Incentive contracts when agents distort probabilities[*]

Víctor González-Jiménez[†]

March 17, 2021

Abstract

This paper shows that stochastic contracts are powerful motivational devices when agents distort probabilities. Stochastic contracts allow the principal to target probabilities that, when distorted by the agent, enhance the motivation to exert effort on the delegated task. This novel source of incentives is absent in traditional contracts. A theoretical framework and an experiment demonstrate that stochastic contracts targeting small probabilities, and thus exposing the agent to a large degree of risk, generate higher performance levels than traditional contracting modalities. This result contradicts the standard rationale that optimal contracts should feature a tradeoff between insurance and efficiency. This unintuitive finding is attributed to probability distortions caused by likelihood insensitivity—cognitive limitations that restrict the accurate evaluation of probabilities.

JEL Classification : C91, C92, J16, J24.
Keywords: Contracts, Risk Attitude, Incentives, Probability Weighting, Experiments.

[*]I am indebted to Wieland Müller, Charles Noussair, Patricio Dalton, Aurelien Baillon, Robert Dur, Eeva Mauring, Cole Williams, Geoffrey Castillo, and Stephanie Koornneef for 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

A fundamental result from incentive theory is that a principal, who cannot observe the agent's action to exert effort, motivates the agent with a contract that offers higher payments in exchange for higher performance levels (Holmstrom, 1979). That solution to the principal's problem features a tradeoff between efficiency, eliciting high effort, and insurance, protecting the agent from risk. Recent literature exploring the influence of behavioral biases on contract design shows that simple and costless extensions to this traditional solution can enhance motivation. For instance, contracts in which wage-irrelevant production goals are specified by the principal generate high effort levels when agents have reference-dependence preferences (Corgnet et al., 2018).[1]

In this paper, I show that stochastic contracts can more effectively incentivize agents who suffer from probability distortion as compared to traditional contracting modalities. Abundant empirical evidence from the literature of decision theory shows that individuals, when making decisions under risk, tend to overweight small probabilities and underweight intermediate to large probabilities (Abdellaoui, 2000, Gonzalez and Wu, 1999, Wu and Gonzalez, 1996, Tversky and Fox, 1995, Tversky and Kahneman, 1992). With a contract that exposes the agent to different degrees of additional risk, the principal not only activates these probability distortions, but also targets those probabilities that, when distorted by the agent, enhance effort exertion. Since these incentives are absent in traditional contracts, because they seek to expose the agent to as little risk as possible, stochastic contracts have the potential to generate greater output at no extra cost for the principal.

I consider a simple version of stochastic contracts in which the agent obtains one of two possible outcomes: a monetary compensation that depends on performance on the delegated task, and a lump-sum payment that does not depend on performance. When offered this contract, the agent faces the risk that effort does not count toward compensation. Throughout, it is assumed that the principal can introduce the desired amount of risk in the agent's environment by adjusting the probability that the performance-contingent compensation is paid. Therefore, under full commitment, i.e. the principal credibly commits to pay the outcomes specified in the contract and to respect the underlying stochastic rule, the agents' decision about how much effort to exert not only depends on the monetary incentives offered by the contract, but also, and more importantly, on the perceived probability that effort affects the expected compensation. Thus, in this setting, the principal problem is, monetary

[1] Other examples of these simple extensions are contracts that include contests for status in the organization when individuals exhibit a preference for status (Besley and Ghatak, 2008, Auriol and Renault, 2008, Moldovanu et al., 2007) and contracts that allow time inconsistent agents to set personal production targets (Kaur et al., 2015).

incentives apart, to choose the probability that best motivates the agent.

To understand how stochastic contracts can outperform traditional contracts, consider a setting in which both contracts are cost-equivalent for the principal. That is, the *expected* monetary compensation associated to supplying any level of output when agents work under the probability contract is 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 both contracts and will exert the same effort. 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 stochastic contract motivates agents to exert higher effort. The underlying reason for such boost in labor supply is that this distortion of probabilities inflates the agents' perceived benefits of supplying higher levels of effort.

