Incentive contracts when agents distort probabilities*

Victor Gonzalez-Jimenez†

February 3, 2020

Abstract

I propose incentive contracts designed to take advantage of the regularity that individuals distort probabilities. These contracts incorporate risk in the agent's compensation. In addition, they include the distinctive feature that the principal can adjust their riskiness by specifying the probability that the agent's compensation depends on his performance on the delegated task. This feature allows the principal to target probabilities that enhance the motivation to perform the task when distorted by the agent. A theoretical framework and an experiment demonstrate that the proposed contracts generate higher performance than more traditional contracts. However, the specified probability is critical to achieve this result. Small probabilities indeed generate higher performance, whereas medium or high probabilities do not generate performance differences. I show that probability distortions caused by likelihood insensitivity—cognitive limitations restricting the accurate evaluation of probabilities—and not those generated by optimism, explain these findings.

JEL Classification : C91, C92, J16, J24.
Keywords: Contracts, Rank-Dependent Utility, Probability Weighting, Experiments.

*I am indebted to Wieland Müller, Patricio S. Dalton, Charles Noussair, Aurelien Baillon, Peter Wakker, Robert Dur, Eeva Mauring, Cole Williams, Geoffrey Castillo, and Stephanie Koornneef for their valuable comments. I am grateful to participants in seminars at Tilburg University, Maastricht University, University of Vienna, and Vienna University of Economics and Business. I am also thankful for comments received at the Foundations of Utility and Risk conference at York, the ESA World Meeting 2018, and the Bogota Experimental Economics Conference 2018. The experiments reported in this paper were funded by the Center for Economic Research (CentER) at Tilburg University.

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

1. Introduction

The traditional focus of the theory of incentives has been the optimal implementation of monetary transfers. A standard result of this literature is that a principal, who lacks knowledge about the agents' abilities on the delegated task, can profitably incentivize labor supply with a menu of contracts offering an array of carefully designed monetary transfers (Laffont and Tirole, 1993, Myerson, 1981, Green and Laffont, 1977). Recent literature also highlights the favorable influence of non-monetary incentives on labor supply. Production targets set by the principal (Kaur et al., 2015, Corgnet et al., 2015, Gómez-Miñambres, 2012), contests for status in the organization (Ashraf et al., 2014, Bandiera et al., 2013, Moldovanu et al., 2007), and environments that incorporate and intensify peer pressure (Mas and Moretti, 2009, Falk and Ichino, 2006) are proven to further enhance labor supply in settings where optimal monetary incentives might be already at work.

In this paper, I introduce a contracting modality featuring a novel class of non-monetary incentives. In particular, I propose a type of contracts that seeks to enhance labor supply by taking advantage of the regularity that individuals distort probabilities. Empirical evidence from the literature of decision-making shows that individuals, when making decisions under risk, tend to overweight small probabilities and underweight moderate to large probabilities (Abdellaoui, 2000, Gonzalez and Wu, 1999, Wu and Gonzalez, 1996, Tversky and Fox, 1995, Tversky and Kahneman, 1992). With these contracts the principal not only activates these probability distortions, but they also enable her to target those distortions of probability that enhance labor supply. Since these incentives are absent in more traditional contracting modalities, the contracts studied in this paper have the potential to generate greater output at no extra cost for the principal.

The proposed contracts, also referred as *probability contracts* through the paper, have the following structure. Agents' payment is stochastic and can take two possible values: a monetary compensation that increases with individual performance on the delegated task and no performance-contingent payment. The distinctive feature of the contracts is that the principal can determine the probability that the performance-contingent compensation is paid, and this decision is made before the task is performed. In a setting with full commitment, that is a setting whereby the principal credibly commits to pay the outcome determined by the stochastic rule of the contracts, the agents' decision about how much output to supply not only depends on the included monetary incentives, but also, and more importantly, on their perceived probability that performance on the delegated task affects their compensation.

To understand how the proposed contracts can outperform more traditional ones, consider a setting in which both types of contracts are cost-equivalent for the principal. This implies that the expected monetary compensation associated to supplying some level of output when agents work under a probability contract is exactly equal to the monetary compensation given to agents when the same level of output is supplied under a traditional contract. Expected value maximizers will be equally motivated under the two contracts and, as a result, will work to supply the same amount of output. However, when the assumption that agents perceive probabilities accurately is relaxed, and instead it is assumed that they overweight the probability that the performance-contingent outcome realizes, the offered probability contract motivates agents to supply more output. The underlying reason for this boost in labor supply is that this specific distortion of probabilities inflates the agents' perceived benefits of supplying higher levels of output.

A simple theoretical framework serves two purposes. First, it presents the necessary conditions guaranteeing the main result of the paper, which is that probability contracts generate higher output than traditional contracts at no extra cost for the principal. Second, it provides a set of predictions that are empirically tested with a laboratory experiment.

The sufficient conditions guaranteeing the main result of the paper are that: i) the representation of the agents' risk preferences admits probability distortions and ii) the principal implements these contracts with a probability that is strongly overweighted by agents, in order to induce risk seeking attitudes and, consequently, a taste for risky compensation schemes. Condition i) is ensured characterizing the agents' risk preferences with rank-dependent utility (Quiggin, 1982). Thus, probability distortions are modeled by means of probability weighting functions that transform cumulative probabilities. Condition ii) depends on the shape of the agents' probability weighting function. Specifically, weighting functions with global concavity or with an inverse-S shape generate intervals where probabilities are overweighted. The existence of intervals where probabilities are overweighted along with a novel condition that I call *bounded concavity of the weighting function* ensure the existence of probabilities that, when targeted by the principal with the proposed contracting device, enhance labor supply. Failing to comply with any of these conditions yields the opposite result: probability contracts generate lower output than traditional contracts. These antagonistic results of the model develop into the main testable predictions.

The model also introduces a taxonomy of probability distortions taken from the literature of decision theory (Abdellaoui et al., 2011, Wakker, 2010, Tversky and Wakker, 1995). In particular, I classify probability distortions as caused either by motivational factors, namely *optimism* or *pessimism* toward risk, or by *likelihood insensitivity*, which refers to the cognitive inability of individuals to accurately evaluate probabilities (Tversky and Wakker, 1995). I show that optimism or likelihood insensitivity suffice to guarantee the efficiency of the proposed contracting modality. Thus, when implementing this contract modality, principals can contract with agents who exhibit optimism toward risk and/or with agents who exhibit limitations in their perception of probabilities.

A controlled laboratory experiment demonstrates that the proposed contracts, when implemented with probability $p = 0.10$, yield higher effort in a real-effort task than a cost-equivalent piece-rate. In contrast, I find that probability contracts implemented with larger probabilities, namely $p = 0.30$ or $p = 0.50$, yield no differences in performance as compared to a cost-equivalent piece-rate. The experiment also features an elicitation of the utility and probability weighting functions of subjects. These data show that subjects display an average weighting function with a strong inverse-S shape and optimism. I demonstrate that this pattern of probability distortion is able to explain the findings of the real-effort part of the experiment. In addition, the data demonstrate that likelihood insensitivity, rather than optimism, generates the documented difference in performance between the probability contract implemented with $p = 0.10$ and the cost-equivalent piece-rate.

While the proposed contracts seem abstract, their practical implementation features well-known tools of contract theory and personnel economics. For instance, the principal can implement them using incentives schemes that feature monetary bonuses paid in the contingency that a milestone is reached. When the principal is able to set the milestone, allowing her to specify the probability that the bonus is achieved, and when she is able to adjust the magnitude of the bonus, the beneficial properties of the contracts proposed in this paper apply. Also, the principal can implement these incentives offering compensation plans that include stock options. When these plans are designed in a way that the event of dividend payment realizes with a small probability, the advantageous properties of the probability contracts apply. I provide a detailed explanation of these applications and provide some more in the last section of the paper.

This paper contributes to several strands of literature. Its theoretical and empirical results add in multiple ways to the literature of behavioral contract theory (See Koszegi (2014) for a review). The main contribution to this literature is the result that stochastic contracts that allow the principal to adjust the degree of risk faced by agents can be optimal when agents transform probabilities. While the optimality of stochastic contracts has been put forward in other settings, such as when agents exhibit aspiration levels (Haller, 1985) or when agents are highly loss averse (Herweg et al., 2010), I am the first to show that they are desirable when agents exhibit probability weighting.

To my knowledge only Spalt (2013) has studied optimal incentives design under probability weighting. The most relevant distinction with respect to that paper is that I focus on the agents' incentive compatibility constraint. That is, I analyze the labor supply incentives that result from exposing agents to different degrees of risk, with the proposed contracting modality, and establish the optimal degree of risk that the principal should implement. Spalt's (2013) analysis does not consider these incentives and ignores such constraint. Another relevant difference is that I propose a general type of contracts that introduces risk in the agents' compensation. They can be brought to practice in multiple ways, among which, but not only restricted to, implementing compensation plans featuring stock options.[1]

Additionally, some theoretical studies show that in settings of hidden action the principal can derive rents when contracting with optimistic agents (De La Rosa, 2011, Gervais et al., 2011, Santos-Pinto, 2008). The rationale behind these findings is that optimistic agents are willing to accept contracts offering high-powered monetary incentives in the contingency that favorable but unlikely outcomes realize, and also excessively low-powered monetary incentives when likely but unfavorable outcomes realize. This paper demonstrates theoretically and empirically that optimism is a sufficient but not a necessary condition to ensure the efficiency of the proposed contracts. That is because in the absence of optimism but in the presence of likelihood insensitivity, the existence of a region of probabilities that the principal can profitably target with these contracts is guaranteed. Moreover, the existence of these profitable probability distortions is ensured even when agents display strong pessimism toward risk—the extreme opposite of optimism. Thus, likelihood insensitivity—the cognitive component of probability weighting—- emerges as a novel deviation from standard preferences that can be exploited by the principal and that is complementary to optimism.[2] Another distinction with respect to these studies is that I model probability distortions through weighting functions that transform cumulative probabilities. The advantages of this representation are that probability distortions are endowed with the choice primitives of rank-dependence and also that probability distortions are directly linked to the risk attitudes of the decision-maker (Abdellaoui, 2002, Quiggin, 1982). Thus, with this characterization I avoid some limitations in

[1] Other crucial differences with respect to Spalt (2013) are the following. First, Spalt (2013) demonstrates the profitability of compensation plans with stock options using a calibration exercise that employs parameters estimated in classical experiments. I demonstrate theoretically, that is using analytical solutions, that the type of stochastic contracts that I propose is more efficient than more traditional contracting modalities. Also, my experiment is designed to directly link the subjects' performance under probability contracts to their risk preferences. This feature of the experimental design allows me to cleanly establish whether their own probability weighting, and not other factors, is driving the result that the proposed contracts can generate greater performance. Second, I propose a decomposition of probability distortions. Probability distortions emerge from optimism/pessimism or from likelihood insensitivity. This decomposition allow me to examine the role of these components on the performance of the proposed contracting modality. Not only is Spalt (2013) silent about this distinction, but his results are restricted to one family of probability weighting functions, namely Tversky and Kahneman (1992), which cannot separate these components and does not have the flexibility to capture all possible shapes of probability weighting.

[2] Moreover, I show that the principal can use probability contracts to exploit optimistic agents through a completely different channel. Specifically, the stochastic nature of these contracts ensures that optimistic agents inflate the expected returns of supplying high levels of output, even when the agents' expected pay is kept constant for all the possible probabilities that the principal can specify.

existing literature. For instance, that agents display excessive optimism about the realization of a favorable outcome but they are still required to display risk aversion. Also, I do not require ambiguity (unknown probabilities) to accommodate optimistic or pessimistic beliefs using the assumption of heterogeneous priors.

The results of this study also contribute to the literature of decision theory in multiple ways. First, I elicit probability weighting functions using the two-step method developed by Abdellaoui (2000) and find that subjects in the experiment exhibit an inverse-S shaped probability weighting function. This result corroborates the common finding in the literature (See Wakker (2010) pp. 204 for an extensive list of references). Furthermore, the experimental results illustrate the importance of using parametric functions that can separate likelihood insensitivity from optimism. I use different methods proposed by Wakker (2010) to isolate these two components of probability weighting and show that they contribute unequally to the efficiency of the contract. Finally, the model highlights the importance of other properties of the curvature of the weighting function. I show that agents who exhibit *bounded concavity of the weighting function* and who overweight some non-empty interval of probability, exhibit, regardless of the concavity of their utility curvature, a region of probabilities that can be profitably targeted by the principal with the proposed contracts.

2. The model

The theoretical framework considers a principal (she) who delegates a task to a representative agent (he). The agent's decision consists on supplying a level of output $y \in [0, \bar{y}]$. This decision depends on the disutility associated to producing output as well as on the incentives included in a take-it or leave-it contract offered by the principal.

I first assume that the agent experiences costs from producing output, which captures the notion that working on the task requires attention, persistence, and effort. I model these costs through the function $c(y, \theta)$, an increasing, two-times-differentiable and strictly convex function.

Assumption 1 (A1). *$c(y, \theta)$ is a C^2 function with $c_y(0, \theta) = 0$ $c_y(y, \theta) > 0$, $c_{yy}(y, \theta) > 0$, and $c_{y\theta}(y, \theta) < 0$ for all y,θ.*

The last expression in Assumption 1 entails that an agent with higher ability displays flatter cost functions, making it is less costly for him to deliver higher levels of output as compared to agents with lower ability.

The agent's ability, θ, is unknown to the principal at the moment in which a contract is offered. This introduces asymmetric information between the principal and the agent. Moreover, the supplied level of output, y, is ex-post observable and, as a consequence, contractible. That is, the principal could offer a contract that ties the agent's compensation to the level of output that has been supplied. The focus of the model is thus settings in which these "performance-pay" contracts can be implemented. In particular, the model compares the incentives generated by different types of performance-pay contracts, among them the proposed probability contracts. The next subsection describes the contracting modalities studied in the paper.

2.1. The contracts

2.1.a. *Piece-rate contracts.–* I use piece-rate contracts as the benchmark. These contracts have the property that equally-sized increments of output taking place at any point in the output space are rewarded with the same monetary compensation. When offered this contract, the agent is incentivized to supply output up to a level at which supplying extra labor to raise output by some arbitrarily small $\epsilon > 0$ becomes excessively costly. As shown by Gibbons (1998), piece-rates in static settings with adverse selection are optimal. That is because they induce the agent to choose the optimal amount of output given his ability, via the the *taxation principle* (Guesnerie, 1995).[3]

Formally, the agent is offered a schedule $W(a, y) = ay$, where $a > 0$ represents a monetary quantity. When offered $W(a, y)$, the agent derives utility from the monetary rewards included in the contract. I assume that such utility can be represented by $b(y)$, an increasing and two-times continuously differentiable function:

Assumption 2 (A2). *$b(y)$ is a C^2 function with $b(0) = 0$ and $b_y(y) > 0$ for all y.*

Note that I do not impose assumptions on the sign of the second derivative of the function $b(y)$. This is because the results of the model will be evaluated under the two signs that $b_{yy}(y)$ can attain.

All in all, the agent's utility when offered $W(a, y) = ay$ can be written as

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

The agent maximizes utility by supplying an output level y^* that satisfies the following

[3] Specifically, piece-rates can compose a menu of linear contracts forming the upper envelope of the optimal incentive contract obtained through the truthful revelation mechanism. Thus, an agent of ability θ self-selects to the best-fitting linear contract by choosing $y(\theta)$ that entails receiving a wage $w(y(\theta)$.

first-order condition:[4]

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

Equation (2) shows higher ability on the task, θ, makes it less costly for an agent to deliver high levels of output. Also, higher monetary incentives, a, increase the marginal utility of output. Hence, increments in these variables motivate the agent to supply higher output levels when working under a piece-rate contract.

2.1.b. *Probability contracts.–* Alternatively, the principal could offer probability contracts. These contracts also offer a monetary compensation that depends on y, but, unlike the piece-rate, the performance-contingent compensation is not given with certainty. Instead, the agent receives such compensation with a probability $p \in (0, 1]$ that is chosen by the principal. As a consequence, the principal has two channels to motivate the agent: i) through the monetary rewards given in exchange of the level of output that is supplied and ii) through changes in the likelihood that such rewards are indeed paid.

Formally, probability contracts can be represented as a lottery $V(A, y) = (Ay, p; 0, 1 - p)$, where $A > 0$ represents a monetary quantity. The timing of these contracts is as follows. The principal moves first choosing $p \in (0, 1]$. After this choice is made, p is communicated to the agent before he makes a decision about the level of y to be supplied. Next, the agent chooses y. Finally, when the contracted work-span concludes, a random device to which the principal commits determines whether or not the agent's compensation depends on the supplied level of y.[5]

When offered $V(A, y)$, the utility of an agent whose risk preferences can be characterized by expected utility theory (EUT, henceforth) becomes:

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

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

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

[4]A necessary condition for (2) to attain a maximum of $U(y)$ is $a^2 b_{yy}(ay) < c_{yy}(y, \theta)$. This implies that $-c_{yy}$ is more concave than b_{yy}. Since the focus of the model is on the incentives that probability contracts generate at the intensive margin rather than at the extensive margin, I impose this assumption wherever necessary.

[5]An alternative representation of this contract is $V(A, y) = (B + Ay, p; B, 1 - p)$ for some fixed-payment $B > 0$. This representation is more feasible implementation of these contracts inasmuch as they yield some non-zero base-pay. The last section of the paper presents implementations of probability contracts, using this type of incentive schemes.

Equation (4) shows that a higher probability, p, generates higher output, y^{**}. Therefore, an agent is motivated to deliver more output when the likelihood of obtaining a performance-contingent payment increases. Also, equation (4) shows that, as with piece-rate contracts, higher ability on the task, θ, and higher monetary incentives, A, yield higher y^{**}.

2.1.c. *The probability contracts and agents who distort probabilities.–* So far I have assumed that when incentivized with $V(A, y)$, the agent evaluates probabilities accurately. That is equivalent to say that his risk preferences can be characterized according to EUT. In this subsection, I relax this assumption and let the agent distort probabilities systematically as suggested by empirical evidence from the literature of decision-making (Abdellaoui, 2000, Gonzalez and Wu, 1999, Wu and Gonzalez, 1996, Tversky and Fox, 1995, Tversky and Kahneman, 1992). As it will become evident with Proposition 1 and Proposition 2, whether an agent distorts or not probabilities is key to the effectiveness of the proposed contract.

I model probability distortions through probability weighting functions. In particular, assume that the agent transforms the probability $p \in [0, 1]$ using a weighting function $w(p)$ that satisfies the following properties.

