Probability Distortions as Incentives*

Victor H. Gonzalez-Jimenez †

March 19, 2018

Abstract

This paper introduces a novel incentive scheme designed to take advantage of the regularity that individuals distort probabilities. Under the proposed incentive scheme, a worker is incentivized to perform a productive task with a lottery that pays with some probability a monetary compensation based on her performance or nothing at all. The principal is able to choose this probability and she makes this decision before the worker performs the task. Thus, her choice could influence the worker's motivation. A theoretical framework and a laboratory experiment demonstrate that this incentive scheme outperforms standard performance-pay schemes that deliver, on expectation, the same monetary incentives. However, the probability at which the scheme is implemented is critical to its effectiveness. A small probability of performance compensation (10%) leads to higher performance than a standard performance-pay scheme, whereas medium and high probabilities (33.3 % and 50%) yield no differences. I present evidence demonstrating that the degree at which individuals overweight small probabilities drives this performance boost.

JEL Classification : C91, C92, J16, J24.

Keywords: Incentives, Experiments, Probability Weighting Functions, Rank-dependent Utility.

---

*I am indebted to the helpful comments made by Patricio S. Dalton, Wieland Mueller, Charles Noussair, Arthur Schram, Goerg Kirchsteiger, Jan Potters, Robert Dur, Sandra Rincon, and Stephanie Koornneef. I am grateful to participants in seminars at Tilburg University, University of Vienna and the Bogota Experimental Economics Conference for their helpful comments. I thank the Center for Economic Research (CentER) at Tilburg University for funding the experiments reported in this paper.

†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

Understanding the scope of incentives is of crucial interest to economists. Incentives, when sufficiently powerful, could modify the pursuits of individuals. The seminal work by Holstrom implements this idea in the context of a principal-agent relationship, wherein a principal uses monetary incentives efficiently to motivate her workers to exert costly effort on a profitable project (Holmstrom, 1999; Holmstrom and Milgrom, 1991; Holmstrom, 1979). This finding lead to a broad literature that studies the optimal design of contracts (see Gibbons and Roberts (2013) for a review).

Although the traditional focus of contracting has been the efficient usage of monetary incentives, recent literature has highlighted the motivational effect of non-monetary incentives. Tools such as production goals provided by the employer (Corgnet et al., 2015; Gómez-Miñambres, 2012), status contests within the organization (Ashraf et al., 2014; Bandiera et al., 2013; Besley and Ghatak, 2008; Auriol and Renault, 2008), and peer-surveiled environments (Falk and Ichino, 2006) have been proven to motivate the worker even in situations in which monetary incentives are already at work. These incentives are attractive for the principal inasmuch as their power does not rely on the monetary incentives that they deliver, but on their capacity to take advantage of cognitive and social biases to enhance the motivation of the individual to perform a task.[1] Hence, by implementing non-monetary incentives a principal could achieve targeted production levels at lower costs.

In this paper I introduce a novel incentive scheme which effectiveness stems from the non-monetary incentives that it delivers. In the incentive scheme, addressed as the "probability contract", a worker is incentivized to perform a task in each of the contracted periods with a lottery. The lottery pays with some probability a monetary compensation based on the worker's performance on the task in that period, or pays nothing. The employer is able to choose the probability that performance in each of the periods is paid, and she makes this decision before the first contracted period starts. Therefore, the choice of the employer could potentially influence the performance levels delivered by the worker.

The timing of the probability contract is as follows. First, the principal chooses the probability that performance in each of the contracted periods counts toward compensation. This choice is then communicated to the worker before she performs the productive task. Subsequently, the worker performs the task in each of the contracted periods. Finally, when all the periods have elapsed, a fraction of periods that fulfil the probability specified by the employer

[1] For instance, letting the worker set a production goal induces psychological losses from falling short to achieve such target, which boosts the exertion of effort. This reflects Heath et al. (1999)'s idea that goals are reference points and can induce loss aversion Kahneman and Tversky (1979). Similarly, tournaments and status-based incentives exploit the individuals' preferences for social status (Besley and Ghatak, 2008; Moldovanu et al., 2007; Auriol and Renault, 2008).

are selected at random. The worker is then compensated for her performance in those periods.

Apart from offering monetary incentives, the probability contract aims to motivate the worker by exploiting the behavioral regularity that individuals distort probabilities systematically (Prelec, 1998; Wu and Gonzalez, 1996a; Tversky and Fox, 1995; Tversky and Kahneman, 1992). This feature makes the proposed incentive scheme more attractive for an employer than standard incentive schemes that only rely on the provision of monetary incentives. To see how, consider a principal implementing the probability contract with a small probability. A worker that overweights small probabilities perceives this probability to be larger, and this motivates her to exert higher effort as compared to the situation in which she evaluated these probabilities accurately. The aggregation of such effect over several periods yields higher productivity than standard performance-pay incentive schemes offering, on expectation, the same monetary incentives. This is because in standard performance-pay schemes performance is constantly evaluated and there is no room for the distortion of probabilities. Indeed, a theoretical model and a laboratory experiment corroborate that implementing the contract with small probabilities, which are overweighted, yield higher performance than a cost-equivalent piece rate.

A theoretical framework studies the incentives delivered by the probability contract. I compare the behavior of the worker when offered the probability contract to a situation in which she is offered a piece rate contract. To simplify the analysis I let these two contracts yield, on expectation, the same monetary incentives, so that they only differ on the frequency at which performance is evaluated. If the preferences of the worker under uncertainty could be represented by expected utility, the risk attitudes of the worker, as captured by the curvature of the utility function, determine the effectiveness of the probability contract. For instance, risk averse workers dislike the uncertainty introduced by the probability contract, and hence would perform worse as compared to the piece rate. This result holds for all the possible probability choices of the principal. However, if workers have preferences under uncertainty that admit a distortion probabilities as captured by a probability weighting function with an inversed s-shape, then the probability at which this contract is implemented affects this comparison. Low probabilities may yield higher performance under the probability contract, whereas larger probabilities may yield lower performance. This is because the overweighting of small probabilities acts as a risk-seeking mechanism and the underweighting of large probabilities as a risk-aversion mechanism. Therefore, a worker with a concave utility function could deliver higher performance under the probability contract if she is offered a probability with a sufficiently large overweighting.

A controlled laboratory experiment is designed to test the predictions of the theoretical framework. The experiment consists of two parts. The first part features the subjects working on an real effort task and getting compensated for their performance in such task. The specific monetary incentives offered to the participants vary with the treatment assignment. A subject could be assigned to one out of four treatments: i) A piece rate, paying a monetary amount for every correctly solved task that she delivers, ii) a probability contract with low probability of evaluation, in which a period is paid with 10% probability, iii) a probability contract with medium probability of evaluation, in which a period is evaluated with 30% probability, and iv) a probability contract with high probability of evaluation, in which a period is evaluated with 50% probability.[2] I calibrate the parameters of the experiment such that, on expectation, a subject that achieved certain performance level gets, on expectation, the same monetary payoffs across the four treatments.[3] The second part of the experiment, featured subjects expressing their preference between lotteries from a list of paired lotteries. This part of the experiment is designed to elicit the subjects' utility functions and their probability weighting functions, separately. To that end, I use the two-step method proposed by Abdellaoui (2000), which has the advantage of not imposing functional forms of the utility in the design of the lotteries, something that certainty equivalents and probability equivalents do.

The results of the first part of the experiment demonstrate that the probability contract with low probability of evaluation yields higher performance in the task than the rest of the contracts. A subject assigned to such treatment outperforms any subject assigned to a different treatment by on average 0.5 standard deviations. This result goes in line with the prediction of the model wherein subjects distort probabilities according to an inverted s-curve. The data of the second part of the experiment show that the majority of the subjects exhibit linear utility functions over monetary outcomes.The data also show that subjects distort probabilities systematically: they overweight small probabilities, but, in stark contrast with most of the literature on probability judgements, they also overweight moderate probabilities and do not underweight large probabilities. Finally, I show that the degree at which subjects distort probabilities captures the totality of the treatment effect. The performance boost that arises from being assigned to the probability contract with low probability appears once the subject overweights small probabilities, and the higher is the subject's overweighting of small probabilities, the higher is her performance.

This paper contributes to several strands of literature. First, it adds to the literature of behavioral contract theory (See Koszegi (2014) for a review). A number of studies have explored the distortions of probabilities to explain insurance behavior (Ania, 2006; Wakker et al., 1997), financial investment and derivatives pricing (Polkovnichenko and Zhao, 2013; Barberis and Huang, 2008), and tax evasion (Dhami and Al-Nowaihi, 2007). To my knowledge this is the

[2]These percentages are chosen based on Prelec (1998)'s probability weighting function, according to which individuals overweight probabilities over the probability interval $p \in [0, .33)$ and underweight probabilities for the interval $p \in (0.33, 1]$. The point $p = 0.33$ is the point at which the individuals evaluate the probability accurately.

[3]I achieve this by changing the monetary payment offered for performance. Particularly, I multiply the piece rate offered in the piece rate treatment by the inverse of the fraction of the periods that are evaluated in each of the treatments. For instance, in the treatment in which only 10% of the periods are evaluated, I multiply the piece rate by 10.

first paper to employ probability distortions to motivate effort. However, the idea that a principal can exploit the cognitive biases of an employee is not new, and has been studied in other situations. For instance, DellaVigna and Malmendier (2004) and Heidhues and Koszegi (2010) study how firms can design contracts to take advantage of the agents' time inconsistency, Fehr et al. (2007), Goette et al. (2004), and Herweg and Mueller (2010) study the design of incentives when agents exhibit loss aversion, and Sandroni and Squintani (2007) studied the design of contracts when agents are overconfident.

Second, this paper contributes to the literature that investigates the effect of incentives on performance in a controlled laboratory environment (Charness and Kuhn, 2011). Particularly, I test the motivational effect of the proposed incentive scheme, the probability weighting contract, using a between subjects design. This is akin to an ideal situation where an employer is able to offer different contracts to his employees, to evaluate which one incentivizes them best. Other papers have studied in the laboratory the motivational effects of tournament incentives (van Dijk et al., 2001; Nalbantian and Schotter, 1997; Bull et al., 2016), high and low powered piece rates (Gneezy and Rustichini, 2000; Ariely et al., 2009), and bonuses (Falk et al., 2008; Fehr et al., 2007; Fehr and Goette, 2004; Nalbantian and Schotter, 1997), among others.

The results of this study also add to the methodological literature that studies the proper incentivization of subjects in the laboratory. The probability contract is based on Holt (1986)'s random-lottery incentive system. In a recent paper **?** has shown that under modest conditions this incentive system is incentive compatible. Among others, this incentive system avoids income effects, which belongs according to healy **?** to a complementarity at-the-top. This feature yields robust estimates of the subjects' risk preferences (Baltussen et al., 2012; Lee, 2008; Hey and Lee, 2005; Cubitt et al., 1998). The results of my paper present a caveat for the implementation of the random-lottery system when subjects work on real effort tasks: this mechanism leverages higher performance than an equivalently powered piece rate when subjects face a small probability that a round is chosen for performance payment.

Finally, my results contribute to the literature that investigates the properties of the probability weighting function (van de Kuilen and Wakker, 2011; Bruhin et al., 2010; Abdellaoui et al., 2008; Fehr-Duda et al., 2006; Etchart-vincent, 2004; Abdellaoui, 2000; Gonzalez and Wu, 1999; Prelec, 1998; Wu and Gonzalez, 1996a; Tversky and Fox, 1995; Tversky and Kahneman, 1992). I study the shape of the probability weighting functions using the two step method by Abdellaoui (2000). However, in my setting the prizes of the lotteries reflect their earnings in the real effort task, which addresses a possible house money effect (Ackert et al., 2006). I find empirical evidence suggesting that individuals overweight small probabilities. In contrast with previous studies subjects also overweight medium-sized probabilities and do not underweight large probabilities. These results correspond to an inversed s-shaped probability weighting function with an elevation that is higher than previously documented.

2 The model

The theoretical framework compares the incentives delivered by the probability contract and those delivered by a standard pay-for-performance incentive scheme. Even though I assume that both incentives schemes are, on expectation, equally costly for the principal, they differ on the frequency in which output is evaluated. Under the standard performance-pay scheme output is constantly evaluated and the worker is compensated according to the output level that she delivers in all periods. In contrast, in the probability contract thee employer chooses the frequency at which output is evaluated and the worker receives a compensation based on her performance in the periods that were chosen for evaluation.

The model considers a worker with a time horizon of T periods $t = 0, 1, ..., T$. In each period t, the worker makes a decision about the production output, $y_t \geq 0$, that she delivers to the employer. This decision depends on the monetary rewards provided by the contract that is chosen by the principal. The following subsection introduces the two contracts that are studied in this paper.

2.1 The contracts

A performance-pay contract: piece rate contract

To represent standard performance-pay incentive schemes I use a piece rate. A piece rate offers constant marginal monetary incentives, which means that the worker is incentivized to produce an additional unit of output irrespective of how many units of output she may have already delivered. This characteristic represents better the notion of performance-pay than other schemes such as lump-sum bonuses for reaching a target, which does not offer monetary incentives to the worker for all possible output levels.