A simple theoretical framework serves two purposes. First, it pins down the conditions guaranteeing the main result of the paper, which is that stochastic contracts can generate higher output than more traditional contracts at no extra cost for the principal. When the agent's weighting function attains a lower-bound, representing the agent's risk attitude from utility curvature brought to the probability space, the principal is better of offering stochastic contracts with small probabilities. Second, it provides a set of predictions that are empirically tested with a laboratory experiment.

A controlled laboratory experiment demonstrates that stochastic contracts, when implemented with probability $p = 0.10$, yield higher performance in an effort intensive task as compared to a cost-equivalent piece-rate. In contrast, I find that stochastic 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 linear utilities and an average weighting function with inverse-S shape. I demonstrate that this pattern of probability distortion explains the treatment effects found in the effort task. In addition, the data show that probability distortions due to *likelihood insensitivity*, which refers to the cognitive inability of individuals to accurately evaluate probabilities (Tversky and Wakker, 1995), explain the difference in performance between the stochastic contract implemented with $p = 0.10$ and the cost-equivalent piece rate.

While stochastic contracts seem abstract and might be for instance prohibited in countries where gambling is forbidden, their incentives can be brought to practice using standard tools of personnel economics. For instance, when output is stochastic, bonus contracts offering a large reward for attaining some production target introduce risk in the agent's environment. When the principal is able to adjust this contract such that the probability of attaining the production target is sufficiently overweighted by the worker, the incentives presented in this paper apply. I provide a detailed explanation of this application 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, when agents distort probabilities, stochastic contracts are preferable to standard traditional contracts. This is at odds with standard results from the theory of incentives stating that the principal faces a trade-off between incentives and insurance. I show that the principal can provide incentives by exposing the agent to large amounts of risk. While the optimality of stochastic contracts has been put forward in other settings, such as multitasking environments (Ederer et al., 2018), when agents exhibit aspiration levels (Haller, 1985), or when agents are loss averse (Herweg et al., 2010), I am the first to theoretically and empirically show that they are desirable when agents exhibit probability distortions.

To the best of my knowledge only Spalt (2013) has studied optimal contract 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 study the incentives that result from offering stochastic contracts with different degrees of risk and establish the degree of risk that best motivates the agent. Spalt's (2013) analysis does not consider these incentives and, due to the setting studied in that paper, ignores such constraint. Another relevant difference is that I study the incentives of a general class of contracts introducing risk in the agents' compensation. These incentives can be brought to practice in multiple ways, using compensation plans with stock options.[2]

Finally, the results of this study also contribute to the literature of decision theory. To the best of my knowledge I am the first to provide applications of probability weighting elicitation techniques to the context of incentives. Furthermore, the experimental results illustrate the importance of using parametrized probability weighting functions that can separate likelihood insensitivity from optimism. I use the different methods proposed by Wakker (2010) and applied in Abdellaoui et al. (2011) to isolate these two components of probability weighting, and show that they contribute unequally to the effectiveness of the stochastic contracts.

[2]Other crucial difference with respect to Spalt (2013) is that he analyzes the profitability of compensation plans with stock options using a calibration exercise that employs parameters estimated in classical experiments. Instead, I demonstrate analytically that stochastic contracts are more effective in motivating agents than more traditional contracting modalities. Also, my experiment is designed to directly link the subjects' performance under probability contracts to their risk preferences. This experimental design feature allows me to cleanly establish whether the subjects' probability weighting function, and not other factors, drives the result that the proposed contracts can generate greater performance.

2. The model

Consider a principal (she) who delegates a task to an agent (he). The agent's decision consists on exerting a level of effort $e \in [0, \bar{e}]$ on the task. This decision depends on the disutility associated to exerting effort, as well as on the monetary incentives included in a take-it or leave-it contract that is offered by the principal before the agent makes the decision to exert effort.

It is assumed that the agent experiences marginally increasing disutility with higher levels of effort. Specifically, the disutility from effort is represented by the cost function $c(e)$ a continuously differentiable, strictly increasing, and convex function.

