain of losses, respectively, is used for the estimations. Standard errors in parenthesis. *** denotes significance at the 0.01 level, ** denotes significance at the 0.05 level, * denotes significance at the 0.1 level.

6 What drives the treatment effect?

According to the theoretical framework, agents with a linear utility function and who overweight small probabilities deliver more output under the probability contract if the implemented probability is largely overweighted. In this section, I demonstrate that the higher performance exhibited by subjects assigned to LowPr is indeed driven by their perception of probabilities.

To examine the role of probability distortions in explaining the treatment differences, I include variables capturing the perception of probabilities of a subject in model (2) from Table 3. Additionally, to account for risk attitudes stemming from the subject's utility curvature, I also include variables capturing the shape of the utility function of a subject. [23] A first approach to this analysis is to include in the model a variable that indicates whether a subject has a probability weighting function with Lower Subadditivity (LS). After all, subjects with LS overweight the probability $p = 1/6$ and are thus more likely to perform better under LowPr than under the other treatments.

Table 11 presents the estimates of the OLS regressions. Models (1) and (2) show that the coefficient associated with LowPr remains significant. Model (2) shows that the interaction between LowPr and LS is significant and positive. Thus, a subject with a weighting function with LS and who is assigned to LowPr exhibits higher performance than a subject assigned to Piecerate. However, a subject without LS and who is assigned to LowPr could still exhibit higher output than a subject assigned to Piecerate. Hence, the treatment effect can be partially explained by whether a subject exhibits LS or not. A shortcoming of this analysis is that it cannot inform us about the degree to which subjects overweighted small probabilities. This is a relevant shortcoming since according to the theoretical framework, an agent needs to exhibit a sufficiently large degree of probability overweighting to exhibit risk-seeking attitudes that generate higher output under LowPr.

If the effectiveness of the probability contract depends on the degree to which subjects distort probabilities, a better analysis must instead include variables capturing the degree to which subjects distort probabilities. For that purpose I construct the variable "Overweight" which features the sum $\sum_{j=i}^{5} w(p_i) - p_i$. If the mechanism underlying the treatment effect is the degree to which subjects overweight probabilities, then introducing this variable in the statistical model, along with an interaction with LowPr, should capture the totality of the treatment effect. Models (3) and (4) in Table 11 show that the coefficient of LowPr remains significant. Also model (4) shows that the interaction between LowPr and Overweight is not significant. Thus, this variable representing probability overweighting is unable to explain the treatment effect. Appendix F shows that the sum $\sum_{j=1}^{2} w(p_j) - p_j$ is also unable to explain the

[23] Note that I do not use model (3) from Table 3 to avoid confounds between the elicited risk preferences and self-reported risk attitudes.

treatment effects,

The main disadvantage of using Overweight as a measure of probability distortions, is that it cannot inform us about whether a subject distorts probabilities due to his limited capacity to understand likelihoods, which is a cognitive component, or whether he is optimistic or pessimistic, which is a motivational component (See Wakker (2010) pp. 205 for a complete discussion). Thus, this variable is unable to distinguish a pessimistic subject, who always gives more probability weights to the lowest payoff of a lottery, from a subject who underweights large and medium-ranged probabilities. This confound is even more problematic when analyzing data that is aggregated across subjects. For instance, in a sample with a majority of subjects being pessimistic, but also with some subjects strongly overweighting small probabilities due to likelihood insensitivity, the variable Overweight, when aggregated across subjects, might lead to the conclusion that, on average, there are no probability distortions in the sample.

Given the nature of the proposed contract, I am interested in establishing whether a subject's sensitivity to likelihoods is driving the treatment effect. To understand whether this cognitive channel of probability distortion is the determinant of the treatment effect, I construct indexes that separate the motivational and cognitive components of probability weighting. As recommended by Wakker (2010), due to their parsimony and easy interpretation, I use the estimates of the probability weighting function when it is assumed that this function follows the neo-additive structure. Thus, for each subject, I estimated

$$w(p_j) = c_i + s_i p_j + \epsilon_j.$$

The estimate $\hat{s}$ is an index of the subject's sensitivity to probabilities, which I call "ACurvature". Higher values of this index imply more responsiveness to likelihoods. Furthermore, the difference between the intercept of the weighting function at $p = 0$ and the intercept of this function at $p = 1$, $\hat{c} - (1 - \hat{c} - \hat{s})$, is an index of optimism which I denominate "Optimism".

I included these two indexes in the statistical model along with their interaction with the treatment indicators. Model (6) in Table 11 shows that the assignment to LowPr, alone, does no longer yield higher performance on the real-effort task and that the coefficient associated with the interaction between ACurvature and LowPr is significant and positive. In contrast, the interaction between Optimism and LowPr is not significant. These findings imply that the treatment effect is explained by the interaction between the assignment to the treatment and the index representing the subject's responsiveness to probabilities.

My interpretation of these results is that a subject with a weighting function exhibiting a degree of sensitivity to probabilities and a high elevation, experiences more severe overweighting of small probabilities than a subject with a weighting function with the same elevation but who is less responsive to probabilities. Figure 1 illustrates this situation. The figure shows that a

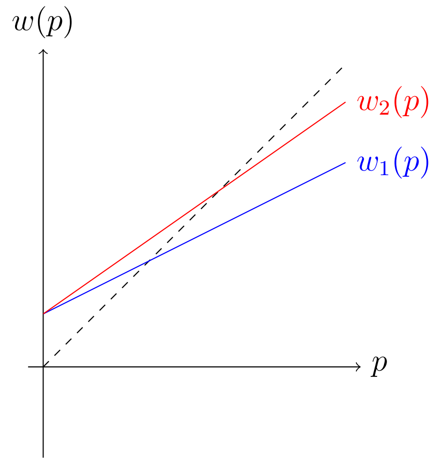

Figure 1: Lower curvature implies more severe overweighting of small probabilities

subject with $w_2(p)$ exhibits more severe overweighting of small probabilities and overweights more probabilities than a subject with $w_1(p)$. Note that the property of high elevation is found on average for all the regressions presented in Table 10. Hence, the intuition behind these estimates is that the higher is ACurvature for a subject, which at the same time entails more severe overweighting of small probabilities, the higher becomes his performance when assigned to LowPr. Moreover, being assigned to LowPr alone, no longer yields higher performance as compared to the piecerate.

A possible explanation for these findings is that the subjects' assignment to the treatments affects the way in which they evaluate probabilities in the second part of the experiment. For instance, subjects assigned to LowPr may exhibit more probability overweighting than subjects in Piecerate since they were exposed to risk in the first part of the experiment. If this were the case, the analysis presented in this section would not be suggestive of a mechanism explaining the treatment effect, but a consequence of the treatment itself. To rule out this possibility, it suffices to show that the degree to which subjects distort probabilities is equivalent across the treatments. The linear regression $w(p) = c + s_k(p)$ with a different k for each treatment shows that there is no evidence to reject the null hypothesis that $s_1 = s_2 = s_3 = s_4$ ($\chi^2(3) = 80, p = 0.848$). Moreover, the degree of overweighting of probabilities captured by Overweight, is on average the same for LowPr and MePr ($t(81) = -0.820, p = 0.414$), LowPr and HiPr($t(86) = -0.539, p = 0.591$), and LowPr and Piecerate ($t(85) = -0.088, p = 0.929$). The same conclusion is reached with the variable Overweight$_S$.

Finally, the findings that the probability contract yields higher performance than the piece rate and that this improvement is explained by the degree to which subjects overweight probabilities, support the assumption that individuals make the decision about how much output to deliver as if each round was considered in isolation. If this were not the case, and instead subjects made the decision about how much output to deliver in all rounds at the outset of the experiment, then we should observe no performance differences across treatments, since under such condition none of the treatments induces risk, implying no difference in non-monetary incentives across the treatments.[24] This result is in line with the findings by Hey and Lee (2005), and Cubitt et al. (1998).

7 Conclusion

This paper introduced a novel incentive scheme designed to take advantage of the behavioral regularity that individuals distort probabilities. A theoretical framework and a laboratory experiment demonstrated that the proposed contract yields higher output than a standard piece rate when i) both contracts offer similar monetary returns for performance, and ii) the probability contract is implemented with a probability that is sufficiently overweighted by the agent. Additionally, I show that the pattern used by individuals to distort probabilities and the degree to which probabilities are distorted affect the efficiency of the contract.

The non-monetary incentives of the contract, crucial to achieving its efficiency, stem from a well-established regularity of human behavior and are therefore available to the principal across different populations of agents. Hence, principals interested in novel ways to motivate labor supply, and willing to use performance-pay schemes, should consider implementing this contract or variants of it that preserve its fundamental property: to incentivize effort by means of inducing probability distortions.

At this point, I would like to discuss additional applications of the probability contract. Perhaps the most straightforward application, is monitoring. Consider a setting in which the principal could choose among different monitoring technologies that allow her to find out the amount of effort that was exerted by an agent on a task. On the one hand, more advanced technologies, and also more expensive ones, allow her to be more precise. On the other hand, cheaper technologies have a component of randomness in their effort assessment. In such a setup, the principal has the choice between choosing the expensive option and compensate the agent exclusively based on his exerted effort, or, in light of the findings of this paper, choosing a cheaper technology that features a probability of accurate monitoring that is overweighted by the agent, which in turn motivates him to supply more output.