Consider a situation in which the worker is offered a piece rate contract. Formally, the worker is offered a wage scheme $W_t = ay_t$, where $a > 0$ represents the monetary reward for every unit of output that she delivers. I assume that the worker has a utility for monetary rewards $b(\cdot)$ that is an increasing and two-times continuously differentiable function.

Assumption 1: $b(\cdot)$ is a C^2 function with $b(0) = 0$ and $b_y(\cdot) > 0$.

Note that I do not make any assumptions over the sign of the second derivative of the utility for monetary rewards. The reason for this is that the results of the model are going to be evaluated under the two different signs that this derivative can attain.

The worker also experiences disutility from producing output in each period, this reflects the idea that working on the task requires attention, persistence, and effort, which can be depleting. I model this disutility through the function $c(y_t, \theta)$ an increasing, twice-differentiable and strictly convex function.

Assumption 2: $c(y_t, \theta)$ is a C^2 function with $c_y(0, \theta) = 0$ $c_y(y_t, \theta) > 0$, $c_{yy}(y_t, \theta) > 0$, and $c_{y\theta}(y_t, \theta) < 0$.

The parameter $0 < \theta \leq 1$ represents the worker's ability. The last expression in Assumption 2 captures that higher ability levels allow the worker to have a flatter cost function and achieve higher production levels.

All in all, the worker's utility in period t can be written as

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

The worker maximizes her utility function by delivering a production level y_t^* that satisfies the following first order condition[4]

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

Equation (2) shows that the optimal production level, y^*, increases with the agent's abilities θ, the assumption that $c_{y\theta}(y_t, \theta) < 0$ guarantees this comparative static. Moreover, higher powered incentives a lead to higher output provision from the part of the agent y_t^*, which is guaranteed by the assumption that ensures that y^* is a maximum of the agent's program, this is that $b(\cdot)_{yy} < c_{yy}(y, \theta)$. Hence, higher abilities make it less costly for a worker to deliver higher levels of output and higher monetary incentives increase the marginal utility of an additional unit of output, which motivates the worker to deliver higher output levels.

For illustrative purposes, consider the functional forms $c(y_t, \theta) = \frac{(y_t/\theta)^2}{2}$ and $b(ay_t) = \frac{(ay_t)^{1-\gamma}}{1-\gamma}$, with $-1 < \gamma < 1$. For these forms the optimal output level chosen by the worker has the closed-form solution $y_t^* = \left(\frac{\theta^2}{a^{\gamma-1}}\right)^{\frac{1}{1+\gamma}}$, that exhibits the positive effects of higher powered monetary incentives and higher abilities on the task on production.

The probability contract

Consider now the situation in which the worker is given the probability contract. This incentive scheme also offers a monetary compensation based on the worker's performance on the task. However, in contrast to the piece rate contract, performance in a period is paid with some probability. The employer is able to choose this probability, and thus has the potential

[4]A necessary condition for (2) to attain a maximum of the function $U(y_t)$ is that $a^2 b_{yy} < c_{yy}(y, \theta)$. This means 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.

to influence the decision of the worker on how much output to deliver.

The contract stipulates the employer moving first at the beginning of the productive horizon. Her choice consists of selecting $\tau \in \{1, 2, ..., T\}$ that represents a fraction of all the contracted periods that she would like to evaluated. This choice is then communicated to the worker before she engages in the productive task. Hence, the worker's choice about how much effort to deliver in each period, is influenced by the probability that performance in a period counts towards her earnings, $p \equiv \frac{\tau}{T}$. Finally, after all the work periods have elapsed, the number of periods satisfying τ are chosen randomly.

The proposed compensation scheme can be written as

$$V_t = \begin{cases} By_t \text{ if period } t \text{ is chosen for evaluation,} \\ 0 \text{ if period } t \text{ is not chosen for evaluation.} \end{cases}$$

Where $B > 0$ represents the monetary compensation offered to the worker for each unit of output that she delivers if period t is chosen for evaluation. From the perspective of the worker in each period, facing this incentive scheme is akin to face the lottery $(By_t, p; 0)$. Her choice when faced with this lottery is to maximize her utility by choosing y_t.

Under this contract, the worker's expected utility at period t is

$$E(U_t(y_t)) = pb(By_t) - c(y_t, \theta). \tag{3}$$

The worker maximizes her expected utility by 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 the higher is the probability that a period is chosen p, the higher is the output delivered by the worker. Intuitively, a worker is motivated to deliver more output if the period in which she is working is more likely to be chosen for compensation. Also, as in the analysis of the incentives delivered by the piece rate contract, I find that higher abilities on the task, θ, and higher monetary incentives B yield higher optimal output.

As an illustration, consider the functional forms $c(y_t, \theta) = \frac{(y_t/\theta)^2}{2}$ and $b(By_t) = \frac{(By_t)^{1-\gamma}}{1-\gamma}$ with $-1 < \gamma < 1$. The optimal output level under these specifications has the closed form solution $y_t^{**} = \left(\frac{\theta^2 p}{B^{\gamma-1}}\right)^{\frac{1}{1+\gamma}}$, which depicts the positive effect of higher abilities of the worker, higher monetary incentives, and higher probabilities of output evaluation on production.

The probability contract and workers who distort probabilities

So far, I have implicitly assumed that the worker evaluates probabilities accurately and the results of the previous analyses hinge on this assumption. In this subsection, I let the worker distort probabilities systematically as it is suggested by empirical evidence (Bleichrodt and Pinto, 2000; Abdellaoui, 2000; Gonzalez and Wu, 1999; Wu and Gonzalez, 1996a; Tversky and Fox, 1995; Tversky and Kahneman, 1992). As it will be later evident, whether a worker distorts or not the probability that a period t is chosen is key to the effectiveness of the probability contract.

Assume that a worker weights the probabilities p that output is evaluated in a period t, using the probability weighting function $w(p)$ that satisfies the following properties.

Assumption 3 $w(p)$ is a 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 $w(1) = 1$.
- $\lim_{p\to 0} \frac{w(p)}{p} = \infty$ and $\lim_{p\to 1} \frac{1-w(p)}{1-p} = \infty$
- There exists a $\tilde{p} \in (0, 1)$ such that $w(\tilde{p}) = \tilde{p}$.
- $w(p) > p$ if $p \in [0, \tilde{p})$ and $w(p) < p$ if $p \in (\tilde{p}, 1]$.
- There exists a unique $\hat{p} \in (0, 1)$ such that $w_{pp}(p) < 0$ if $p \in [0, \hat{p})$ and $\pi_{pp}(p) > 0$ if $p \in (\hat{p}, 1]$.

According to Assumption 3, the probability weighting function $w(p)$ is a two-times continuously differentiable function that maps the unit interval into itself and exhibits a positive slope everywhere. This function infinitely-overweights infinitesimal probabilities and infinitely-underweights near-one probabilities. Moreover, it contains three fixed points: one at $p = 0$, another at $p = 1$, and an interior fixed point $\tilde{p} \in (0, 1)$. Additionally, the function has an inversed-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 that $\tilde{p} = \hat{p}$, characteristic of early axiomatizations of probability weighting functions, and I instead let these two values differ.

The assumption that workers evaluate probabilities according to the probability weighting function $w(p)$ is not a sufficient representation of the worker's preferences when they face the uncertainty posed by the probability contract. This is because a probability weighting function itself is not a theory of risk and needs to be embedded within other theories.Therefore, I assume that the preferences of the worker are represented either by rank-dependent utility (Quiggin, 1982), (RD from here onward) or by cumulative prospect theory (Tversky and Kahneman, 1992), (CPT from here onward). These theories accommodate probability weighting functions but differ in the assumptions made over the utility that the individual derives from monetary outcomes. Under RD the worker's utility for money can be represented by the function $b(\cdot)$ with the additional assumption that this function is concave. The rank-dependent utility of the worker, who distorts probabilities according to $w(p)$ and who faces the probability contract is similar to equation (3), with the differences that p is replaced by $w(p)$ and that that $b_{yy} \leq 0$.[5] Therefore, the worker maximizes the rank-dependent expected 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 that presented in previous analyses: The output level chosen by the worker increases with higher skills θ and with higher monetary incentives B. Also, higher probabilities of evaluation p yield higher output, but in comparison to the case in which the worker 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 in the interval $p \in (\hat{p}, 1]$.

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

Alternatively, under CPT the worker evaluates the monetary outcomes of the contract around a reference point $r > 0$. This process is represented by the value function $v(y, r)$ that has the following properties.

Assumption 4: $v(y, r)$ is the piecewise function $v(y,r) = \begin{cases} b(By_t - r) \text{ if } pBy_t \geq r, \\ -\lambda b(r - By_t) \text{ if } pBy_t < r. \end{cases}$, with $r > 0$, $\lambda > 1$, $b(0) = 0$, $b_y(By - r) > 0$ for all $y \geq 0$, $b_{yy}(By - r) < 0$ for any $\beta y > r$ and $b_{yy}(By - r) > 0$ for any $\beta y < r$.

[5] According to RD 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 $\pi_n = w(p_n), \pi_{n-1} = w(p_{n-1} + p_n) - w(p_n)), ..., \pi_1 = 1 - \sum_j^n$. The ordered outcomes of the lottery enter the utility function through a 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, simply because in our setting a worker faces in each period the lottery $(p, By_t; 1 - p, 0)$ so that the decision weight is $\pi = w(p) - w(0)$, which becomes $\pi = w(p)$ since $w(0) = 0$.

The reference point r typically represents a monetary amount that the worker expects to receive, a monetary amount that was received in the past, or an amount of money that the worker has at period (Koszegi and Rabin, 2006; Kahneman et al., 1991). The value function specifies that monetary changes above that reference point, lead to marginally decreasing utility increments and monetary increments below that reference point lead to marginally increasing utility increments. In other words, the worker's value function is concave in gains and convex in losses, where losses and gains are relative to the reference point r. Additionally, the worker is loss averse which means that she looms losses larger than gains. This is represented by the parameter $\lambda > 1$ which enters the utility function only in the domain of losses.[6] The utility of the worker with CPT preferences is equal to

$$U(y, r) = w(p)v(By, r) + w(1 - p)v(0, r) - c(y, \theta). \tag{6}$$

There are three underlying assumptions about this utility function that are worth being discussed. First, the curvature of the utility for money under RD and CPT are represented by the same function $b(\cdot)$. This suggests that the functional form that captures the tastes for monetary outcomes is the same irrespective of whether these outcomes are evaluated around a non-zero reference point or in absolute terms. Even though this may be a stringent assumption, it simplifies the comparative analyses of the contracts. One could think about a power function evaluating the preferences for By_t under RD and $By_t - r$ under CPT. Second, I assume that the function that captures the curvature is the same for gains and for losses, this means that loss aversion, as represented by the parameterλ, is the same for large and small monetary outcomes. This assumption is made in the literature by Wakker (2010) and Köbberling and Wakker (2005) to avoid analytical problems of loss aversion that arise when different curvatures are used for gains and losses. Finally, I assume that the probability weighting function is the same for gains and losses. This allows be to abstract from situations of probabilistic loss aversion, and let the parameter λ to be only source of loss aversion.

The worker with CPT preferences provides a level of output, y_C^{**} that satisfies the following system of equations

[6]According to CPT an individual facing a lottery $(x_1, p_1; x_2, p_2; ...; x_n, p_n)$ ranks the outcomes in an increasing arrangement $x_1 < x_2 < ... < r < ... < x_n$ and evaluates the outcomes of the lottery relative to r through the function $v(y, r)$. The lottery outcomes $r < ... < x_n$ enter as gains and the outcomes $x_1 < ... < r$ enter as losses. The individual assigns decision weights to each of these outcomes according to their rank in the following way $\pi_n = w(p_n), \pi_{n-1} = w(p_{n-1} + p_n) - w(p_n)), ..., \pi_{r+1} = 1 - \sum_{j=r+1}^{n} p_j$ for gains and $\pi_1 = w(p_1), \pi_2 = w(p_1 + p_2) - w(p_1)), ..., \pi_{r-1} = 1 - \sum_{j=r-1}^{n} p_j$ for losses. In our model, the decision-weights are equivalent to the probability weighting functions, this is because in our setting a worker faces in each period the lottery $(p, By_t; 1 - p, 0)$ so that the decision weight is $\pi = w(p) - w(0)$, which becomes $\pi = w(p)$ since $w(0) = 0$.

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

$$B(1 - w(1-p))\lambda b_y(r - By_C^{**}) - c_y(y_C^{**}, \theta) = 0 \text{ if } pBy_t < r. \tag{8}$$

Let us first consider the case in which the worker is in the domain of gains. According to equation (7) output is increasing on the monetary incentives offered by the contract and on the abilities of the worker. Also, as in the case of RD preferences, the probability that a period is evaluated increases output in a non-linear way, with probability increments within the region $p \in (0, \hat{p})$ yielding smaller increases in output than equally large probability increments within the region $p \in (\hat{p}, 1)$.