Assumption 1. *$c(e)$ is a $\mathcal{C}^2$ function with $c'(e) > 0$, $c''(e) > 0$, and $c(0) = 0$.*

Moreover, I posit that the level of effort translates into output in a deterministic way.

Assumption 2. *$y = f(\theta, e) = \theta e$ for all $\theta \in [0, 1]$.*

The parameter $\theta \in [0, 1]$, included in the production function, captures the agent's ability on the task. Hence, Assumption 2 states that agents with higher ability deliver can higher levels of output as compared to agents with lower ability without having to exert higher effort.

Settings whereby the link between effort, output, and ability is deterministic have been labeled as "false moral hazard" in the literature (See (Laffont and Martimort, 2002) Ch. 7.2). The rationale for investigating the effectiveness of the stochastic contracts in this setting, rather than using a standard moral hazard environment, is that it allows me to establish whether introducing risk in the agent's environment is beneficial in a setting in which he otherwise would not have to face any risk. In an extension of the model, presented in Appendix B, I consider a setting in which output is stochastic. That appendix shows that the main result presented in this section holds in a setting in which the relationship between output y and effort e is not deterministic.

To incentivize the agent to exert high effort, the principal offers the agent a contract including a transfer $t(y)$. I assume that the monetary incentives included in the contract enter the agent's utility through the function $b\left(t(y)\right)$ about which I make the following assumption:

Assumption 3. *$b(t)$ is a $\mathcal{C}^2$ function with $b(0) = 0$ and $b'(t) > 0$.*

Note that Assumption 3 does not impose restrictions on the sign of the second derivative of $b\left(t(y)\right)$. That is because the results of the model will be evaluated under the two signs that this derivative attains.

We study two types of contracts: deterministic contracts and stochastic contracts. As a benchmark of deterministic contracts, I use piece-rates. Formally, a piece-rate is $t_d(y) := ay$, where $a > 0$ represents a monetary quantity.[3] The present model can be thought of a principal wanting to switch from linear compensation contracts to stochastic contracts. While piece-rates and linear contracts may be optimal under rather specific conditions, this type of incentive scheme is prevalent in the theory of incentives and is broadly used in practice.

All in all, the agent's utility when offered the contract t_d is:

$$U(t_d) = b(ay) - c(e). \tag{1}$$

Alternatively, the principal can incentivize the agent to work on the task using a stochastic contract. The considered stochastic contract also offers a monetary compensation that depends on the agent's level of output, but, unlike the piece-rate, this compensation is not given with certainty. Instead, the agent receives such compensation with probability $p \in (0, 1]$, chosen ex-ante by the principal. As a consequence, the principal has two channels to motivate the agent with the probability contract: i) via the monetary rewards given in exchange of the level of output that is supplied and ii) via changes in the likelihood that such rewards are indeed paid.

Formally, the stochastic consists of the lottery-like compensation schedule $t_s(y) := (Ay, p; 0, 1 - p)$, where $A > 0$ represents a monetary quantity.[4] The timing of the contract is as follows. The principal moves first choosing $p \in (0, 1]$ and A. After this choice is made, p and A are communicated to the agent before he makes a decision about the level of e to be exerted. Next, the agent chooses e. Finally, when the contracted work-span concludes, a random device to which the principal credibly commits determines whether or not the agent's compensation depends on the supplied level of y.

I assume that the agent's risk preferences are characterized by rank-dependent utility (RDU, henceforth) (Quiggin, 1982). These risk preferences are descriptively valid inasmuch as they capture probability distortions (Abdellaoui, 2000, Gonzalez and Wu, 1999, Wu and Gonzalez, 1996, Tversky and Fox, 1995, Tversky and Kahneman, 1992). Moreover, this preference representation entails that risk attitude is not only determined by the curvature of the function b, but also by the agent's sensitivity to the probabilities. Sensitivity to

[3] A more general representation consists of the linear contract $t_s = F + Ay$, for some fixed-payment $F \geq 0$. This representation is more realistic inasmuch as it leaves some non-zero base-pay to the agent. Linear contracts not only consider piece-rate contracts but also fixed salary contracts. Since F does not generate any incentives to provide effort to the agent, the normalization $F = 0$ is considered.