Assumption 3 (A3). *A probability weighting function is* $w(p) : [0, 1] \to [0, 1]$ *with:*

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

According to Assumption 3, the probability weighting function $w(p)$ is an increasing and two-times continuously differentiable function that maps the unit interval onto itself. The weighting function contains *at least* two fixed-points: one at $p = 0$ and another one at $p = 1$. Furthermore, $w(p)$ can exhibit three different shapes: a concave shape if $\hat{p} = 1$, a convex shape if $\hat{p} = 0$, and an inverse-S shape, that is first concave and then convex, whenever $\hat{p} \in (0, 1)$. The latter shape generates an additional interior fixed-point, $\tilde{p} \in (0, 1)$.[6]

I consider a theory of risk that incorporates distortions of probabilities through probability weighting functions, namely Rank-Dependent Utility (RDU, henceforth) (Quiggin, 1982). When offered $V(A, y)$, an agent with RDU preferences ranks the outcomes of the offered contract in terms of their desirability. Thus, the payment Ay is ranked above the zero-payment outcome. This ranking affects the agent's decision weights, which are the subjective probabilities assigned to each outcome. Specifically, the agent's decision weight associated to getting paid according to his own performance is $\pi(Ay) = w(p) - w(0) = w(p)$ and the decision weight of getting the zero payment is $\pi(0) = w(1) - w(p) = 1 - w(p)$.[7] Thus, the rank-dependent expected utility of the agent when offered $V(A, y)$ is:

$$RDU(y) = w(p)b(Ay) - c_y(y, \theta). \tag{5}$$

The rational agent with rank-dependent preferences maximizes utility choosing the production level, y_R^{**} that satisfies the following first-order condition:

$$w(p)Ab_y(Ay_R^{**}) - c_y(y_R^{**}, \theta) = 0. \tag{6}$$

The influence of the parameters of the model on the optimal production level, y_R^{**} , is similar as those in previous analyses. In particular, the output level supplied by the agent increases with higher skills θ and with higher monetary incentives A. In addition, higher p generates higher output levels. However, in contrast to the case in which the agent has EUT preferences, the influence of p on y_R^{**} is non-linear: a probability increment in the interval $p \in [0, \hat{p})$ yields smaller output increments than an equally-sized probability increments in the interval $p \in (\hat{p}, 1]$.

Another theory of risk that incorporates distortions of probabilities through probability weighting is Cumulative Prospect Theory (CPT, henceforth) (Tversky and Kahneman, 1992). CPT is a more descriptive version of RDU. An agent with CPT preferences also ranks the outcomes of the contract in terms of their desirability and assigns decision weights, composed of transformed probabilities, accordingly. However, unlike RDU preferences, the agent with

[6] As noted by Wakker (2010), a weighting function with *cavexity*, that is first concave and then convex, does not necessarily ensure the existence of the interior point. However, the assumptions $\lim_{p \to 0^+} w_p(p) = \infty$ if $\hat{p} > 0$ and $\lim_{p \to 1^-} w_p(p) = \infty$ if $\hat{p} < 1$ along with *cavexity* guarantee the existence of an interior fixed point $\tilde{p} \in (0, 1)$.

[7] More generally an individual with RDU preferences facing a lottery $(x_1, p_1; x_2, p_2; ...; x_n, p_n)$ ranks the prizes of the lottery in an increasing arrangement $x_1 < x_2 < ... < x_n$ and assigns decision weights to each of these outcomes according to their rank in the following way $\pi_n = w(p_n), \pi_{n-1} = w(p_{n-1} + p_n) - w(p_n), ..., \pi_1 = 1 - \sum_j^n w(p_n)$. The ordered outcomes of the lottery enter the utility function through the function $b(\cdot)$ which is typically assumed to be either concave or linear.

CPT preferences displays different risk attitudes for the domain of gains, that is all outcomes of the contract ranked above some reference point $r \geq 0$, and the domain of losses, that is all outcomes of the contract ranked below the reference point $r \geq 0$. Despite this difference, the incentives generated by probability contracts are similar to those of an agent with RDU preferences, since rank-dependence is an integral part of CPT. In the interest of space, I relegate the formal description of CPT preferences and the analysis of the incentives produced by the studied contracts on agents with CPT preferences to Appendix B.

2.1.d. *Probabilistic risk attitudes and their decomposition.–* Characterizing the agent's risk attitudes with either RDU or CPT preferences introduces *probabilistic risk attitudes* (Wakker, 2001, Tversky and Wakker, 1995, Wakker, 1994). That is, the agent's attitudes toward risk are not only determined by his sensitivity toward outcomes, as assumed by EUT, but also by his sensitivity toward probabilities. Therefore, the agent's global risk attitudes are determined by the interaction between these two psycho-physical reactions. Note that as much as risk attitudes generated by sensitivity toward outcomes is captured by the shape of the basic utility function, probabilistic risk attitudes are captured by the shape of the probability weighting function. In particular, concave weighting functions generate probabilistic risk-seeking while convex weighting functions generate probabilistic risk aversion.

Probabilistic risk attitudes are relevant to the analysis of incentives included in probability contracts for two reasons. First, as it will be demonstrated later on, probability contracts have the advantage over piece-rates that they enable the principal to induce probabilistic risk attitudes in the agent, which enhance labor supply. In particular, when the principal implements probability contracts with a probability that induces strong probabilistic risk-seeking, she can obtain higher output levels than those that would have been obtained with piece-rates and this boost in labor supply comes at no extra cost for the principal. Second, I investigate the specific attribute of probabilistic risk attitudes that guarantees this relevant result. To that end, I decompose probabilistic risk attitudes into two components that will be described below, and examine how each of them affects the agent's motivation when he works under $V(A, y)$.

The following decomposition of probabilistic risk attitudes is based on Abdellaoui et al. (2011), Wakker (2010), and Wakker (2001). The first component of probabilistic risk attitudes captures deviations from EUT through either pessimism or optimism toward risk. These two motivational factors affect probability evaluations because of an irrational belief that unfavorable outcomes, in the case of pessimism, or favorable outcomes, in the case of optimism, happen more often (Wakker, 2010). Pessimism is represented with a convex weighting function, which assigns large weights to worst-ranked outcomes and small weights to best-ranked outcomes. This representation has been the convention in the early theoretical literature on rank-dependence (Yaari, 1987, Chew et al., 1987). Conversely, optimism is represented with a concave weighting function, which assigns large decision weights to best-ranked outcomes and small decision weights to worst-ranked outcomes. Figure 1, presents an example of pessimism and optimism.

The second component of probabilistic risk attitudes is likelihood insensitivity (Abdellaoui et al., 2011, Wakker, 2001, Tversky and Wakker, 1995). This component captures the notion that individuals distort probabilities because they are not sufficiently sensitive towards changes in intermediate probabilities and are overly sensitive to changes in extreme probabilities (Wakker, 2001). Hence, probabilities are also distorted because of cognitive and perceptual limitations that impede the accurate evaluation of probabilities. An extreme characterization of likelihood sensitivity is a weighting function $w(p)$ with the properties of Assumption 3 and with the additional restrictions $\tilde{p} = 0.5$, $\lim_{p\to 0^+} w(p) = 0.5$, and $\lim_{p\to 1^-} w(p) = 0.5$. Such a weighting function entails that the agent accurately discriminates certainty and impossibility, but assigns $p \approx 0.5$ to uncertain events. An opposing characterization is that of an agent who is fully sensitive to probabilities. Such agent's probability weighting function, ignoring the potential presence of optimism and pessimism, can be represented with $w(p) = p$. Figure 2 presents examples of different degrees of likelihood insensitivity.

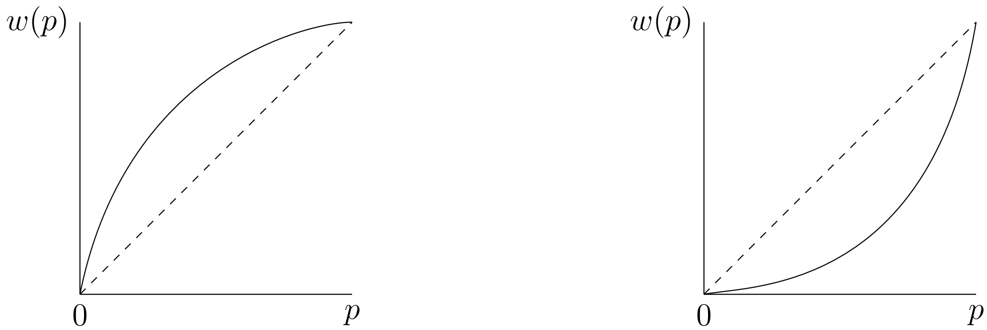
These two components of probabilistic risk attitudes are formally defined next:[8]

Definition 1: Pessimism (Optimism). *For a fully likelihood sensitive agent, pessimism (optimism) is represented by $w(p)$ with the properties of A3 and $\hat{p} = 0$ ($\hat{p} = 1$).*

Definition 2: Likelihood insensitivity. *For an agent without optimism or pessimism, likelihood insensitivity is represented with a $w(p)$ with the properties of A3 and $\tilde{p} = 0.5$.*

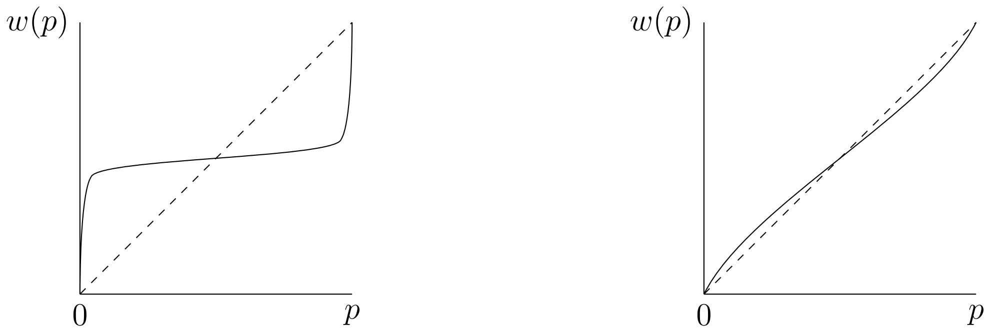
When optimism or pessimism and likelihood insensitivity coexist, they generate probabilistic risk attitudes that can be characterized by a weighting function with an inverse-S shape. The location of the interior fixed-point of this weighting function depends on whether the agent displays pessimism or optimism. A pessimist agent who is also likelihood insensitive, exhibits a $w(p)$ with an interior fixed-point in the interval $\tilde{p} \in (0, 0.5)$. On the other hand, an optimist agent who is also likelihood insensitive has a weighting function with an interior fixed point $\tilde{p} \in (0.5, 1)$. Also, note that Assumption 3 does not impose $\tilde{p} = \hat{p}$, which is a

[8] These two phenomenon have been addressed in the psychological literature as *curvature* and *elevation* (Gonzalez and Wu, 1999). I instead use the jargon used in economics. Pessimism was defined early on by Yaari (1987) and likelihood insensitivity was defined later by Tversky and Wakker (1995).



(a) Example of optimism (b) Example of pessimism

Figure 1: Motivational sources of probability distortion



(a) Example of extreme likelihood insensitivity (b) Example of moderate likelihood insensitivity

Figure 2: Cognitive sources of probability distortion

property of early representations of probability weighting functions. Instead, these two values could differ. The location of $\tilde{p}$ with respect to $\hat{p}$ is informative of the strength of optimism and pessimism with respect to likelihood insensitivity.[9]

When comparing $V(A, y)$ and $W(a, y)$, special focus is given to the roles of likelihood insensitivity and optimism. These two components of probabilistic distortion yield different requirements with regard to the implementation of the proposed contracts. In particular, if likelihood insensitivity leads to higher output supply when $V(A, y)$ is offered, then the higher performance obtained with this type of contracts is due to cognitive limitations that can be inherent to human nature and that can be readily available to the principal. Instead, if optimism yields that $V(A, y)$ generates higher output, then the principal needs to contract with agents that are optimistic when facing risk. This is a stringent requirement, given the abundant evidence that individuals are mostly averse to risk, and it can leverage additional restrictions on the principal's program (Spinnewijn, 2013)

2.2. Contract comparisons

In this subsection, I compare the two considered types of contracts with respect to the output levels that they generate. An alternative analysis, that not only incorporates the agent's decision about how much output to supply when different contracts are offered, i.e. the agent's incentive compatibility constraint, but also the agent's participation constraint as well as the objective function of the principal is presented in Appendix C. Such analysis yields similar conclusions regarding the optimal implementation of the contract: the principal is better off implementing probability contracts, as long as they are implemented with a probability that is severely overweighted by the agent.

To facilitate the comparisons between the contracts, I make an assumption about the monetary incentives offered by both types of contracts. In particular, I assume that probability contracts offer, on expectation, the same monetary rewards as piece-rates. Formally, let $A = \frac{a}{p}$, so that $E(V(a, y, p)) = ay = W(a, y)$. This equivalence allows me to focus on the incentives produced by probability contracts implemented with different p and how these incentives compare to those generated by piece-rates offering similar monetary payment.[10]

[9] When the convexity associated with pessimism outweighs the concavity of likelihood sensitivity in the interval $p \in (0, 0.5)$, the weighting function is convex at a larger range of probabilities and $\tilde{p} > \hat{p}$. Alternatively, when the concavity associated with optimism outweighs the convexity of likelihood sensitivity in the interval $p \in (0.5, 1)$, the weighting function is concave at a larger range of probabilities and $\tilde{p} < \hat{p}$. These conclusions are based on an additive representation of likelihood insensitivity and pessimism (optimism) in the weighting function. I assume such representation given their independence (Wakker, 2010).

[10] A consequence of this assumption is that probability contracts nest piece-rates. Specifically, when the principal chooses to compensate output constantly, this is as $p \to 1$, then $A \approx a$. Conversely, in a setting in which the principal decides to evaluate output with little frequency, this is as $p \to 0^+$, the monetary

The assumed equivalence is incorporated in the first-order conditions presented in equations (4) and (6). Therefore, these conditions can be rewritten as

$$ab_y\left(\frac{a}{p}y^{**}\right) - c_y(y^{**}, \theta) = 0, \tag{7}$$

and

$$\frac{w(p)}{p}ab_y\left(\frac{a}{p}y_R^{**}\right) - c_y(y_R^{**}, \theta) = 0, \tag{8}$$

respectively.

Since p is now included in the basic utility of the agent and the behavior of the function $b(y)$ as p changes is unknown, I introduce additional assumptions regarding the basic utility function that are presented next:

Auxiliary assumption (A4). $b(y,p)$ is a C^2 function with $b(0,p) = 0$, $b_y(y,p) > 0$, $b_{yp}(y,p) > 0$ if $b_{yy}(y,p) < 0$, and $b_{yp}(y,p) < 0$ if $b_{yy}(y,p) > 0$ for all y and p.

Next to all conditions stated in Assumption 2, it is assumed that $b_y(y,p)$ is increasing in p whenever $b_{yy}(y,p) < 0$. The intuition behind this condition is that the monetary reward $A = \frac{a}{p}$ becomes smaller with larger probabilities, which, for an agent with a basic utility function that is concave on y, implies that larger probabilities increase the marginal utility of supplying extra output. In contrast, for an agent with a basic utility that is convex on y, larger probabilities, which entail lower values of A, imply lower increments of marginal utility when higher levels of output are supplied.

We are now in a position to compare the considered contracts. I begin comparing y^* and y^{**} from equations (2) and (7). These two levels of output correspond to an agent whose risk preferences can be represented by EUT when he is offered piece-rate contracts and when he is offered probability contracts, respectively. To build intuition about how these production levels compare, consider the functional forms $c(y,\theta) = \frac{(y/\theta)^2}{2}$ and $b(y,p) = \frac{(\frac{ay}{p})^{1-\gamma}}{1-\gamma}$ with $\gamma \in (-1,1)$. Under these functional forms I obtain $y^* = \left(\frac{\theta^2}{a^{\gamma-1}}\right)^{\frac{1}{1+\gamma}}$ and $y^{**} = \left(\frac{\theta^2 p}{A^{\gamma-1}}\right)^{\frac{1}{1+\gamma}}$. Hence, it is evident that a necessary condition for $y^{**} > y^*$ is $\gamma \in (-1,0)$, which implies that, under the considered functional forms, an agent needs to be have a convex basic utility, and thus display risk seeking attitudes, to supply higher output under probability contracts.

The result of the example above suggests that the effectiveness of the proposed contracts

incentives for delivering one additional unit of output become large enough in order to compensate the agent for the low frequency at which output is paid.

depends on the agent's risk attitudes, which, for the case of EUT, are exclusively determined by the curvature of $b(y,p)$. Proposition 1 generalizes this result.

Proposition 1. *Under Assumptions A1, A2, and A4, for an agent with any* $\theta \in (0,1)$ *and weighting function* $w(p) = p$*, then* $y^{**} > y^*$ *for all* $p \in (0,1]$ *if* $b_{yy}(y,p) \geq 0$.

Proof. See Appendix A.

■

Proposition 1 states that when the agent's risk preferences can be represented by EUT, the curvature of $b(y,p)$ is critical for the efficiency of probability contracts. When $b_{yy}(y,p) < 0$ the agent experiences disutility from the risk that is introduced by the stochastic nature of the probability contracts, leading him to supply lower output than if he were offered a piece-rate. In contrast, if $b_{yy}(y,p) > 0$, the agent experiences utility from the risk that is introduced by these contracts, which motivates him to supply more output as compared to the case in which piece-rates were offered. These results hold for any probability that can be specified by the principal.

Next, I compare y^* and y_R^{**} from equations (2) and (8). These two output levels correspond to an agent with risk preferences represented by RDU when he is offered piece-rate contracts and when he is offered probability contracts, respectively.

Again, I build intuition assuming $c(y,\theta) = \frac{(y/\theta)^2}{2}$ and $b(y,p) = \frac{(\frac{ay}{p})^{1-\gamma}}{1-\gamma}$ with $\gamma \in (-1,1)$. These functional forms yield $y^* = \left(\frac{\theta^2}{a^{\gamma-1}}\right)^{\frac{1}{1+\gamma}}$ and $y_R^{**} = \left(\frac{\theta^2 w(p)}{A^{\gamma-1}}\right)^{\frac{1}{1+\gamma}}$. Note that as long as $\frac{w(p)}{p^{1-\gamma}} > 1$, then $y_R^{**} \geq y^*$. Hence, the comparison between these two output levels is determined by γ, as in Proposition 1,and also by p, which is chosen by the principal. Thus, given some curvature level γ, the principal can ensure $y_R^{**} \geq y^*$ by choosing an appropriate p. To illustrate how the principal's choice can influence the comparison between y_R^{**} and y^*, consider first an agent with $\gamma = 0$. For him $y_R^{**} \geq y^*$ is ensured whenever $\frac{w(p)}{p} > 1$, which for an agent with a $w(p)$ such that $\hat{p} > 0$, holds for any interval $p \in (0,\tilde{p})$. Instead, choosing $p \in [\tilde{p},1]$ yields $y_R^{**} < y^*$ regardless of whether $w(p)$ displays $\hat{p} > 0$ or $\hat{p} = 0$. Consider now agents displaying $\gamma > 0$. For them the condition ensuring $y_R^{**} \geq y^*$ becomes more stringent the larger γ is. To see how, note that $\frac{w(p)}{p^{1-\gamma}} > 1 \iff \frac{ln(w(p))}{ln(p)} < (1-\gamma)$, where it is evident that for larger values of γ, $|ln(w(p))|$ is required to be smaller than $|ln(p)|$. In other words, in face of stronger concavity of the utility function, larger degrees of overweighting of probabilities are needed to ensure $y_R^{**} \geq y^*$.

Proposition 2 generalizes the aforementioned rationale and also demonstrates that an agent who distorts probabilities according to a weighting function with $\hat{p} > 0$, can deliver higher performance when incentivized with probability contracts.