When the worker is in the domain of losses the majority of the parameters of the model have a similar influence on output. For instance, higher abilities, higher incentives, and higher probabilities of evaluation increase the output level that is delivered by the worker. Moreover, higher values of the loss aversion parameter $\lambda > 1$ yield higher output levels, thus the worker delivers higher output to avoid experiencing losses.

Finally, the effect of the reference point r is ambiguous for both domains. Although a higher reference point shifts to the right the curves $Bw(p)b_y(By - r)$ and $B(1 - w(1-p))\lambda b_y(r - By)$, the intersection between any of these two curves and the marginal cost of production determines whether higher reference points yield higher output. To see this consider a reference point of zero, for this case only the curve $Bw(p)b_y(By - r)$ is relevant and the crossing point of this curve with the marginal cost of effort determines output. Increasing r leads to higher output as $Bw(p)b_y(By - r)$ shifts to the right, but by doing so $B(1 - w(1-p))\lambda b_y(r - By)$ becomes relevant. As soon as the marginal cost of effort intersects with the curve $B(1 - w(1-p))\lambda b_y(r - By)$, raising r could be counterproductive.The intuition is that higher reference points yield higher output up to a point after which they become unattainable and demotivate the worker. This is a well known regularity of endogenous reference points (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 this analysis, I let the two contracts deliver, on expectation, the same monetary incentives. This allows me to focus on the motivational effect of the probability contract as compared to the piece rate when different probabilities p are chosen.

Formally, let $B = a\frac{T}{t}$, so that the expected earnings from the probability contract$E(V_t) =$

pBy_t become $E(V_t) = pa\frac{T}{t}y_t = ay_t$, which is equivalent to the earnings offered by the piece rate contract W_t. Note that when the employer chooses to evaluate output constantly, this is $\lim_{\tau \to T} B = a$. Additionally, when the employer decides to evaluate output with very little frequency the compensation received by the worker under the probability contract becomes large enough so that the identity $E(V_t) = W_t$ is maintained.

With this cost-equivalence in mind, I rewrite the first order conditions describing the optimal output levels delivered by the 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

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

$$a\frac{1-(w(1-p))}{p}\lambda b_y(r - \frac{a}{p}y_C^{**}) - c_y(y_C^{**}, \theta) = 0 \text{ if } ay_t < r. \tag{12}$$

Let us start comparing y^* and y^{**} from equations (2) and (9). This comparison features a worker that evaluates probabilities accurately when she works under the probability contract. To build intuition about how these two production levels compare, l study the case in which the functional forms $c(y_t, \theta) = \frac{(y_t/\theta)^2}{2}$ and $b(By_t) = \frac{(By_t)^{1-\gamma}}{1-\gamma}$ are assumed. Under these conditions, a necessary condition for the inequality $y^{**} > y^*$ to hold is $\gamma \in (-1, 0]$, which means that the worker is required to be risk-seeking to deliver higher production levels under the probability contract. In contrast, a risk-averse worker delivers higher production levels under the piece rate. This analysis suggests that the effectiveness of the probability contract depends on the worker's risk preferences. Proposition 1 generalizes this result using the general functions $b(\cdot)$ and $c(y_t, \theta)$.

Proposition 1: For a worker with ability level $\tilde{\theta} \in (0, 1)$, then $y^{**} \geq y^*$ if $b_{yy}(\cdot) \geq 0$.

Proof. Suppose that $y^{**} < y^*$. For a worker with $\tilde{\theta}$ and since $c_{yy}(\theta, y_t) > 0$, this inequality can be written as $c_y(y^{**}, \tilde{\theta}) < c_y(y^*, \tilde{\theta})$. Using equations (2) and (9), we get

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

Note that for any $p \in [0, 1)$ and $y_t > 0$, then $\frac{ay_t}{p} > ay_t$. Assuming $b_{yy}(\cdot) \geq 0$, inequality (13) cannot hold and we have reached a contradiction.Then it must be that $y^{**} \geq y^*$ if $b_{yy}(\cdot) \geq 0$. Similarly, $y^{**} \geq y^*$ does not hold for $b_{yy} < 0$, then it must be that $y^{**} < y^*$ if $b_{yy} < 0$. Finally, note that for $p = 1$ the contracts are identical in terms of the incentives they yield, so in this case the inequality $y^{**} \geq y^*$ is binding.

□

Proposition 1 shows that when the worker does not distort probabilities, the curvature of $b(\cdot)$ is essential for the effectiveness of the probability contract. A risk averse worker, this is a worker with $b_{yy}(\cdot) < 0$, dislikes the uncertainty that this contract introduces and delivers lower performance as compared to a piece rate. In contrast, a risk seeking and a risk-neutral worker deliver higher performance under the probability contract, this is because the risk-seeking worker likes the uncertainty introduced by this contract and because the risk-neutral worker derives higher marginal utility from the monetary incentives offered by the probability contract. Also, it is important to emphasize that the employer's choice about the probability at which the probability contract is implemented p does change the result that for a population of workers that are predominantly risk averse, this contract yields lower output. This result constitutes the first prediction of the model.

Prediction 1: Risk averse workers deliver lower output under the probability contract. This result is independent of the employer's choice p.

Let us now focus on the comparison between y^* and y_R^{**} from equations (2) and (10). This comparison features a worker whose preferences under uncertainty can be represented by RD preferences. To build intuition, I first study how the production levels under the two contracts compare when the functional forms $c(y_t, \theta) = \frac{(y_t/\theta)^2}{2}$ and $b(By_t) = \frac{(By_t)^{1-\gamma}}{1-\gamma}$, are assumed. For this specific case and in contrast to the result from Proposition 1, I find that a worker with risk averse preferences, this is a worker with $0 < \gamma < 1$, could perform better under the probability weighting contract. This is because the sufficient condition for the inequality $y_R^{**} \geq y^*$ to hold is $\frac{w(p)}{p^{1-\gamma}} > 1$, which depends on the curvature of the utility function γ but also on $w(p)$ and p, which is chosen by the employer. To understand the importance of the employer's choice p, consider first a worker with $\gamma = 0$. For this worker, the inequality $y_R^{**} \geq y^*$ holds when $\frac{w(p)}{p} > 1$, which is induced when the employer chooses $p \in (0, \tilde{p})$. Hence, even though these two contracts deliver on expectation the same monetary incentives, the overweighting of the probability that a period t is chosen motivates the worker to deliver higher output in the probability contract.

As the worker becomes more risk averse, as captured by the curvature of her utility function, this is as $\gamma \to 1$, the necessary condition $\frac{w(p)}{p^{1-\gamma}} > 1$ becomes more stringent. In such cases, to guarantee $y_R^{**} \geq y^*$, the employer needs to choose probabilities that induce larger degrees of overweighting of probabilities, which occurs at $p \in (0, \tilde{p})$. [7]

Proposition 2 generalizes the result that workers that distort probabilities according to $w(p)$, deliver higher performance under the probability contract, even when they exhibit some degree of risk aversion that stems from the curvature of $b(\cdot)$.

Proposition 2: For a worker with an ability level $\tilde{\theta} \in (0,1)$ and risk aversion $|b_{yy}(\cdot)| \leq C$ for $C < \infty$, then $y_R^{**} \geq y_t^*$ if $p \in (p_m, \tilde{p})$ with $p_m = argmax_p(w(p) - p)$.

Proof. Suppose that $y_R^{**} < y^*$, $b_{yy}(\cdot) < 0$, and $\tilde{\theta} \in (0,1)$. Since $c_y(\tilde{\theta}, y) > 0$, $y_R^{**} < y^*$ can be written as $c_y(y_R^{**}, \tilde{\theta}) < c_y(y^*, \tilde{\theta})$. Using equations (2) and (10) we get

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

Proposition 1, states that (13) holds for any $p \in (0,1)$ if $b_{yy}(\cdot) < 0$. For $p \in (0, \tilde{p})$, we have $\frac{w(p)}{p} > 1$, which yields $y_R^{**} > y^{**}$. Thus, even if $b_{yy}(\cdot) < 0$, inequality (14) does not hold for all $p \in (0,1]$. Thus, for $p \in (0, \tilde{p})$, then $y_R^{**} > y^*$ even if $b_{yy}(\cdot) < 0$.

Next, I show that for $p \in (0, \tilde{p})$, $y_R^{**} > y^*$ does not hold for all $b(\cdot)$ with $b_{yy}(\cdot) < 0$. Since $0 \leq w(p) \leq 1$ and $0 < p \leq 1$, then $\exists M < \infty$ such that $\frac{w(p)}{p} \leq M$. However, $\nexists P < \infty$ such that $|b_{yy}(\cdot)| \leq P$, so even if $p \to 0$ is implemented, for sufficiently large $b_{yy}(\cdot) < 0$, then $y_R^{**} > y^*$ does not hold. Since $b(\cdot)$ is C^2, then $\exists C < 0$ such that $b_{yy}(\cdot) = C$ that induces $y_R^{**} = y^*$ when $p \to 0$ is implemented.

Which probabilities induce $y_R^{**} > y^*$? if $b_{yy}(\cdot) < C$, a sufficient condition ensuring $y_R^{**} > y^*$ must guarantee$\frac{\partial\left(\frac{w(p)}{p} b_y(\frac{ay_R^{**}}{p}\right)}{\partial p}) > 0$. This condition can be written as

$$\left(\frac{w(p) - w_p(p)p}{p}\right) b_y\left(\frac{ay_R^{**}}{p}\right) - \frac{w(p)}{p} y_R^{**} a b_{yy}\left(\frac{ay^{**}}{p}\right) > 0, \tag{15}$$

and holds if $w(p) > w_p(p)p$. Note that $\frac{w(p)}{p} > 1$ for $p \in (0, \tilde{p})$. Hence, It suffices to find a subset of $p \in (0, \tilde{p})$ where $w_p(p) < 1$. Given that $w_p(p_m) = 1$ for $p_m \equiv argmax_p(w(p) - p)$ and

[7]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 in this specific case happens for smaller probabilities since they ensure that $ln(p)$ attains larger negative values than $ln(w(p))$.

$w_{pp} < 0$ for $p \in (0, \tilde{p})$ then $p \in (p_m, \tilde{p})$ guarantees $y_R^{**} > y^*$ when $b_{yy}(\cdot) < C$

□

Proposition 2 demonstrates that the employer could be better off using the probability contract even when the curvature of the utility of the worker induces risk aversion. To achieve this result the employer is required to choose a probability p that induces a sufficiently large overweighting of probabilities in the worker. The more concave is the worker's utility, the larger is the probability overweighting that needs to be induced by the employer. However, this trick does not work for a worker with a very concave utility function, since her curvature outweighs any feasible level of overweighting of probabilities.

The intuition of this result is that, as pointed out by Tversky and Kahneman (1992), the systematic overweighting of probabilities acts as a risk seeking mechanism; by perceiving small probabilities to be larger, subjects choose more often lotteries that are unlikely to happen. Analogously, workers that otherwise would have delivered similar production levels under the two contracts, experience a production boost under the probability contract, because they perceive that the probability that a period is evaluated to be larger than it actually is. Whether the risk-seeking effect stemming from the overweighting of probabilities outweighs the risk aversion from the curvature of the function $b(\cdot)$, depends on two things: the degree at which subjects overweight probabilities, which can be manipulated by the employer with the choice of p, and the risk aversion that stems from the curvature of the function $b(\cdot)$. The results that the employer can motivate the risk averse using the probability contract constitutes Prediction 2.

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

Let us now study the case in which the worker's preferences can be represented through CPT. These preferences represent a worker that evaluates monetary outcomes as gains or losses around a reference point $r > 0$ and who in the presence of uncertainty, distorts probabilities according to $w(p)$. I compare the output delivered by this worker under the piecerate and under the probability weighting contract. The optimal output delivered by this worker under the piece rate contract is given by y_C^*, that satisfies the system of equations

$$ab_y(ay_C^* - r) - c_y(y_C^*, \theta) = 0 \text{ if } ay_t > r, \tag{16}$$

$$\lambda ab_y(r - ay_C^*) - c_y(y_C^*, \theta) = 0 \text{ if } ay_t < r. \tag{17}$$

An underlying assumption of this utility, is that the worker exhibits the same reference point across the two contracts. This may seem an stringent assumption, but given that our contracts pay on expectation the same, do not feature a performance target, do not have a bonus, and do not elicit an expectation from the part of the worker, it makes sense to assume that a worker has the same reference point under the two contracts [8].

The optimal output delivered by the worker under the probability weighting contract y_C^{**} is given by equations (11) and (12). Lemma 1, shows that in the domain of gains, the conditions presented in Proposition 2 guarantee that the probability contract yields higher output .

Lemma 1: For a worker with ability level $\tilde{\theta} \in (0, 1)$, with preferences over monetary outcomes represented by $v(r, y)$, with $r > 0$, $\lambda > 1$, and for $pBy > r$, then $y_C^{**} \geq y_C^*$ if $p \in (p_m, \tilde{p})$ and $b_{yy}(\cdot) < C$ for some $C < \infty$.