[4] A more general representation of the stochastic contract is $t_s = (F + Ay, p; F, 1-p)$ for some fixed-payment $F \geq 0$. This representation is more realistic inasmuch as it leaves some non-zero base-pay to the agent. Throughout the paper I use the normalization $F = 0$.

probabilities is captured in the model by a probability weighting function, $w(p)$, with the following properties:

Assumption 4. *A probability weighting function is* $w(p) : [0,1] \to [0,1]$ *such that:*

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

According to Assumption 4, $w(p)$ is an increasing and two-times continuously differentiable function that maps the unit interval onto. The probability 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 $\tilde{p} = 1$, a convex shape if $\tilde{p} = 0$, and an inverse-S shape if $\tilde{p} \in (0,1)$. The latter shape generates an additional interior fixed-point, $\hat{p} \in (0,1)$.[5]

All in all, the rank-dependent expected utility of the agent when offered t_s is:

$$RDU(t_s) = w(p)b(Ay) - c(e). \tag{2}$$

Notice that when $w(p) = p$, RDU collapses to expected utility theory (EUT, from here onward). An agent with risk preferences characterized by EUT exhibits the following expected utility when working under t_s:

$$\mathbb{E}\left(U(t_s)\right) = pb(Ay) - c(e). \tag{3}$$

Another theory of risk that incorporates distortions of probabilities through probability weighting functions is Cumulative Prospect Theory (CPT, henceforth) (Tversky and Kahneman, 1992). CPT is a more descriptive version of RDU. An agent with CPT preferences also exhibits sensitivity to probabilities. However, the agent with CPT preferences also displays relativistic perception of outcomes with respect to a *reference point*. In the interest of space, I relegate the formal description of CPT preferences and the analysis of the incentives produced by the proposed contracts on agents with CPT preferences to Appendix C.

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

2.1. Probabilistic risk attitudes and their decomposition

This subsection can be omitted if the reader is acquainted with the concepts of likelihood insensitivity (Tversky and Wakker, 1995), and pessimism and optimism toward risk (Yaari, 1987, Abdellaoui, 2002). As previously mentioned, characterizing the agent's risk preferences with RDU introduces *probabilistic risk attitudes* (Wakker, 1994). These attitudes are the influence of the agent's sensitivity to probabilities on his global risk attitude.

To precisely investigate the way in which probabilistic risk attitude affects the effectiveness of stochastic contracts, I distinguish between two components of probability distortion. This decomposition is based in Wakker (2010). The first component captures *motivational* deviations from EUT stemming from pessimist or optimist attitudes toward risk. These factors affect probability evaluations because of the agent's irrational belief that unfavorable outcomes, in the case of pessimism, or favorable outcomes, in the case of optimism, realize more often. Pessimism is represented with a convex weighting function and optimism is represented with a concave weighting function. Figure 1 presents graphical examples of optimism and pessimism.

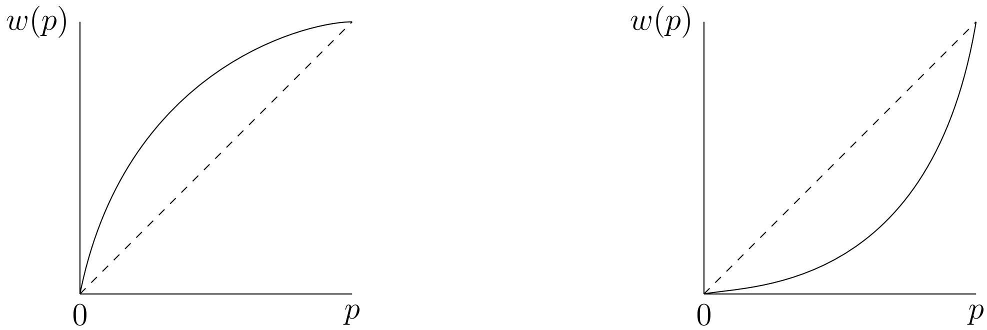
Definition 1. *Pessimism (optimism) is characterized by a probability weighting function* $w(p)$ *with the properties of Assumption 4 and* $\tilde{p} = 0$ ($\tilde{p} = 1$).