Proposition 2. *Under Assumptions A1, A2, A3, and A4, for an agent with any* $\theta \in (0, 1)$ *with weighting function displaying* $\tilde{p} > 0$ *and* $|w_{pp}(p)| < \infty$*, there exists* $p^* \in (0, \tilde{p})$ *such that* $y_R^{**} > y^*$ *if* $p < p^*$*.*

Proof. See Appendix A.

■

When the agent's risk preferences are represented by RDU and he has a weighting function that generates overweighting of probabilities for some non-empty interval, the principal can obtain higher output levels using probability contracts. However, to achieve this result the principal needs to target probabilities that are severely overweighted by the agent. This ensures that the probabilistic risk seeking attitudes induced by her choice outweigh potential risk averse attitudes generated by the shape of the agent's basic utility, guaranteeing global risk seeking attitudes. These risk attitudes imply a preference for stochastic contracts over piece-rates and, as a consequence, that more output is supplied under the proposed contracting modality.

A technical requirement included in Proposition 2 is that the agent's weighting function displays $|w_{pp}(p)| < \infty$. Intuitively, it is required that the agent's experienced sensitivity to changes in probability is bounded. This requirement guarantees that whenever there is an interval at which probabilities are overweighted, there must exist probabilities therein that, when included in the contract, induce global risk seeking attitudes. A family of weighting functions that complies with this requirement is the *constant relative sensitivity* proposed by Abdellaoui et al. (2010) when their index of relative sensitivity is bounded. The axiomatic foundations of that family of weighting functions, their empirical validity, and empirical evidence of the boundedness of their curvature are also presented in Abdellaoui et al. (2010).

Appendix B shows that agents with CPT preferences are also motivated to supply more output when incentivized with probability contracts. It is straightforward to show that probability contracts have the same motivational effect on agents with CPT preferences who are in the domain of gains as that exhibited by RDU agents. Moreover, when in the domain of losses, the three components shaping risk attitudes in CPT agents, namely loss aversion, probability risk attitudes, and utility curvature, enhance performance when the proposed contracts are offered. That is because the convex utility curvature and the probabilistic risk attitudes induced by the principal both generate risk seeking, inducing a taste for stochastic contracts. In addition, loss aversion motivates the agent to work harder on the task in order to move past his reference point.[11]

Next I comment on the role of likelihood insensitivity and optimism in Proposition 2. Note that a crucial requirement in Proposition 2 is that $w(p)$ exhibits $\tilde{p} > 0$, which ensures that probabilities are overweighted by the agent over some non-empty interval. Corollary 1 and Corollary 2 show that optimism and likelihood insensitivity are sufficient conditions to guarantee the existence of such interval.

Corollary 1. *Optimism in the absence of likelihood insensitivity guarantees Proposition 2.*

Proof. See Appendix A. ∎

Corollary 2. *Likelihood insensitivity in the presence of optimism or pessimism guarantees Proposition 2.*

Proof. See Appendix A. ∎

To understand the intuition of Corollary 1 note that optimism is a sufficient condition to guarantee Proposition 2 since, absent likelihood insensitivity, it generates $\tilde{p} = 1$ and, as a consequence, entails that the whole probability interval is overweighted. In contrast, when the agent exhibits pessimism, absent likelihood insensitivity, the entire probability interval is underweighted and Proposition 2 cannot hold. Therefore, more optimistic agents are more likely to be motivated by the incentives offered by the stochastic contract.

Corollary 2 shows that likelihood insensitivity, in the absence of pessimism or optimism, entails having a weighting function with a fixed-point $\tilde{p} = 0.5$ and, thus, that the interval of probabilities $p \in (0, 0.5)$ is overweighted. It is straightforward to show that when likelihood insensitivity coexists with optimism, the existence of an interval where probabilities are overweighted is still ensured. The most relevant implication of Corollary 2 is that even if the agent exhibits extreme pessimism, the existence of a non-empty interval where probabilities are overweighted is guaranteed by likelihood insensitivity. The independence of pessimism

[11] This seems at odds with Herweg et al. (2010) and de Meza and Webb (2007) who show that in a setting of hidden action, loss averse agents require contracts entailing low risk—binary contracts whereby payment is for some intervals insensitive to performance. However, note that in this framework performance is deterministic and the only source of uncertainty is that introduced by the contract. Thus, to avoid the disutility from being in the domain of losses, the agent can work harder on the task to past her reference point. This is not possible in the moral hazard context of Herweg et al. (2010) and de Meza and Webb (2007).

and likelihood insensitivity, as proposed by Wakker (2010), and the continuity of $w(p)$ ensure this result.

Finally, Corollary 2 has a relevant implication for the applicability of the proposed contracting modality. In particular, it states that the principal does not necessarily need to contract with a pool of agents that are overly optimistic about the risk that is implied by these contracts to ensure their effectiveness. It has been shown that such screening process can complicate the principal's program Spinnewijn (2013). Instead, cognitive factors that impede the accurate evaluation of probabilities also guarantee the efficiency of these contracts. Therefore, likelihood insensitivity emerges a deviation from standard preferences that can be targeted by the principal using the type of stochastic contracts proposed here.

To summarize, Proposition 1 and Proposition 2 yield opposing results. As shown throughout this section the theory used to characterize the agent's risk preferences as well as the probability that governs probability contracts are crucial in generating this antagonism. The specific predictions derived from this theoretical framework are presented in Section 4, after the experimental design and procedures have been described.

3. Experimental Method

The experiment was conducted at Tilburg University's CentERlab in April 2017. The participants were all students at the university and were recruited using an electronic system. The 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 are presented in Appendix D.

The experiment consisted of two parts. Upon arrival, participants were informed that their earnings from either part one or those from part two would become their final earnings and that this would be decided by chance at the end of the experiment. In the first part of the experiment subjects performed a task that demanded effort and attention. The task consisted of summing five two-digit numbers.[12] Each summation featured randomly drawn numbers by the computer, ensuring similar levels of difficulty among participants. When a participant knew the answer to the numbers that appeared in his screen, he submitted it using the computer interface. Immediately after submission, a new summation appeared on

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

the computer screen and the participant was invited to solve it. In total, subjects had 10 rounds of four minutes each to complete as many summations as they could.

There were four treatments that differed in the incentives given to subjects to perform the task. Subjects were randomly assigned to one of these four treatments. The baseline treatment is *Piecerate*. Subjects assigned to this treatment were paid 0.25 Euros for every correctly solved summation. The other treatments also offered monetary rewards that depended on individual performance of subjects on the task. However, in these treatments correct summations only in some of the rounds, chosen at random at the end of the experiment, counted toward their earnings. These treatments were designed to represent probability contracts implemented with different probabilities. In particular, the treatments *LowPr*, *MePr* and *HiPr* featured a low, a medium, and a high probability, respectively, that performance in a round counted toward the subjects' earnings. In LowPr one round was randomly chosen at the end of the experiment and performance in that round was paid. Similarly, in MePr and HiPr, three and five rounds, respectively, were randomly chosen at the end of the experiment and only performance in those rounds was paid. Note that this representation of probability contracts requires that subjects exhibit isolation (Tversky and Kahneman, 1981), which is strongly supported by the literature of experimental economics when these *random incentive devices* are implemented (See for instance Baltussen et al. (2012), Hey and Lee (2005) and Cubitt et al. (1998)).[13] [14]

As in the theoretical framework, the monetary incentives offered in Piecerate, LowPr, MePr and HiPr were calibrated such that subjects faced, on expectation, the same monetary incentives across treatments. For instance, a subject assigned to LowPr received 2.50 Euros for each correct summation in the round that was chosen for compensation, which was tenfold of what a subject assigned to Piecerate earned for each correctly solved summation. This difference in monetary payments exactly accounts for the difference in the probability that performance in a round is paid across the treatments. Similarly, subjects assigned the MePr and HiPr treatments received a compensation of 0.85 and 0.50 Euros, respectively, for each correctly solved summation in the rounds that were randomly chosen for compensation.[15]

[13]Isolation in this setting refers as to when a subject chooses the amount of effort to exert in a task as if each round was evaluated in isolation. Under this condition, the treatments LowPr, MePr, and HiPr, generate uncertainty about whether the performance to be supplied in the round counts towards performance.

[14]A common misunderstanding regarding the *random lottery incentive* is that the independence axiom is a necessary condition to guarantee that subjects understand the incentives encompassed by this device. First, note that the independence axiom is a sufficient, but it not a necessary condition since it also needs to be complemented with dynamic principles to validate experimental measurement (Baltussen et al., 2012). Second, the condition of isolation (Tversky and Kahneman, 1981) leads to proper experimental measurement when the random incentive scheme is used, even if the independence axiom is not assumed (Baltussen et al., 2012).

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

The probabilities governing the treatments LowPr, MePr and HiPr were chosen according to the common finding in the literature of decision-making: subjects distort probabilities according to an inverse-S shape probability weighting function with an interior fixed point at approximately $p = 0.33$. If subjects in the experiment follow this regularity, then they should, on average, overweight the probability that a round is chosen with 10% chance, underweight the probability that a round is chosen with 50% chance, and approximately evaluate accurately the probability that a round is chosen with 30% chance. Hence, the experiment was designed to observe performance differences across the treatments as long as the non-monetary incentives generated by probability distortions are sufficiently powerful.

Once the last round of the real-effort task was over, participants were asked to state their beliefs about how well they did in the real-effort task. I included this belief elicitation to assess whether subjects anticipated the effect of the treatments on their own performance. A subject received a bonus of one Euro if his answer was exactly equal to the number of correct summations that he performed over the ten rounds. This elicitation was unanticipated and the monetary compensation when the subject provided a correct answer was small as compared to the other sources of earnings in the experiment. These two characteristics ensure incentive compatibility (Schlag et al., 2015).

In the second part of the experiment, the subjects' task was to choose between pairwise lotteries. This part of the experiment was designed to elicit their utility and the probability weighting functions. To elicit these two functions, I used the two-step method developed by Abdellaoui (2000). This method has the advantage of not making assumptions about the way in which subjects evaluate probabilities nor the way in which they evaluate monetary outcomes.[16]

This part of the experiment consisted of 11 decision sets. Decision sets 1 to 6 constitute the first step of Abdellaoui (2000)'s methodology. These decision sets elicit a sequence of outcomes $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ that made the subject indifferent between a lottery $L = (x_{j-1}, 2/3; 0.5, 1/3)$ and a lottery $R = (x_j, 2/3; 0, 1/3)$ for $j = \{1, ..., 6\}$. Indifference was found through bisection. Specifically, a subject was required to express his preference between two initial lotteries. After having made a choice, the outcome x_j of lottery R changed as a function of this choice, such that either the outcome of the chosen lottery was replaced by a less attractive alternative, or the outcome of the not chosen one was replaced by a more attractive alternative, while the other lottery remained the same. When facing the new situation, the subject was invited to make a choice again between the modified lotteries L

Piecerate.

[16] A drawback of this method is that is not incentive compatible when subjects are aware of the chained nature of the questions. I overcome this disadvantage by adding a set of decisions in-between that are not used in the analysis and by randomizing the appearance of the lotteries of decision sets 7 to 11.

and R. This process was repeated four times for each decision set. The left panel of Table 1 presents an example illustrating the bisection procedure for Decision sets 1 to 6.

The lotteries in decision sets 1 to 6 were designed such that the resulting sequence $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ yielded equally spaced utility levels, i.e. $u(x_j) - u(x_{j-1}) = u(x_{j-1}) - u(x_{j-2})$ for all for $j = \{1, ..., 6\}$.[17] The starting point of the program, x_0, was set such that the monetary outcomes used in the lotteries reflected the subject's earnings in the first part of the experiment. Specifically, x_0 was set at 2/5th of what a subject earned in the first part of the experiment. The advantage of using monetary outcomes of similar magnitude as the incentives offered in the real-effort task, is that I can correlate the behavior of the subjects in such task with their elicited preferences. Subjects were not informed about this calibration.

Decision sets 7 to 11 constitute the second step of Abdellaoui (2000)'s methodology. These decision sets were designed to elicit a sequence of probabilities,

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

where $p_{j-1} = \frac{j-1}{6}$ and $j = \{2, ..., 6\}$. These probabilities made the subjects indifferent between the lottery $L = (x_6, w^{-1}(p_{j-1}); x_0, 1 - w^{-1}(p_{j-1}))$ and the degenerate lottery x_{j-1}. Again, indifference was found through bisection, with the probability of lottery L changing as a function of the subject's previous choices. The right panel of Table 1 presents an example illustrating the bisection procedure for these decision sets. The lotteries in decision sets 7 to 11 were designed so that the elicited probabilities yield equally spaced probability weights, i.e. $w(p_j) - w(p_{j-1}) = w(p_{j-1}) - w(p_{j-2})$ for $p_{j-1} = \frac{j-1}{6}$ and $j = \{2, ..., 6\}$.

Once the second part of the experiment was over, subjects were given feedback about their performance in the real-effort task, were told which round(s) counted toward payment if assigned to LowPr, MePr or HiPr, and were informed whether their belief was correct. Also, they were informed about the lottery that was chosen for compensation for the second part of the experiment and its realization. In addition, subjects learned whether part one or part two counted toward their final earnings.

4. Hypotheses

The theoretical model generates a set of hypotheses that are tested with the experiment. The first two hypotheses regard the subjects' performance in the real-effort task. According to Proposition 1, piece-rate contracts outperform the probability contracts when the majority

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

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

#	Lotteries	Interval	Choice	Lotteries	Probability	Choice
1	L=$(1, 0.66; 0.50, 0.33)$ R=$(3.7, 0.66; 0, 0.33)$	[1, 6.4]	L	L=$(x_1, 1)$ R=$(x_6, .50; 1, 0.5)$	$[0, 1]$	L
2	L=$(1, 0.66; 0.50, 0.33)$ R=$(5.05, 0.66; 0, 0.33)$	[3.7,6.4]	R	L=$(x_1, 1)$ R=$(x_6, .75; 1, 0.25)$	$[.5, 1]$	L
3	L=$(1, 0.66; .050, 0.33)$ R=$(4.38, 0.66; 0, 0.33)$	[3.7,5.05]	R	L=$(x_1, 1)$ R=$(x_6, .87; 1, 0.13)$	$[.75, 1]$	R
4	L=$(1, 0.66; 0.50, 0.33)$ R=$(4.04, 0.66; 0, 0.33)$	[3.7,4.38]	L	L=$(x_1, 1)$ R=$(x_6, .81; 1, 0.19)$	$[.75, .87]$	L
5	L=$(1, 0.66; 0.50, 0.33)$ R=$(4.21, 0.66; 0, 0.33)$	[4.04,4.38]	L	L=$(x_1, 1)$ R=$(x_6, .85; 1, 0.15)$	$[.81, .87]$	L
End		$x_1 \in$[4.21, 4.38]			$p_1 \in [85, 87]$	

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

of subjects exhibit risk preferences that can be represented by EUT and have a concave utility curvature, $b_{yy}(y, p) < 0$. When these two conditions are met, subjects assigned to Piecerate exhibit higher performance than subjects assigned to LowPr, MePr, or HiPr. Moreover, since HiPr is less risky than MePr, and MePr is less risky than LowPr, these subjects should display second highest performance under HiPr, followed by performance under MePr, and lowest performance when assigned LowPr.

Hypothesis 0. *Subjects with risk preferences as in EUT and with a concave basic utility function display performance levels that conform to the ranking: Piecerate > HiPr > MePr > LowPr.*

In stark contrast, Proposition 2 demonstrates that probability contracts outperform piece-rate contracts even if most individuals exhibit $b_{yy}(y, p) < 0$. The reason behind such result is that the assumption that agents evaluate probabilities accurately, crucial to obtain Proposition 1, is relaxed and agents can experience probabilistic risk attitudes. Hence, when the principal implements probability contracts with a probability that is sufficiently overweighted by the agent, inducing a taste for stochastic contracts, she obtains higher output than if she used piece-rates.

As mentioned in Section 3, if subjects in the experiment conform to the common finding that $\tilde{p} = 0.33$ and the non-monetary incentives contained in the probability contracts are strong, subjects assigned to LowPr should display higher performance than subjects in Piecerate. Instead, subjects in HiPr should display lower performance than those in Piecerate, and the performance levels between subjects in Piecerate and MePr should be indistinguishable. These comparisons are included in Hypothesis 1.

Hypothesis 1. *Subjects with risk preferences as in RDU and* $\tilde{p} = 0.33$ *exhibit average performance levels that conform to the ranking:* $LowPr > MePr = Piecerate > HiPr$

Empirical support in favor of Hypothesis 1 does not conclusively validate the theoretical model. Because factors other than probabilistic risk attitudes could spawn these performance differences. Hence, if the model is accurate, performance differences between the LowPr and Piecerate and/or between the HiPr and Piecerate should be explained by the subjects' tendency to overweight small probabilities and/or underweight large probabilities.

Hypothesis 2. *Subjects assigned to LowPr (HiPr) who have a weighting function that overweights (underweights) small (large) probabilities exhibit higher (lower) performance with respect to subjects assigned to Piecerate.*

Finally, if Hypothesis 2 holds, I am interested in understanding which component of probabilistic risk attitudes causes the performance differences between the treatments. Corollary 1 predicts that more optimistic agents, independently of whether they are likelihood insensitive or not, should display higher performance when assigned to treatments representing probability contracts.

Hypothesis 3. *Optimistic subjects exhibit higher performance when assigned to the treatments representing the probability contract as compared to subjects in Piecerate.*

Corollary 2 predicts that likelihood insensitive subjects, regardless of whether they also exhibit optimism or pessimism, should display steeper performance differences across the treatments in the direction predicted by Hypothesis 1.

Hypothesis 4. *Likelihood insensitive subjects exhibit steeper performance differences in the direction predicted by Hypothesis 1.*

The accuracy of these hypotheses will be evaluated in light of experimental data in the next sections.

5. Results

5.1. Treatment effects

In this subsection I compare performance in the real-effort task across the treatments. Performance is defined as the total number of correctly solved summations by a subject. Table 2 presents the descriptive statistics of performance by treatment. This table shows that, as predicted by Hypothesis 1, the probability contract with $p = 0.10$ generates higher performance than the piece-rate contract. Specifically, subjects assigned to the LowPr treatment solved on average 20.56 % more summations than subjects assigned to Piecerate ($t(84.454) = 2.361, p = 0.010$).[18] The effect size of this difference in performance is of 0.5 standard deviations which is significant at the 5% confidence level.[19]

Table 2: Descriptive statistics of performance by treatments

Treatment	LowPr	MePr	HiPr	Piecerate	Total
Mean	98.116	87.900	83.75	81.377	87.686
Median	91	87	82.500	77	85
St.dev.	34.659	28.134	24.358	31.684	30.412
N	43	40	44	45	172

In contrast, probability contracts implemented with higher probabilities generate similar average performance as the piece-rate contract. Subjects assigned the MePr treatment solved 87.9 correct summations on average and subjects assigned the HIPR treatment solved 83.7 correct summations on average, neither of which are statistically different from the average number of correct summations solved by subjects assigned to Piecerate.[20] These findings partially support Hypothesis 1, which accurately predicts that MePr induces similar performance as Piecerate, but incorrectly predicts that HiPr generates lower performance than Piecerate.