Proof. The proof is similar to that of Proposition 2, with the only difference that $r > 0$ is at both sides of the inequality. □

As in the case of RD preferences, the employer could be better off offering the probability contract if she implements it with a probability that induces a sufficiently large overweighting of probabilities in the worker. The overweighting of probabilities works as risk-seeking mechanism, motivating the worker to work harder than if she had evaluated these probabilities accurately. Again, the employer is required to induce larger overweighting of probabilities when the worker exhibits more concavity in the value function for the domain of gains. However, there is a certain level of concavity for the domain of gains after which the employer is better off offering the piece rate.

Let us now analyze the behavior of the worker when she is in the domain of losses. Lemma 2 shows that the probability contract yields higher output if it is implemented with any probability that does not induce an underweighting of probabilities in the worker.

Lemma 2: For a worker with ability level $\tilde{\theta} \in (0, 1)$, with preferences for monetary outcomes represented by $v(r, y)$, with $r > 0$, $\lambda > 1$, and for $pBy < r$, then $y_C^{**} \geq y_C^*$ for any $p \in (0, \tilde{p}]$.

Proof. Suppose that $y_C^{**} < y_C^*$, for $\tilde{\theta} \in (0, 1)$ and since $c_y(\tilde{\theta}, y) > 0$, $y_C^{**} < y_C^*$ can be rewritten

[8] 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 qualitative results presented in this section

as $c_y(y_C^{**}, \tilde{\theta}) < c_y(y^*, \tilde{\theta})$. Using equations (12) and (18) we get,

$$\frac{(1 - w(1-p))}{p} b_y(r - \frac{a}{p} y_C^{**}) < b_y(r - a y_C^*). \tag{18}$$

For any $p \in (0, 1)$ and $y_t > 0$, then $r - \frac{a y_t}{p} > r - a y_t$. Given that $b_{yy}(\cdot) \geq 0$ for $pBy < r$, then $b_y(r - \frac{a}{p} y_C^{**}) > b_y(r - a y_C^*)$. If $p \in (0, \tilde{p}]$ then $\frac{1 - w(1-p)}{p} \geq 1$ and we have reached a contradiction. Then it must be that $y_C^{**} \geq y_C^*$ at $pBy < r$, if $p \in (0, \tilde{p})$.

□

According to CPT, workers in the domain of losses have a convex value function and because they behave as if they like the uncertainty induced by the probability contract, they work harder under this contract than under the piece rate. However, to maintain this effect it is required to implement the contract with probabilities that do not act as a risk-aversion mechanism, which could potentially outweigh the risk seeking behavior of the workers. In other words, as long as the employer implements the contract with $p \in (0, \tilde{p}]$, we should observe higher output from the part of the workers under the probability contract.

Lemma 1 and 2 show that the conditions presented in Proposition 1 suffice to motivate the worker to work harder under the probability contract even when her preferences are represented by CPT. On the one hand, the employer needs to set a probability that induces a sufficiently high overweighting of probabilities when the worker is in the domain of gains. On the other hand, the employer needs to set a probability that does not induce an underweighting of probabilities if the worker is in the domain of losses. This result constitutes Prediction 2a.

Prediction 2a: Workers with a non-zero reference point and that distort probabilities systematically deliver higher output under the probability contract when the choice of p yields a sufficiently large overweighting of probabilities.

With these competing predictions in mind, I run a controlled laboratory experiment in which subjects work 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 a high, medium, and a low probability of evaluation. According to Prediction 1, the frequency at which performance is evaluated is irrelevant. If subjects do not distort probabilities, their risk preferences determine the effectiveness of the probability contract. In such case, a pool of subjects that are predominantly risk averse, is going to display lower performance in the task under the probability contract. Nonetheless, if subjects distort probabilities, we can expect the difference in probabilities at which the probability contract is implemented to matter: subjects assigned to the treatment with low performance evaluation must exhibit higher performance than those assigned the high performance evaluation.

Throughout this section I focused on the worker's incentive compatibility constraint when she faces a probability contract as compared to the case in which she faces a piece rate. In Appendix A I show a detailed analysis of the solution of the employer's program when the agent has RD preferences . The main finding of this analysis is that the risk neutral principal also chooses a low probability $p \in (0, \tilde{p})$ in order to induce a sufficiently large overweighting of probability. However, since the employer's problem also takes into account the possibility that the worker may not accept this contract, with the participation constraint, the employer is often forced to implement the piece rate.

3 Experimental Method

The experiment was conducted at CentERLAB in Tilburg University in April 2017. The participants were recruited using an electronic system and all of them were students at the university. The dataset 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 can be found in Appendix B.

The experiment consisted of two parts. Upon arrival, participants were informed that their earnings in part one or those from part two become their definitive final earnings and that this was decided by chance at the end of the experiment. In the first part of the experiment, subjects performed a task required their effort and attention. The task consisted on adding the sum of five two-digit numbers.[9] Each summation featured randomly drawn numbers, guaranteeing similar levels of difficulty for all participants. When a subject knew the answer to the summation, she submitted her answer using the computer interface, and a new summation task appeared in their screen. Subjects had 10 rounds of 4 minutes each to complete as many summations as they could. In other words, they had 40 minutes to do their best in the task.

There were four treatments that differed in the way in which monetary incentives were given to the subjects. The participants were randomly assigned to one of these four treatments. The baseline treatment is *Piecerate* in which subjects were paid 0.25 euros for each correctly solved summation. The other three treatments also provided monetary incentives on the basis of individual performance on the task; however, in these treatments a randomly chosen subset of rounds was chosen at the end of the experiment and only performance in that subset counted

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

towards their earnings.

The treatments *LowPR*, *MePR* and *HiPR* feature a low, a medium, and a high probability, respectively, that performance in a round was selected for payment. Specifically, in LowPR one round out of the ten was randomly chosen and performance in that specific round counted towards the subject's earnings. Similarly, in MePR and HiPR, three and five rounds, respectively, out of ten rounds were randomly chosen and counted towards the subject's earnings. These numbers were chosen based on the probability weighting function derived by Prelec (1998). According to this probability weighting function, subjects overweight probabilities when they face a probability of 10%, they underweight probabilities when they face a probability of 50%, and they evaluate probabilities accurately when they face a probability of 33.3%.

As in the theoretical framework, the monetary compensation in LowPR, MePR and HiPR, was calibrated in such a way that subjects, on expectation, earn the same amount across all treatments. For instance, a subject assigned the LowPR received 2.50 Euros for each correctly solved summation in the period that was chosen for performance evaluation. This compensation is ten times more what a subject assigned the Piecerate earns. This monetary difference makes up for the low frequency in which output is evaluated in LowPR. Similarly, subjects assigned the MePR and HiPR treatments received a compensation of 0.85 and 0.50 Euros per solved task, respectively.[10]

Once the last round of the real effort task was over, the participants were asked to state their beliefs about how well they did in the first part of the experiment. This elicitation is used to investigate whether subjects anticipated the effect of the treatment on their own performance. Participants received a bonus of one Euro if their answer was exactly equal to the number of correct summations throughout the ten rounds. This belief elicitation was unanticipated by the subjects and the monetary compensation for accurate answers is small compared to other sources of earnings in the experiment. As explained by Schlag et al. (2015), these two characteristics are desirable inasmuch as they ensure incentive compatibility and do not allow subjects to hedge against their own performance in the task(Blanco et al., 2010).

The second part of the experiment was designed to elicit the subjects utility and probability weighting functions. The experimental method used in this part of the experiment is based on the two-step method developed by Abdellaoui (2000). Specifically, subjects faced 11 decision sets in which they needed to specify their preference between two lotteries. Decision sets 1 to 6 were designed to elicit a subject's utility function, whereas decision sets 7 to 11 were designed to elicit the probability weighting function. The participants knew that at the end of the experiment one of their decisions was chosen at random and that the lottery realization of that decision determined their earnings in this part of the experiment.

In decision sets 1 to 6, participants faced an interval of monetary outcomes $[x_{i-1}, x_{i-1} + \epsilon]$,

[10]These compensations correspond to 3.3 times and two times what a subject receives in Piecerate.

where $i = \{1, 2, 3, 4, 5, 6\}$ indicates a decision set. In each decision set, the subject had to choose between a lottery L and lottery R. Lottery L offered the lower bound of the interval, x_{i-1}, occurring with probability $p = 2/3$, and 0.50 Euros with probability 1/3. Lottery R offered the midpoint of the interval, $x_{i-1} + \frac{\epsilon}{2}$, with probability $p = 2/3$. Decision sets featured five iterations each, the choice of the subject in the first iteration determined the lotteries that she faced in the second iteration. For instance, if a subject chooses L in the first iteration, the interval of monetary outcomes in the second iteration becomes $[x_{i-1} + \frac{\epsilon}{2}, x_{i-1} + \epsilon]$, which has the midpoint of the interval of monetary outcomes in the first iteration as lower bound. However, if the subject chooses R the interval of monetary outcomes becomes $[x_{i-1}, x_{i-1} + \frac{\epsilon}{2}]$, which features the midpoint of the previous iteration as upperbound. Therefore, the lotteries L and R in the second iteration depend on the choice of the subject in the first iteration.

This process is repeated until the fifth iteration is achieved. The midpoint of the last iteration becomes the indifference point x_i over the interval of monetary outcomes. $[x_{i-1}, x_{i-1} + \epsilon]$. After the fifth repetition, a new decision set begins whereby x_i becomes the lower bound of the interval of monetary outcomes. This program stops once the sequence $x_1, ..., x_6$ is obtained. The lower bound in the first decision set, x_0, was set at one Euro, ϵ was set equal to 2/5 of the subject's earnings in the first part of the experiment. This calibration allows me to obtain data about the subject's preferences over the monetary incentives that were at stake in the first part of the experiment. Table 1 presents an example that illustrates the procedure followed by the program for a subject that earned 16 Euros in the first part of the experiment, and who chooses L in repetitions 1, 4, and 5.

In decision sets 7 to 11 participants had to choose between the sure outcome L which offered x_i with $i = 1, 2, 3, 4, 5, 6$ and the prospect R which offered x_6 with probability p and x_0 with probability $1 - p$. The program finds the p that makes the subject indifferent between these two lotteries. The process is analogous to that used to elicit the sequence $x_1, ..., x_6$. In the first iteration the probability was set at $p = 0.50$, the midpoint of the probability interval $[0, 1]$. The subject's choice had an influence in the probabilities that were presented in the next iteration. Choosing L leads makes the lower bound of the probability interval for the next iteration equal to 0.5, so that the interval becomes $[0.5, 1]$. Choosing R makes the upper bound of the probability interval for the next iteration equal to 0.5 and the interval becomes $[0, 0.5]$. In the next iteration, the subject had to choose again between L and R but with a different probability that governs the prospect R.This process is repeated five times and the midpoint of the last repetition is the subject's probability p_i that makes the subject indifferent between R and L. In each decision set, a different x_i with $i = 1, 2, 3, 4, 5, 6$ was used, so that the sequence $p_1, ..., p_5$ is obtained. Table 1 also shows an example of the program used to elicit these probabilities.

Once the second part of the experiment was over, subjects were presented with feedback

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

Iteration	Alternatives	Interval	Choice	Alternatives	Probabilities	Choice
1	L=(x_0, 0.66; 50), R=(3.7, 0.66)	[1, 6.4]	L	L=(x_1, 1), R=(x_6, 50;)	[0, 100]	L
2	L=(x_0, 0.66; 50), R=(5.05, 0.66)	[3.7,6.4]	R	L=(x_1, 1), R=(x_6, 75)	[50, 100]	L
3	L=(x_0, 0.66; 50), R=(4.38, 0.66)	[3.7,5.05]	R	L=(x_1, 1), R=(x_6, 87)	[75, 100]	R
4	L=(x_0, 0.66; 50), R=(4.04, 0.66)	[3.7,4.38]	L	L=(x_1, 1), R=(x_6, 81)	[75, 87]	L
5	L=(x_0, 0.66; 50), R=(4.21, 0.66)	[4.04,4.38]	L	L=(x_1, 1), R=(x_6, 85)	[81, 87]	L
End		x_1 ∈[4.21, 4.38]			p_1 ∈[85, 87]	

about their performance in the real effort task, the rounds that counted towards payment if assigned to LowPR, MePR or HiPR and whether their belief was correct. Also, subjects were informed about the decision set that was chosen for compensation for the second part of the treatment, its realization, their experimental earnings for this second part of the experiment, and their definitive experimental earnings. Finally. the 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 self reported measures of self-efficacy and a self-reported measure of mathematical abilities. Appendix B presents the questionnaire.

4 Treatment effects

4.1 Performance

I first compare the average performance delivered 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, which suggest that a probability weighting contract with a probability of 10 % delivers higher performance than the piece rate contract. Particularly, a subject assigned to the LowPR treatment solves on average 20.56 % more summations as compared to a subject assigned the Piecerate treatment (t(84.454)=2.361, p=0.010).

In contrast, the probability contract implemented with higher probabilities deliver similar average performance to the Piecerate contract. A subject assigned the MePR treatment achieved on average 87.9 correct summations, and a subject assigned the HIPR treatment achieved on average 83.7 correct summations, which are both not statistically different from the average correct summations under Piecerate, 81.37 summations.[11] A Bonferroni correction for multiple testing yields that the null hypothesis that LowPR and Piecerate yield the same

[11] The t-tests of these comparisons are (t(83)=1.005, p=.159) and (t(82.44)=-0.386, p=.692), 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 delivered by each subject.

performance is still rejected.