Another application is bonuses rewarding the achievement of performance targets. Consider a setting in which the principal sets a milestone or performance target to the agent. Conditional on reaching such target, the agent is compensated with a performance-based compensation. In a setting in which the principal and the agent know that the performance target is attained

[24]Even in a setting in which the mapping from effort to performance is not deterministic, we should observe no differences in performance across treatments since the rate of errors are equally distributed across treatments due to the randomization.

Table 11: The mechanism driving the treatment effects

	(1) Perfomance	(2) Perfomance	(3) Perfomance	(4) Perfomance	(5) Perfomance	(6) Perfomance
LowPr	15.277**	16.385**	16.110**	22.014**	15.544**	10.918
	(6.744)	(7.554)	(6.834)	(8.956)	(6.801)	(6.654)
MePr	1.466	1.423	1.725	1.619	-0.351	-0.044
	(6.166)	(6.163)	(6.191)	(6.240)	(6.237)	(6.285)
HiPr	1.917	2.011	1.830	1.476	0.945	0.994
	(5.538)	(5.577)	(5.493)	(5.474)	(5.412)	(5.460)
LowSub	2.161	2.940				
	(4.999)	(5.806)				
LowPr*LowSub		16.776*				
		(9.320)				
Overweight			0.003	0.001		
			(0.004)	(0.004)		
LowPr*Overweight				0.008		
				(0.006)		
Acurvature					2.614**	2.157*
					(1.025)	(1.117)
LowPr*Acurvature						5.389***
						(1.824)
Optimism					3.409	2.403
					(2.768)	(3.136)
LowPr*Optimism						6.828
						(4.862)
Self-efficacy	0.740	0.738	0.833	0.724	0.648	0.676
	(1.190)	(1.194)	(1.183)	(1.165)	(1.160)	(1.187)
Task difficulty	-6.713***	-6.690***	-6.823***	-6.744***	-6.875***	-6.812***
	(1.943)	(1.961)	(1.935)	(1.922)	(1.944)	(1.961)
Math skills	2.814**	2.825**	2.723**	2.868**	2.493*	2.502*
	(1.295)	(1.296)	(1.279)	(1.274)	(1.282)	(1.306)
Gender	-2.486	-2.486	-2.873	-2.787	-1.976	-2.208
	(4.593)	(4.608)	(4.626)	(4.629)	(4.543)	(4.786)
Mixed Utility	-11.432	-11.499	-11.232	-12.981	-11.709	-12.032
	(11.015)	(11.022)	(10.930)	(10.883)	(11.129)	(11.259)
Convex Utility	9.475	9.105	11.657	17.242	9.826	10.566
	(12.431)	(12.140)	(12.137)	(12.742)	(12.006)	(12.096)
Linear Utility	-8.943	-8.983	-10.643	-11.620	-9.377	-9.296
	(9.402)	(9.399)	(9.600)	(9.654)	(9.518)	(9.595)
Constant	83.581***	83.302***	87.942***	86.687***	83.665***	84.269***
	(13.731)	(13.943)	(14.505)	(14.495)	(13.891)	(14.015)
Adj. R^2	0.265	0.278	0.284	0.294	0.270	0.286
N	172	172	172	172	172	172

Note: This table presents the estimates of the Ordinary Least Squares regression of the model $Performance_i = \beta_0 + \beta_1 LowPr + \beta_2 LowPr * LS + \beta_3 MePr + \beta_4 HiPr + \beta_5 LS + Controls'\Gamma + \epsilon_i$, with $E(\epsilon|MePr, LowPr, HiPr, Controls, LS) = 0$, which are presented in columns (1) and (2). The OLS estimates of the model $Performance_i = \beta_0 + \beta_1 LowPr + \beta_2 LowPr * Overweight \beta_3 MePr + \beta_4 HiPr + \beta_5 Overweight + Controls'\Gamma + \epsilon_i$, with $E(\epsilon|MePr, LowPr, HiPr, Controls, Overweight) = 0$ which are presented in columns (3) and (4). Also, the OLS estimates of the model $Performance_i = \beta_0 + \beta_1 LowPr + \beta_2 LowrPr * Acurvature + \beta_3 LowrPr * Optimism + \beta_4 MePr + \beta_5 HiPr + \beta_6 * Acurvature + \beta_7 optimism + Controls'\Gamma + \epsilon_i$, with $E(\epsilon|MepR, LowPr, HiPr, Controls, Optimism, Acurvature) = 0$ "Performance" is the number of correctly solved sums in the first part of the experiment, "LowPr", "MePr" and "HiPr" are dummy variables that capture whether the subject was assigned to the treatment with low, medium or high probability of outcome evaluation, respectively. "LS" captures whether a subject has a probability weighting function with lower subadditivity. "Overweight" captures the general probability overweighting of a subject through the index $\sum_{j=1}^{5} w(p_j) - p_j$. For each subject the maximum likelihood estimation of the regression $w(p) = c + sp$ was performed with truncation at $w(0)$ and $w(1)$. "Optimism' ' is the index $\frac{2c+s}{2}$ and "ACurvature" is the index c. The controls considered in this model are "Gender" which captures the gender of the participant, "Belief " which captures the performance belief of the subject, "Math Skills " which captures the self-reported mathematical skills of the subject, "Task Difficulty " which captures the self-reported difficulty to perform the task. Robust standard errors in parenthesis. *** denotes significance at the 0.01 level, ** denotes significance at the 0.05 level, * denotes significance at the 0.1 level.

with some probability, the principal could set a high target, entailing a low probability of achievement, along with a high compensation for performance. According to my findings, this practice would take advantage of the probability perception of the agent, who overweights the probability that he could attain the milestone. Such probability distortion generates higher performance than if he was given a low target combined with a low compensation.

Finally, I would like to state that this study has limitations that should be addressed in future research. I discuss two of the most relevant limitations. First, even though there are obvious advantages of using controlled laboratory environments to compare the effectiveness of incentive schemes (Charness and Kuhn, 2011), these advantages come at the cost of external validity. A more general understanding of the motivational effect of the probability contract, requires performing a similar test in a situation that features longer working periods, more powerful monetary incentives, more meaningful tasks and a more natural setup for subjects. Field experiments incorporating these characteristics are an ideal tool to evaluate the external validity of the findings reported in this paper.

Second, some findings from the literature of decision-making suggest that individuals perceive probabilities differently when they are described, i.e. probabilities from description, as compared to situations where probabilities are experienced (Hertwig et al., 2004; Hau and Pleskac, 2008). If this is the case, then the probability contract could have ambiguous effects on performance in settings where the contract is implemented repeatedly. Specifically, in first implementations the predictions and results of the paper hold since the probabilities of the contract are described. However, in subsequent implementations the agent has gained experience about the probabilities that govern the contract. This experience might lead to lower performance if the probability weighting function is transformed in a way such that it becomes convex in the region of probabilities that is implemented by the principal. Studying the properties of the probability contract in a setup that admits repeated implementations could shed light on the efficiency of the contract when the probabilities governing the contract are not exclusively described.

References

Abdellaoui, M. (2000). Parameter-Free Elicitation of Utility and Probability Weighting Functions. *Management Science*, 46(11):1497–1512.

Abdellaoui, M., Baillon, A., Placido, L., and Wakker, P. P. (2011). The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation. *American Economic Review*, 101(2):695–723.

Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference Points and Effort Provision. *American Economic Review*, 101(2):470–492.

Ashraf, N., Bandiera, O., and Lee, S. S. (2014). Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior and Organization*, 100:44–63.

Azrieli, Y., Chambers, C., and Healy, P. J. (2018). Incentives in experiments: A theoretical analysis. *Journal of Political Economy*, 126(4):1472–1503.

Bandiera, O., Barankay, I., and Rasul, I. (2013). Team Incentives: Evidence From a Firm Level Experiment. *Journal of the European Economic Association*, 11(5):1079–1114.

Berns, G. S., Capra, M. C., Chappelow, J., Moore, S., and Noussair, C. N. (2008). Nonlinear Neurobiological Probability Weighting Functions For Aversive Outcomes. *Neuroimage*, 39(4):2047–2057.

Blanco, M., Koch, A. K., Engelmann, D., and Normann, H.-T. (2010). Belief Elicitation in Experiments: Is there a Hedging Problem ? *Experimental Economics*, 13(4):412–438.

Bleichrodt, H. and Pinto, L. (2000). A Parameter-Free Elicitation of the Probability Weighting Function in Medical Decision Analysis. *Management Science*, 46(11):1485–1496.

Bruhin, A., Fehr-Duda, H., and Epper, T. (2010). Risk and Rationality: Uncovering Heterogeneity in Probability Distortion. *Econometrica*, 78(4):1375–1412.

Buser, T., Niederle, M., and Oosterbeek, H. (2014). Gender, competition and career choices. *The Quarterly Journal of Economics*, 129:1409–1447.