Among the treatments representing the probability contract, the LowPr generates greater average performance. This treatment generates 17% higher average performance than HiPr ($t(75.215) = 2.232, p = 0.014$), and 11% higher average performance than MePr

[18] A Wilcoxon-Mann-Whitney test also rejects the null hypothesis of no performance difference between Piecerate and LowPr ($z = 2.634, p < 0.01$)

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

[20] The t-tests of these comparisons are ($t(83) = 1.005, p = 0.159$) and ($t(82.44) = -0.386, p = 0.692$), respectively. Wilcoxon-Mann-Whitney tests of these comparisons yield ($z = 1.321, p = 0.186$) and ($z = -0.895, p = 0.3710$, respectively.

($t(79.575) = 1.478, p = 0.0716$).[21] Therefore, statistical inference using pairwise testing suggests that LowPr generates highest average performance, while the other three treatments produce similar performance.

An analysis based on regressions performed at the individual level leads to the same conclusions. I regress each subject's performance is regressed on treatment dummies, dummies that capture different shapes of the utility function as well as dummies that capture different shapes of the weighting function. Specifically, utility functions are classified as having linear, concave, convex, or mixed shape. Details of this classification are provided in Appendix F.[22] Also, the probability weighting functions are classified as displaying lower subadditivity (LS, henceforth) and/or upper subadditivity (US, henceforth). A weighting function with LS assigns larger decision weights to best-ranked outcomes than to middle-ranked outcomes. A weighting function with US assigns larger decision weights to worst-ranked outcomes than middle-ranked outcomes.[23] An alternative classification used for weighting functions indicates the strength of the possibility effect relative to the certainty effect. The variable "Possibility" takes a value of one if the possibility effect is stronger than the certainty effect and zero otherwise.[24] Details of the classifications used for probability weighting functions are provided in Appendix G.

Table 3 presents the regressions estimates. For all specifications, the coefficient associated to assignment to LowPr is significant and positive at the 5% significance level, which corroborates the aforementioned result that subjects assigned to that treatment display higher average performance than subjects assigned to Piecerate, the benchmark treatment of the regression. Similarly, the coefficient of LowPr is significantly higher than the estimate associated with HiPr ($F(1, 159) = 6.58$) and significantly higher than that associated to MePr ($F(1, 159) = 6.02$). Thus, among the studied contracts, the LowPr produces the highest performance.

An alternative explanation to these results is that LowPr generates higher performance because it circumvents income effects (See Azrieli et al. (2018) and Lee (2008)). In contrast, these effects are present in Piecerate and this is what demotivates subjects assigned to this treatment to perform the task, especially toward the last rounds of the experiment. This

[21] Wilcoxon-Mann-Whitney tests of these differences yield ($z = 1.96, p = 0.049$) and ($z = 1.035, p = 0.07$), respectively. In addition, the effect sizes of these differences are of 0.4805 standard deviations and 0.322 standard deviations, respectively. Both of which are significant at the 10 % level.

[22] In short, a variable $\Delta_j^{''} \equiv (x_j - x_{j-1}) - (x_{j-1} - x_{j-2})$ for $j = 2, 3, 4, 5, 6$, is constructed for each subject. A subject is classified as having linear utility if most values $\Delta_j^{''}$ are close to zero, concave utility if most values $\Delta_j^{''}$ are positive, convex utility if most values $\Delta_j^{''}$ are negative, and mixed utility otherwise.

[23] In short, a subject in the experiment exhibits LS when $w^{-1}\left(\frac{1}{6}\right) < w^{-1}\left(\frac{2}{6}\right) - w^{-1}\left(\frac{1}{6}\right)$. A subject exhibits US when $1 - w^{-1}\left(\frac{5}{6}\right) < w^{-1}\left(\frac{5}{6}\right) - w^{-1}\left(\frac{4}{6}\right)$.

[24] A subject has a possibility effect that is stronger than the certainty effect when $1 - w^{-1}\left(\frac{5}{6}\right) < w^{-1}\left(\frac{1}{6}\right)$.

Table 3: Regression of performance on treatments

	(1) Performance	(2) Performance	(3) Performance	(4) Performance
LowPr	16.739**	16.558**	16.001**	16.526**
	(7.090)	(7.508)	(7.532)	(7.589)
MePr	6.522	6.714	6.335	6.585
	(6.487)	(6.610)	(6.677)	(6.724)
HiPr	2.372	1.684	1.616	0.758
	(5.985)	(5.888)	(6.308)	(6.016)
Concave		14.359	15.067	15.090
		(9.401)	(9.529)	(9.681)
Convex		7.623	8.527	7.185
		(10.109)	(10.469)	(10.513)
Mixed		3.864	3.698	4.259
		(6.625)	(6.699)	(6.785)
US			0.904	
			(5.183)	
LS			2.924	
			(5.053)	
Possibility				4.901
				(7.637)
Certainty				7.062
				(7.791)
Constant	81.378***	79.819***	78.497***	74.667***
	(4.726)	(5.025)	(5.242)	(7.371)
R^2	0.045	0.062	0.065	0.064
Observations	172	172	172	172

Note: This table presents the estimates of the Ordinary Least Squares regression of the model $Performance_i = \gamma_0 + \gamma_1 LowPr + \gamma_2 MePr + \gamma_3 HiPr + Controls'\Lambda + \epsilon_i$, with $E(\epsilon | MepR, LowPr, HiPr, Controls) = 0$. "Performance" is the number of correctly solved sums solved by a subject in the first part of the experiment, "Piecerate", "LowPr", "MePr" and "HiPr" are binary variables that indicate if a subject was assigned to the treatment offering a piece-rate or probability contracts implemented with low, medium or high probability, respectively. "Piecerate" is the benchmark of the regression. 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.

line of thought also suggests that LowPr generates higher performance than MePr and HiPr because, by paying less rounds of the real-effort task, it is more effective in minimizing income effects. However, the data show that significant performance differences across the treatments emerge in the first round, when income effects are absent. A regression of performance in a given round on treatment dummies, round dummies, and relevant controls is estimated with standard errors clustered at the individual level. The estimates of this regression show that in the first round subjects assigned to LowPr achieve 1.6 higher average summations as compared to subjects in Piecerate ($p = 0.019$).[25]

While I find compelling evidence of performance differences performance across the contracts, the data on subjects' beliefs suggest that they are statistically indistinguishable across treatments. This finding implies that subjects in the experiment did not anticipate the motivational effect of the probability contract implemented with probability $p = 0.10$. Appendix E presents a detailed analysis of these data and a broader discussion of the implications of this result.

All in all, the data on performance in the real-effort task invalidate Hypothesis 0 and partially support Hypothesis 1. However, recall that Hypothesis 1 was structured around the common finding that individuals overweight all probabilities up to $p = 0.33$ and underweight all probabilities thereafter. Instead, the analyses of the data show that subjects in the experiment overweighted on average the probability $p = 0.10$ and evaluated approximately accurately the probabilities $p = 0.3$ and $p = 0.5$. In the next section, I show that subjects indeed display an average probability weighting function with such shape and, thus, that can accommodate these findings.

5.2. Probability weighting functions

In this subsection, I analyze the data of the second part of the experiment. These data feature the subjects' choices between pairwise lotteries, which were designed to elicit the subjects' utility and probability weighting functions. The analysis of these data show that subjects display an average probability weighting function with a strong inverse-S shape and with optimism. This shape induces a strong overweighting of small probabilities as well as moderate underweighting of large probabilities.

As explained in section 3, the second part of the experiment consisted of 11 decision sets. Decision sets 1 to 6 were designed to elicit the sequence $\{x_1, x_2, x_3, x_4, x_5, x_6\}$, representing the subjects' preferences over monetary consequences of similar magnitude as the stakes offered in the real-effort task. Different analyses of the data demonstrate that the majority of

[25]Subjects in LowPr also exhibit higher average performance as compared to subjects in MePr ($\chi(1)^2 = 2.21, p = 0.069,$ one tailed) and subjects in HiPr ($\chi(1)^2 = 5.04, p = 0.026$)

subjects exhibit linear utility functions over monetary consequences, which is in line with the findings of Wakker and Deneffe (1996) and Abdellaoui (2000), as well as with the critique put forward by Rabin (2000). Given this result and since the main focus of the paper is on probability weighting functions and their influence on the efficiency of the probability contact, I relegate the complete analysis of these data to Appendix E.

Decision sets 7 to 11 of the second part of the experiment were designed to elicit the sequence

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

with $p_{j-1} = \frac{j-1}{6}$ and $j = \{2, ..., 6\}$. These data are analyzed to examine how subjects evaluated probabilities. To that end, I perform regressions at the individual level that relate the elicited probabilities to the probability weights that they map. The rationale for using regressions as the main analysis of these data is that i) they provide a good indication of the average degree of probability distortion in the experiment, ii) the resulting estimates can be used to compare the degree of probability distortion in the experiment to those reported in previous studies, and iii) with the resulting estimates one can construct indexes of likelihood insensitivity and optimism, which according to Corollary 1 and Corollary 2 are critical to understand the determinants behind the efficiency of the probability contracts.[26] Alternative analyses of these data, including non-parametric analyses as well as analyses of the data performed at the individual level, are presented in Appendix G.

To perform the regressions, I assume the most well-known parametric functions in the literature. Specifically, I use the neo-additive probability weighting function (Chateauneuf et al., 2007), Tversky and Kahneman's (1992) probability weighting function, Prelec's (1998) two-parameter probability weighting function, and Goldstein and Einhorn's (1987) log-odds probability weighting function. Using different parametric functions ensures robustness, i.e. the results of this analysis do not stem from the underlying assumptions of a particular functional form.

Table 4 presents the regression estimates. Panel 1 presents the estimates of a truncated regression of the neo-additive functional, $w(p) = c + sp$.[27] The resulting estimates display $\hat{c} > 0$ and $\hat{c} + \hat{s} < 1$, which imply that subjects on average overweighted small probabilities and underweighted large probabilities.[28] Furthermore, $\hat{c}$ and $\hat{s}$ are larger and smaller, respectively than the estimates reported in Abdellaoui et al. (2011), suggesting that subjects in the

[26]Comparisons across studies must be taken with a grain of salt inasmuch as resulting differences cannot only be attributed to differences in preferences, but also to the different stakes and methods used to elicit risk preferences.

[27]The assumed truncation at the extremes, $w(0)$ and $w(1)$, provides the estimation with the flexibility to admit weighting functions with S-shape.

[28]These conclusions also hold when the regression is estimated without truncation.

experiment exhibited higher degrees of optimism toward risk as well as higher degrees of likelihood insensitivity.

A more traditional parametric representation of the probability weighting function was proposed by Tversky and Kahneman (1992), which relates probabilities and their associated weights through the non-linear function $w(p) = \frac{p^{\psi}}{(p^{\psi}+(1-p)^{(1-\psi)})^{\frac{1}{\psi}}}$. The second panel of Table 4 shows that the non-linear least squares method generates the estimate $\hat{\psi} = 0.59$, which is lower than those reported in previous studies. Specifically, classical experiments report estimates in the range 0.60-0.75 (Bleichrodt and Pinto, 2000, Abdellaoui, 2000, Wu and Gonzalez, 1996, Tversky and Kahneman, 1992). Therefore, subjects in my experiment display a weighting function with more severe overweighting of small probabilities.

A severe disadvantage of Tversky and Kahneman's (1992) weighting function is that both likelihood insensitivity and optimism/pessimism influence ψ, so their effect on probabilistic risk attitudes cannot be separated. To overcome such disadvantage, I also assumed the log-odds weighting function proposed by Goldstein and Einhorn (1987), $w(p) = \frac{\delta p^{g}}{\delta p^{g}+(1-p)^{g}}$, which can, up to some extent, separate these two components. The estimates of the non-linear least squares regression are presented in Panel 3. The magnitude of $\hat{g}$ indicates that the average weighting function has a strong inverse-S shape and the magnitude of $\hat{\delta}$ indicates optimism, on average. These coefficients are lower and higher, respectively, than those found in previous studies (Bruhin et al., 2010, Bleichrodt and Pinto, 2000, Abdellaoui, 2000, Gonzalez and Wu, 1999, Wu and Gonzalez, 1996, Tversky and Fox, 1995). Thus, subjects in the experiment had an average weighting function with more likelihood insensitivity and more optimism than previously documented.

Lastly, I also assume Prelec (1998)'s probability weighting function with two parameters, $w(p) = exp(-\beta(-ln(p)))^{\alpha}$. This parametric functional also separates, up to some extent, optimism from likelihood insensitivity. Panel 4 presents the estimates of a non-linear least squares regression. The estimate $\hat{\alpha}$, which is statistically lower than one, entails that the average probability function has a strong inverse-S shape. Moreover, the estimate $\hat{\beta}$, which is also statistically lower than one, entails that subjects display optimism on average. Previous estimations of this probability weighting function report larger values of α and β (Murphy and Ten Brincke, 2018, Haridon et al., 2018, Fehr-duda, 2012, Abdellaoui et al., 2011, Bleichrodt and Pinto, 2000). Hence, subjects display an average probability weighting function with a stronger inverse-S shape and more optimism than documented by previous studies.

Altogether, these estimations lead to the conclusion that subjects display an average probability weighting function with a strong inverse-S shape and optimism. The coexistence of these properties produces pattern of probability distortion whereby small probabilities are strongly overweighted and large probabilities are moderately underweighted. Such

Table 4: Parametric estimates of the weighting function

	(1)	(2)	(3)
Panel 1: Neo-additive (truncated)			
$w(p) = c + s * p$			
$\hat{c}$	0.194 ***(0.021)	0.228*** (0.024)	0.155*** (0.024)
$\hat{s}$	0.566*** (0.035)	0.463*** (0.037)	0.686*** (0.044)
Log-Likelihood	220.288	75.200	166.842
Panel 2: Tversky & Kahneman (1992)			
$w(p) = \frac{p^{\psi}}{(p^{\psi}+(1-p)^{(1-\psi)})^{\frac{1}{\psi}}}$			
$\hat{\psi}$	0.598*** (0.016)	0.597*** (0.012)	0.785*** (0.037)
Adj. R^2	0.838	0.827	0.866
Panel 3: Goldstein and Einhorn (1987)			
$w(p) = \frac{\delta p^g}{\delta p^g+(1-p)^g}$			
$\hat{g}$	0.281***(0.025)	0.196***(0.027)	0.426*** (0.042)
$\hat{\delta}$	0.921***(0.020)	0.892*** (0.029)	0.982***(0.032)
Adj. R^2	0.863	0.845	0.888
Panel 4: Prelec (1998)			
$w(p) = exp(-\beta(-ln(p)))^{\alpha}$			
$\hat{\alpha}$	0.284***(0.025)	0.143*** (0.025)	0.357*** (0.033)
$\hat{\beta}$	0.841***(0.015)	0.596***(0.024)	0.944*** (0.019)
Adj. R^2	0.864	0.907	0.851
N	860	304	550
Ref.Point	No/Zero	Beliefs	Beliefs
Domain	No/Gains	Gains	Losses

Note: This table presents the estimates of the probability weighting function when parametric estimates are assumed. Panel 1 presents the maximum likelihood estimates of the equation $w(p) = c + s(p)$ when truncation at $w(p) = 0$ and at $w(p) = 1$ is assumed. Panel 2 presents the non-linear least squares estimation of the function $w(p) = \frac{p^{\psi}}{(p^{\psi}+(1-p)^{(1-\psi)})^{\frac{1}{\psi}}}$. The third panel presents the non-linear least squares estimates of the parametric form $w(p) = \frac{\delta p^g}{\delta p^g+(1-p)^g}$. The last panel presents the non-linear least squares estimates of the function $w(p) = exp(-\beta(-ln(p)))^{\alpha}$. The first column in all the panels presents the estimates when all the data are used. The second and third columns present the estimations when it is assumed that Beliefs is the reference point and only data for the domain of gains and the domain of losses, respectively, is used for the estimations. 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.

pattern accommodates the findings of the first part of the experiment, namely that LowPr generates higher output than Piecerate, and that HiPr, MePr, and Piecerate produce similar performance. In the next section, I conclusively demonstrate that the elicited probability weighting functions of subjects explain the treatment effects. In addition, the next section investigates the contribution of optimism and likelihood insensitivity to the treatment effects.

Finally, I performed the aforementioned estimations accounting for the possibility that subjects have CPT preferences with an expectations-based reference point (Koszegi and Rabin, 2006). In particular, I assumed the reference point to be the monetary equivalent of each subject's belief in the first part of the experiment. Outcomes above this reference point belong to the domain of gains, while outcomes below this reference point belong to the domain of losses. I perform separate regressions for each domain. Estimates of these regressions are presented in columns (2) and (3) of Table 4. All in all, I find that for all considered parametric forms and for both domains, subjects have weighting functions with inverse-S shapes and optimism. As a consequence, the results presented in this section are robust to the assumption that subjects' preferences can be represented by CPT with an expectations-based reference point.

5.3. Likelihood insensitivity and the treatment effect

This subsection reconciles the results of the two parts of the experiment. First, I present empirical evidence supporting Hypothesis 2. That is, I demonstrate that the higher average performance of subjects assigned to LowPr is caused by their tendency to overweight small probabilities. Second, I show that likelihood insensitivity, alone, explains the treatment effects documented in Section 5. This result validates Hypothesis 4.

To empirically verify the validity of Hypothesis 2, I extend the statistical models presented in Table 3 by including interactions between the variable indicating assignment to LowPr and variables that capture the shape of the probability weighting function of a subject. I focus on variables that indicate whether a subject exhibits a weighting function with overweighting of small probabilities. Specifically, I focus on the variables LS, Possibility, and $\text{Overweight}_{p=\frac{1}{6}}$. The last variable takes a value of one if a subject overweights the probability $p = \frac{1}{6}$ and zero otherwise. The other two variables were already described in Section 5.1. The variables relate in the following way: a subject for who LS takes a value of one, necessarily overweights the probability $p = \frac{1}{6}$ and might exhibit a possibility effect that is stronger than the certainty effect. Similarly, a subject for who Possibility takes a value of one surely overweights $p = \frac{1}{6}$ and is likely to exhibit LS.

I first consider the binary variable LS. Column (1) in Table 5 presents the OLS estimates of the extended regression. I find that subjects assigned to LowPr who have weighting functions with lower subadditivity display an average performance level that is significantly higher than that of subjects in Piecerate. In contrast, subjects assigned to LowPr with weighting functions without lower subadditivity display an average performance level that is statistically indistinguishable to that of subjects in Piecerate. Thus, only subjects with a weighting function assigning larger decision weights to small probabilities relative to the weights assigned to medium-ranged probabilities display higher performance levels when assigned to LowPr and, as a result, exhibit (pronounced) treatment effects.

Next I investigate the role of the variable $\text{Overweight}_{p=\frac{1}{6}}$ in explaining the treatment effects. Note that subjects who overweight $p = \frac{1}{6}$ necessarily overweight $p = \frac{1}{10}$ and should, in light of the theoretical model, display a treatment effect when assigned to LowPr. The estimates of the model, presented in column (2) of Table 5, confirm this conjecture. In particular, subjects working under LowPr who overweighted the probability $p = \frac{1}{6}$ exhibit higher average performance than subjects in Piecerate. Moreover, these subjects also exhibit a steeper treatment effect than that of subjects who were assigned to LowPR but who did not overweight the probability $p = \frac{1}{6}$.