Among the three probability contracts, the LowPR delivers higher average performance. This contract yields 17% higher average performance when compared to the HiPR treatment (t(75.215)= 2.232 ,p=0.014), and on average leads to 11% more correctly solved summations than the MePR contract (t(79.575)= 1.478, p=0.0716). A Bonferroni correction for multiple testing allows me to reject the null hypothesis that LowPR=HIPR, but not the hypothesis that LowPR=MePR. All in all, the analysis of the descriptive statistics of performance demonstrates that LowPR delivers higher average performance as compared to the Piecerate, and HiPR treatments.[12]

To control for additional factors that may influence this result other than the treatment assignment, I regressed the subject's performance on treatment dummies, 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. The results from the regression confirm the aforementioned findings. First, the coefficient associated to LowPR treatment is significant and positive, which supports the result that a subject assigned to this treatment delivers higher average performance as compared to the benchmark of the regression, the Piecerate treatment. Additionally, the estimate of LowPR is significantly higher than the estimates of HiPR treatment (F(1,159)=6.58) and the MePR treatment (F(1,159)=6.02), thus among the probability contracts, the LowPR yields the highest performance.

These results support Prediction 2 and Prediction 2a of the theoretical framework: if sub-

[12] An alternative analysis is to look at average performance per round. Overall, I find the same qualitative results as looking at the aggregate of correct summations. I find evidence to reject the null that LowPR and Piecerate yield on average the same performance (p=0.010). Moreover, I find evidence to reject the null that LowPR and HIPR yield on average the same performance (p=0.013) but no evidence to reject the null that LowPR and MePR yield on average the same performance once the Bonferroni correction is performed, (p=0.07). I find no evidence to reject the null that MePR and HiPR deliver the same average performance by round (p=0.47) and that Piecerate delivers the same average performance than MePR and HiPR, (p=0.69) and (p=0.32), respectively.

jects distort probabilities in a systematic way, then the probability at which the probability contract is implemented is essential for the effectiveness of this contract. A probability contract implemented with low probability yields higher performance than a cost-equivalent piece rate and also higher performance than probability contracts implemented higher probability of performance evaluation. Section 5 investigates how these subjects evaluate probabilities.

4.2 Performance Beliefs

The previous subsection presented evidence that the probability contract implemented with low probability yields higher performance than a piece rate contract and higher performance than probability contracts implemented with higher probability. In this subsection, I investigate whether participants anticipate these performance differences, as reflected by their beliefs about their own performance in the task. If participants anticipate that a contract with low probability of evaluation improves their performance on the task, then we can conclude that the psychological incentives delivered by each of the treatments are internalized by the participants.

Table 4 presents the descriptive statistics of the performance beliefs by treatment. I find no belief differences between treatments, which supports the notion that subjects do not anticipate the relationship between the distortion of probabilities, the psychological incentives offered by the contracts, and task performance.The average beliefs of the subjects in LowPR, MePR, and HiPR are not statistically different to those elicited by subjects assigned to Piecerate.[13] Moreover, there is no empirical evidence of a difference in performance beliefs between the three probability weighting contracts: There is no significance difference in 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 contracts (t(83.241)=1.1885,p=.1190)

To account for factors that may be driving this result other than the treatment assignment, I regressed an individual's performance beliefs on treatment dummies, 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 of the statistical model. The coefficients associated to the MePR, HiPR and LowPR are not significant, which demonstrates that there are no statistical differences between the beliefs of the subjects in each of these treatments and the beliefs of the subjects assigned the Piecerate treatment. Additionally, there is no evidence of a statistical difference between the LowPR and the MePR coefficients (F(1,160)=0.29), and the LowPR and HiPR coefficients (F(1,160)=0.09).

The result that subjects do not internalize the incentives delivered by LowPR could explain why in practice some of the incentives using probabilistic evaluations are not taking advantage of

[13] 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: Treatment Effects

	(1) Performance	(2) Performance	(3) Performance
LowPR	16.74**	17.46**	13.95**
	(7.090)	(7.042)	(5.674)
MePR	6.522	5.779	-0.763
	(6.487)	(6.643)	(4.678)
HiPR	2.372	0.927	-0.170
	(5.985)	(6.117)	(4.270)
Risk-General		-1.771*	-2.129**
		(0.995)	(0.859)
Risk-Occupation		0.230	-0.0568
		(1.056)	(0.769)
Risk Health		-0.584	-0.567
		(1.049)	(0.811)
Risk Drive		2.028**	1.893**
		(0.952)	(0.816)
Belief			0.382***
			(0.063)
Self-efficcacy			0.190
			(0.879)
Task-Difficulty			-3.980***
			(1.441)
Math Skills			2.240**
			(1.100)
Gender			-4.947
			(4.038)
Constant	81.38***	84.69***	56.69***
	(4.726)	(7.573)	(7.866)
R^2	0.045	0.073	0.500
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 and high probability of outcome evaluation, respectively. The controls considered in this model are "Gender" a variable that captures the gender of the participant, "Belief " captures the performance belief of the subject, "Math Skills " which captures the self-reported mathematical skills of the subject, "Task Difficulty " 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 experimental treatment. Beliefs are the subjects' prediction of the number of correct summations solved in the first part of the experiment.

the regularity that individuals distort probabilities. Either the principal himself is not aware of the possibility that these psychological incentives are powerful enough to boost performance or she is more sophisticated and believes that workers are not going to incorporate these incentives in their belief system, and therefore into their performance in the task.

5 The elicitation of utilities and probability weighting functions

The result that the probability contract implemented with a low probability of evaluation outperforms the piece rate contract is in line with Prediction 2 and Prediction 2a. However, a key assumption underlying these two predictions is that the worker distorts probabilities in a systematic way, overweighting small probabilities and underweighting moderate to large probabilities. In this section I investigate whether subjects exhibit such a systematic distortion of probabilities and whether subjects have risk attitudes that may explain these results. I use data from the second part of the experiment which features the subjects' preferences over lotteries. The lotteries are designed to elicit risk attitudes and probability weighting functions without imposing any functional assumption. The data suggest that subjects have linear utility functions and that they overweight small probabilities.

The shape of the utility functions

I first investigate the shape of the utility function of subjects. The decision sets 1 to 6 from the second part of the experiment elicit the sequence $\{x_1, x_2, x_3, x_4, x_5, x_6\}$, which represents the subjects' preferences over monetary outcomes. I am mainly interested in two characteristics of this sequence: the sign of its slope and its curvature. To study these characteristics I construct two measures, the difference of the sequence, $\Delta'_i \equiv x_i - x_{i-1}$, for $i = 1, ...6$, and the second difference of the sequence, $\Delta''_i \equiv \Delta'_i - \Delta'_{i-1}$ for $i = 2, ..., 6$. The sign of Δ'_i as i increases

Table 5: Performance Beliefs and treatment Effects

	(1) Beliefs	(2) Beliefs	(3) Beliefs
MePR	8.847	8.251	2.041
	(9.055)	(9.259)	(9.053)
LowPR	10.683	9.604	6.861
	(8.977)	(9.205)	(8.507)
HiPR	.845	2.66	-5.71
	(8.452)	(8.685)	(8.183)
Risk-General		0.734	-0.792
		(1.384)	(1.301)
Risk-Occupation		0.037	-0.322
		(1.521)	(1.381)
Risk Health		0.995	1.518
		(1.529)	(1.563)
Risk Drive		0.293	0.360
		(1.413)	(1.408)
Self-efficacy			-0.720
			(1.546)
Task-Difficulty			-5.897**
			(2.664)
Math Skills			4.255**
			(1.856)
Gender			14.770**
			(5.857)
Constant	73.178***	65.963***	59.362***
	(6.461)	(10.589)	(13.877)
R^2	0.014	0.023	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 and high probability of outcome evaluation, respectively. The controls considered in this model are "Gender" a variable that captures the gender of the participant, "Belief " captures the performance belief of the subject, "Math Skills " which captures the self-reported mathematical skills of the subject, "Task Difficulty " 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.

determines the sign of the slope of the sequence, which captures whether a subject has a preference for larger monetary outcomes. Moreover, the sign of Δ_i'' as i increases determines the curvature of these preferences, which determines the shape of the utility function of the subject. For instance, a subject with an increasing sequence $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ and negative values of Δ_i'', has a preference for larger monetary amounts and experiences smaller utility increments with larger monetary amounts, this is equivalent to say that this subject has a concave utility function.

I provide a classification of the subjects according to the curvature of their utility function. Given that a subject has multiple Δ_i'''s and that subjects they could make mistakes in the experiment, this classification is based on the sign that has the most occurrence. Particularly, subjects with at least three positive Δ_i''s are classified as having a convex curvature. Subjects with at least three negative Δ_i''s are classified as exhibiting a concave curvature. Subjects with three or more Δ_i''s that are not significantly different from zero are classified as having linear preferences. Finally, subjects with a utility function that cannot be classified as concave, convex, or linear, are classified to have a mixed utility. To statistically asses the sign of a Δ_i'', I use a confidence interval. The confidence interval was constructed using the standard deviation of Δ_i'' for $i = 1, ...5,$, which was then multiplied by the factors 0.64 and -0.64. This confidence interval is such if Δ_i'', for some i, follows a normal distribution, then 50% of the data would be contained within the interval.[14]

As it is so far described, this classification does not take into account the possibility that the preferences of the subjects can be described by CPT and that they make decisions around a non-zero reference point. To account for this, I perform an additional classification that accounts for the possibility that subjects have a non-zero reference point. I use the subject's beliefs about her own performance in the real effort task translated into monetary earnings as reference point. This reference point is expectation-based as in Koszegi and Rabin (2006) and captures the subjects' expectations about their earnings in the first part of the experiment (Abeler et al., 2011; Pokorny, 2008).This non-zero reference point is addressed as "Beliefs" wherever necessary. The additional classification captures the shape of the utility function of a subject either in the domain of gains or in the domain of losses, the specific domain depends on where the majority of a subject's Δ_i''s lie. With this analysis I can investigate whether the majority of subjects classified as having concave and convex utilities are in fact in the domain of gains and in the domain of losses, respectively.

We are now in the position to investigate the properties of the utility functions. I find that

[14]Different confidence intervals were used in this analysis, in which the standard deviation of Δ_i'' for some i was 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 very different from the results that are reported in the paper: the majority of subjects exhibit a linear utility function. However, these more stringent intervals yield less subjects exhibiting a mixed utility function, and more subjects exhibiting a linear utility function.

Table 6: Classification of Subjects According to Curvature

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

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 6 presents the classification of the subjects in the experiment according to the curvature of their preferences. The results suggest that the majority of subjects exhibit linear utility functions. Particularly, I find that 77% of the subjects have linear utility functions, while the rest of the subjects have mixed utility functions (13% of the subjects), and convex utility functions (7% of the subjects). Of the subjects classified as having mixed utility functions only 6, which is 3 % of the subjects, presented δ_i''s suggesting diminishing sensitivity. 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 convex utility functions ($p<0.01$, one sided test).

When the analysis incorporates a non-zero reference point, a similar qualitative conclusion is reached: the majority of the subjects exhibit a linear utility function. I observe that in the domain of gains 65 % of the subjects have linear utilities, and in the domain of losses 98% of the subjects exhibit linear utilities. The number of subjects classified as having a linear utility functions is higher than the number of subjects within any other classification for both domains. I observe no patterns indicating that subjects exhibit concave and convex utilities in the domain of gains and in the domain of losses, respectively.

The result that more than two-thirds of the subjects exhibit linear utility functions is at odds with the principle of diminishing sensitivity, a key property of 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 in the second part of the experiment requires large monetary outcomes to obtain utility functions with pronounced curvature. Therefore, one of the advantages of the experimental design, that it elicits the utility function of a subject using lottery outcomes that reflect the monetary stakes present in the first part of the experiment, is also the reason that strong evidence for diminishing sensitivity may not be observed.

Until now I have focused on a classification of the subjects based on individual data. To understand how these results aggregate, I analyze the values of the sequence $x_1, ..., x_6$ when averaged across subjects. Table 7 presents the average values of x_i, the standard deviation of x_i, and the average values of Δ'_i.These values indicate that x_i is increasing with i, suggesting that on average 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 linear shape decreases when the amount of money is sufficiently large to allow for diminishing sensitivity. This result is also obtained by Abdellaoui (2000).

The same analysis is performed separately for the domain of gains and the domain of losses when Beliefs is assumed to be the reference point. Table 7 shows that subjects exhibit a preference for larger monetary amounts in both domains. Moreover, diminishing sensitivity appears to manifest differently between domains, with subjects exhibiting more of it in the domain of gains. However, concluding that that the curvature between domains is different on the basis of this finding may be incorrect, given that the lotteries in the experiment only admit positive prizes, leaving no room for the subjects to exhibit as much sensitivity in the domain of losses as in the domain of gains. On the other hand, Abdellaoui (2000) reports a similar finding when the lottery prizes in the domain of losses cover a range of similar length as the range covered by the lotteries in the domain of gains.