Camerer, C. F. and Ho, T.-h. (1994). Violations of the Betweenness Axiom and Nonlinearity in Probability. *Journal of Risk and Uncertainty*, (8):167–196.

Charness, G. and Kuhn, P. (2011). Lab labor: What can labor economists learn from the lab? *Handbook of Labor Economics*, 4(Part A):229–330.

Chateauneuf, A., Eichberger, J., and Grant, S. (2007). Choice under Uncertainty with the Best and Worst in Mind: Neo-additive Capacities. *Journal of Economic Theory*, 137(1):538–567.

Corgnet, B., Gómez-Miñambres, J., and Hernán-Gonzalez, R. (2015). Goal Setting and Monetary Incentives : When Large Stakes Are Not Enough. *Management Science*, 61(12):2926–2944.

Cubitt, R. P., Starmer, C., and Sugden, R. (1998). On the Validity of the Random Lottery Incentive System. *Experimental Economics*, 1:115–131.

Dalton, P. S., Ghosal, S., and Mani, A. (2016). Poverty and Aspirations Failure. *Economic Journal*, 126:165–188.

DellaVigna, S. and Malmendier, U. (2004). Contract Design and Self-Control: Theory and Evidence. *The Quarterly Journal of Economics*, 119(May):353–402.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2012). The intergenerational transmission of risk and trust attitudes. *Review of Economic Studies*, 79(November 2011):645–677.

Etchart-Vincent, N. (2004). Is Probability Weighting Sensitive to the Magnitude of Consequences? An Experimental Investigation on Losses. *The Journal of Risk and Uncertainty*, 28(3):217–235.

Falk, A. and Ichino, A. (2006). Clean Evidence on Peer Effects. *Journal of Labor Economics*, 24(1):39–57.

Fehr, E., Klein, A., and Schmidt, K. M. (2007). Fairness and Contract Design. *Econometrica*, 75(1):121–154.

Fehr-duda, H. (2012). Probability and Risk : Foundations and Economic Implications of Probability-Dependent Risk Preferences. *Annual Review of Economics*, 4(1):567–593.

Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Gibbons, R. (1987). Piece-Rate Incentive Schemes. *Journal of Labor Economics*, 5(4):413–429.

Gibbons, R. and Roberts, J. (2013). *The Handbook of Organizational Economics*. Number 3. Princeton Unversity Press, Princeton.

Goette, L., Huffman, D., and Fehr, E. (2004). Loss aversion and labor supply. *Journal of the European Economic Association*, 2(2-3):216–228.

Goldstein, W. and Einhorn, H. (1987). Expression theory and the preference reversal phenomena. *Psychological Review*, 94(2):236.

Gómez-Miñambres, J. (2012). Motivation through goal setting. *Journal of Economic Psychology*, 33(6):1223–1239.

Gonzalez, R. and Wu, G. (1999). On the Shape of the Probability Weighting Function a. *Cognitive Psychology*, 166:129–166.

Haridon, O., Aycinena, D., Vieider, F. M., Belianin, A., and Bandur, A. (2018). Off the Charts : Massive Unexplained Heterogeneity in a Global Study of Ambiguity Attitudes. *Review of Economics and Statistics*, 100(4):664–677.

Hau, R. and Pleskac, T. J. (2008). The Description-Experience Gap in Risky Choice : The Role of Sample Size and Experienced Probabilities. *Journal of Behavioral Decision Making*, 21:493–518.

Heath, C., Larrick, R. P., and Wu, G. (1999). Goals as reference points. *Cognitive psychology*, 38(1):79–109.

Heidhues, B. P. and Koszegi, B. (2010). Exploiting Naïvete about Self-Control in the Credit Market. *The American Economic Review*, 100(5):2279–2303.

Hertwig, R., Barron, G., Weber, E. U., and Erev, I. (2004). Decisions From Experience and the Effect of Rare Events in Risky Choice. *Psychological Science*, 15(8):534–539.

Herweg, F. and Mueller, D. (2010). Binary payment schemes: Moral hazard and loss aversion. *The American Economic Review*, 100(5):2451–2477.

Hey, J. D. and Lee, J. (2005). Do subjects separate (or are they sophisticated)? *Experimental Economics*, 8:233–265.

Holmstrom, B. (1979). Moral hazard and observability. *Bell Journal of Economics*, 10(1):74–91.

Holmstrom, B. (1999). Managerial Incentive Problems: A Dynamic Perspective. *The Review of Economic Studies*, 66:169–182.

Holmstrom, B. and Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics and Organization*, 7(January 1991):24–52.

Holt, C. (1986). Preference Reversals and the Independence. *The American Economic Review*, 76(3):508–515.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1991). Anomalies The Endowment Effect , Loss Aversion , and Status Quo Bias. *Journal of Economic Perspectives*, 5(1):193–206.

Koszegi, B. (2014). Behavioral Contract Theory. *Journal of Economic Literature*, 52:1075–1118.

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

Lazear, E. P. (1986). Salaries and Piece Rates. *The Journal of Business*, 59(3):405.

Moldovanu, B., Sela, A., and Shi, X. (2007). Contests for Status. *Journal of Political Economy*, 115(2):338–363.

Murphy, R. O. and Ten Brincke, R. H. W. (2018). Hierarchical Maximum Likelihood Parameter Estimation for Cumulative Prospect Theory: Improving the Reliability of Individual Risk Parameter Estimates. *Management Science*, 64(1):308–326.

Niederle, M. and Vesterlund, L. (2007). Do Women Shy Away From Competition? Do Men Compete Too Much? *Quarterly Journal of Economics*, 122(3):1067–1101.

Pokorny, K. (2008). Pay-but do not pay too much. An experimental study on the impact of incentives. *Journal of Economic Behavior and Organization*, 66:251–264.

Prelec, D. (1998). The Probability Weighting Function. *Econometrica*, 66(3):497–527.

Quiggin, J. (1982). A theory of anticipated utility.

Rabin, M. (2000). Risk Aversion and Expected-Utility theory : A Calibration Theorem. *Econometrica*, 68(5):1281–1292.

Sandroni, B. A. and Squintani, F. (2007). Overconfidence , Insurance , and Paternalism. *The American Economic Review*, 97(5):1994–2004.

Spalt, O. (2013). Probability Weighting and Employee Stock Options. *The Journal of Financial and Quantitative Analysis*, 48(4):1085–1118.

Starmer, C. and Sugden, R. (1991). Does the Random-Lottery Incentive System Elicit True Preferences ? An Experimental Investigation Author. *The American Economic Review*, 81(4):971–978.

Tversky, A. and Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review*, 102(2):269–283.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Wakker, P. and Deneffe, D. (1996). Eliciting von Neumannâ€“Morgenstern Utilities When Probabilities Are Distorted or Unknown. *Management Science*, 42(8):1131–1150.

Wakker, P. P. (2010). *Prospect Theory for Risk and Ambiguity*. Cambridge university press.

Wu, G. and Gonzalez, R. (1996). Curvature of the Probability Weighting Function. *Management Science*, 42(12):1676–1690.

Wu, G., Heath, C., and Larrick, R. (2008). A prospect theory model of goal behavior. *Mimeo*.

Appendix A: Proofs

Proposition 1

Proof. Suppose that $y^{**} \leq y^*$. For an agent with $\tilde{\theta}$, and since $c_y(\theta, y) > 0$ and $c_{yy}(\theta, y) > 0$, then $c_y(y^{**}, \tilde{\theta}) \leq c_y(y^*, \tilde{\theta})$. Using equations (2) and (9), it is possible to further rewrite the initial inequality as

$$b_y\left(\frac{ay^{**}}{p}\right) \leq b_y(ay^*). \tag{13}$$

Since $\frac{ay}{p} > ay$ for any $p \in (0, 1)$ and any $y \in (0, \bar{y}]$, then $b_y(\frac{ay}{p}) > b_y(ay)$ if $b_{yy}(\cdot) \geq 0$, which contradicts (13). This contradiction is also reached when $y^{**} < y^*$. To see how, consider $y^{**} = 1$ and $y^* = 2$, the inequality $\frac{a}{p} < 2a$ does not hold for $p \in (0, \frac{1}{2})$. Thus, $b_y(\frac{a}{p}) > b_y(2a)$ if $b_{yy}(\cdot) \geq 0$ and $p \in (0, \frac{1}{2})$, which contradicts (13). In general, one could find sufficiently small p to ensure that $y^{**} < y^*$ if $b_{yy}(\cdot) \geq 0$ is contradicted. Then, it must be that $y^{**} > y^*$ if $b_{yy}(\cdot) \geq 0$. Following a similar procedure it is possible to show that $y^{**} \geq y^*$ cannot hold if $b_{yy}(\cdot) < 0$ and that it must be that $y^{**} < y^*$ if $b_{yy}(\cdot) < 0$.