It will be useful to determine when an agent suffers from stronger degrees of optimism or pessimism than others. The following definition, due to Yaari (1987) formalizes these comparisons.

Definition 2. *An agent* i *with weighting function* $w(p)_i$ *is more optimistic (pessimistic) than an agent* j *with weighting function* $w(p)_j$ *if* $w(p)_i = \psi(w(p)_j)$, *with* $\psi : [0, 1] \to [0, 1]$ *a strictly increasing, continuous, and concave (convex) function.*

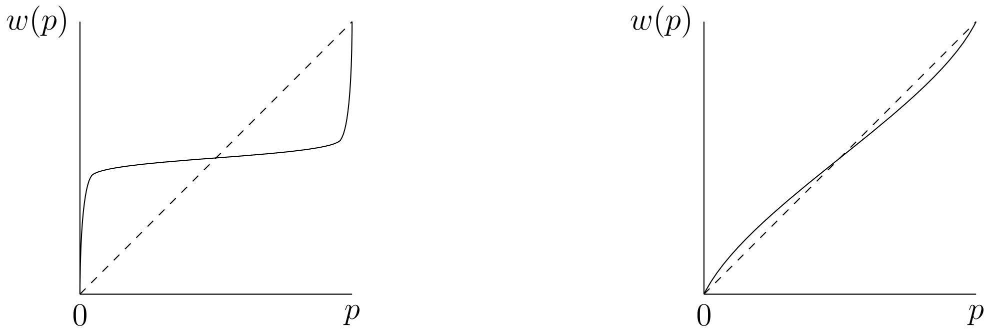
The second component is likelihood insensitivity (Tversky and Wakker, 1995, Wakker, 2001). This component captures the notion that individuals distort probabilities because they are not sufficiently sensitive towards changes in intermediate probabilities and are overly sensitive to changes in extreme probabilities. This deviation from EUT is due to cognitive and perceptual limitations. An extreme characterization of likelihood sensitivity is a step-shaped probability weighting function assigning $w(p) \approx 0.5$ to all probabilities $p \in (0, 1)$, i.e. $w(p)$ with the properties of Assumption 4 and $\hat{p} = 0.5$, $\lim_{p\to 0^+} w(p) = 0.5$, and $\lim_{p\to 1^-} w(p) = 0.5$. An opposing characterization to likelihood insensitivity is that of an agent who is fully sensitive to probabilities, $w(p) = p$. Figure 2 presents graphical examples of different degrees of likelihood insensitivity.[6]

[6] These two phenomena have been addressed in the psychological literature as *curvature* and *elevation* (Gonzalez and Wu, 1999). I instead use the jargon used in economics.



(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

Definition 3. *Likelihood insensitivity is characterized with a probability weighting function $w(p)$ with the properties of Assumption 4 and $\tilde{p} = 0.5$.*

It will become useful to understand the conditions for which an agent suffers from stronger likelihood insensitivity. The following definition based on Tversky and Wakker (1995)'s subadditivity provides us with such comparative.

Definition 4. *An agent i with weighting function $w(p)_i$ is more likelihood insensitive than an agent j with weighting function $w(p)_j$ if $w(p)_i = \phi(w(p)_j)$, where $\phi : [0, 1] \to [0, 1]$ a strictly increasing, continuous, and subadditive function.*

The co-existence of optimism or pessimism, and likelihood insensitivity generates probabilistic risk attitudes that can represented with a probability weighting function with an inverse-S shape. The location of the interior fixed-point, $\hat{p}$ of such weighting function depends on whether the agent displays pessimism or optimism. For instance, a pessimist agent who is also likelihood insensitive, exhibits a $w(p)$ with an interior fixed-point located in the interval $\hat{p} \in (0, 0.5)$.

When comparing $t_s(y)$ and $t_d(y)$ special focus is given to the roles of likelihood insensitivity and optimism. These two components yield different requirements with regard to the implementation of the contracts. In particular, if likelihood insensitivity leads to higher effort. when $t_s(y)$ is offered, then the higher performance obtained with this type of contracts is due to cognitive limitations that can be inherent to the agent's perception of probability, and that can be readily available to the principal. Instead, if optimism yields that $t_s(y)$ 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 generally averse to risk, and thus pessimistic.