To gain a better understanding of the aforementioned results, I assume two parametric forms of the utility function, a CRRA and a power utility function, and estimate their parameters using a non-linear least squares regression. Table 8 presents the estimates of the regression, which indicate that the average utility function is linear. For instance, when the CRRA functional form is assumed the estimate is $\gamma = 0.005$ which is not significantly different from zero. The linearity of the average utility function is consistent with the large proportion of subjects that are classified as having a linear utility function as shown by Table 6 and the modest increments that the averaged series x_i exhibits as i increases, presented in Table 7.

I account for the possibility that subjects' make choices using Beliefs as a reference point by performing the regression separately for the domain of gains and for the domain of losses. In the domain of losses the estimated coefficients indicate linearity, which is in line with the results when the whole data is analyzed and the results from Tables 6 and 7. However, in the domain of gains subjects exhibit a modest degree of risk-seeking, which is at odds with the principle of diminishing sensitivity from CPT. An explanation for this result is that the assumed non-zero reference point classifies 12 out of 13 of the subjects with convex utility functions in the domain

Table 7: 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.

of gains, who drive this result. In the absence of these subjects, I also reach the conclusion of linearity for the domain of gains.

The shape of the probability weighting functions

The choice sets 7 until 11 described in Section 3 were designed to elicit the sequence of probabilities p_1, p_2, p_3, p_4, p_5 that make the subjects indifferent between lottery R with prizes x_6 and x_0 and a sure outcome L that pays x_i. I use these data to investigate the perception of probabilities of the subjects. I present an analyzis of individual data in which subjects are classified according to the shape of their probability weighting function. This analysis is based on Bleichrodt and Pinto (2000). To perform this classification I first construct the variable $\partial^j_{j-1} \equiv \frac{p_j - p_{j-1}}{prob_j - prob_{j-1}}$ which captures the average slope of the probability weighting function between probabilities j and $j-1$, and the variable $\bigtriangledown^j_{j-1} \equiv \partial^j_{j-1} - \partial^{j-1}_{j-2}$, which captures the change in the average slope between successive probabilities. The sign of $\bigtriangledown^j_{j-1}$ as j varies determines whether subjects have concave, convex, or probability weighting functions with lower subadditivity, this is that near zero probability intervals have higher impact than middle probability intervals, and upper subadditivity, that near one probability intervals have higher impact than middle probability intervals.

A subject had a probability weighting function with lower subadditivity if $\bigtriangledown^{0.33}_{0.16}$ was negative, and a subject had a probability weighting function with upper subadditivity if $\bigtriangledown^{1}_{0.83}$ was positive. Moreover, as in the analysis of utility functions, I account for choice error in the

Table 8: Parametric Estimates of the utility function

CRRA Utility $\frac{(x_{i-1}+\frac{\epsilon}{2})^{1-\gamma}}{1-\gamma}$			
$\hat{\gamma}$	.005 (.003)	-.054 *** (.006)	-.014 (.026)
Adj. R^2	0.922	0.887	0.303
N	1032	287	745
Power Utility $(x_{i-1}+\frac{\epsilon}{2})^{\phi}$			
$\hat{\phi}$	.994 (.002)	1.048 *** (.005)	1.012 (.022)
Adj. R^2	0.923	0.862	0.303
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 $\frac{(x_{i-1}+\frac{\epsilon}{2})^{1-\gamma}}{1-\gamma}$ 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.

experiment by classifying subjects according to the sign that has the most incidence. Hence, a subject had a concave function if at least three $\bigtriangledown^{j}_{j-1}$ were negative and he did not exhibit upper subadditivity and a convex function if at least three $\bigtriangledown^{j}_{j-1}$ were positive and he did not exhibit lower subadditivity.

Table 8 shows that 50% of the subjects exhibit upper subadditivity and 61% of the subjects exhibit lower subadditivity, but only 31% of the subjects present probability weighting functions with both characteristics. This suggests that more subjects present the possibility effect than the certainty effect and that it is likely that a subject with the possibility effect may not present the certainty effect. I also find that 22% and 27% of the subjects exhibit concave and convex probability weighting functions respectively. These results contradict those of Bleichrodt and Pinto (2000) who finds that only 15% of subjects have concave and convex utility functions.

An alternative analysis confirms the aforementioned results. A sign test shows that the number of subjects displaying $p_1 - 1/6 > 0$, in total 117 subjects, is significantly higher than those presenting $p_1 - 1/6 < 0$. Suggesting than on average subjects exhibit the possibility effect. Moreover, there is no significant difference between the number of subjects presenting $1 - p_5 > 1/6$, which is in total 83 subjects, and the number of subjects presenting $1 - p_5 < 1/6$, which suggest that on average subjects did not display the certainty effect in the data.

Table 10 provides the medians and means of $p_1, p_2, p_3, p_4,$ andp_5 when averaged over subjects. Also, one-tailed sign tests testing the sign of the difference between the elicited probability and

Table 9: Classification of subjects according to the shape of p_1, p_2, p_3, p_4, p_5

Reference Point	Domain	Convex	Concave	Lower Sub.	Upper Sub.	Lower and Upper
No/Zero	No/Gains	48	38	106	86	51
Beliefs	Gains	22	18	43	56	26
Beliefs	Losses	26	20	50	43	25

Note: This table presents the subjects classification of the subjects according to the shape of their probability weighting function. Subjects are classified as having a lower subadditive, upper subadditive, convex, concave, or linear probability weighting function. 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 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.

the probability that it maps, i.e. $p_1 - 1/6$, are presented in Table 11. These two tables suggest that when all the data is used, participants overweight small probabilities. Particularly, the probability $\frac{1}{6}$ is on average perceived by the subjects to be .306 and the probability $\frac{2}{6}$ is on average perceived to be .448. Moreover, in contrast with most of the literature in probability judgments and in accordance with the analysis of the individual data, I find no evidence that subjects underweight large probabilities.

When this analysis is performed incorporating the non-zero reference point, I find that subjects overweight the probability $\frac{1}{6}$ in both domains, but the probability $\frac{2}{6}$ is only overweighted in the domain of gains. Moreover, I find no evidence that subjects underweight large probabilities. These results suggest that, in stark contrast with the literature of probability judgements, the probability weighting function in the domain of gains, as defined by Beliefs, has higher elevation than the probability weighting function in the domain of losses.

To conclude this section, I perform parametric estimations of the average probability weighting function. I estimate the parameters of the probability weighting functions proposed by Tversky and Kahneman (1992) and by Prelec (1998). This exercise allows me to compare the degree of probability distortion among the subjects in my experiment with that of previous studies and to analyze the characteristics of the probability weighting functions across domains. Table 12 presents the estimates of the non-linear least squares regression for each parametric specification. Overall, I find point estimates indicating an overweighting of probabilities over a larger probability domain than previously documented. For instance, when the probability weighting function is assumed to have the functional form proposed by Tversky and Kahneman (1992), the estimate for the whole sample is $\hat{\psi} = 0.804$, which is significantly larger than the point estimates found in the literature which are of the order of 0.60 to 0.70 (See Bleichrodt and Pinto (2000), Abdellaoui (2000),Wu and Gonzalez (1996b), Camerer and Ho (1994), and

Table 10: Medians and Means of $\{p_1, p_2, p_3, p_4, p_5\}$

p rob.	Mean	Median	Mean	Median	Mean	Median
1/6	.306	.234	.412	.390	.249	.234
2/6	.448	.406	.553	.578	.383	. 359
3/6	.524	.484	.582	.609	.492	.484
4/6	.638	.671	.677	.765	.628	.640
5/6	.781	.859	.807	.890	.763	.828
Ref.Point	No/Zero		Beliefs		Beliefs	
Domain	No/Gains		Gains		Losses	
N	173		79		93	

Note: This table presents the average and median of the sequence p_1, p_2, p_3, p_4, p_5 elicited in the second part of the experiment with the decision sets seven until eleven. Column 1 presents the actual probabilities. Columns 2 and 3 present the mean and median of the sequence of p_1, p_2, p_3, p_4, p_5 using all the data. Columns 4 and 5 present the mean and median of the sequence of probabilities when the reference point is assumed to be tthe monetary equivalent of a subject's beliefs about her performance. Column 4 present p_1, p_2, p_3, p_4, p_5 when the lottery prizes are above this reference point. Column 5 present p_1, p_2, p_3, p_4, p_5 when the lottery prizes are below this reference point.

Table 11: Counts of $p_i - prob > 0$ and $p_i - prob < 0$

$p_i - prob$	>0	<0	>0	<0	>0	<0
$p_1 - 1/6$	117***	55	61***	18	65**	39
$p_2 - 2/6$	105***	67	55***	24	56	48
$p_3 - 3/6$	79	93	40	39	45	59
$p_4 - 4/6$	87	85	44	35	56	48
$p_5 - 5/6$	89	83	49 **	30	46	58
Ref.Point	No/Zero		Beliefs		Beliefs	
Domain	No/Gains		Gains		Losses	
N	173		42		130	

Note: This table presents the number of positive and negative events for the expression $p_i - prob_i$ for $prob_i = 1/6, 2/6, 1/2, 4/6, 5/6$. The significance of the count is tested with one tailed sign tests. Columns 2 and 3 present the mean and median of the sequence of p_1, p_2, p_3, p_4, p_5 using all the data. Columns 4 and 5 present the mean and median of the sequence of probabilities when the reference point is assumed to be the monetary equivalent of a subject's beliefs about her performance. Column 4 present p_1, p_2, p_3, p_4, p_5 when the lottery prizes are above this reference point. Column 5 present p_1, p_2, p_3, p_4, p_5 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.

Table 12: Parametric Estimates of the weighting function

T& K (1992)			
$\hat{\psi}$	.804*** (.044)	.811*** (.047)	.785*** (.037)
Adj. R^2	0.832	0.808	0.866
N	860	395	520
Prelec (1998)			
$\hat{\alpha}$	.660***(.037)	.602*** (.050)	.654*** (.049)
$\hat{\beta}$	.792 ***(.032)	.689 ***(.044)	.844***(.031)
Adj. R^2	0.832	0.83	0.852
N	860	430	430
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 presents the estimates of the equation $p = \frac{prob^{\psi}}{(prob^{\psi}+(1-prob)^{(1-\psi)})^{\frac{1}{\psi}}}$.The lower panel presents the estimates of the equation $p = exp(-\beta(-ln(prob)))^{\alpha}$. 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.

Tversky and Kahneman (1992)).This larger estimate implies that the overweighting of probabilities occurs over the interval $p \in [0, 0.42)$. This result is robust to performing the analysis including the non-zero reference point.

Table 12 also presents the estimates of the probability weighting function when the functional form is assumed to be Prelec (1998)'s two-parameter function. The estimate $\hat{\alpha}$, which is statistically lower than one for all specifications, suggests that the average probability function is first concave and then convex, i.e. has an inverse S-shape.Moreover, the estimate $\hat{\beta}$, which is also statistically lower than one for all specifications, suggests that the point of inflection, the point after which the probability weighting function becomes convex, lies above the 45 degree line of the space $(prob, p)$. Previous estimations of this probability weighting function have found similar values of α, but considerably larger values of β (Fehr-Duda and Epper, 2012; Bleichrodt and Pinto, 2000). This suggests that subjects overweighted probabilities over the interval $p \in [0, 0.5]$ and exhibited a probability weighting function with an elevation that is higher than previously documented. Finally, Table 11 show that, in stark contrast with most of the results in the literature of probability judgements, subjects exhibit higher elevation in the domain of gains than in the domain of losses when Beliefs is assumed to be the reference point.

6 What drives the treatment effect?

The previous two sections demonstrate that i) the probability contract outperforms the piece rate contract when the principal implements it with a low probability of evaluation, ii) the subjects' overweight small and moderate probabilities, and iii) the subjects have linear utility functions. The aim of this section is to investigate whether the subjects' probability distortions are driving the treatment effect. According to the theory presented in Section 2, the subjects' systematic overweighting of small probabilities is the reason behind the higher average performance in LowPR. Indeed, the results suggest that the degree at which subjects overweight small probabilities, alone, explains the totality of the treatment effect from being assigned to LowPR.

To understand the determinants behind the effectiveness of the probability contract , I add to the regression presented in section 4 two variables and their interactions: A variable that captures the extent at which the subject overweights probabilities and variables that capture the curvature of the subject's utility function. The variables representing the curvature of the utility function are binary variables indicating whether the subject's utility function can be clasified as concave, convex, linear or mixed, according to the classification presented in the previous section.