Furthermore, column (3) in Table 5 presents the results of a regression that evaluates the role of Possibility in explaining the treatment effects. The estimates entail that subjects assigned to LowPr and for who Possibility takes a value of one display a significant treatment effect. Instead, subjects assigned to LowPr and who do not exhibit the possibility effect do not display performance differences as compared to subjects in Piecerate. This finding suggests that subjects with concave weighting functions or strong inverse-S shaped weighting functions, who are more likely to exhibit a strong possibility effect, attain higher performance levels when assigned LowPr.

To summarize, the results presented in Table 5 provide robust empirical evidence that subjects with weighting functions that induce overweighting of small probabilities display pronounced treatment effects. These findings support Hypothesis 2.

We are now in a position to investigate the influence of likelihood insensitivity and optimism on performance. To that end, I first classify subjects according to their degree of likelihood sensitivity and optimism. Following Wakker (2010) and Abdellaoui et al. (2011), I estimate for each subject, i, the following neo-additive functional:

$$w(p_{ij}) = c_i + s_i p_{ij} + \epsilon_i.$$

The magnitude of the estimate $\hat{s}_i$ indicates subject's i sensitivity to probabilities. If $\hat{s}_i < 1$, this subject is not sufficiently responsive to changes in probabilities and is classified as likelihood insensitive. Instead, if $\hat{s}_i > 1$, this subject is too sensitive to changes in probabilities and is classified as likelihood sensitive. I find that 102 subjects in my sample are likelihood insensitive and 61 subjects are classified as likelihood sensitive. Importantly, the degree of likelihood insensitivity and that of likelihood sensitivity are balanced across treatments.

In addition, the magnitude of $\hat{c}_i$ and that of $\hat{c}_i + \hat{s}_i$ determine the degree of optimism of subject i. Whenever $\hat{c}_i > 0$ and $\hat{c}_i - \hat{s}_i < 1$, the subject assigns large weights to best-ranked outcomes and small decision weights to worst-ranked outcomes, and, as a consequence, exhibits optimism. Alternatively, if $\hat{c}_i < 0$ and $\hat{c}_i - \hat{s}_i > 1$, the subject exhibits pessimism. I find that 80 subjects in my sample display optimism while 32 subjects display pessimism. Again, the degrees of optimism and pessimism are balanced across treatments.

Binary variables capturing these classifications are added to the regressions presented in Table 3. I also include interactions between assignment to LowPr and the variables indicating whether a subject is likelihood insensitive and whether a subject exhibits optimism. These interactions allow me to evaluate the strength of the treatment effect among likelihood insensitive and/or optimistic subjects.

The resulting estimates are presented in columns (1) and (2) of Table 6. All in all, I find empirical support for Hypothesis 4. Specifically, I find that likelihood insensitive subjects assigned to LowPr display higher average performance as compared to subjects assigned to Piecerate. In contrast, subjects assigned to LowPr and who were not classified as likelihood insensitive did not exhibit performance differences with respect to the baseline treatment. These findings support the theoretical result that likelihood insensitivity ensures the efficiency of probability contracts when they are implemented with a small probability. In addition, the data show that subjects displaying optimism and who were assigned to LowPr exhibit average performance levels that are statistically indistinguishable from those of subjects in Piecerate. Therefore, optimism, alone, is unable to explain the treatment effects documented in Section 5. This result invalidates Hypothesis 3.

For the sake of robustness, I also estimate the parameters of the weighting functions proposed by Prelec (1998) and Goldstein and Einhorn (1987) for each subject. As mentioned before, these functions contain, each, two parameters. One parameter mainly influences likelihood insensitivity and the other mainly influences optimism. On the basis of the magnitude of these parameters, I classify subjects according to their optimism towards risk as well as according to their sensitivity toward probabilities. I include these alternative classifications in a regression to evaluate the strength of the treatment effects among likelihood insensitive subjects and among optimistic subjects.

Columns (2) and (3) in Table 6 present the results of the regressions. Altogether, the regression estimates corroborate the aforementioned results and, thus, the empirical validity of Hypothesis 4 but not that of Hypothesis 3. In particular, I find that likelihood insensitivity explains the documented performance differences between LowPr and Piecerate. Only subjects classified as likelihood insensitive and assigned to LowPr exhibit higher average performance levels as compared to subjects in Piecerate. Moreover, optimistic subjects assigned to LowPr do not exhibit performance differences with respect to subjects in the benchmark treatment.

6. Discussion and Conclusion

This paper introduced a novel type of stochastic contracts designed to take advantage of the behavioral regularity that individuals distort probabilities. A theoretical framework and a laboratory experiment demonstrated that these contracts generate higher labor supply than standard piece rates. However, to achieve this result the principal is required to offer them with a small probability that the performance-contingent payment realizes. This implementation of the proposed contracts induces risk seeking in agents and, as a consequence, a preference for risky compensation schemes. I show that the agents' insensitivity to likelihoods, the cognitive component of probability distortion, guarantees this result.

In terms of its implementation, the contract can be brought to practice in multiple ways.

- The most straightforward implementation of the proposed contracts is through incentive schemes that offer monetary bonuses contingent on the achievement of a milestone. In particular let the contract offered to the agent be $V = (B + A, p; B, 1 - p)$ where B is a fixed salary and A is a bonus that is paid when the agent attains a target or goal $0 < g < \bar{y} = 1$. This setting relaxes the assumption that θ is known by the agent. Instead, let $\theta \sim Unif[0, 1]$ and this information is known to both agent and principal. The agent with RDU preferences maximizes $U = \int_0^1 b(B) - c(y, \theta)w(f(\theta))d\theta + \int_0^g b(A)w(f(\theta))d\theta = b(B) + w(1 - g)b(A) - c(y, \theta)$. Thus, the principal can choose a challenging target g, entailing a low probability of achievement, and combine it with a considerably large monetary bonus A. Such *long-shot* incentive scheme yields, as the findings of this paper show, higher performance than if the agent were offered the monetary equivalent of the contract using piece-rates.

- Another practical application of these contracts considers compensation plans with option stocks. At time $t = 0$, the principal can offer a contract $V = B + n_0$ consisting of a fixed salary B and n_0 options. The option stocks have maturity $t = T$ and strike price K. The random price of the stock is P_T and is drawn from the continuous probability distribution $F(P_T|y)$. Also, assume weak first order stochastic dominance, that is $F(P_T|y') \geq F(P_T|y'')$ for any $\bar{y} > y'' > y' > 0$. This means that higher output entails a higher or equal likelihood that the price of the stock is higher. When

Table 5: The influence of probability overweighting on the treatment effects

	(1) Performance	(2) Performance	(3) Performance
LowPr*Mechanism	29.055**	17.418*	21.821**
	(12.056)	(10.302)	(9.822)
Mechanism	1.834	3.031	0.089
	(7.380)	(6.005)	(7.394)
LowPr	7.248	17.459**	2.745
	(7.538)	(8.601)	(7.848)
MePr	6.582	6.543	6.852
	(6.748)	(6.577)	(6.760)
HiPr	1.320	2.067	0.330
	(6.235)	(5.985)	(5.967)
US	4.654		
	(6.275)		
BOTH	-9.950		
	(10.072)		
Certainty			6.233
			(7.222)
Concave	16.373*	14.570	15.431
	(9.225)	(9.460)	(9.735)
Convex	10.449	7.656	8.525
	(12.795)	(9.999)	(8.363)
Mixed	3.740	4.064	4.465
	(6.854)	(6.749)	(6.851)
Constant	79.035***	78.899***	77.935***
	(5.115)	(5.471)	(6.793)
Mechanism	LS	$\text{Overweight}_{p=\frac{1}{6}}$	Possibility
R^2	0.089	0.063	0.08
Observations	172	172	172

Note: This table presents the estimates of the Ordinary Least Squares regression of the model $Performance_i = \gamma_0 + \gamma_1 LowPr + \gamma_2 LowPr * Mechanism + \gamma_3 Mechanism + \gamma_4 MePr + \gamma_5 MePr + \gamma_6 HiPr + Controls'\Gamma + \epsilon_i$, with $E(\epsilon_i|MePr, LowPr, HiPr, Controls, Mechanism) = 0$. "Performance" is the number of correctly solved sums solved by a subject in the first part of the experiment, "Piecerate", "LowPr", "MePr" and "HiPr" are binary variables that indicate if a subject was assigned to the treatment offering a piece-rate or probability contracts implemented with low, medium or high probability, respectively. "Piecerate" is the benchmark of the regression. In column (1) Mechanism is equal to "LS" a binary variable that takes a value of one if a a subject has a weighting function with lower subadditivity and zero otherwise. In column (2) Mechanism is equal to "$\text{Overweight}_{p=\frac{1}{6}}$" a binary variable that takes a value of one if a subject overweights the probability $p = \frac{1}{6}$. In column (3) Mechanism is equal to "Possibility " a binary variable takes a value of one if a subject has a weighting function with the possibility effect being stronger than the certainty effect and zero otherwise. 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 6: The influence of likelihood insensitivity and optimism on the treatment effects

	(1) Performance	(2) Performance	(3) Performance	(4) Performance
LowPr*Likelihood ins.		24.182**	27.209**	23.654**
		(10.625)	(12.607)	(11.947)
LowPr*Optimism		-3.112	17.081	-3.594
		(9.958)	(16.404)	(32.031)
LowPr	15.636**	6.812	10.249	5.154
	(6.522)	(11.115)	(13.613)	(11.850)
MePr	4.912	4.657	4.699	4.243
	(6.501)	(6.525)	(6.616)	(6.696)
HiPR	0.398	0.100	1.061	1.492
	(6.324)	(6.350)	(6.410)	(6.373)
Likelihood ins.	12.809**	12.155*	11.167	7.641
	(6.277)	(7.150)	(9.458)	(9.620)
Optimism	-9.574*	-12.281*	11.532	-14.163
	(5.589)	(6.743)	(14.330)	(30.600)
Pessimism	11.900	10.974	10.370	-13.416
	(7.288)	(7.355)	(14.029)	(30.561)
Mixed	4.037	3.698	6.144	5.299
	(6.764)	(6.789)	(7.065)	(7.047)
Convex	6.374	4.344	10.026	10.645
	(17.822)	(18.033)	(18.842)	(18.684)
Concave	12.721	12.417	15.238*	13.829
	(8.673)	(8.700)	(8.848)	(8.815)
Constant	75.564***	77.548***	56.395***	83.116***
	(6.304)	(6.579)	(16.747)	(31.805)
R^2	0.101	0.108	0.094	0.100
Observations	172	172	172	172
Likelihood ins.	$\hat{s} < 1$	$\hat{s} < 1$	$\hat{\alpha} < 1$	$\hat{g} < 1$
Optimism	$\hat{c} > 0$ and $\hat{s} - \hat{c} < 1$	$\hat{c} > 0$ and $\hat{s} - \hat{c} < 1$	$\hat{\beta} < 1$	$\hat{\delta} > 1$

Note: This table presents the estimates of the Ordinary Least Squares regression of the model $Performance_i = \gamma_0 + \gamma_1 LowPr{*}Likelihoodins. + \gamma_2 LowPr{*}Optimism + \gamma_3 LowPr + \gamma_4 MePr + \gamma_5 HiPr + \gamma_6 Likelihoodins. + \gamma_7 Optimism + Controls'\Gamma + \epsilon_i$, with $E(\epsilon_i | MePr, LowPr, HiPr, Piecerate, Optimism, Likelihoodins., Controls, Mechanism) = 0$. "Performance" is the number of correctly solved sums solved by a subject in the first part of the experiment, "Piecerate", "LowPr", "MePr" and "HiPr" are binary variables that indicate if a subject was assigned to the treatment offering a piece-rate or probability contracts implemented with low, medium or high probability, respectively. "Piecerate" is the benchmark of the regression. "Likelihood ins." is a binary variable that takes a value of one if the subject is likelihood insensitive and zero otherwise. "Optimism' is a binary variable that takes a value of one if the subject displays optimism and zero otherwise. 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.

offered this compensation scheme, the agent with RDU preferences maximizes $U = b(B) + \int_0^{\bar{y}} b(n_0 max(P_T - K, 0))dw(F(P_T|y)) - c(y, \theta)$. First, note that as long as the probability that the constituting stock options yield dividends is overweighted, these compensation packages are accepted by the agent. It has been shown that this is true even in the extreme case in which $F(P_T|y)$ is right-skewed (Spalt, 2013). Moreover, these compensation packages generate higher labor supply when the agent overweights the probability of dividend yield. That is because the contribution of the agent's output to the probability of dividend yield will be jointly overweighted with the probability of dividend yield, which inflates the perceived benefits of supplying more output.

- Another implementation of the contract is auditing. Consider a setting in which the principal can choose among different auditing technologies. On one hand, more advanced technologies, and also more expensive ones, allow her to constantly observe the agent's performance. With these expensive technologies the principal can exactly link the agent's compensation to his performance on the task. Let the performance-pay contract in the case of perfect auditing be $W = ay$, where $a > 0$. On the other hand, cheaper technologies give rise to stochastic processes and, as a consequence, that the exact labor supply of an agent is observed with some probability $p \in (0, 1)$. Let the performance pay of those schemes be $V = (p, Ay; 1 - p, B)$, so that B is paid when no auditing takes place and Ay determines the agent's payment in the case of auditing. The results of this paper show that when facing such trade off, the principal can choose the cheaper technology and combine it with a large performance-contingent pay— say, at least an $\tilde{A} > 0$ such that $\tilde{A}py + (1 - p)B = ay$ for any $y > 0$, $B > 0$, and $a > 0$. The results of the paper show that if an agent with RDU preferences overweights the probability of being audited, p, he will be motivated him to supply more output than under the perfect auditing technology.

The present study has several limitations that open avenues for future research. First, even though there are obvious advantages of using controlled laboratory environments, such as the possibility of offering different incentive contracts to similarly skilled individuals to evaluate their effectiveness and the possibility of perfectly observing labor supply, these advantages come at the cost of external validity. A more general understanding of the incentives included in the probability contracts requires performing complementary tests in situations that feature longer working periods, more powerful monetary incentives, more meaningful tasks, and a more natural setup for subjects. Field experiments designed to incorporate these characteristics while retaining the possibility of establishing causal inference are ideal to examine the external validity of the results reported in this paper.

Second, I consider a setting of risk. A less restrictive version of these contracts considers ambiguity, i.e. events realizing with unknown probabilities. Adapting the contracts to ambiguity would ease upon their implementation. That is because the principal does not need to implement the contract using events that realize with known probability. These events are rare and at times need to be artificially implemented, which could restrict the credibility of the principal's commitment. Moreover, recent research suggests that when making decisions under uncertainty, individuals display more insensitivity toward ambiguity than toward risk (Baillon et al., 2018, Abdellaoui et al., 2011). Therefore, the effectiveness of the proposed contracts can be further enhanced when the favorable outcome is indexed to rare events of unknown probability.

References

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

Abdellaoui, M. (2002). A Genuine Rank-Dependent Generalization of the von Neumann-Morgenstern Expected Utility Theorem. *Econometrica*, 70(2):717–736.

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

Abdellaoui, M., Haridon, O. L., and Zank, H. (2010). Separating curvature and elevation : A parametric probability weighting function. *Journal of Risk and Uncertainty*, 41:39–65.

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

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

Baillon, A., Huang, Z., Selim, A., and Wakker, P. P. (2018). Measuring Ambiguity Attitudes for All (Natural) Events. *Econometrica*, 86(5):1839–1858.

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*, 15:418–443.

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.

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

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

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

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

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

Chew, S. H., Karni, E., and Safra, Z. (1987). Risk aversion in the theory of expected utility with rank dependent probabilities. *Journal of Economic theory*, 42(2):370–381.

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

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

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

De La Rosa, E. L. (2011). Overconfidence and Moral Hazard. *Games and Economic Behavior*, 73(2):429–451.

de Meza, D. and Webb, D. C. (2007). Incentive design under loss aversion. *Journal of the European Economic Association*, 5(1):66–92.

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

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

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

Gervais, S., Heaton, J. B., and Odean, T. (2011). Overconfidence , Compensation Contracts , and Capital Budgeting. *The Journal of Finance*, 66(5):1735–1777.

Gibbons, R. (1998). Incentives in Organizations. *The Journal of Economic Perspectives*, 12(4):115–132.

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

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

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

Green, J. and Laffont, J. J. (1977). Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods. *Econometrica*, 45(2):427–438.

Guesnerie, R. (1995). *A Contribution to the Pure Theory of Taxation*. Cambridge University Press.

Haller, H. (1985). The principal-agent problem with a satisficing agent. *Journal of Economic*

Behavior & Organization, 6(4):359–379.

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

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

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

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

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.

Kaur, S., Kremer, M., and Mullainathan, S. (2015). Self-Control at Work. *Journal of Political Economy*, 123(6):1227–1277.

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.

Laffont, J. J. and Tirole, J. (1993). *A Theory of Incentives in Procurement and Regulation*. MIT-Press, Cambridge(Mass.).

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.

Mas, A. and Moretti, E. (2009). Peers at Work. *American Economic Review*, 99(1):112–45.

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

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

Myerson, R. (1981). Optimal Auction Design. *Mathematics of Operations Research*, 6(1):58–73.

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.

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

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

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

Santos-Pinto, L. (2008). Positive Self-Image and Incentives in Organisations. *Economic Journal*, 118(531):1315–1332.

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.

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

Spinnewijn, J. (2013). Insurance and Perceptions: How to Screen Optimisms and Pessimists. *Economic Journal*, 123(569):606–633.

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

Tversky, A. and Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, 211(4481):453–458.

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

Tversky, A. and Wakker, P. P. (1995). Risk Attitudes and Decision Weights. *Econometrica*, 63(6):1255–1280.

Wakker, P. P. (1994). Separating Marginal Utility and Probabilistic Risk Aversion. *Theory and Decision*, 36(1):1–44.

Wakker, P. P. (2001). Testing and Characterizing Properties of Nonadditive Measures through Violations of the Sure-Thing Principle. *Econometrica*, 69(4):1039–1059.

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

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

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

Yaari, M. E. (1987). The Dual Theory of Choice under Risk. *Econometrica*, 55(1):95–115.

Appendix A: Proofs

Proposition 1

Proof. Suppose that $y^{**} \leq y^{*}$. For an agent with arbitrary ability $\theta = \tilde{\theta}$, and since $c_y(\theta, y) > 0$ and $c_{yy}(\theta, y) > 0$, it must be that $c_y(y^{**}, \tilde{\theta}) \leq c_y(y^{*}, \tilde{\theta})$. Using equations (2) and (7), it is possible to rewrite this inequality as:

$$b_y\left(\frac{ay^{**}}{p}\right) \leq b_y(ay^{*}). \tag{9}$$

Since $\frac{ay}{p} > ay$ for any $p \in (0,1)$ and any $y \in (0, \bar{y}]$, then $b_y(\frac{ay}{p}) > b_y(ay)$ if $b_{yy}(y,p) \geq 0$, which contradicts (9). This contradiction is also reached when the strict inequality $y^{**} < y^{*}$ is considered. To see how, let $y^{**} = 1$ and $y^{*} = 2$. The inequality $\frac{a}{p} < 2a$, corresponding to (9), does not hold for $p \in (0, \frac{1}{2})$. Then it must be $b_y(\frac{a}{p}) > b_y(2a)$ if $b_{yy}(y,p) \geq 0$ and $p \in (0, \frac{1}{2})$, which contradicts (9). In general, it is possible to find sufficiently small p ensuring a contradiction of $y^{**} < y^{*}$ whenever $b_{yy}(y,p) \geq 0$. Hence, it must be that $y^{**} > y^{*}$ if $b_{yy}(y,p) \geq 0$.