2.2. Contract comparisons

In this subsection the considered contracts are compared with respect to the output levels that they generate. To facilitate these comparisons, I make an assumption about the monetary incentives offered by both contracts. In particular, I assume that probability contracts offer, on expectation, the same monetary rewards as piece-rates. Formally, let $A \equiv \frac{a}{p}$, so that $\mathbb{E}(t_s) = ay = t_d$. This equivalence allows me to focus on the incentives produced by probability contracts implemented with different probability p and how these incentives compare to those generated by piece rates offering a similar monetary payment. Notice that a consequence of this assumption is that stochastic contracts nest piece-rates, i.e. as $p \to 1$, then $A \to a$.

Proposition 1 shows that the effectiveness of the probability contract relative to the piece-rate contract depends on the agent's risk attitudes. The proofs of the main results of the paper are relegated to Appendix A.

Proposition 1. *Under Assumptions 1-4,*

(i) for an EUT agent, $w(p) = p$, the stochastic contract implemented with $p < 1$ generates lower output than the piece rate if $b''(t) < 0$.

(ii) for an RDU agent, the stochastic contract implemented with small probabilities generates higher output than the piece rate if $w(p) \geq p \exp\left(\int_p^1 \frac{\rho\left(\frac{ay}{\mu}\right)}{\mu} d\mu\right)$, where $\rho\left(\frac{ay}{p}\right) := -\frac{\frac{ay}{p} b''\left(\frac{ay}{p}\right)}{b'\left(\frac{ay}{p}\right)}$.

Proposition 1 part (i) formalizes the conventional wisdom that introducing risk is counterproductive when the agent's utility exhibits diminishing returns. In such case, the principal would be better off incentivizing the agent with the piece-rate contract. That the agent exhibits EUT preference is fundamental to this result.

Part (ii) of Proposition 1 presents a more interesting result. It states that under RDU preferences, the stochastic contract elicits higher performance if the agent's probability weighting function exhibits a segment where probabilities are sufficiently overweighted. Specifically, probability overweighting should attain the lower-bound $\frac{w(p)}{p} > \exp\left(\int_p^1 \frac{\rho\left(\frac{ay}{\mu}\right)}{\mu} \mathrm{d}\mu\right)$ at some $p \in (0,1)$. That lower-bound reflects the risk attitudes emerging from utility curvature brought at the probability space. If such a probability segment exists, the principal needs to choose a small probability to induce strong risk seeking attitudes in the agent, which in turn generate a preference for risky contracts over deterministic contracts.

The most prominent proposals of probability weighting functions, e.g. Prelec (1998), Goldstein and Einhorn (1987), and Tversky and Kahneman (1992), exhibit extreme sensitivity to small probabilities.[7] That is they assume that $\lim_{p\to 0} \frac{w(p)}{p} = \infty$. Therefore, if $\exp\left(\int_p^1 \frac{\rho\left(\frac{ay}{\mu}\right)}{\mu} \mathrm{d}\mu\right)$ is bounded as $p \to 0$, then the condition in Proposition 1 part (ii) holds at small probabilities for those probability weighting functions and the principal is better off choosing $t_s(y)$ with small probabilities.

I provide intuition of Proposition 1 part (ii) with the following examples.

Example 1: Linear utility. Let $b''(\cdot) = 0$. Then $\rho\left(\frac{ay}{\mu}\right) = 0$ and the lower-bound from Proposition 1 (ii) becomes $w(p) \geq p$. In words, the principal should target probabilities that are overweighted by the agent to guarantee that the stochastic contract generates higher output than the piece rate. To make things more concise, consider Prelec (1998)'s weighting function $w(p) = \exp\left(-\beta(-ln(p))^{\alpha}\right)$. Suppose first that $\alpha = 1$, then $w(p) = p^{\beta}$, a power function. The requirement $p^{\beta} \geq p$ holds for all $p \in (0,1)$ if $\beta < 1$; the weighting function needs to be concave. Suppose instead that $\beta = 1$, then $w(p) = \exp\left(-(-ln(p))^{\alpha}\right)$. The restriction $w(p) > p$ holds in the segment $p \in (0, \frac{1}{e})$ as long as $\alpha < 1$.