Moreover, I perform the analysis with different variables that represent different measures of the subjects' overweighting of probabilities. The first variable aims to capture a subject's overweighting of small probabilities. The variable "Overweight$_s$ features the sum of the differences $p_1 - \frac{1}{6}$ and $p_2 - \frac{2}{6}$. If the only mechanism underlying the treatment effect is the subjects' overweighting of small probabilities, then introducing this variable, along with its interaction with the treatment assignment indicator "LowPR", should capture the totality of the treatment effect of being assigned to LowPR. Alternatively, if the treatment effect is due to mechanisms other than the overweighting of small probabilities, then even after introducing this variable in the analysis, the binary variable indicating the assignment to LowPR must remain significant just as in the regression presented in Section 4. A second variable aims to represent the subjects' overweighting of moderate probabilities, "Overweight$_M$" is a variable that features the sum of the differences $p_3 - \frac{1}{2}$ and $p_4 - \frac{2}{3}$. Introducing this variable in the analysis corresponds to the finding that subjects overweight moderate probabilities. However, the "LowPR" treatment was designed to exploit the regularity that individuals overweight small probabilities and its effectiveness should not depend on probability distortions of moderate probabilities, I therefore expect that this variable is not able to capture the treatment effect. Finally, the variable "Overweight$_L$" features the sum of the differences $p_4 - \frac{2}{3}$ and $p_5 - \frac{5}{6}$. I perform the analysis using this variable in the spirit of a falsification test: the effect of being assigned to "LowPR" should not depend on this variable given the empirical evidence that subjects evaluate large probabilities accurately and that according to the theory, the effectiveness of LowPR does not depend on the subjects evaluation of large probabilities.

Table 12 presents the estimates of the regression. The benchmark of this regression is a subject assigned to the piece rate treatment. Additionally, note that this regression assumes that the subject's reference point is at zero, which in the light of the evidence presented in the previous section is not a problematic since the conclusions from section 5 are robust to the inclusion of non-zero reference points, i.e. overweighting of small and moderate probabilities and linear utility functions. Columns (1) and (2) present the estimates of the model when"Overweight$_S$" is included in the analysis, columns (3) and (4) present the estimates of the model when the measure "Overweight$_2$" is included, and columns (5) and (6) use "Overweight$_L$" as the variable for probability overweighting.

The estimates in columns (1) and (2), show that the interaction between the degree at which subjects overweight small probabilities and the assignment to LowPR is significant and positive. Moreover, the assignment to LowPR alone does not yield higher performance as compared to the benchmark. This suggests that the assignment to the LowPR is effective in boosting performance when the subject overweights small probabilities. Higher overweighting of probabilities leads to higher performance under the contract in LowPR. This finding is robust to the inclusion of the shape of the subjects' utility function. In contrast to columns (1) and (2), the coefficient "LowPR" in columns (3) and (4), remains significant with the inclusion of the interaction between "Overweighting" and "LowPR". This suggests that even though subjects overweight moderate probabilities, it this is not the channel through which the treatment effect takes place. As expected, a similar conclusion can be drawn from columns (4) and (5): The subjects' distortion of large probabilities, is not the mechanism through which assignment to LowPR improves performance as compared to the piece rate.

A possible driver of these results is that my design relies on an specific order in which the subjects perform the tasks and it may be that the treatment assignment in the first part of the experiment has an effect on the way in which subjects evaluate probabilities in the second part of the experiment. For instance, subjects assigned to LowPR may overweight more small probabilities than subjects in Piecerate, since they have been exposed to some uncertainty in the first part of the experiment. If this was the case, the analysis presented in this section is not suggestive of a mechanism that explains the treatment effect, but a consequence of the treatment itself. It suffices to show that the degree at which subjects distort probabilities is equivalent across the treatments to disregard this possibility. Indeed, I find that the degree at which subjects overweight small probabilities is, on average, the same across treatments. The degree of overweighting of probabilities captured by Overweighting$_S$ is on average the same for subjects in Piecerate and LowPR (t(84.631)=-0.796,p=0.427), as well as for subjects in Piecerate and HiPR (t(85.432)=-1.094,p= 0.276), and for subjects in Piecerate and MePR (t(82.642)=0.344,p=0.731). The same conclusion is reached when the analysis is performed with the variables Overweighting_M and Overweighting_L. This result validates the result of this subsection that the overweighting of small probabilities explains the higher performance of subjects assigned to LowPR.

7 Conclusion

This paper proposed and tested a novel incentive scheme designed to take advantage of the behavioral regularity that subjects overweight small probabilities. I find that the contract yields higher output than a standard piece rate compensation when i) both contracts are cost-equivalent, and ii) the contract is implemented with a probability of performance evaluation of one-tenth. Moreover, the data suggest that the degree at which subjects overweight small probabilities is fundamental for the efficiency of this contract.

The finding that the probability contract is more effective than a cost-equivalent piecerate makes the proposed scheme attractive for employers that already employ standard pay-for-performance compensations. However, it remains undiscussed the way in which such an incentive scheme should be implemented. My proposal is that this contract is implemented in the form of a bonus, rather than becoming the base pay of the worker. For instance , after the achievement of certain performance threshold in a month, the worker can be compensated according to her performance in a randomly chosen period after this threshold has been reached. There are two reasons for this proposal: First, if the worker's preferences for monetary incentives, can be represented by cumulative prospect theory, their probability distortion may depend on a non-zero reference point. The employer can fix this reference point, by introducing a performance threshold after which production is evaluated and paid according to the probability weighting contract. Second, the present study does not take into consideration other important determinants of work performance such as the effect of the proposed incentive scheme on the workers' well-being and/or their intrinsic motivation to work. Therefore, if the proposed contract is harmful to these two relevant determinants, its counterproductive effects are less detrimental to productivity when it is introduced as a production bonus, rather than a base-pay.

This study has limitations that could be addressed in future research. First, even though there are obvious advantages of using controlled laboratory environments to test the effectiveness of incentive schemes (Charness and Kuhn, 2011), these advantages come at the cost of external validity. To study in further detail the motivational effect of the probability contract incentive scheme, it is necessary to perform a test in a situation with longer working periods, more powerful monetary incentives, and more meaningful tasks. Field experiments that incorporate these characteristics seem an ideal tool to evaluate the external validity of this

Table 13: Probability overweighting and treatment effects

	(1) Perfomance	(2) Perfomance	(3) Perfomance	(4) Perfomance	(5) Perfomance	(6) Perfomance
LowPR	8.291	8.425	14.167**	14.071**	15.249***	15.173**
	(6.053)	(6.199)	(5.611)	(5.875)	(5.798)	(6.072)
MePR	-2.094	-2.417	-0.674	-0.626	0.754	0.692
	(5.849)	(5.795)	(4.640)	(4.668)	(4.711)	(4.725)
HiPR	-3.007	-3.410	0.012	-0.198	-0.087	-0.203
	(5.523)	(5.449)	(4.217)	(4.195)	(4.090)	(4.111)
LowPR*Overweight	26.881***	26.706***	9.790	9.759	12.267	11.852
	(8.681)	(9.272)	(11.587)	(11.735)	(8.538)	(8.690)
MePR*Overweight	7.175	8.411	15.382	15.892	14.162	13.565
	(9.304)	(9.338)	(11.248)	(11.273)	(8.935)	(9.170)
HiPR*Overweight	13.472*	14.181*	15.129*	15.015*	-3.223	-2.744
	(8.681)	(9.272)	(9.088)	(8.852)	(12.751)	(12.215)
Overweight	-17.116***	-17.750***	-12.385**	-12.792**	-8.942*	-9.040*
	(5.512)	(5.529)	(6.265)	(6.096)	(5.352)	(5.119)
Belief	0.384***	0.378***	0.381***	0.374***	0.379***	0.372***
	(0.061)	(0.062)	(0.064)	(0.064)	(0.063)	(0.064)
Self-efficcacy	0.111	0.219	0.327	0.416	0.276	0.357
	(0.927)	(0.936)	(0.890)	(0.895)	(0.884)	(0.901)
Task-Difficulty	-4.079***	-4.103***	-4.084***	-4.124***	-3.946***	-4.031***
	(1.430)	(1.457)	(1.441)	(1.455)	(1.409)	(1.431)
Math Skills	2.196**	2.102*	2.239**	2.185**	2.454**	2.395**
	(1.107)	(1.116)	(1.099)	(1.106)	(1.091)	(1.100)
Risk-General	-2.451***	-2.539***	-2.276**	-2.358***	-2.238**	-2.309**
	(0.912)	(0.923)	(0.893)	(0.902)	(0.882)	(0.897)
Risk-Occupation	0.376	0.397	0.196	0.165	0.046	0.041
	(0.790)	(0.836)	(0.812)	(0.866)	(0.790)	(0.854)
Risk Health	-0.699	-0.707	-0.533	-0.565	-0.542	-0.574
	(0.853)	(0.842)	(0.826)	(0.822)	(0.829)	(0.824)
Risk Drive	1.821**	1.871**	1.792**	1.802**	1.964**	1.955**
	(0.827)	(0.834)	(0.883)	(0.879)	(0.849)	(0.855)
Gender	-4.739	-4.646	-5.403	-5.173	-5.008	-4.838
	(4.142)	(4.250)	(4.097)	(4.269)	(4.129)	(4.337)
Concave		6.132		5.890		5.619
		(7.843)		(8.056)		(7.745)
Convex		3.037		8.626		7.854
		(8.430)		(9.216)		(9.345)
Linear		0.601		-0.029		0.655
		(4.784)		(5.083)		(5.206)
Constant	61.716***	61.548***	56.072***	56.569***	53.620***	53.655***
	(7.587)	(9.034)	(7.203)	(9.309)	(7.462)	(9.446)
Used Variable	Overweight$_S$	Overweight$_M$	Overweight$_M$	Overweight$_L$	Overweight$_L$	Overweight$_L$
R^2	0.525	0.527	0.511	0.514	0.510	0.512
N	172	172	172	172	172	172

Note: This table presents the estimates of the Ordinary Least Squares regression of the model $Performance_i = beta_0 + \beta_1 MePR + \beta_2 LowPR + \beta_3 HiPR \beta_4 MePR * Overweight + \beta_5 LowPR * Overweight + \beta_6 HiPR * Overweight + \beta_7 Overweight + Controls'\Gamma + \epsilon_i$, 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 and high probability of outcome evaluation, respectively. 'Overweight" represents one of the three measures of probability overweighting. Models (1) and (2) use the overweighting variable constructed as $p_1 - p_2 - 0.166 - 0.33$. Models (3) and (4) use the overweighting variable constructed as $p_3 - p_4 - 0.5 - 0.66$. Models (5) and (6) use the overweighting variable constructed as $p_5 - p_4 - 0.66 - 0.833$. The controls considered in this model are "Gender" a variable that captures the gender of the participant, "Belief " captures the performance belief of the subject, "Math Skills " which captures the self-reported mathematical skills of the subject, "Task Difficulty " 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.

contract.

Second, some findings from the literature of decision-making suggest that individuals distort probabilities in a different manner when situations involve probabilities that are experienced as compared to situations in which these probabilities are described Hertwig et al. (2004); Hau and Pleskac (2008). If this is the case, the probability contract could have ambiguous effects on performance in a dynamic setting: Once the contract is implemented the worker may be more efficient, in this case the probability that a period is chosen is descriptive and the worker distorts probabilities as predicted by our model. However, in the following implementations of the contract the worker has gained experience about the probability that a period is chosen for evaluation, which leads to an underweighting small probabilities and underperformance. Studying the dynamic aspect of this contract could shed light on this possibility.

Third, it could be argued that the efficiency of the probability contract stems from the salience of the monetary payment offered in each treatment. Since each treatment delivers, on expectation, the same monetary incentives, the probability contract implemented with a small probability includes a large piece rate for each unit of production that is evaluated. Subjects that do not take into account the stochastic nature of the contract may be motivated by the salience of the large piece rate and work harder in the task. Even though in section 6 I presented evidence that the degree at which subjects overweight probabilities explains the totality of the treatment effect, it may be that subjects who overweight small probabilities are also prone to the saliency of this piecerate. The experimental design that I present in this paper cannot disentangle these two effects. Future experiments should be aimed at understanding whether this saliency is also a relevant determinant of the effectiveness of the contract.

Finally, the present research is oblivious about the effect of the probability contract on the worker's well-being and the intrinsic motivation to perform their job. The effectiveness of the probability contract requires research about the impact of the proposed compensation on the workers' satisfaction at work and their motivation to perform their job-related tasks.

References

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

Abdellaoui, M., Bleichrodt, H., and L'Haridon, O. (2008). A Tractable Method to Measure Utility and Loss Aversion Under Prospect Theory. *Journal of Risk and Uncertainty*, 36(245).

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

Ackert, L. F., Charupat, N., Church, B. K., and Deaves, R. (2006). An experimental examination of the house money effect in a multi-period setting. *Experimental Economics*, 9(1):5–16.

Ania, A. B. (2006). Optimal demand management policies with probability weighting. *Working Paper*, (February):1–17.

Ariely, D., Gneezy, U., Loewenstein, G., and Mazar, N. (2009). Large Stakes and Big Mistakes. *The Review of Economic Studies*, 76:451–469.

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.

Auriol, E. and Renault, R. (2008). Status and Incentives. *The RAND Journal of Economics*, 39(1):305–326.

Baltussen, G., Post, G. T., van den Assem, M. J., and Wakker, P. P. (2012). Random incentive systems in a dynamic choice experiment. *Experimental Economics*, 15(3):418–443.