□

Proposition 2

Proof. Consider first the interval $p \in (\tilde{p}, 1)$. Since $\frac{w(p)}{p} < 1$ if $p \in (\tilde{p}, 1)$ and $b_{yy}(\cdot) \leq 0$, then $\frac{w(p)}{p} b_y(\frac{ay}{p}) < b_y(ay)$ for any $y \in (0, \bar{y}]$. Moreover, for arbitrary $\epsilon > 0$, $\frac{w(p)}{p} b_y(\frac{a(y+\epsilon)}{p}) < b_y(ay)$ holds. Suppose that $y_R^{**} > y^{**}$. For an agent with $\tilde{\theta}$, and since $c_y(\theta, y) > 0$ and $c_{yy}(\theta, y) > 0$, then $c_y(y_R^{**}, \tilde{\theta}) > c_y(y^{**}, \tilde{\theta})$. Equations (9) and (10), entail that the previous inequality can be written as $\frac{w(p)}{p} b_y(\frac{ay_R^{**}}{p}) > b_y(ay^{**})$, which contradicts that $\frac{w(p)}{p} b_y(\frac{a(y+\epsilon)}{p}) < b_y(ay)$ for arbitrary $\epsilon > 0$. Then it must be that $y_R^{**} < y^{**}$ if $p \in (\tilde{p}, 1)$ and $b_{yy}(\cdot) \leq 0$. In addition, Proposition 1 demonstrates that $y^{**} < y^*$ if $b_{yy}(\cdot) < 0$ for any $p \in (0, 1)$. Therefore, $y_R^{**} < y^{**} < y^*$ if $p \in (\tilde{p}, 1)$ and $b_{yy}(\cdot) < 0$.

Now, consider the interval $p \in (0, \tilde{p}]$. Note that $\frac{w(p)}{p} \geq 1$ if $p \in (0, \tilde{p}]$ and $b_{yy}(\cdot) \leq 0$, then $\frac{w(p)}{p} b_y(\frac{ay}{p}) < b_y(ay + \epsilon)$ or $\frac{w(p)}{p} b_y(\frac{ay}{p}) \geq b_y(ay + \epsilon)$ are possible for any $\epsilon \geq 0$. Thus, either $y_R^{**} \geq y^* > y^{**}$ or $y^* > y_R^{**} \geq y^{**}$ if $p \in (0, \tilde{p})$ and $b_{yy}(\cdot) < 0$. Suppose that $y^* \geq y_R^{**}$. For an agent with $\tilde{\theta}$, and due to $c_{yy}(\tilde{\theta}, y) > 0$ and $c_y(\tilde{\theta}, y) > 0$, this inequality can be written as $c_y(y^*, \tilde{\theta}) \geq c_y(y_R^{**}, \tilde{\theta})$. Moreover, using equations (2) and (10) the initial inequality can be further rewritten as

$$b_y(ay^*) \geq \frac{w(p)}{p} b_y\left(\frac{ay_R^{**}}{p}\right). \tag{14}$$

Let us analyze what happens at the extremes of the considered probability interval. Since $\frac{w(p)}{p} = 1$ at $p = \tilde{p}$, then $ay < \frac{a(y+\epsilon)}{\tilde{p}}$ for any $y > 0$ and arbitrary $\epsilon \geq 0$. Moreover, given that

$b_{yy}(\cdot) \leq 0$, (16) holds at $p = \tilde{p}$. Nonetheless, (14) does not hold as as $p \to 0^+$. To evaluate $\lim_{p\to 0^+} \frac{w(p)}{p} b_y\left(\frac{ya}{p}\right)$, which yields an indeterminate form, use L'Hospital's rule as follows

$$\lim_{p\to 0^+} \frac{w(p)}{p} b_y\left(\frac{ya}{p}\right) = \lim_{p\to 0^+} \frac{d\left(\frac{w(p)}{p}\right)}{dp} \Big/ \frac{1}{\frac{d\left(b_y\left(\frac{ya}{p}\right)\right)}{dp}} = \lim_{p\to 0^+} \frac{pw_p(p) - w(p)}{\frac{b_{yp}(ay/p)(ay)}{b_y(ay/p)^2}} = \infty.$$

The last equality is due to $\lim_{p\to 0^+} b_{yp}(ay/p) = 0$, $\lim_{p\to 0^+} \frac{1}{b_y(ay/p)} = \infty$, and $\lim_{p\to 0^+} w_p(p)p = \infty$.[25] Hence, (16) does not hold as $p \to 0^+$ and it must be that $y^* < y_R^{**}$.

I compute $\frac{\partial\left(\frac{w(p)}{p} b_y\left(\frac{ay}{p}\right)\right)}{\partial p} = \frac{(pw_p(p) - w_p(p))}{p^2} b_y\left(\frac{ay}{p}\right) - 2\frac{w(p)}{p^3} ay b_{yp}\left(\frac{ay}{p}\right)$ to analyze the behavior of the right hand side of (14) as p increases over $p \in (0, \tilde{p})$. Suppose that $\frac{\partial\left(\frac{w(p)}{p} b_y\left(\frac{ay}{p}\right)\right)}{\partial p} > 0$, which implies

$$\frac{pw_p(p)}{w(p)} > 1 + \frac{b_{yp}\left(\frac{ay}{p}\right)\frac{ay}{p}}{b_y\left(\frac{ay}{p}\right)}. \tag{15}$$

For $p \in (0, \tilde{p})$ and $\tilde{p} > \hat{p}$, the left hand side of (15) decreases since $\frac{\partial\left(\frac{w(p)}{p}\right)}{\partial p} > 0$ and $w_{pp}(p) < 0$. Therefore, the largest value that $\frac{pw_p(p)}{w(p)}$ attains must be as $p \to 0^+$. I evaluate the limit $\lim_{p\to 0^+} \frac{pw_p(p)}{w(p)}$ using L'Hospital's rule as follows,

$$\lim_{p\to 0^+} \frac{pw_p(p)}{w(p)} = \lim_{p\to 0^+} \frac{\frac{dp}{dp}}{\frac{d\left(\frac{w(p)}{w_p(p)}\right)}{dp}} = \lim_{p\to 0^+} \frac{1}{1 - \frac{w(p)w_{pp}(p)}{w_p(p)^2}}.$$

Which implies that $\lim_{p\to 0^+} \frac{pw_p(p)}{w(p)} \in (0, 1)$ and contradicts (15) given that $b_{yp}(ay) > 0$ is assumed. Therefore, $\frac{\partial\left(\frac{w(p)}{p} b_y\left(\frac{ay}{p}\right)\right)}{\partial p} < 0$.

Given that $\lim_{p\to 0^+} \frac{w(p)}{p} b_y\left(\frac{ya}{p}\right) = \infty$, $b_y(ay) > b_y\left(\frac{ay}{p}\right)$ holding at $p = \tilde{p}$, and $\frac{\partial\left(\frac{(w(p)}{p} b_y\left(\frac{ay}{p}\right)\right)}{\partial p} < 0$, then the existance of a $p^* \in (0, \tilde{p})$ such that (14) holds with equality is guaranteed. When $\tilde{p} < \hat{p}$, p^* is unique since the left hand side of (14) is always decreasing over $p \in (0, \tilde{p})$. However, when $\tilde{p} \geq \hat{p}$, the existence of p^* is still guaranteed, since $w(p)$ is always first concave and then convex, but p^* is not unique since the left hand side of (14) could not be decreasing over $p \in (\hat{p}, \tilde{p})$. For the latter case we refer to p^* as the smallest possible value that makes (16) bind. Hence, $y^* \geq y_R^{**}$ cannot hold if $p \in (0, p^*)$ and instead it must be that $y_R^{**} > y^*$.

$\square$

[25] Note that $\lim_{p\to 0^+} w_p(p)p$ is also an indeterminate form. I use a L'Hospital's rule again to evaluate this limit: $\lim_{p\to 0^+} w_p(p)p = \frac{\frac{dp}{dp}}{\frac{d\left(\frac{1}{w_p(p)}\right)}{dp}} = \lim_{p\to 0^+} \frac{-1}{\frac{w_{pp}(p)}{(w_p(p))^2}} = \infty$. The last equality is due to $w_{pp}(p) < 0$, $\lim_{p\to 0^+} \frac{1}{w_p(p)} = 0$ and the assumption that $\lim_{p\to 0^+} \frac{w_{pp}(p)}{w_p(p)} \in \Re$.

Lemma 1

Proof. The proof is similar to that of Proposition 2, with the difference that $\frac{r}{p} > 0$ and r are subtracted from the utility of the agent.

$\square$

Lemma 2

Proof. Suppose that $y_C^{**} \leq y_C^*$. Since $c_y(\tilde{\theta}, y) > 0$ and $c_{yy}(\tilde{\theta}, y) > 0$, then $c_y(y_C^{**}, \tilde{\theta}) \leq c_y(y_C^*, \tilde{\theta})$. Using equations (12) and (18), as well as the duality $z(p) = 1 - w(1-p)$, the initial inequality can be further rewritten as,

$$\frac{(1 - w(1-p))}{p} b_y\Big(\frac{r}{p} - \frac{a}{p} y_C^{**}\Big) \leq b_y(r - a y_C^*). \tag{16}$$

Since $\frac{r}{p} - \frac{ay}{p} > r - ay$ for any $y > 0$, any $r > 0$ such that $y < r/a$ holds, and any $p \in (0,1)$, and given that $b_{yy}(\cdot) \geq 0$ for $y < r/a$, then $b_y(\frac{r}{p} - \frac{a}{p}y) > b_y(r - ay)$. For arbitrary $\epsilon > 0$, $b_y(\frac{r}{p} - \frac{a}{p}(y+\epsilon)) > b_y(r - ay)$ holds. Moreover, $\frac{1-w(1-p)}{p} \geq 1$ for $p \in (0, \tilde{p}]$. These two statements contradict (18). Hence, it must be that $y_C^{**} > y_C^*$ for $y < r/Bp$ if $p \in (0, \tilde{p})$.