Following a similar procedure it is possible to demonstrate that $y^{**} \geq y^{*}$ cannot hold if $b_{yy}(\cdot) < 0$ and that it must be that $y^{**} < y^{*}$ if $b_{yy}(y,p) < 0$.

■

Proposition 2

Proof. First, consider the interval $p \in (\tilde{p}, 1)$ and $b_y(y,p) \leq$. Since $\frac{w(p)}{p} < 1$ whenever $p \in (\tilde{p}, 1)$, then $\frac{w(p)}{p} b_y(\frac{ay}{p}) < b_y(ay)$ for any $y \in (0, \bar{y}]$ and $a > 0$. Moreover, the inequality $\frac{w(p)}{p} b_y(\frac{a(y+\epsilon)}{p}) < b_y(ay)$ holds for arbitrary $\epsilon > 0$. Then it must be that $y_R^{**} < y^{**}$ if $p \in (\tilde{p}, 1)$ and $b_y(y,p) \leq$. Suppose instead that $y_R^{**} > y^{**}$. For an agent with arbitrary ability $\theta = \tilde{\theta}$, and since $c_y(\theta, y) > 0$ and $c_{yy}(\theta, y) > 0$, then it must be that $c_y(y_R^{**}, \tilde{\theta}) > c_y(y^{**}, \tilde{\theta})$. I rewrite this inequality using equations (7) and (8) as $\frac{w(p)}{p} b_y(\frac{ay_R^{**}}{p}) > b_y(ay^{**})$. Note that the previous inequality contradicts the conclusion that $\frac{w(p)}{p} b_y(\frac{a(y+\epsilon)}{p}) < b_y(ay)$ for arbitrary $\epsilon > 0$. Then it must be that $y_R^{**} < y^{**}$ if $p \in (\tilde{p}, 1)$ and $b_{yy}(y,p) \leq 0$. Given that Proposition 1 demonstrates that $y^{**} < y^{*}$ if $b_{yy}(y,p) < 0$ for any $p \in (0,1)$, then the ranking $y_R^{**} < y^{**} < y^{*}$ holds if $p \in (\tilde{p}, 1)$ and $b_{yy}(y,p) < 0$.

Next, consider the interval $p \in (0, \tilde{p}]$ and $b_y(y,p) \leq$. Since $\frac{w(p)}{p} \geq 1$ whenever $p \in (0, \tilde{p}]$, then both $\frac{w(p)}{p} b_y(\frac{ay}{p}) < b_y(ay + \epsilon)$ and $\frac{w(p)}{p} b_y(\frac{ay}{p}) \geq b_y(ay + \epsilon)$ are feasible for any $y \in (0, \bar{y}]$. Therefore, the ranks $y_R^{**} \geq y^{*} > y^{**}$ and $y^{*} > y_R^{**} \geq y^{**}$ are feasible if $p \in (0, \tilde{p})$ and $b_{yy}(y,p) \leq 0$. Suppose that $y^{*} \geq y_R^{**}$. For an agent with arbitrary ability $\theta = \tilde{\theta}$, and due to

$c_{yy}(\tilde{\theta}, y) > 0$ and $c_y(\tilde{\theta}, y) > 0$, it must be that $c_y(y^*, \tilde{\theta}) \geq c_y(y_R^{**}, \tilde{\theta})$. Equations (2) and (8) can be used to rewrite this inequality as:

$$b_y(ay^*) \geq \frac{w(p)}{p} b_y\left(\frac{ay_R^{**}}{p}\right). \tag{10}$$

The validity of the inequality in (10) is first studied at the extremes of the interval, that is at $p \to 0^+$ and at $p = \tilde{p}$. Since $\frac{w(p)}{p} = 1$ at $p = \tilde{p}$, $ay < \frac{a(y+\epsilon)}{\tilde{p}}$ holds for any $y \in (0, \bar{y}]$ and arbitrary $\epsilon \geq 0$. This together with $b_{yy}(\cdot) \leq 0$ yields that Equation (10) holds at $p = \tilde{p}$.

Since $\lim_{p\to 0^+} \frac{w(p)}{p} b_y\left(\frac{ya}{p}\right)$ yields an indeterminate form due to $\lim_{p\to 0^+} b_y\left(\frac{ay}{p}\right) = 0$, I use L'Hospital's rule to evaluate the limit as follows:

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

The last equality is due to $\lim_{p\to 0^+} \frac{w(p)}{p} = \infty$ and $b_{yp}(\frac{ay}{p}) > 0$. Note that $\lim_{p\to 0^+} w_p(p)p$ is also an indeterminate form. I evaluate this expression with L'Hospital's rule as follows:

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

Where the last inequality holds due to $|w_{pp}(p)| < \infty$. Therefore, the inequality in (10) does not hold as $p \to 0^+$ and it must be that $y^* < y_R^{**}$ whenever $p \to 0^+$.

Next, I study the behavior of the right hand side of (10) with changes of p in $p \in (0, \tilde{p})$. To that end, I compute the derivative $\frac{\partial\left(\frac{w(p)}{p} b_y\left(\frac{ay}{p}\right)\right)}{\partial p} = \frac{(pw_p(p) - w(p))}{p^2} b_y(\frac{ay}{p}) - \frac{w(p)}{p^3 ay} b_{yp}(\frac{ay}{p})$. Suppose that $\frac{\partial\left(\frac{w(p)}{p} b_y\left(\frac{ay}{p}\right)\right)}{\partial p} > 0$, then changes in the considered interval satisfy:

$$\frac{w_p(p)}{\frac{w(p)}{p}} > 1 + \frac{b_{yp}(\frac{ay}{p})\frac{ay}{p}}{b_y(\frac{ay}{p})}. \tag{11}$$

For $p \in (0, \tilde{p})$ and $\tilde{p} > \hat{p}$, the left hand side of (11) decreases in $p \in (0, \tilde{p})$ since $\frac{\partial\left(\frac{w(p)}{p}\right)}{\partial p} > 0$ and $w_{pp}(p) < 0$. This implies that the largest value that $\frac{pw_p(p)}{w(p)}$ attains is at $p \to 0^+$, this value is evaluated with L'Hospital's rule:

$$\lim_{p\to 0^+} \frac{pw_p(p)}{w(p)} = \lim_{p\to 0^+} \frac{w_{pp}(p)p}{w_p(p)} + 1 = 1$$

Hence, the inequality in (11) is cannot hold and instead it must be that $\frac{\partial\left(\frac{w(p)}{p} b_y\left(\frac{ay}{p}\right)\right)}{\partial p} < 0$ in

$p \in (0, \tilde{p})$.

Since $\lim_{p \to 0^+} \frac{w(p)}{p} b_y(\frac{ya}{p}) = \infty$, $b_y(ay^*) > b_y(\frac{ay_R^{**}}{p})$ at $p = \tilde{p}$, $\frac{\partial\left(\frac{(w(p)}{p} b_y(\frac{ay}{p})\right)}{\partial p} < 0$, and the continuity of $w(p)$ and $b(y,p)$, the existence of a $p^* \in (0, \tilde{p})$ such that equation (10) holds with equality is guaranteed. When $\tilde{p} < \hat{p}$, p^* is unique since the left hand side of (11) is always decreasing over the interval $p \in (0, \tilde{p})$. Whenever $\tilde{p} \geq \hat{p}$, the existence of p^* is also guaranteed, since $w(p)$ is always first concave and then convex, but p^* is not unique since the left hand side of (11) is not decreasing in the interval $p \in (\hat{p}, \tilde{p})$. For the latter case, I refer to p^* as the smallest possible value that makes (10) bind. Hence, $y^* \geq y_R^{**}$ cannot hold if $p \in (0, p^*)$ and instead it must be that $y_R^{**} > y^*$.

■

Lemma 1

Proof. I first show that optimism is a necessary condition for $\hat{p} > 0$. Let $O(p)$ a weighting function with the properties given in Assumption 3 and $\hat{p} = \{0, 1\}$. Let $L(p)$ a weighting function with the properties given in Assumption 3 and $\hat{p} = \tilde{p} = 0.5$. Suppose that $w(p) \equiv O(p)$. Then, by construction, $\hat{p} = \{0, 1\}$ and the existence of a non-empty interval $(0, \hat{p})$ is guaranteed if $O(p)$ displays $\hat{p} = 1$, that is in the presence of optimism.

Next, I show that likelihood insensitivity is a sufficient condition to generate a non-empty interval $(0, \hat{p})$. Let $w(p) \equiv O(p) + L(p)$, with $O(p)$ such that $\hat{p} = 0$. Note that $\lim_{p \to 1^-} w_p(p) = \lim_{p \to 1^-} L_p(p) + O_p(p) = \infty$ since $\lim_{p \to 1^-} L_p(p) = \infty$ and $\lim_{p \to 1^-} O_p(p) = \infty$. In addition, for $p \in (0.5, 1]$, then $w_{pp}(p) > 0$ since $L_{pp}(p) + O_{pp}(p) > 0$ and at exactly $p = 0.5$ we have $w_p(0.5) = L_p(0.5) + O_p(0.5) = 0 + k$ for some $k < \infty$ due to $\lim_{p \to^- 1} O_p(1) = \infty$, $\lim_{p \to 0^+} O_p(p) = 0$, $O_{pp}(p) < 0$ and the continuity of $O_p(p)$ that guarantee $|O_p(0.5)| < \infty$. Moreover, $\lim_{p \to 0^+} w_p(p) = \lim_{p \to 0^+} L_p(p) + O_p(p) = \infty$. Hence, due to the continuity of $L_p(p)$ and $O_p(p)$, it must be that $L_{pp}(p) + O_{pp}(p) < 0$ for some probabilities in the interval $p \in (0, 0.5]$. This implies the existence of an interior fixed-point $\hat{p} \in (0, 1)$ whereby $w_{pp}(\hat{p}) = 0$.

Following a similar procedure it is straightforward to show the existence of $\hat{p} \in (0, 1)$ when $O(p)$ is a weighting function with the properties of Assumption 3 and $\hat{p} = 1$. Therefore, regardless of $sgn(O_{pp}(p))$, the existence of a non-empty interval $(0, \hat{p})$ where probabilities are overweighted is guaranteed whenever $w(p) = L(p) + O(p)$.

■

Appendix B: Agents with CPT preferences

In this Appendix, I analyze the incentives that the two contracts generate when agents have risk preferences characterized by CPT. I find that under mild additional conditions, the result stated in Proposition 2 holds and probability contracts can generate higher output than piece-rates. This result is not surprising since CPT is a more descriptive version of RDU that maintains the property that decision weights are composed by probability distortions of cumulative probabilities. Hence, the principal is also able to induce probabilistic risk attitudes which allow the principal to generate a taste for risky compensation contracts in the agent.

Agents with CPT preferences evaluate outcomes relative to a reference point $r > 0$. Outcomes below these reference point are coined *losses* and outcomes above it are *gains*. Typically, r represents a monetary amount that the worker expects to receive (Koszegi and Rabin, 2006) or a monetary amount that she owns (Kahneman et al., 1991). The novelty of CPT is that the agent can exhibit different risk preferences across these domains. Formally, the evaluation of outcomes when $V(A, y)$ is offered, and the equivalence $A = \frac{a}{p}$ is assumed, is captured by the value function $v(y, r)$ with the following properties:

Assumption B.1 (CPT value function): $v(y, r)$ is the piecewise function,

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

With $r \geq 0$, $\lambda > 1$, $b(0, p) = 0$, $b_y(y, p) \geq 0$ for all $y \in [0, \bar{y}]$, $b_{yy}(y, p) < 0$ if $\frac{ay}{p} > r$, and $b_{yy}(y, p) > 0$ if $\frac{ay}{p} < r$.

Note that for the domain of gains the value function is concave while in the domain of losses this function is convex. This implies that risk aversion or risk seeking attitudes are generated by the relativity of an outcome with respect to r.

Additionally, the worker is loss-averse, which means that for him losses loom larger than gains. This is represented by the parameter $\lambda > 1$ which enters the value function only for the domain of losses. Loss aversion in this setup can be understood as the agent wanting to be compensated for a loss amounting $u(q)$ with an indemnity $\lambda u(-q)$.

The CPT agent, as the RDU agent, transforms the probabilities associated to outcomes of the contract with a weighting function. However, unlike RDU, CPT is rich enough to admit different transformations of probabilities for losses and gains. Let $w(p)$ as in Assumption 3 be the weighting function used to transform probabilities of outcomes classified as gains. The transformation of probabilities associated to outcomes classified as losses is done through $z(p)$. Let $w(p)$ and $z(p)$ relate through the duality $z(p) = 1 - w(1-p)$. This implies that decision weights that result from ordering the outcomes according to a rank from most-desirable to least-desirable is equivalent to the decision weights that result from ranking outcomes from least-desirable to most-desirable.[29]

All in all, the utility of the agent with CPT preferences when working under $V(y,p)$ is equal to:

$$CPT(y,r) = \begin{cases} w(p)v(y,r) + (1-w(p))v(0,r) - c(y,\theta), \text{ if } \frac{ay}{p} \geq r = 0, \\ w(p)v(y,r) + z(p)v(0,r) - c(y,\theta), \text{ if } \frac{ay}{p} \geq r > 0, \\ z(p)v(y,r) + w(1-p)v(0,r) - c(y,\theta), \text{ if } r > \frac{ay}{p} > 0. \end{cases} \tag{12}$$

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

The agent with CPT preferences supplies a level of output y_C^{**} satisfying the following system of equations

$$\frac{a}{p}w(p)b_y\Big(\frac{ay_C^{**}}{p} - r\Big) - c_y(y_C^{**},\theta) = 0, \text{ if } \frac{ay}{p} \geq r, \tag{13}$$

$$\frac{a}{p}z(p)\lambda b_y\Big(r - \frac{ay_C^{**}}{p}\Big) - c_y(y_C^{**},\theta) = 0, \text{ if } \frac{ay}{p} < r, . \tag{14}$$

Let us first consider the case $\frac{ay}{p} \geq r$. According to equation (13), output increases with the monetary incentives offered by the contract a and with the ability of the agent θ. Also, higher probabilities increase output in a non-linear way, with equally-sized probability increments within the region $p \in (0, \hat{p})$ yielding smaller increases than equally-sized probability increments in the region $p \in (\hat{p}, 1)$.

As shown by Equation (14), when $\frac{ay}{p} < r$, the parameters of the model have a similar influence on the agent's decision. Specifically, higher ability, higher monetary incentives, and higher probabilities increase output. In addition, higher values of the loss aversion parameter,

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

λ, yield higher output. This comparative static captures the notion that an agent is willing to supply more output to avoid experiencing losses.

Finally, the effect of a higher reference point r on output is ambiguous. Higher reference points shift the schedules $\frac{a}{p}w(p)b_y\Big(\frac{ay_C^{**}}{p} - r\Big)$ and $\frac{a}{p}z(p)\lambda b_y\Big(r - \frac{ay_C^{**}}{p}\Big)$ to the right. This is suggestive of higher reference points generating higher output. However, the solution to equations (13) and (14) features multiple equilibria, with equilibria at low levels of output that become lower as r increases. Thus, when the agent supplies output according to these low-output equilibria, higher reference points yields lower output.[30] The intuition of this result is that higher reference points can generate higher output up to a level after which they become unattainable and demotivate the agent. This is a well-known regularity in the literature of endogenous reference points (Dalton et al., 2016, Corgnet et al., 2015, Heath et al., 1999).

Next, consider the case of an agent with risk preferences as in CPT receiving $W(a, y)$. Even though this contract does not contain risk, which allows me to disregard the presence of probability weighting, the agent still makes decisions relative to a reference point (Tversky and Kahneman, 1991). It is assumed that an agent has the same reference point across the two contracts. This assumption is sensible inasmuch as both incentive contracts pay on expectation the same monetary amount and because the agent owns the same amount of money before being offered any of the contracts.

Therefore, when offered the piece-rate the agent with riskless prospect theory preferences (Tversky and Kahneman, 1991) supplies output according to y_C^* satisfying the following system of first-order conditions

$$ab_y(ay_C^* - rp) - c_y(y_C^*, \theta) = 0, \text{ if } ay \geq rp \tag{15}$$

$$\lambda ab_y(rp - ay_C^*) - c_y(y_C^*, \theta) = 0, \text{ if } ay < rp. \tag{16}$$

Note that the reference point r becomes rp under the piece-rate due to the equivalence $Ap = a$, made in Section 2 and which we maintain here.

We are now in a position to compare the two contracts. Let us first consider the case

[30] To see how, consider first $r = 0$. For such case only the curve capturing the marginal utility of gains, $\frac{a}{p}w(p)b_y(\frac{ay}{p} - r)$, is relevant to our case since only this curve attains non-negative values of output. The unique crossing point of $\frac{a}{p}w(p)b_y(\frac{ay}{p} - r)$ with $c_y(y, \theta)$ determines the optimal output level. Now, consider a small increment of $r = \epsilon$ for sufficiently small $\epsilon > 0$. Since higher values of r shift $\frac{a}{p}w(p)b_y(\frac{ay}{p} - r)$ to the right, the crossing point between this curve and $c_y(y, \theta)$ also shifts to the right. However, now the curve $\frac{a}{p}z(p)\lambda b_y\big(r - \frac{ay}{p}\big)$ can also attain positive values of output. Note that since $b_{yy}(y, p) > 0$ if $\frac{ay}{p} < r$ there are multiple crossings between $\frac{a}{p}z(p)\lambda b_y\big(r - \frac{ay}{p}\big)$ and $c_y(y, \theta)$. One of these crossings is always smaller than the other, which, due to shift to the right of $\frac{a}{p}z(p)\lambda b_y\big(r - \frac{ay}{p}\big)$ as r increases, becomes lower with higher reference points.

$\frac{ay}{p} \geq r$. Lemma B.1 demonstrates that Proposition 2 ensures that probability contracts induces higher output than piece-rates.

Lemma B.1: For an agent with $\tilde{\theta} \in (0,1)$, who evaluates outcomes using $v(r,y)$ from Assumption B.1, and who transforms probabilities with $w(p)$ from Assumption 3 and $|w(p)| < \infty$, then $y_C^{**} \geq y_C^{*}$ if $p \in (0,p^*]$ and $ay \geq rp$.