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.

Barberis, N. and Huang, M. (2008). Stocks as Lotteries : The Implications of Probability Weighting for Security Prices. *American Economic Review*, 98(5):2066–2100.

Besley, B. T. and Ghatak, M. (2008). Status Incentives. *American Economic Review*, 98(2):206–211.

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, J. 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.

Bull, C., Schotter, A., and Weigelt, K. (2016). Tournaments and Piece Rates : An Experimental Study. *Journal of Political Economy*, 95(1):1–33.

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.

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(2):115–131.

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

Dhami, S. and Al-Nowaihi, A. (2007). Why do people pay taxes? Prospect theory versus expected utility theory. *Journal of Economic Behavior and Organization*, 64(1):171–192.

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., Fehr, E., and Fischbacher, U. (2008). Testing theories of fairness-Intentions matter. *Games and Economic Behavior*, 62:287–303.

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

Fehr, E. and Goette, L. (2004). Do Workers Work More if Wages Are High ? Evidence from a Randomized Field Experiment. *The American Economic Review*, 97(1):298–317.

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

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

Fehr-Duda, H., De Gennaro, M., and Schubert, R. (2006). *Gender, financial risk, and probability weights*, volume 60.

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

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

Gneezy, U. and Rustichini, A. (2000). Pay Enough or Don't Pay at All. *Quarterly Journal of Economics*, 115(3):791–810.

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

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.

Hau, R. and Pleskac, T. J. (2008). The Description â€" 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(3):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 Axiom. *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.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–291.

Köbberling, V. and Wakker, P. P. (2005). An index of loss aversion. *Journal of Economic Theory*, 122:119–131.

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.

Lee, J. (2008). The effect of the background risk in a simple chance improving decision model. *Journal of Risk and Uncertainty*, 36(1):19–41.

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

Nalbantian, H. R. and Schotter, A. (1997). Productivity Under Group Incentives : An Experimental Study. *American Economic Review*, 87(3):314–341.

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.

Polkovnichenko, V. and Zhao, F. (2013). Probability weighting functions implied in options prices. *Journal of Financial Economics*, 107(3):580–609.

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

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3(4):323–343.

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

Schlag, K. H., Tremewan, J., and van der Weele, J. J. (2015). A penny for your thoughts: a survey of methods for eliciting beliefs. *Experimental Economics*, 18:457–490.

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.

van de Kuilen, G. and Wakker, P. P. (2011). The Midweight Method to Measure Attitudes Toward Risk and Ambiguity. *Management Science*, 57(May 2017):582–598.

van Dijk, F., Sonnemans, J., and van Winden, F. (2001). Incentive Systems in a Real Effort Experiment. *European Economic Review*, 45(2):187–214.

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

Wakker, P. P., Thaler, R. H., and Tversky, A. (1997). Probabilistic Insurance. *Journal of Risk and Uncertainty*, 15:7–28.

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

Wu, G. and Gonzalez, R. (1996b). 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: The employer's choice

In the main body of the paper, I show that the probability contract yields higher performance than a cost-equivalent piece rate whenever the worker distorts probabilities according to a probability weighting function that overweights small probabilities and the employer chooses a probability that induces a sufficiently powerful overweighting of probabilities. The purpose of this appendix is to show that when implementing the probability contract, the risk neutral employer finds it optimal to choose a probability that induces high overweighting of probabilities.

In this section I assume that workers distort probabilities systematically according to $w(p)$ as defined in Section 2. This case is interesting, since, as shown in Section 2, it is the constitutes the case in which the choice of p has an influence on the worker's behavior. Moreover, I assume that workers have linear utility over monetary outcomes, which is a result that is supported by the results of Section 5.5 and simplifies greatly this analysis. Note that under this assumptions, risk aversion is present in this model and it is implied by the curvature of the probability weighting function.

Let us start the analysis by setting up the employer's program, which consists on minimizing the compensation offered to the worker, subject to the individual rationality constraint and the incentive compatibility constraint. This program can be formally written as

$$\begin{aligned} &\underset{p}{\text{Min}} && By_t p, \\ &\text{subject to} && i) \quad \underset{y}{\text{argmax}}\; w(p)By_t - c(\theta, y_t), \\ & && ii) \quad w(p)By_t - c(\theta, y_t) \geq 0. \end{aligned}$$

As in Section 2, I employ the cost equivalence $B = \frac{a}{p}$ that equalizes, in expectation, the performance based incentives of the piece rate and the probability contract. Thus, the Lagrangian of this problem can be written as

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

The first order condition of the Lagrangian with respect to p is

$$\frac{\partial \mathcal{L}}{\partial p} : \quad (\lambda_1 + \lambda_2 y_t)\frac{a}{p} \left(w_p(p) - \frac{w(p)}{p} \right) = 0. \tag{20}$$

When either the IC or IR constraints bind, this is when $\lambda_1 > 0$ or $\lambda_2 > 0$ and $y_t > 0$, the solution of the Lagrangian is given by the fixed-point p^*,

$$\left\{p^* \in (0,1) : p^* = \frac{w(p^*)}{w_p(p^*)}\right\}. \tag{21}$$

To understand the properties of p^*, define $g(p) \equiv \frac{w(p)}{w_p(p)}$. Note that $g(0) = 0$ since $\lim_{p\to 0^+} w_p(p) = \infty$, $g(1) = 0$ due to $\lim_{p\to 1^-} w_p(p) = \infty$, and $g(\hat{p}) = \infty$ since $\lim_{p\to\hat{p}} w_p(p) = 0$. Additionally, $g(p)$ is increasing for the interval $p \in [0, \hat{p})$, given that $w_p(p) > 0$ for $p \in [0, \hat{p})$, and decreasing for the interval $p \in [1, \hat{p}]$, since $w_p(p) < 0$ for $p \in [1, \hat{p}]$.

The existence of the interior fixed point $p^* = g(p^*)$ is guaranteed since p is a linear and increasing function in the unit interval, with values at the extreme points $p = 0$ and $p = 1$, and $g(p)$ is an increasing function in the interval $p \in [0, \hat{p})$ with values $g(0) = 0$ and $g(1) = 0$ at the extreme points. Then it must be that $g(p)$ is decreasing somewhere in the interval $p \in (\hat{p}, 1]$, which entails that p and $g(p)$ intersect somewhere $p \in (0, 1)$. [15]

Finally, I investigate whether p^* is a solution for all the values of p by studying the shape of the Lagrangian over the entire probability domain. The second order condition of the Lagrangian is,

$$\frac{\partial^2 \mathcal{L}}{\partial p^2} : \qquad -(\lambda_1 + \lambda_2 y_t)\frac{a}{p}\left(w_{pp}(p) - \frac{2w_p(p)}{p} + \frac{2w(p)}{p^2}\right). \tag{22}$$

Note that equation (13), when evaluated at p^*, becomes positive if and only if $w_{pp}(p) < 0$. Hence p^* is a unique solution of the program for $p \in (0, \hat{p})$. However, when $w_{pp}(p) > 0$ the second order condition evaluated at p^* is negative, this implies that the Lagrangian is concave and attains a minimum value at the extreme values of the domain for which $w_{pp}(p) > 0$, hence either at $p = \hat{p}$ or $p = 1$. Among others, this implies that over the interval $p \in [0, 1]$ there are multiple solutions, which is a property that stems from the shape of the probability weighting function.

A natural question that arises at this point is, what guarantees that $p^* = g(p^*)$ happens at $p^* \in (0, \hat{p})$? A sufficient condition is that $w_{ppp}(p) < 0$ for any $p \in (0, \hat{p})$, which means that the probability weighting function, $w(p)$ starts linear for near-zero probabilities and becomes more concave as p approaches $\hat{p}$. Hence, the degree of probability overweighting is large for near-zero probabilities and due to the increasing concavity of the function becomes smaller as p increases. The key property, is that this condition guarantees the strict convexity of $g(p)$ for the interval $p \in (0, \hat{p})$, which together with the properties that $g_p(p) > 0$, $g(0) = 0$, and $g(\hat{p}) = \infty$ for the interval $p \in (0, \hat{p})$, guarantee that $p^* \in (0, \hat{p})$. It is straightforward to show at near-zero probabilities then $p > g(p)$, due to $g(0) = 0$ and the convexity of $g(p)$. Moreover, when $p \to \hat{p}$ then $g(p) > p$ since $g(\hat{p}) = \infty$ and $p \in (0, \hat{p}]$. Hence at some point in the interval

[15] 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 $w_{pp}(p)>0$.

it must be that $p^* = g(p^*)$.

In what is left of this Appendix, I describe the optimal solution. Thus far, we assumed a probability weighting function that allows for $\hat{p} \neq \tilde{p}$, which complicates the description and intuitive nature of the optimal solution since it may be that p^* does not lead to an overweighting of probabilities when $\hat{p} > \tilde{p}$. Additionally, for the case in which $\hat{p} < \tilde{p}$, there is a convex part of the probability weighting function where probabilities are overweighted. I make an effort to describe thoroughly the optimal action of the employer when he faces these cases.

Let us focus on the case in which the employer faces a worker with $\hat{p} < \tilde{p}$, which represents a worker with a probability weighting function with high elevation. As mentioned above, for the interval $p \in (0, \hat{p})$ the optimal solution p^* can be implemented for this worker. However, for the interval $p \in [\hat{p}, 1]$, p^* cannot be a solution and instead, the extreme values of the interval $p \in \{\hat{p}, 1\}$ provide constitute the solution to the problem. Which of these two values would be chosen by the employer? Note that when evaluated at $p = \hat{p}$, the IC and IR constraints become larger than when evaluated at $p = 1$. This happens since $p = \hat{p}$ induces an overweighting of probabilities that guarantees $\frac{w(p)}{p} > 1$. Thus, the employer chooses $p = \hat{p}$ since this probability yields higher values of the constraints IC and IR, which at the same time yield the lowest value of the Lagrangian.

Let us now turn to the case in which the employer has a worker with $\hat{p} > \tilde{p}$, which represents a worker with a probability weighting function with low elevation. Again, for the interval $p \in (0, \hat{p})$ the optimal solution p^* can be implemented. However, note that there is a possibility that this solution does not yield an overweighting of probabilities. This represents the case of an employer who is better off inducing a p that overweights probabilities, but that understands that these contracts are not going to be accepted by the employee. Hence, if he is forced to implement a contract within $p \in (0, \hat{p})$, the employer ends-up implementing a probability contract that does not induce overweighting of probabilities. Moreover, for the interval $p \in [\hat{p}, 1]$, the solution of the program is $p = 1$ since any other value of $p \in [\hat{p}, 1]$ yields $\frac{w(p)}{p} < 1$ which leads to lower values of the IC and IR constrains than those achieved by $p = 1$.

All in all, the solutions to the minimization program for the principal can be described by

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

Note that there are two possibilities for the principal. First, for a worker with a probability weighting function with high elevation, the employer could implement the inflection point probability, $p = \hat{p}$, which is the probability that induces the highest overweighting of probabilities in the convex domain of the probability weighting function or the interior solution $p = p^*$. Second, for a worker with a probability weighting function with low elevation, the employer implements the piece rate contract $p = 1$ or the interior solution p^*.

Finally, let us explore what would do when he is forced to choose a unique option. In a setting in which the employer is able to implement any $p \in (0, 1]$ and that he faces a worker with $\hat{p} < \tilde{p}$, which $p \in \{p^*, \hat{p}\}$ would he implement? This answer depends on the curvature $w(p)$. Note first that it must be that $p^* < \hat{p}$, otherwise p^* is not a solution. Second, a function $w(p)$ with $w_{ppp}(p) < 0$, this is one which overweights close to zero probabilities at a very large margin and for which this margin decreases as p increases, then p^* is chosen. This is because at p^* the worker overweights probabilities at a larger margin. However, for a function with the opposite characteristic, that starts with a moderate overweighting of probabilities and achieves the maximum overweighting at the inflection point $\hat{p}$, then $\hat{p}$ is chosen.

Alternatively, suppose the employer is able to choose freely $p \in (0, 1]$ and that he faces a worker with $\hat{p} > \tilde{p}$. Which $p \in \{p^*, 1\}$ would he implement? In this case he would choose $p = 1$ whenever $p^* \in [\tilde{p}, \hat{p}]$, since for this case p^* induces an underweighting of probabilities in the worker, which yields lower production levels than those achieved when the employer implements a piece rate contract. Moreover, when $p^* \in [0, \tilde{p}]$ then p^* induces an overweighting of probabilities in the worker, which yields higher production levels than a piece rate.

Appendix B: 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 entirely on your decisions and effort. 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 randomly chosen 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 a given time span of 10 rounds, and each round lasts for 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.

The previous examples show what you have to do in this part of the experiment. The only thing left to do is to specify how you will earn money.

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 all summations will count towards your earnings at a rate of 25 cents each. Once you understood these instructions Press ”OK”. When everyone is ready we will start with this part of the experiment.

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 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 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 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 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 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 be faced 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, different the monetary prizes can change and you ought to make a choice again One of the eleven choices will be randomly picked and will count towards your earnings for this part of the experiment. You will be faced with one example next.

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"