$\square$

Appendix B: Agents with CPT preferences

This analysis compares the output delivered by this agent when he works under the piece rate as compared to the situation in which he works under the probability contract. I assume that when offered the piece rate, this agent has preferences as in CPT for riskless choice (Tversky and Kahneman, 1991). Hence, under the piece rate, the optimal output delivered by the agent y_C^* satisfies the following system of first-order conditions

$$ab_y(ay_C^* - r) - c_y(y_C^*, \theta) = 0, \text{ if } y \geq r/a, \tag{17}$$

$$\lambda ab_y(r - ay_C^*) - c_y(y_C^*, \theta) = 0, \text{ if } y < r/a. \tag{18}$$

These first-order conditions as well as those in equations (11) and (12), illustrates a crucial assumption of the presen t analysis which is that an agent has the same reference point across the two contracts. This may seem an stringent assumption, but given that the considered contracts pay on expectation the same monetary amounts, do not feature a performance goal that could act as a reference point, do not include a bonus, and do not elicit an expectation on the part of the agent, it is not a unlikely assumption to make.[26]

Let us first consider the case in which the agent's production locate him in the domain of gains. Lemma 1 demonstrates that the condition over the principal's choice from Proposition 2 guarantees that the probability contract induces higher output.

Lemma 1: For an agent with ability $\tilde{\theta} \in (0, 1)$, with preferences for monetary outcomes represented by the value function $v(r, y)$, with $r > 0$, and $\lambda > 1$, and who evaluates probabilities using $w(p)$ and $z(p)$, then $y_C^{**} \geq y_C^*$ for the domain $y \geq \frac{r}{a}$ if $p \in (0, p^*]$.

As in the case of RDU preferences, the principal could be better off offering the probability contract if it is implemented with a probability that induces a sufficiently large degree of overweighting of probabilities in the agent. This is because the overweighting of small probabilities acts as a risk-seeking mechanism, which, when sufficiently large, could outweigh the risk-averse attitudes generated by the shape of the value function in the domain of gains. These risk-seeking attitudes motivate the agent to work harder on the task under the probability contract as compared to a situation in which she was offered the piece rate.

Let us consider the case in which the agent's output level locate him in the domain of losses. Again, the principal could incentivize the agent to work harder under the probability contract and the conditions in Proposition 2 guarantee this result.

[26] It is straightforward to see that relaxing this assumption and letting $r_{Prob} \geq r_{PR}$, with r_{Prob} the reference point of under the probability contract and r_{PR} the reference point under the piece rate, yields the same results.

Lemma 2: For an agent with ability $\tilde{\theta} \in (0, 1)$, with preferences for monetary outcomes represented by the value function $v(r, y)$, with $r > 0$ and $\lambda > 1$, and who evaluates probabilities using $w(p)$, then $y_C^{**} > y_C^*$ for $y < \frac{r}{a}$ if $p \in (0, p^*]$.

According to CPT, the agent exhibits a convex value function in the domain of losses. Such a curvature induces risk-seeking attitudes in the agent, which incentivize higher output supply under the probability contract. To maintain these favorable risk attitudes, the principal should avoid choosing probabilities that induce risk-averse attitudes. This could be done implementing any $p \in (0, \tilde{p}]$, which is guaranteed when $p \in (0, p^*)$.

Appendix C: The principal's choice

In the main body of the paper I demonstrate that for an agent who distorts probabilities systematically, the probability contract could lead to higher output than a cost-equivalent piece rate. A necessary condition for this result is that the principal anticipates the shape of the probability weighting function of the agent, and implements the contract using a probability that is sufficiently overweighted by the agent. The purpose of this appendix is to show that this conclusion is also reached in an analysis that focuses on the decision of the principal. This analysis is more complete inasmuch as it not only focuses on the incentive compatibility constraint of the agent.

Throughout this appendix I assume that agents distort probabilities systematically using $w(p)$ which has the properties described in Assumption 3 in Section 2. Note that this constitutes the interesting case given that under this assumption, the choice of p has an influence on the agent's behavior. Moreover, I assume that agent has a linear basic utility, which is a result that is supported by the empirical results of Appendix E and previous utility elicitations using lotteries with small stakes (See Wakker & Deneffe (1996) and Rabin (2000)). Finally, I assume that the agent evaluates outcomes with respect to final positions of the asset and not relative to a reference point, this is done to simplify the analysis. Together, these assumptions entail that the risk attitudes of the agent do not stem from the curvature of his basic utility, but from the curvature of the probability weighting function. In other words, I assume that the agent's risk preferences are represented by RDU preferences with $b_{yy}(\cdot) = 0$.

Let us start by setting up the principal's program. The principal's objective is to minimize the compensation offered to the agent, subject to the participation constraint of the agent, which entails that the principal needs to offer a contract ensuring that is going to be accepted by the worke, and the incentive compatibility constraint, which entails that the contract offered by the principal ensures that the agent delivers an optimal amount of output. This program can be formally written as

$$\begin{aligned}
&\underset{p}{\text{Min}} && Byp, \\
&\text{subject to} && IC: \quad \underset{y}{\text{argmax}}\; w(p)By - c(\theta, y), \\
& && PC: \quad w(p)By - c(\theta, y) \geq 0.
\end{aligned}$$

As in Section 2, I employ the cost equivalence $B = \frac{a}{p}$ that equalizes the expected value of the extrinsic incentives offered by the piece rate with those offered by the probability contract.

The Lagrangian of the principal's problem is

$$\mathcal{L} = ay - \lambda_1 \left(w(p)\frac{a}{p} - c_y(y, \theta) \right) - \lambda_2 \left(w(p)\frac{ya}{p} - c(y, \theta) \right). \tag{19}$$

The first-order condition of the Lagrangian with respect to p, representing the optimal choice of the risk-neutral principal, is

$$\frac{\partial \mathcal{L}}{\partial p}: \quad -\lambda_1 \left(w_p(p)\frac{a}{p} - w(p)\frac{w(p)a}{p^2} \right) - \lambda_2 \left(w_p(p)\frac{ay}{p} - w(p)\frac{ay_t}{p^2} \right). \tag{20}$$

Some rewritting of (20) yields,

$$\frac{\partial \mathcal{L}}{\partial p}: \quad (\lambda_1 + \lambda_2 y_t)\frac{a}{p}\left(w_p(p) - \frac{w(p)}{p} \right) = 0. \tag{21}$$

Notice from (21) that when either the IC or PC constraints bind, this is if $\lambda_1 > 0$ or if $\lambda_2 > 0$, and naturally when $y > 0$, the solution of the Lagrangian is given by the following probability p^{**},

$$\left\{ p^{**} \in (0,1) : p^{**} = \frac{w(p^{**})}{w_p(p^{**})} \right\}. \tag{22}$$

At this point we would like to understand the properties of the solution p^{**}. Let us start by studying the existence and uniqueness of this solution. Start by defining $g(p) \equiv \frac{w(p)}{w_p(p)}$. Note that $g(0) = 0$ since $\lim_{p\to 0^+} w_p(p) = \infty$ and $w(0) = 0$, also note that $g(1) = 0$ since $\lim_{p\to 1^-} w_p(p) = \infty$ and $w(1) = 1$, finally note that $g(\hat{p}) = \infty$ since $\lim_{p\to \hat{p}} w_p(p) = 0$. Moreover, $g(p)$ is increasing for the interval $p \in [0, \hat{p})$, given that $w_{pp}(p) > 0$ in $p \in [0, \hat{p})$, and alternatively $g(p)$ is decreasing for the interval $p \in [1, \hat{p}]$, since $w_{pp}(p) < 0$ in $p \in [1, \hat{p}]$.