Proof. Suppose instead that $y_C^{**} < y_C^{*}$ for all $p \in (0,1]$. For an agent with $\theta = \tilde{\theta}$ it must be that $c_y(y_C^{**},\tilde{\theta}) < c_y(y_C^{*},\tilde{\theta})$. Using equations (13) and (15), the assumed inequality can be rewritten as

$$\frac{a}{p}w(p)b_y\Big(\frac{ay_C^{**}}{p} - r\Big) < ab_y(ay_C^{*} - rp). \tag{17}$$

For any $y \in (0,\bar{y}]$, $p \in (0,1)$, and $a > 0$, and whenever $\frac{ay}{p} \geq r$, the inequality $\frac{ay}{p} - r > ay - rp$ holds. According to Assumption B.1, $b_{yy}(y,p) < 0$ if $\frac{ay}{p} > r$, then $b_y\left(\frac{ay}{p} - r\right) < b_y(ay - rp)$. Similarly, $b_y\left(\frac{ay}{p} - r\right) < b_y(a(y+\epsilon) - rp)$ holds for arbitrary $\epsilon > 0$. Hence, the inequality in (17) holds for any $p \in [\tilde{p},1]$ since for this interval $\frac{w(p)}{p} < 1$ and $b_y\left(\frac{ay}{p} - r\right) < b_y(a(y+\epsilon) - rp)$ letting $y_C^{**} = y$ and $y_C^{*} = y_C^{**} + \epsilon$.

Next, consider $p \in (0,\hat{p}]$. Since $b_y\left(\frac{ay}{p} - r\right) \leq b_y(a(y+\epsilon) - rp)$ for arbitrary $\epsilon > 0$ remains valid, but now $\frac{w(p)}{p} > 1$, then both $y_C^{**} < y_C^{*}$ and $y_C^{**} \geq y_C^{*}$ are feasible. As $p \to \tilde{p}$ then $\frac{w(\tilde{p})}{\tilde{p}} = 1$ and the inequality in (17) holds. Instead, as $p \to 0^+$ this inequality does not hold. To see how note that $\lim_{p\to 0} ab_y(ay_C^{*} - rp) = y^* \leq \bar{y}$ so the right hand side of (17) is identical to the left hand side of (10) as $p \to 0^+$. More importantly, $\lim_{p\to 0^+} \frac{a}{p}w(p)b_y\Big(\frac{ay_C^{**}}{p} - r\Big) = \infty$, using L'Hospital rule in the same way as in the proof of Proposition 2. Hence, the inequality in (17) does not hold as $p \to 0^+$ since.

Since the behavior of $\frac{a}{p}w(p)b_y\Big(\frac{ay}{p} - r\Big)$ as p changes in $p \in (0,\tilde{p})$ is the same as that of $\frac{a}{p}w(p)b_y\Big(\frac{ay}{p}\Big)$, contained in the right hand side of (10), then the existence of a $p^* < \tilde{p}$ such that choosing any $p \in (0,p^*)$ contradicts the inequality in (17) is guaranteed. Thus, it must be that if $p \in (0,p^*)$ is chosen by the principal then $y_C^{**} > y_C^{*}$.

■

As in the case of RDU preferences, the principal could be better off offering probability contracts as long as she implements it with a probability that induces a sufficiently large degree of overweighting of probabilities. This is because the overweighting of small probabilities acts as a risk-seeking mechanism, which, when sufficiently strong, could outweigh the risk-averse attitudes generated by concavity of the value function in the domain of gains. These global risk-seeking attitudes motivate the agent to work harder on the task under probability contracts.

Now consider $\frac{ay}{p} < r$. Again, the result from Proposition 2 can be guaranteed but now under less stringent conditions. Lemma B.2 demonstrates this result and presents these milder conditions.

Lemma B.2: For an agent with $\tilde{\theta} \in (0, 1)$, who evaluates outcomes with $v(r, y)$ from Assumption B.1, and who transforms probabilities using $z(p) \equiv 1 - w(1 - p)$ with $w(p)$ as in Assumption 3 and $|w_{pp}(p) < \infty|$, then $y_C^{**} > y_C^{*}$ if $p \in (0, \tilde{p}]$ and $\frac{ay}{p} < r$.

Proof. Suppose instead that $y_C^{**} \leq y_C^{*}$. Since $c_y(\tilde{\theta}, y) > 0$ and $c_{yy}(\tilde{\theta}, y) > 0$, then it must be that $c_y(y_C^{**}, \tilde{\theta}) \leq c_y(y_C^{*}, \tilde{\theta})$. Using equation (14), equation (16), and the duality $z(p) = 1 - w(1 - p)$, this inequality can be rewritten as:

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

The inequality $r - \frac{ay}{p} > rp - ay$ holds for any $y \in (0, \bar{y}]$, $r > 0$, $a > 0$, and $p \in (0, 1)$ whenever $r > \frac{ay}{p}$. According to Assumption B.1, for this domain $b_{yy}(\cdot) \geq 0$, then $b_y(r - \frac{a}{p}y) \geq b_y(rP - ay)$. This inequality continues to hold whenever $b_y(r - \frac{a}{p}(y + \epsilon)) > b_y(rp - ay)$ for arbitrary $\epsilon > 0$. Moreover, $\frac{1 - w(1 - p)}{p} \geq 1$ for $p \in (0, \tilde{p}]$. Together, these two facts lead to the conclusion that inequality in (18) not to hold. Hence, it must be that $y_C^{**} > y_C^{*}$ if $p \in (0, \tilde{p})$ and $\frac{ay}{p} < r$.

■

An agent with CPT preferences exhibits a convex value function in the domain of losses. This curvature generates risk-seeking attitudes, which favor labor supply under the probability contract. To maintain these favorable risk attitudes, the principal should avoid choosing probabilities that induce risk-averse attitudes. This could be done implementing any $p \in (0, \tilde{p}]$. Since $p^{*} < \tilde{p}$ the result in Lemma B.2 is guaranteed when $p \in (0, p^{*})$.

Appendix C: The principal's problem

The purpose of this Appendix is to complement the analysis presented in Section 2. The present analysis incorporates the incentive compatibility constraint of the agent, as in the main body of the paper, but also the participation constraint of the agent. We focus on the decision of the principal when these constraints are taken into account. The solution of the principal's program corroborates the result contained in Proposition 2. Specifically, the principal prefers probabilities that are strongly overweighted by the agent, which generate a taste for risky incentive contracts and, thus, incentivize higher labor supply than if the piece-rate were given.

In this Appendix I assume that the agent has risk preferences as in RDU. That is, he distorts cumulative probabilities using the weighting function $w(p)$ described by Assumption 3. In addition, I assume throughout that the agent has a linear basic utility $b_{yy}(y,p) = 0$. This result is empirically supported by Appendix E and previous utility elicitations using lotteries with small stakes (See Wakker & Deneffe (1996) and Abdellaoui (2000)). These two assumptions entail that the risk attitudes of the agent are uniquely determined by the probabilistic risk attitudes of the agent. Hence, the choice of p can drastically change the attitudes of the agent toward risky or riskless contracts.

I focus in a setting in which $V(A, y)$ is the only contracting option and the equivalence $A = \frac{a}{p}$ is assumed. Thus, when introducing risk is disadvantageous, the principal can choose $p = 1$ and the agent is compensated according to a piece-rate. Finally, it is also assumed that the principal is an expected value maximizer.

All in all, the principal's objective is to minimize the amount paid when using $V(A, y)$, subject to the participation constraint and the incentive compatibility of the agent. This problem is formally described by the following program:

$$\begin{aligned} &\underset{p}{\text{Min}} && Ayp, \\ &\text{subject to} && IC: \quad \underset{y}{\text{argmax}}\; w(p)Ay - c(\theta, y), \\ & && PC: \quad w(p)Ay - c(\theta, y) \geq 0. \end{aligned} \tag{19}$$

The solution to the principal's problem is presented in Proposition C.1. This solution features a choice that depends on the probability interval as well as on the location of $\hat{p}$ with respect to $\tilde{p}$ in the agent's weighting function.

Proposition C.1 *The solution to the program in Equation (19) is*

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

Where p^{**} is the fixed-point $p^{**} = \frac{w(p^{**})}{w_p(p^{**})}$, $w(p)$ is a weighting function with the properties in Assumption 3 and $|w_{pp}(p)| < \infty$.

Proof. The Lagrangian of the program in Equation (19) can be set as:

$$\mathcal{L} = ay - \lambda_1 \left(w(p)\frac{a}{p} - c_y(y, \theta) \right) - \lambda_2 \left(w(p)\frac{ya}{p} - c(y, \theta) \right). \tag{20}$$

The first-order condition of the Lagrangian is

$$\frac{\partial \mathcal{L}}{\partial p}: \quad -\lambda_1 \left(w_p(p)\frac{a}{p} - w(p)\frac{w(p)a}{p^2} \right) - \lambda_2 \left(w_p(p)\frac{ay}{p} - w(p)\frac{ay_t}{p^2} \right). \tag{21}$$

Rewritting of (21) yields,

$$\frac{\partial \mathcal{L}}{\partial p}: \quad (\lambda_1 + \lambda_2 y_t)\frac{a}{p} \left(w_p(p) - \frac{w(p)}{p} \right) = 0. \tag{22}$$

Notice from (22) that when $\lambda_1 > 0$ or if $\lambda_2 > 0$ the solution of the Lagrangian is given by the fixed-point $p^{**} = \frac{w(p^{**})}{w_p(p^{**})}$.

Let $g(p) \equiv \frac{w(p)}{w_p(p)}$. Then, $g(0) = 0$ since $\lim_{p \to 0^+} w_p(p) = \infty$ and $w(0) = 0$. Also, $g(1) = 0$ since $\lim_{p \to 1^-} w_p(p) = \infty$ and $w(1) = 1$. Similarly, $g(\hat{p}) = \infty$ since $\lim_{p \to \hat{p}} w_p(p) = 0$. Moreover, note that $g(p)$ is increasing in $p \in [0, \hat{p})$, due to the fact that $w_{pp}(p) > 0$ in $p \in [0, \hat{p})$, and $g(p)$ is decreasing in $p \in [1, \hat{p}]$, due to the fact that $w_{pp}(p) < 0$ in $p \in [1, \hat{p}]$.[31]

These properties of $g(p)$ guarantee the existence of $p^{**} = g(p^{**})$ in $p \in (0, 1)$. To see how, note that p is an increasing function that attains a minimum at $p = 0$ and a maximum at $p = 1$. Moreover, $g(p)$ attains values $g(0) = 0$, $g(1) = 0$, and $g(\hat{p}) = \infty$. Then, due to the C^2 continuity of $w(p)$, it must be that p and $g(p)$ intersect at some $p \in (0, 1)$.

The intersection $p^{**} = g(p^{**})$ exists at $p \in (0, \hat{p})$. Note that $g(0) = 0$ so that $g(p)$ and p have the same departing point. Since $\lim_{p \to 0^+} g_p(p) = 1 - \frac{w(p)w_{pp}(p)}{(w_p(p))^2} = 1$, due to $|w_{pp}| < \infty$ and $\lim_{p \to 0^+} w_p(p) = \infty$, the slopes of $g(p)$ and p are equal as $p \to 0^+$. Given that $w_{pp}(p) < 0$,

[31] 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.

small increments of p away from zero yield that $g_p(p) = 1 - \frac{w(p)w_{pp}(p)}{(w_p(p))^2} > 1$, implying that $g(p) < p$ for small values of p. Since $g(\hat{p}) = \infty$, then it must be that these functions intersect in some $p \in (0, \tilde{p}]$.

To investigate if p^{**} is a solution to the program in (19) for all values of p, I analyze the shape of the Lagrangian. The second-order condition of the Lagrangian in (20) is,

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

Equation (23) becomes positive when evaluated at p^{**} if $w_{pp}(p) < 0$. Hence p^{**} is a solution of the program as long as $p \in (0, \hat{p})$. In contrast, if $w_{pp}(p) > 0$, the second order condition in (23) becomes negative, implying that the objective function attains a minimum value at one of the extremes, $p = \hat{p}$ or $p = 1$. Hence, unless $p^{**} = 1$ or $p^{**} = \hat{p}$, there exists multiple solutions to the principal's problem.

Finally, note that according to Assumption 3, $w(p)$ might display $\hat{p} \neq \tilde{p}$. Let $\hat{p} < \tilde{p}$. For $p \in (0, \hat{p})$, the solution $p = p^{**}$ can be implemented. Instead,for $p \in [\hat{p}, 1]$ the solution is either $p = \hat{p}$ or $p = 1$. At $p = \hat{p}$, the IC and PC constraints of Equation (19) become larger than at $p = 1$, which implies that at $p = \hat{p}$ the lowest value of the Lagrangian is achieved. Thus, the principal chooses $p = \hat{p}$ if $p \in (\hat{p}, 1)$.

Let now $\hat{p} > \tilde{p}$. Again, for $p \in (0, \hat{p})$, $p = p^{**}$ is implemented even though it might imply $\frac{w(p)}{p} < 1$. Moreover, for the interval $p \in [\hat{p}, 1]$, the solution to (19) is $p = 1$ since $p = \hat{p}$ yields $\frac{w(p)}{p} < 1$ which leads to lower values of the IC and PC constrains than those implied by $p = 1$. ■

Proposition C.1 presents multiple solutions to the principal's program. This is because $w(p)$, which determines the shape of the global utility function, can be concave, convex, or both, that is first concave and then concave. When concavity and convexity coexist, an interior as well as a knife-edge solution to the principal's program are generated.

Specifically, for $p \in (0, \hat{p}]$ the principal sets $p = p^{**}$. When the weighting function of the agent displays $\tilde{p} > \hat{p}$, this solution always induces an overweighting of probabilities in the agent, which generates a global preference for risk, enhancing the agent's performance when he works under the probability contract.

The same conclusion is achieved when $\tilde{p} < \hat{p}$ whenever $p^{**} \in (0, \tilde{p})$. However, in this case implementing the contract with p^{**} can generate global risk aversion whenever $p^{**} \in [\tilde{p}, \hat{p}]$, which demotivates labor supply. The intuition behind this result is that when the principal is restricted to the interval $p < \tilde{p}$, she incentivizes low labor supply choosing $p^{**} \in [\tilde{p}, \hat{p}]$, but she is better off implementing such a contract than one with a lower probability that will not be accepted by the agent.

Additionally, for $p \in (\hat{p}, 1]$ the optimal choice is $p = \{\hat{p}, 1\}$. When $\hat{p} < \tilde{p}$, setting $p = \hat{p}$ is preferred since this could induce overweighting of this probability in the agent, which incentivizes high labor supply. Instead, when $\hat{p} > \tilde{p}$, setting $p = 1$ is preferred, since implementing any $p \in (\hat{p}, 1]$ would induce underweighting of probabilities in the agent, demotivating labor supply.

To conclude this Appendix, I find that the principal's optimal choice consists on multiple solutions that either induce overweighting of probabilities in the agent or imply implementing the piece-rate contract.[32] This solution corroborates the result in Proposition 2; whenever possible the principal chooses a probability that generates a taste for risky incentive contracts, that enhance labor supply when the agent works under a probability contract with $p < 1$.

However, the solution presented in this appendix, as well as that presented in Proposition 2 requires strong conditions. The most noteworthy is that the principal has a detailed knowledge of the agent's weighting function. Future analyses of probability contracts should focus on relaxing this assumption and the consequences that this relaxation imposes on the principal's choice when contracting with the probability contract.

[32] Note that for the special case in which $\tilde{p} > \hat{p}$ and $p^{**} \in [\tilde{p}, \hat{p}]$, the principal can also choose $p = 1$ when she is not restricted to choose in $p \in (0, \hat{p})$. A simple profit maximization argument entails that $p = 1$ is preferred to $p =^{**}$.

Appendix D: Instructions

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

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

Part one

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

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

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

[Examples displayed]

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

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

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

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

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

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

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

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

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

Part two

In this part of the experiment your task is to choose among two possible alternatives. Your earnings on this part of the experiment depend on how good your choices are.Particularly, you will face with 11 decision sets. In each of these sets you need to choose between the option L, that delivers a fixed amount of money, and the option R that is a lottery between two monetary amounts. Each decision set contains six choices.

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

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

Survey

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

- People can behave differently in different situations. How would you rate your willingness to take risks while driving? Rate yourself from 0 to 10, where 0 means "unwilling to take any risks" and 10 means "fully prepared to take risks"
- How would you rate your willingness to take risks in financial matters? Rate yourself from 0 to 10, where 0 means "unwilling to take any risks" and 10 means "fully prepared to take risks"
- How would you rate your willingness to take risks with your health? Rate yourself from 0 to 10, where 0 means "unwilling to take any risks" and 10 means "fully prepared to take risks"
- How would you rate your willingness to take risks in your occupation? Rate yourself from 0 to 10, where 0 means "unwilling to take any risks" and 10 means "fully prepared to take risks"
- How would you rate your willingness to take risks in your faith in other people? Rate yourself from 0 to 10, where 0 means "unwilling to take any risks" and 10 means "fully prepared to take risks"

Appendix E: Performance Beliefs

Section 5 presents empirical evidence supporting Hypothesis 1. Specifically, the experimental data suggest that subjects assigned to LowPr display higher average performance than subjects assigned to Piecerate, MePr, or HiPr. The aim of this Appendix is to investigate whether subjects anticipate the effect of the incentives included in probability contracts on their own performance. To that end, I analyze the subjects' beliefs about their own performance in the real-effort task across treatments. If subjects internalize these incentives, their beliefs should reflect the performance differences across the treatments documented in Section 5.

Table 7 presents the descriptive statistics of performance beliefs by treatment. The data suggest that subjects in LowPr, MePr and HiPr display beliefs that are statistically indistinguishable.[33] Likewise, I find that subjects assigned to Piecerate display similar beliefs as those of subjects assigned to LowPr, MepR, and HiPr.[34] These results suggest that subjects did not anticipate the effect that the different probabilities had on their own performance.

Table 7: 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	74.500
St. Dev.	40.864	40.156	36.139	43.318	40.147
N	43	40	44	45	172

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

I also perform a regression analysis to account for factors other than the treatment assignment that could drive these results. In particular, I incorporate variables that capture the shape of the utility function and the shape of the probability weighting function of subjects. Table 11 presents the OLS estimates. The regression estimates corroborate the aforementioned results. Specifically, the coefficients associated with the MePr, HiPr and LowPr treatments are not significant, suggesting no statistical differences between the average beliefs of subjects assigned to those treatments and those of subjects assigned to Piecerate. Furthermore, there is no evidence to reject the null hypothesis that the coefficients associated

[33] The t statistics of these comparisons are: MePr vs. LowPr ($t(80.749) = 0.206, p = 0.837$), HiPr vs. MePr ($t(78.819) = 0.956, p = 0.3418$), and LowPr vs. HiPr treatments($t(83.241) = 1.1885, p = .1190$)

[34] 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.

with LowPr and MePr are equal ($F(1,160) = 0.29$), as well as no evidence to reject the null hypothesis that the coefficients of LowPr and HiPr are equal ($F(1,160) = 0.09$).

All in all, the belief data suggest that subjects do not internalize the non-monetary incentives included in the probability contract. Hence, subjects assigned to LowPr exhibit a steep gap between their beliefs and their actual performance on the task. I conjecture that such gap can be explained in light of the findings of Berns et al. (2008), who show that distortions of probabilities primarily involves the perceptual stage of the cognitive process rather than stages of consciousness that allow individuals to internalize such distortions. Thus, without being able to internalize their distorted perception of probabilities, subjects are unlikely to understand how an eventual assignment to LowPr or HiPr can influence their own performance.

Understanding the reasons behind the gap between performance and performance beliefs is beyond the scope of this paper and requires methodologies that allow the researcher to study in more detail the cognitive processes underlying probability judgments.

Finally, note that the observed gap between the subjects' performance and their performance beliefs could explain the scarce occurrence of the probability contract or similar schemes. If a principal does not anticipate the non-monetary incentives of the contract, she may be inclined to choose simpler payment modalities. Furthermore, even if the principal is informed about the way in which agents distort probabilities, she might correctly believe that the agent is not going to internalize the non-monetary incentives included in the probability contract. However, this belief can be erroneously interpreted, leading her to the conclusion that such lack of anticipation yields similar labor supply across the contracts. The importance of this paper is showing not only that the proposed contract is more effective, but also that its incentives are cognitive and unanticipated by the worker.

Table 8: Regression of performance beliefs on treatments

	(1) Beliefs	(2) Beliefs	(3) Beliefs	(4) Beliefs
LowPr	10.683	10.675	10.205	9.098
	(8.977)	(9.366)	(9.434)	(9.336)
MePr	8.847	8.262	7.929	7.651
	(9.055)	(9.285)	(9.327)	(9.214)
HiPr	0.845	-0.140	0.348	-3.373
	(8.452)	(8.294)	(8.427)	(8.452)
Concave		26.405*	24.950*	24.873*
		(14.354)	(14.470)	(13.872)
Convex		10.174	9.955	4.495
		(17.330)	(18.148)	(16.378)
Linear		5.057	3.726	
		(8.177)	(8.251)	
LS			3.408	
			(6.883)	
US			-3.818	
			(6.159)	
Possibility				16.649
				(11.371)
Certainty				17.551
				(11.135)
Constant	73.178***	67.485***	69.583***	57.489***
	(6.461)	(8.764)	(9.755)	(11.152)
R^2	0.014	0.037	0.041	0.050
Observations	172	172	172	172

Note: This table presents the estimates of the Ordinary Least Squares regression of the model $Belief_i = \theta_0 + \theta_1 LowPr + \theta_2 MePr + \theta_3 HiPr + Controls'\Lambda + \epsilon_i$, with $E(\epsilon | MepR, LowPr, HiPr, Controls) = 0$. "Beliefs" is the predicted number of correctly solved sums by a subject in the first part of the experiment, "LowPr", "MePr" and "HiPr" are dummy variables that capture whether the subject was assigned to the treatment offering the probability contract with low, medium or high probability, respectively. The controls considered in this model are "LS" a binary variable that takes a value of one if a a subject has a weighting function with lower subadditivity and zero otherwise, "US" a binary variable that takes a value of one if a subject has a weighting function with upper subjectivity and zero otherwise, "Possibility" a binary variable takes a value of one if a subject has a weighting function with the possibility effect being stronger than the certainty effect and zero otherwise. "EQ" a binary variable taking a value of one if a subject has a weighting function with the having the same magnitude as the possibility effect and zero otherwise."Concave" a binary variable that takes a value of one if a subject has a concave utility function and zero otherwise, "Convex" a binary variable that takes a value of one if a subject has a convex utility function and zero otherwise, and "Linear" a binary variable that takes a value of one if a subject has a linear utility function and zero otherwise. 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.

Appendix F: Utility functions

This appendix investigates the properties of the utility functions elicited in the second part of the experiment. Decision sets 1 to 6 elicit the sequence of outcomes $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ that characterize the subjects' preference over monetary outcomes. This elicited sequence of outcomes has the property that it ensures equally-spaced utility values, i.e. $u(x_j) - u(x_{j-1}) = u(x_{j-1}) - u(x_{j-2})$.

I analyze two properties of this sequence when is plotted again the utility values that it maps: the sign of the resulting slope and the resulting curvature. Therefore, I construct two variables, let $\Delta'_i \equiv x_j - x_{j-1}$, for $j = 1, ..., 6$ and $\Delta''_j \equiv \Delta'_j - \Delta'_{j-1}$ for $i = 2, ..., 6$. The sign of Δ'_j as j increases determines the sign of the slope, i.e. whether a subject prefers larger monetary outcomes. Similarly, the sign of Δ''_j as j increases determines the utility curvature. For example, a subject with $\Delta'_j > 0$ and $\Delta''_j < 0$ for all $j = 1, ..., 6$ exhibits a preference for larger monetary outcomes and experiences smaller utility increments with larger monetary outcomes, this is equivalent to say that this subject has a concave utility function.

I classify the participants of the experiment according to the curvature of their utility function. Since I have multiple observations for each subject and it was possible that subjects made mistakes, this classification is based on the sign of Δ''_j with the most occurrence. Specifically, a subject with at least three negative Δ''_js was classified as having a convex utility, a subject with at least three positive Δ''_js had a concave utility and subject with three or more Δ''_is had a linear utility. A subject with a utility function that cannot be classified as concave, convex, or linear, had a mixed utility. Furthermore, to statistically asses the sign of a Δ''_j, I construct confidence intervals around zero. In particular, I multiply the standard deviation of each Δ''_j by the factors 0.64 and -0.64. Thus, if Δ''_j follows a normal distribution, 50% of the data should lie within the confidence interval.[35]

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

[35] Alternative confidence intervals were also constructed. These confidence intervals also used the standard deviation of a Δ''_j which was multiplied by different factors. For instance, 1 and -1, 1.64 and -1.64, and 2 and -2. The qualitative results of these analyses are not different from the main result that the majority of subjects exhibit a linear utility function. This is not surprising inasmuch as these confidence intervals are more stringent and yield less subjects classified as having a mixed utility function and more subjects exhibiting a linear utility function.

subjects classified as having mixed utility functions, only 6 (3 % of subjects) presented Δ_j''s that suggest a utility function that is first convex and then concave. A proportions test suggest that the proportion of subjects with linear utility is significantly larger than 50% ($p<0.001$). Moreover, this test also yields that the proportion of subjects having linear functions is significantly larger than the proportion of subjects with mixed utility ($p<0.001$) and concave utility ($p<0.001$).

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

Table 9 also presents the results of the aforementioned analysis when it is assumed that subjects have CPT preferences with an expectations-based reference point. As in the main body of the paper, I assume that this reference point is the monetary equivalent of each subject's belief. Consequences above this reference point are gains and consequences below it are losses. This alternative analysis also leads to the conclusion that the majority of the subjects exhibit a linear utility function. Specifically, I find that 65 % of the subjects have linear utilities in the domain of gains and 98% of the subjects exhibit linear utilities in the domain of losses.

To understand how the aforementioned results aggregate, I analyze the sequence $\{x_1, ..., x_6\}$ when each outcome is averaged across subjects. Table 10 presents the descriptive statistics of the aggregated outcomes. I find that the average outcome x_j is increasing with j, corroborating the average taste for larger monetary outcomes. Moreover, the column displaying the average values of the variable Δ_j' shows that as j increases, the increments of x_j become larger. Thus, the tendency of the average utility function to display linearity ceases as the monetary outcomes contained in the lotteries become larger. In fact, for large values of x_j the average utility function displays some concavity. This result is also found by Abdellaoui (2000).

In addition, I assume two well-known families of utility functions and estimate their parameters using non-linear least squares. Specifically, I assume a power utility, belonging to the CRRA family of utility functions, and an exponential function, belonging to the CARA family of utility functions. Table 11 presents the estimates of the regressions. For

Table 9: Classification of subjects according to utility curvature

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

Note: This table presents the subjects classification according to the shape of their utility function. Subjects are classified as having a convex, concave, linear or mixed utility function based on the sign of Δ_j with more occurrence. 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 two parametric specifications I find that the average utility function of the subjects is approximately linear. For instance, when the power utility function $u(x) = x^{\phi}$ is assumed, the estimate of the parameter is $\phi = 0.995$. This conclusion is consistent with the large proportion of subjects that were classified as having a linear utility function and the modest increments that the averaged outcomes x_j exhibit as j increases presented in Table 10.

These analyses are also performed under the assumption that subjects have CPT preferences with an expectation-based reference point. According to Table 10, there subjects exhibit an average preference for larger monetary amounts in both domains. Also, the descriptive statistics suggest a decreasing tendency of the utility function to be linear as the amount of money becomes larger in the domain of gains and lower in the domain of losses; In the domain of gains the utility function of the average subject tends to concavity, while in the domain of losses the function tends to convexity.

Furthermore, the data suggest that diminishing sensitivity manifests at different degrees across the domains, with subjects exhibiting more diminishing sensitivity in the domain of gains than in the domain of losses. This difference is explained by fact that only positive outcomes were used to elicit the sequence $\{x_1, x_2, x_3, x_4, x_5, x_6\}$. This leaves little room for subjects to exhibit as much sensitivity in the domain of losses as in the domain of gains. Note that I chose to elicit preferences using only positive outcomes since the second part of the experiment was designed to understand the subjects' risk preferences over the monetary incentives at stake in the first part of the experiment. A more complete analysis of diminishing sensitivity across domains, and of risk preferences in general, requires lotteries featuring negative outcomes.

Table 10: Aggregate results x_1,x_2,x_3,x_4,x_5, and x_6

j	x_j	Δ'_j	x_j	Δ'_k	x_j	Δ'_j
1	2.579 (1.990)	1.579	3.761(4.037)	3.037	1.576 (0.548)	0.576
2	4.573 (4.445)	1.993	8.167 (5.226)	4.129	2.167(0.931)	0.590
3	6.684 (6.792)	2.110	12.545(7.564)	4.378	2.761(1.280)	0.593
4	9.179 (9.420)	2.495	17.812 (9.826)	5.266	3.515 (1.800)	0.754
5	11.773 (11.880)	2.594	23.156(11.598)	5.344	4.353 (2.589)	0.837
6	14.379 (14.418)	2.605	28.400 (13.608)	5.243	5.287 (3.727)	0.934
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 $\Delta'_j = x_j - x_{j-1}$. 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 $\Delta'_j = x_j - x_{j-1}$ 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 $\Delta'_j = x_j - x_{j-1}$ for values below Beliefs for each subject.

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

All in all, the data suggest that subjects have linear utility functions. This result is robust to the assumption that subjects have CPT preferences with an expectations-based reference point. This is not a surprising finding given the magnitude of the stakes used to elicit these risk preferences. Furthermore, the conclusion that the utility function is linear implies that probability risk attitudes fully determine the subjects' risk attitudes. Hence, performance differences across treatments must be explained by probability distortions rather than by the curvature of the basic utility.

Table 11: Parametric estimates of average utility function

Exponential (CARA) $1 - exp(-\gamma(x_{j-1} + \frac{\epsilon}{2}))$			
$\hat{\gamma}$	0.977 (0.001)	0.946 (0.001)	1.337 (0.001)
Adj. R^2	0.922	0.887	0.303
N	1032	412	619
Power Utility (CRRA) $(x_{j-1} + \frac{\epsilon}{2})^{\phi}$			
$\hat{\phi}$	0.995 (0.001)	0.992 (0.001)	1.035 (0.007)
Adj. R^2	0.925	0.971	0.756
N	1032	412	619
Ref.Point	No/Zero	Beliefs	Beliefs
Domain	No/Gains	Gains	Losses

Note: This table presents the estimates of the non linear least squares regression. The upper panel assumes that the parametric form $1 - exp(-\gamma(x_{j-1} + \frac{\epsilon}{2}))$ and the lower panel assumes the parametric form $(x_{j-1} + \frac{\epsilon}{2})^{\phi}$. 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.

Appendix G: Individual analysis of probability weighting functions

This appendix presents alternative analyses of the probability weighting functions. Decision sets 7 to 11 included in the second part of the experiment were designed to elicit the subjects' weighting functions. In the main body of the paper, I present parametric analyses of the average data. In this appendix, I present non-parametric analyses of these data performed at the individual level.

The first analysis seeks to classify each subject according to the shape of their weighting function and is borrowed from Bleichrodt and Pinto (2000). Five possible shapes of the weighting function were considered. A subject could display a weighting function with the properties of lower subadditivity (LS), upper subadditivity (US) or both properties. These properties compare the behavior of the weighting functions at the extremes to their behavior at intermediate probabilities. Moreover, a subject could display a concave or a convex shape. These shapes characterize the weighting function over the entire probability interval.

To perform the classification, I constructed the variable $\partial_{j-1}^{j} \equiv \frac{w(p_j)-w(p_{j-1})}{w^{-1}(p_j)-w^{-1}(p_{j-1})}$, which captures the average slope of the probability weighting function between the probabilities j and $j-1$. I also constructed $\triangledown_{j-1}^{j} \equiv \partial_{j-1}^{j} - \partial_{j-2}^{j-1}$, which represents the change of the average slope of the weighting function between successive probabilities.

The sign of $\triangledown_{0.16}^{0.33}$ and that of $\triangledown_{0.83}^{1}$ are computed for each subject. If a subject exhibits $\triangledown_{0.16}^{0.33} < 0$, then his probability weighting exhibits LS. In other words, his probability weighting function assigns larger weights to small probabilities than to medium-ranged probabilities. Moreover, if a subject has $\triangledown_{0.83}^{1} > 0$, then his probability weighting function exhibits the property of US. That is, his weighting function assigns larger weights to large probabilities than to medium-ranged probabilities. These variables were used in the main body of the paper to understand the effect of these properties in the treatment effects.

In addition, I examine the sign of $\triangledown_{j-1}^{j}$ as j increases to examine the global shape of the weighting function. A subject was classified as having a concave weighting function if at least three (out of five) $\triangledown_{j-1}^{j}$ had a negative sign and he did not exhibit US. Alternatively, a subject had a convex probability weighting function if at least three (out of five) $\triangledown_{j-1}^{j}$ were positive and he did not exhibit LS. Note that these classifications allow for the possibility of response error.

Table 12 presents the results of the individual classification. I find that 57 % of subjects exhibit LS, 75% of subjects exhibit US and 44% of subjects display probability weighting functions with both LS and US. Therefore, most subjects in the experiment had weighting functions that yield overweighting of small probabilities or underweighting of large probabilities. Also, almost half of subjects exhibit probability weighting functions that assign large weights to small and large probabilities. However, that these proportions are considerably lower than those reported by Bleichrodt and Pinto (2000).

Moreover, I find that 39% of the subjects exhibit convex weighting functions and only 13% of the subjects exhibit concave weighting functions. Thus, more subjects in the experiment were pessimistic than optimistic. Furthermore, the proportion of subjects in the experiment with either concave or convex probability weighting functions is higher than that reported by Bleichrodt and Pinto (2000), who finds that only 15% of the subjects have probability weighting functions with either of these shapes.

Table 12: Classification of subjects according to the shape of their weighting function

Reference Point	Domain	Convex	Concave	LS	US	LS & US
No/Zero	No/Gains	68	23	98	129	76
Beliefs	Gains	29	9	49	63	38
Beliefs	Losses	39	14	49	66	38

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

I perform an alternative classification of LS and US also proposed by Bleichrodt and Pinto (2000). This classification considers weights assigned to extreme probabilities. In particular a subject has a weighting function with LS if $w^{-1}\left(\frac{1}{6}\right) < 0.16$. Similarly, a subject has a weighting function with US if $1 - w^{-1}\left(\frac{5}{6}\right) < 0.16$. This alternative classification of LS and US is admittedly less accurate than the initially proposed classification. The reason is that assigning weights to small and large probabilities larger than their objective probabilities does not guarantee that the weights assigned to medium-ranged probabilities.

The results of the alternative classification are presented in Table 13. I find that a similar proportion of subjects exhibit US and LS. Specifically, 40.12% of subjects exhibit LS and 38.37% subjects exhibit US. Also, only 20% of subjects exhibit both LS and US. These proportions are considerably lower than those obtained with the initial classification and are also smaller to those reported by Bleichrodt and Pinto (2000).

Another classification of weighting functions considers the strength of the possibility effect relative to the certainty effect. A subject has a possibility effect that is stronger

Table 13: Classification of subjects according to LS, US, or both

Reference Point	Domain	LS	US	Both
No/Zero	No/Gains	55	89	25
Beliefs	Gains	18	49	8
Beliefs	Losses	37	40	17

Note: This table presents the classification of subjects according to the shape of their weighting functions. Subjects are classified as having weighting functions with LS if $w^{-1}\left(\frac{1}{6}\right) < w^{-1}\left(\frac{2}{6}\right) - w^{-1}\left(\frac{1}{6}\right)$. Subjects have weighting functions with US if $1 - w^{-1}\left(\frac{5}{6}\right) < w^{-1}\left(\frac{5}{6}\right) - w^{-1}\left(\frac{4}{6}\right)$. When these two properties hold, subjects are classified in Both.

than the certainty effect when $1 - w^{-1}\left(\frac{5}{6}\right) > w^{-1}\left(\frac{1}{6}\right)$. Table 14 shows that the majority of subjects in the experiment have probability weighting functions with the certainty effect exceeding the possibility effect. This result is in line with the findings of Tversky and Fox (1995). Nevertheless, the proportion of subjects for which Certainty exceeds Possibility is not negligible as it constitutes 32 % of subjects.

Table 14: Classification of subjects according to strength of possibility effect

Reference Point	Domain	Certainty	Possibility	Equal
No/Zero	No/Gains	107	55	10
Beliefs	Gains	57	18	4
Beliefs	Losses	50	37	6

Note: This table presents the classification of subjects according to the strength of the possibility effect with respect to the certainty effect. Subjects are classified Possibility, that is having probability weighting function where the possibility effect exceeds the certainty effect if $1 - w^{-1}\left(\frac{5}{6}\right) > w^{-1}\left(\frac{1}{6}\right)$. Instead, if $1 - w^{-1}\left(\frac{5}{6}\right) < w^{-1}\left(\frac{1}{6}\right)$ subjects were classified certainty. Finally, subjects with $1 - w^{-1}\left(\frac{5}{6}\right) = w^{-1}\left(\frac{1}{6}\right)$were classified Equal.

As in the main body of the paper, I consider the possibility that subjects have CPT preferences with an expectations-based point. All previous analyses are performed under the assumption that the monetary equivalent of a subject's belief in the real-effort task is the reference point.[36] The results of these analyses are also presented in Table 12, Table 13, and

[36]It is important to emphasize that the nature and intuition of the classification under the assumption that Beliefs is the reference point differs from the original classification. The reason for this difference is that the

Table 14. All in all, I find that the aforementioned results are robust to subjects having CPT preferences. In particular for both domains there is a large proportion of subjects with US and/or LS. Also, regardless of the domain, more subjects exhibit weighting functions with the certainty effect being stronger than the possibility effect.

In conclusion, the analyses of the data at the individual level suggest that the majority of subjects have weighting functions with US or LS. Moreover, I find that lower than half of subjects exhibit both properties at the same time, which is a remarkable difference with respect to Bleichrodt and Pinto (2000). Finally, as in Abdellaoui (2000) and Tversky and Fox (1995), I find that the certainty effect is stronger than the possibility effect for a larger share of individuals.

data does not admit enough $\bigtriangledown^{j}_{j-1}$s to analyze the shape of the probability weighting function of a subject for the domain of gains as well as for the domain of losses. Instead, I analyze the shape of a subject's probability weighting function for the domain wherein the majority of his $\bigtriangledown^{j}_{j-1}$s lie. Thus, this analysis could shed light on whether subjects who have most of their choices in the domain of losses exhibit weighting functions of different shape than subjects who have most of their choices in the domain of gains.