The aforementioned properties of $g(p)$ guarantee the existence of the fixed-point $p^{**} = g(p^{**})$ in $p \in (0, 1)$. To see how, note that p is a linear and increasing function in the unit interval with minimum value at $p = 0$ and maximum value at $p = 1$. Moreover, $g(p)$ is an increasing function in the interval $p \in [0, \hat{p})$ with values $g(0) = 0$, $g(1) = 0$, and $g(\hat{p}) = \infty$. In the interval $p \in (\hat{p}, 1]$, $g(p)$ is decreasing while and p is increasing in this interval, then at some probability in this interval p and $g(p)$ intersect.[27]

Let us now investigate whether p^{**} is a solution for the principal's program for all possible values of p. This is done by studying the shape of the Lagrangian over the entire probability support. The second-order condition of the Lagrangian in (19) is,

$$\frac{\partial^2 \mathcal{L}}{\partial p^2}: \quad -(\lambda_1 + \lambda_2 y_t)\frac{a}{p}\left(w_{pp}(p) - \frac{2w_p(p)}{p} + \frac{2w(p)}{p^2} \right). \tag{23}$$

Equation (23), when evaluated at p^{**}, becomes positive if $w_{pp}(p) < 0$. Hence p^{**} is a unique solution of the program for $p \in (0, \hat{p})$. In contrast, if $w_{pp}(p) > 0$ the second order condition evaluated at p^{**} becomes negative, implying that in this probability interval the Lagrangian

[27] Since $g_p(p) = 1 - \frac{w(p)w_{pp}(p)}{(w_p(p))^2}$, a sufficient condition for $g_p(p) > 0$ is that $w_{pp}(p)<0$, and a necessary condition for $g_p(p) < 0$ is that $w_{pp}(p)>0$.

is concave and the objective function attains a minimum value at one of the extreme values, $p \in (\hat{p}, 1]$. Hence, in the whole interval $p \in [0, 1]$ there are multiple solutions. This multiplicity of solutions stems from the assumed shape of the $w(p)$.

Thus far, we have proven the existence of the the solution $p^{**} = g(p^{**})$ over the entire probability interval $p \in (0, 1)$. However, as shown in the previous paragraph p^{**} is the unique solution in $p \in (0, \hat{p})$. Hence, we want to study the conditions guaranteeing that p^{**} exists in $p \in (0, \hat{p})$. A sufficient condition for this is that $w_{ppp}(p) < 0$ in $p \in (0, \hat{p})$. This condition means that $w(p)$ is linear for near-zero probabilities and becomes more concave as p approaches $\hat{p}$. Note if $w_{ppp}(p) < 0$ in $p \in (0, \hat{p})$, then $g(p)$ is strictly convex at $p \in (0, \hat{p})$, which together with the fact that $g(0) = 0$ entails that $p > g(p)$ at near-zero probabilities. In addition, as $p \to \hat{p}$ then $g(p) > p$ since $g(\hat{p}) = \infty$ and $p < \hat{p} < 1$. Hence, in some probability in $p \in (0, \hat{p})$ we have that $p = g(p)$.

In what is left of the Appendix, I describe the optimal solution of the principal's problem. Note that the $w(p)$ allows $\hat{p} \neq \tilde{p}$. This entails that if $\hat{p} > \tilde{p}$, the solution p^{**} may not yield an overweighting of probabilities. Moreover, if $\hat{p} < \tilde{p}$ there is overweighting of probabilities in the convex region of $w(p)$. As it will become evident, these properties, implied by these two cases, are determinant to characterize the solution of the principal's problem. Thus, I describe the solution distinguishing the cases$\hat{p} > \tilde{p}$ and $\hat{p} < \tilde{p}$.

Let us first focus on the case in which the principal faces an agent with a $w(p)$ with $\hat{p} < \tilde{p}$. For the interval $p \in (0, \hat{p})$ the optimal solution p^{**} can be implemented. However, for $p \in [\hat{p}, 1]$ the solution is at one of the extreme values $p \in \{\hat{p}, 1\}$. Which $p \in \{\hat{p}, 1\}$ would be chosen by the principal? to answer that question let us take a look at the behavior of the lagrangian at each of the two values. At $p = \hat{p}$, the IC and PC constraints become larger than at $p = 1$, this entails that$p = \hat{p}$ the lowest value of the Lagrangian is achieved. This result is achieved because $p = \hat{p}$ induces an overweighting of probabilities $\frac{w(p)}{p} > 1$, which at the same time inflates these constraints. Thus, the principal chooses $p = \hat{p}$ if $p \in (\hat{p}, 1)$.

Let us now turn to the case in which the principal has an agent with $w(p)$ such that $\hat{p} > \tilde{p}$. For the interval $p \in (0, \hat{p})$ the optimal solution p^{**} can be implemented. Notice that this solution might not induce an overweighting of probabilities in the agent. The intuition behind this result is that the principal implements p^{**}, even when it does not induce an overweighting of probabilities in the agent, so that the contract is accepted by the agent. This means that the principal is better off implementing a contract that is going to be accepted by the agent, even though the non-monetary incentives that it entails do not yield better outcomes than the piece rate. Moreover, for the interval $p \in [\hat{p}, 1]$, the solution of the program is $p = 1$ since any other value of $p = \hat{p}$ yields $\frac{w(p)}{p} < 1$ which leads to lower values of the IC and PC constrains than those implied by $p = 1$.

All in all, the solution to the minimization program for the principal is given by $p^{\star}$ which can be written down as,

$$p^\star = \begin{cases} p^{**}, \text{ if } p < \hat{p}, \\ \hat{p}, \text{ if } \hat{p} < \tilde{p} \text{ and } p > \hat{p}, \\ 1, \text{ if } \hat{p} > \tilde{p} \text{ and } p > \hat{p}. \end{cases}$$

The principal has two possible optimal actions for each case. He chooses p^{**} irrespective of the location of $\tilde{p}$ relative to $\hat{p}$ as long as p^{**} is a solution. In other words p^{**} is chosen long as $p < \hat{p}$. This action always yields an overweighting of probabilities if $\hat{p} < \tilde{p}$ and yields an overweighting of probabilities for the interval $p \in (0, \tilde{p})$ if $\hat{p} > \tilde{p}$. However, the action $p = p^{**}$ induces underweighting of probabilities in the agent for the interval $p \in [\tilde{p}, \hat{p}]$ if $\hat{p} > \tilde{p}$. This means that agents with a $w(p)$ with low elevation may yield lower output under the probability contract whenever $p^{**} \in [\tilde{p}, \hat{p}]$. As explained above, the intuition of this particular case of the solution is that the principal finds it optimal to implement a contract that is accepted by the agent even when its (non-monetary) incentives may yield lower output than those embedded by a piece rate. Additionally, the principal facing an agent with $w(p)$ with high elevation chooses $p = \hat{p}$ whenever $p > \hat{p}$. This action is chosen since $\hat{p}$ is overweighted by the agent. Finally, for an agent with a $w(p)$ with low elevation the principal chooses $p = 1$ which is the equivalent of the probability contract.

To conclude, I find that the principal implements the probability contract with probabilities that are overweighted by the agent, except for the special case in which the agent has a probability with low elevation and $p^{**} \in [\tilde{p}, \hat{p}]$. Notice that this solution requires that the principal has a detailed knowledge of the probability weighting function of the agent. Future analyses of the probability contract should focus on relaxing this assumption and the consequences that this relaxation imposes on the decision of the principal.

Appendix D: Instructions

This is an experiment in the economics of decision-making. The instructions are simple and if you follow them carefully and make certain decisions, you might earn a considerable amount of money, which will be paid to you via bank transfer at the end of the experiment. The amount of money that you earn will depend on your decisions and effort, and partly on chance. Once the experiment has started, no one is allowed to talk to anybody other than the experimenter. Anyone who violates this rule will lose his or her right to participate in this experiment. If you have further questions when reading these instructions please do not hesitate to raise your hand and formulate the question to the experimenter.

The experiment consists of two parts. Your earnings in part one or part two of the experiment will be chosen at the end of the experiment and become your final earnings. Whether the earnings of part one or the earnings of part two will be your final earnings will be established by roll of a die.

Part one

In this part of the experiment your task is to complete summations. Your earnings in this part of the experiment depend only on the number of correct summations that you deliver. You need to complete as many summations as you can in 10 rounds, each round lasts four minutes. In other words you will have a total of 40 minutes to complete as many summations as you can.

Each summation consists of five-two digit numbers. For example 11+22+33+44+55=? Once you know the answer to the sum of these five two digit numbers, input the answer in the interface, Click OK, and a new set of numbers will appear on your screen.

For your better understanding you will face with two examples next.

[Examples displayed]

The previous examples show what you have to do in this part of the experiment. The only thing left to be explained is to specify how you are going to earn money by completing the summations.

Piecerate Treatment The payment rule: In this part of the experiment each correct summation will add 25 Euro cents to your experimental earnings.

Remember: you have 40 minutes to complete summations, and only correct summations will count towards your earnings at a rate of 25 euro cents each. If you understood these instructions Press "OK". When everyone is ready we will start with this part of the experiment.

LowPr Treatment **The payment rule:** In this part of the experiment 1 out of all the 10 rounds will be randomly chosen. The specific round is chosen at random by the computer at the end of this part of the experiment. This is, once you completed summations in all the 10 rounds, only the correct summations in a randomly chosen round will count towards your earnings at a rate of 250 euro cents per correct summation.

Remember: you have 40 minutes to complete summations, and only correct summations in 1 specific round, chosen randomly by the computer at the end of the experiment, will count towards your earnings at a rate of 250 euro cents each. If you understood these instructions Press "OK". When everyone is ready we will start with this part of the experiment.

MePr Treatment **The payment rule:** In this part of the experiment 3 out of all the 10 rounds will be randomly chosen. The specific rounds are chosen at randomby the computer at the end of this part of the experiment. This is, once you completed summations in all the 10 rounds, only the correct summations in that randomly chosen round will count towards your earnings at a rate of 85 euro cents per correct summation.

Remember: you have 40 minutes to complete summations, and only correct summations in 3 specific rounds, chosen randomly by the computer at the end of the experiment, will count toward your earnings at a rate of 85 euro cents each. If you understood these instructions Press "OK". When everyone is ready we will start with this part of the experiment.

HiPr Treatment **The payment rule:** In this part of the experiment 5 out of all the 10 rounds will be randomly chosen. The specific rounds are chosen at random by the computer at the end of this part of the experiment. This is once you completed summations in all the 10 rounds, only the correct summations in that randomly chosen round will count towards your earnings at a rate 50 euro cents per correct summation.

Remember: you have 40 minutes to complete summations, and only correct summations in 5 specific rounds, chosen randomly by the computer at the end of the experiment, will count toward your earnings at a rate of 50 euro cents each. If you understood these instructions Press "OK". When everyone is ready we will start with this part of the experiment.

Part two

In this part of the experiment your task is to choose among two possible alternatives. Your earnings on this part of the experiment depend on how good your choices are.

Particularly, you will face with 11 decision sets. In each of these sets you need to choose between the option L, that delivers a fixed amount of money, and the option R that is a lottery between two monetary amounts. Each decision set contains six choices.

Be Careful! Every time you make a choice between L and R, the monetary prizes of the options are going to change and you ought to make a choice again. One of your choices will be randomly picked by the computer, will be played and its realization will count towards your earnings for this part of the experiment. You will be faced with one example next.

[Example displayed]

If it is clear what you have to do in this part of the experiment. Press ”OK” to start, once everyone is ready this part of the experiment will begin.

Survey

- Gender:
- Age:
- What is your education level? (Bachelor, Exchange, Pre-Master, Master, PhD):
- What is the name of your program of studies?
- How difficult did you find the task? (where 1 stands for easy and 5 for Very difficult)
- Rate how confident you are that you can do the task good enough so you can be in the top half of performers in this group as of now. (1-Not confident, 10- Very confident)
- Are you any good at adding numbers? (1-Not good at all, 10-Very good)
- Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?
- Rate yourself from 0 to 10, where 0 means "unwilling to take any risks" and 10 means "fully prepared to take risks".
- People can behave differently in different situations. How would you rate your willingness to take risks while driving? Rate yourself from 0 to 10, where 0 means "unwilling to take any risks" and 10 means "fully prepared to take risks"

- How would you rate your willingness to take risks in financial matters? Rate yourself from 0 to 10, where 0 means "unwilling to take any risks" and 10 means "fully prepared to take risks"
- How would you rate your willingness to take risks with your health? Rate yourself from 0 to 10, where 0 means "unwilling to take any risks" and 10 means "fully prepared to take risks"
- How would you rate your willingness to take risks in your occupation? Rate yourself from 0 to 10, where 0 means "unwilling to take any risks" and 10 means "fully prepared to take risks"
- How would you rate your willingness to take risks in your faith in other people? Rate yourself from 0 to 10, where 0 means "unwilling to take any risks" and 10 means "fully prepared to take risks"

Appendix E: The shape of the utility functions

In this appendix, I investigate the shape of the utility function of the participants. Decision sets numbered 1 to 6 elicit the sequence of outcomes $\{x_1, x_2, x_3, x_4, x_5, x_6\}$, which has the characteristic of yielding equally spaced utilities, i.e. $u(x_j) - u(x_{j-1}) = u(x_{j-1}) - u(x_{j-2})$. I am interested in two characteristics of this sequence when each outcome is plotted against the expected value of the lottery from which it was elicited: i) the sign of the slope and ii) the curvature. To study these characteristics, I construct two measures: the difference, $\Delta'_i \equiv x_i - x_{i-1}$, for $i = 1, ...6$, and the second difference, $\Delta''_i \equiv \Delta'_i - \Delta'_{i-1}$ for $i = 2, ..., 6$. The sign of Δ'_i as i increases determines the sign of the slope, i.e. whether a subject prefers larger monetary outcomes. Moreover, the sign of Δ''_i as i increases determines the curvature. For example, a subject with $\Delta'_i > 0$ and $\Delta''_i < 0$ for all i has a preference for larger monetary outcomes and experiences smaller utility increments with larger monetary outcomes, this is equivalent to say that this subject has a concave utility function.

I classify the participants of the experiment according to the curvature of their utility function. Given that a subject has multiple Δ''_i's and that it is possible that he makes mistakes, this classification is based on the sign of Δ''_i with the most occurrence. Thus, a subject with at least three positive Δ''_is was classified as having a convex utility. A subject with at least three negative Δ''_is was classified as exhibiting a concave utility. A subject with three or more Δ''_is that are not significantly different from zero was classified as having a linear utility. Finally, a subject with a utility function that cannot be classified as concave, convex, or linear, was classified as having a mixed utility. To statistically asses the sign of a Δ''_i, I use confidence intervals around zero. The confidence intervals were constructed using the standard deviation of Δ''_i, for each $i = 1, ...5$, which was then multiplied by the factors 0.64 and -0.64. Hence, if Δ''_i follows a normal distribution, 50% of Δ''_is in the data are to be contained within the confidence interval.[28]

The data suggest that all the subjects in the experiment exhibit an increasing sequence $\{x_1, ..., x_6\}$ which denotes, not surprisingly, a generalized preference for larger amounts of money. Table 12 presents the classification of the subjects in the experiment according to the curvature of their utility function. The data suggest that the majority of subjects exhibit linear utility functions. Specifically, 77% of the subjects have linear utility functions, while the rest of the subjects have mixed utility functions (13% of the subjects), and concave utility functions (7% of the subjects). Of the subjects classified as having mixed utility functions, only

[28]Different confidence intervals were used in this analysis. These confidence intervals were constructed using the standard deviation of a Δ''_i and multiplied by other factors, such as 1 and -1, 1.64 and -1.64, and 2 and -2. The qualitative results of these analyses are not different from the main result reported here that the majority of subjects exhibit a linear utility function. Since these confidence intervals are more stringent, they yield less subjects classified as having a mixed utility function and more subjects exhibiting a linear utility function.

Table 12: Classification of Subjects According to Utility Curvature

Reference Point	Domain	Convex	Concave	Linear	Mixed	Total
No/Zero	No/Gains	3	13	133	23	172
Belief	Gains	3	12	43	21	79
Belief	Losses	0	1	90	2	93

Note: This table presents the subjects classification according to the shape of their utility function. Subjects are classified as having a convex, concave, linear or mixed utility function based on the sign of most occurrence of Δ_i. The first row presents the classification when the analysis is performed with all the data. The second and third columns feature the analysis taking into account the Beliefs non-zero reference point, which is the monetary equivalent of a subject's beliefs about her performance in the first part of the experiment. The second row presents the analysis when the monetary outcomes of the lotteries are above the reference point, whereas the third row presents the analysis when the monetary outcomes of the lotteries are below the reference point

6 (3 % of subjects) presented Δ_i''s that suggest a utility function that is first convex and then concave, indicating diminishing sensitivity around some non-zero reference point. A binomial test shows that the number of subjects classified as having linear functions is significantly larger than those classified to have mixed utility functions ($p<0.01$, one sided test) and also larger than subjects classified as having convex utility functions ($p<0.01$, one sided test).

Table 12 also presents the analysis of the curvature of the utility function when it is assumed that Beliefs is the reference point. This alternative analysis also yields that the majority of the subjects exhibit a linear utility function. Specifically, I find that 65 % of the subjects have linear utilities in the domain of gains and 98% of the subjects exhibit linear utilities in the domain of losses. This division of the outcomes into gains and losses around Beliefs does not yield new indications of diminishing sensitivity, since the majority of the data of those subjects classified as having a convex utility function are in the domain of gains.

The result that more than two-thirds of the subjects exhibit linear utility is at odds with the principle of diminishing sensitivity, a key property of cumulative prospect theory (CPT). However, disregarding CPT as a possible representation for the subjects' preferences for money on the basis of this classification may be incorrect. As pointed out by Wakker and Deneffe (1996), the trade-off method, used to elicit $\{x_1, x_2, x_3, x_4, x_5\}$, requires lotteries with large monetary outcomes in order to obtain utility functions with pronounced curvature. Therefore, one of the advantages of the experimental design, that it elicits the utility function and the probability weighting function of a subject over the monetary outcomes at stake in the first part of the experiment, is also the reason that diminishing sensitivity may not be observed.

Until now I have focused on data at the individual level. To understand how these results aggregate, I analyze the values of the sequence $x_1, ..., x_6$ when each x_i is averaged across subjects. Table 13 presents the average of x_i, the standard deviation of x_i, and the average of Δ_i'.These values show that x_i is increasing with i, suggesting that subjects have a preference for larger monetary amounts. Also, the columns containing the average values of Δ_i' show that the increments of x_i become moderately larger as i increases, suggesting that the tendency of the utility function to exhibit a linear shape decreases when the amount of money becomes larger. Since the first step of Abdellaoui's method elicits a sequence of ff equivalents that are equally spaced in terms of utility levels, the values of Δ_i' as i increases show that the utility function of the average subjects tends to become concave. This result is also obtained by Abdellaoui (2000).

To gain a better understanding of the aforementioned results, I assume two parametric families of utility and estimate the parameters of these utility functions using non-linear least squares. Specifically, I assume a power utility, which belongs to the CRRA family of utility functions, and an exponential function, which belongs to the CARA family of utility functions. Table 14 presents the estimates of the regressions. For the two parametric specifications I find that the average utility function of the subjects is linear. For instance, when the power utility function $u(x) = x^\phi$ is assumed, the estimate of the parameter is $\phi = 0.995$ which is statistically not different from one. This conclusion is consistent with the large proportion of subjects that were classified as having a linear utility function, as shown by Table 12, and the modest increments that the averaged series x_i exhibits as i increases presented in Table 13.

These analyses are also performed under the assumption that Beliefs is the reference point. Table 13 shows that subjects exhibit a preference for larger monetary amounts in both domains. I also observe a decreasing tendency of the utility function to be linear as the amount of money becomes larger with respect to the reference point; In the domain of gains the utility function of the average subject tends to concavity, while in the domain of losses the function tends to convexity. Moreover, the data suggest that diminishing sensitivity manifests at different degrees across domains, with subjects exhibiting more diminishing sensitivity in the domain of gains than in the domain of losses. This difference is explained by fact that only positive outcomes were used to elicit the outcomes $\{x_1, x_2, x_3, x_4, x_5, x_6\}$, which leaves no room for the subjects to exhibit as much sensitivity in the domain of losses as in the domain of gains. I chose to elicit preferences using only positive outcomes since the aim of the second part of the experiment was to understand the preferences of the subjects over the stakes offered in the real-effort task. A more complete analysis of diminishing sensitivity across domains, and of preferences in general in the domain of losses under the assumption that subjects have CPT preferences, requires lotteries featuring negative outcomes.

I estimate the parameters of the utility function for each domain assuming a power or an exponential utility function. For the domain of losses, the estimated coefficients suggest linearity, with an estimated coefficient of $\phi = .992$ when the power utility function is assumed. This result is also found for the domain of losses, where the estimation yields $\phi = 1.035$ when

Table 13: Aggregate results x_1,x_2,x_3,x_4,x_5, and x_6

i	x_i	Δ'_k	x_i	Δ'_k	x_i	Δ'_k
1	2.579 (1.99)	1.579	3.761(4.037)	3.037	1.576 (.548)	.576
2	4.573 (4.445)	1.993	8.167 (5.226)	4.129	2.167(.931)	.590
3	6.684 (6.792)	2.110	12.545 (7.564)	4.378	2.761(1.280)	.593
4	9.179 (9.420)	2.495	17.8120 (9.826)	5.266	3.515 (1.800)	.754
5	11.773 (11.880)	2.594	23.156(11.598)	5.344	4.353 (2.589)	.837
6	14.379 (14.418)	2.605	28.400 (13.608)	5.243	5.287 (3.727)	.934
Ref.Point	No/Zero		Beliefs		Beliefs	
Domain	No/Gains		Gains		Losses	

Note: This table presents the average, standard deviations of the sequence $x_1, x_2, x_3, x_4, x_5, x_6$ along with the difference $\Delta'_j = x_i - x_{i-1}$. Standard deviations are presented in parenthesis. Columns 2 and 3 present these statistics when all the data is taken into account. Columns 4, 5, 6, and 7, present these statistics when it is assumed that subjects make decisions around Beliefs as a reference point. Columns 4 and 5 present the mean and median of $x_1, x_2, x_3, x_4, x_5, x_6$ along with $\Delta'_j = x_i - x_{i-1}$ for values above Beliefs for each subject. Columns 6 and 7 present the mean and median of $x_1, x_2, x_3, x_4, x_5, x_6$ along with $\Delta'_j = x_i - x_{i-1}$ for values below Beliefs for each subject.

the power utility function is assumed.

All in all, the data suggest that subjects have linear utility functions. This result is robust to the assumption that subjects have a non-zero reference point following an expectation-based rule. This is not a surprising finding given the stakes offered in the experiment and the discussions about utility curvatures by Wakker and Deneffe (1996) and Rabin (2000). Furthermore, the conclusion that the utility function is linear implies that probability distortions, if any, fully determine the risk attitudes of the subjects in the experiment. Thus, performance differences across treatments must be explained by probability distortions rather than by the curvature of the basic utility.

Table 14: Parametric Estimates of the utility function

Exponential (CARA) $1 - exp(-\gamma(x_{i-1} + \frac{\epsilon}{2}))$			
$\hat{\gamma}$	.977 (.001)	.946 (.001)	1.337 (.001)
Adj. R^2	0.922	0.887	0.303
N	1032	412	619
Power Utility (CRRA) $(x_{i-1} + \frac{\epsilon}{2})^{\phi}$			
$\hat{\phi}$	.995 (.001)	.992 (.001)	1.035 (.007)
Adj. R^2	0.925	0.971	0.756
N	1032	412	619
Ref.Point	No/Zero	Beliefs	Beliefs
Domain	No/Gains	Gains	Losses

Note: This table presents the estimates of the non linear least squares regression. The upper panel assumes that the parametric form $1 - exp(-\gamma(x_{i-1} + \frac{\epsilon}{2}))$ and the lower panel assumes the parametric form $(x_{i-1} + \frac{\epsilon}{2})^{\phi}$. The first column presents uses all the data. The second and third column present the data for the domain of gains and the domain of losses, respectively, when the reference point is Beliefs. Standard errors in parenthesis. *** denotes significance at the 0.01 level, ** denotes significance at the 0.05 level, * denotes significance at the 0.1 level.

Appendix F: Additional regressions

Table 15: The mechanism driving performance

	(1) Perfomance	(2) Perfomance	(3) Perfomance	(4) Perfomance
LowPr	15.807**	12.531*	15.697**	22.228**
	(6.865)	(7.479)	(6.861)	(8.825)
MePr	1.407	1.292	1.521	1.542
	(6.185)	(6.190)	(6.221)	(6.265)
HiPr	1.484	1.279	1.619	1.281
	(5.456)	(5.445)	(5.476)	(5.454)
LS and US	-1.699	-8.092		
	(5.509)	(6.370)		
LowPr*LS and US		20.896**		
		(8.862)		
OverweightS			0.001	-0.001
			(0.001)	(0.001)
LowPr*OverweightS				0.002
				(0.001)
Self efficacy	0.608	0.761	0.688	0.536
	(1.169)	(1.133)	(1.197)	(1.182)
Task difficulty	-6.737***	-6.704***	-6.759***	-6.612***
	(1.945)	(1.915)	(1.954)	(1.930)
Math skills	2.922**	2.927**	2.848**	3.038**
	(1.291)	(1.278)	(1.271)	(1.263)
Gender	-2.350	-2.434	-2.459	-2.321
	(4.591)	(4.621)	(4.635)	(4.624)
Mixed Utility	-11.320	-11.454	-11.487	-12.152
	(11.005)	(11.059)	(11.072)	(10.681)
Convex Utility	9.705	8.454	9.523	16.338
	(11.892)	(13.429)	(12.275)	(13.584)
Linear Utility	-8.548	-8.961	-9.062	-8.987
	(9.471)	(9.550)	(9.422)	(9.190)
Constant	84.332***	84.466***	84.810***	82.302***
	(13.700)	(13.812)	(13.772)	(13.644)
Adj. R^2	0.264	0.273	0.264	0.273
N	172.000	172.000	172.000	172.000

Note: This table presents the estimates of the Ordinary Least Squares regression of the model $Performance_i = \beta_0 + \beta_1 LowPr + \beta_2 LowPr * LSandUS + \beta_3 MePr + \beta_4 HiPr + \beta_5 LSandUS + Controls'\Gamma + \epsilon_i$, with $E(\epsilon|MePr, LowPr, HiPr, Controls, LSandUS) = 0$, which are presented in columns (1) and (2). The OLS estimates of the model $Performance_i = \beta_0 + \beta_1 LowPr + \beta_2 LowPr * OverweightS \beta_3 MePr + \beta_4 HiPr + \beta_5 OverweightS + Controls'\Gamma + \epsilon_i$, with $E(\epsilon|MePr, LowPr, HiPr, Controls, OverweightS) = 0$ which are presented in columns (3) and (4). "Performance" is the number of correctly solved sums in the first part of the experiment, "LowPr", "MePr" and "HiPr" are dummy variables that capture whether the subject was assigned to the treatment with low, medium or high probability of outcome evaluation, respectively. "LS and US" captures whether a subject has a probability weighting function with lower subadditivity and upper subadditivity. "OverweightS" captures the general probability overweighting of a subject through the index $\sum_{j=1}^{2} w(p_j) - p_j$. The controls considered in this model are "Gender" which captures the gender of the participant, "Belief " which captures the performance belief of the subject, "Math Skills " which captures the self-reported mathematical skills of the subject, "Task Difficulty " which captures the self-reported difficulty to perform the task. Robust standard errors in parenthesis. *** denotes significance at the 0.01 level, ** denotes significance at the 0.05 level, * denotes significance at the 0.1 level.