Example 2: CRRA utility. Let $\rho\left(\frac{ay}{\mu}\right) = 1 - k$ where $k \in \mathbb{R}$. The lower-bound from Proposition 1 (ii) becomes $w(p) \geq p \exp\left((1-k)\int_1^p \frac{1}{\mu} \mathrm{d}\mu\right) \Leftrightarrow w(p) \geq p^k$. In words, the stochastic contract should target probabilities that are *sufficiently* overweighted by the agent and that overcome the eventual risk aversion from utility curvature. Consider again Prelec (1998)'s weighting function. Suppose first that $\alpha = 1$, then $w(p) \geq p^k \Leftrightarrow \beta < k$ which entails that the weighting function must be more concave than the utility curvature. Consider next $\beta = 1$. The stochastic contract generates higher performance if $\exp\left(-(-ln(p))^{\alpha}\right) > p^k$. If

[7] Well-known non-continuous probability weighting functions such as those proposed by Chateauneuf et al. (2007) and Kahneman and Tversky (1979) also exhibit extreme sensitivity near impossibility and near certainty. In fact, the non-continuity of these weighting functions stems from the observation that subjects exhibit strong distortions of probability for events with extreme probabilities.

$\alpha < 1$, then the overweighting of probabilities in the segment $p \in (0, \frac{1}{e})$ should be stronger than that implied by p^k.

A common property in the previous examples is that the principal needs to target probabilities that are *sufficiently* overweighted by the agent. In Example 1 this happens at any interval where probabilities are overweighted $w(p) > p$. Instead, in Example 2 this happens at probability segments where the weighting function exhibits more concavity than that implied by the utility curvature in the probability space.[8]

In the rest of this section I investigate how the motivational and cognitive components of probability distortion enhance or inhibit the existence of the interval where probabilities are sufficiently overweighted. The following corollaries show that optimism and/or likelihood insensitivity facilitate the existence of overweighting intervals that can be targeted by the agent.

Corollary 1. *Let* $\exp\left(\int_p^1 \frac{r\left(\frac{ay}{\mu}\right)}{\mu} d\mu\right) < B$ *for some* $B < \infty$. *Stronger optimism in the sense of Definition 2 makes the condition for Proposition 1 part (ii) less stringent.*

To understand Corollary 1 notice that optimism, on its own, implies that the the weighting function is concave and that the whole probability interval is overweighted. Since in tha case the weighting function is concave, small probabilities are more strongly overweighted than intermediate or large probabilities. As optimism becomes stronger, that is as the function becomes more concave, the degree at which small probabilities is overweighted increases. Hence, with stronger optimism it is more likely that the probability weighting function exhibits more concavity than the concavity implied by the utility function when brought to the probability space, and that the condition in Proposition 1 part (ii) holds.

Corollary 2. *Let* $\exp\left(\int_p^1 \frac{r\left(\frac{ay}{\mu}\right)}{\mu} d\mu\right) < B$ *for some* $B < \infty$. *Stronger likelihood insensitivity in the sense of Defintion 4 makes the condition for Proposition 1 part (ii) less stringent.*

Likelihood insensitivity, on its own, implies that a probability interval at small or intermediate probabilities is overweighted. As the agent becomes more likelihood insensitive, that is as his weighting function becomes more subadditive, larger probability weights are given to small probabilities. Making more likely that the concavity of the weighting function at small probabilities is larger than that implied by the concavity of the utility when brought to

[8] An example in which this targeting cannot happen and the condition Proposition 1 (ii) does not hold emerges when b belong to the CARA family. Under that assumption, then the condition to guarantee Proposition 1 (ii) becomes $w(p) \geq p \exp\left(Day\left(\frac{1}{p} - 1\right)\right)$, where $D \in \mathbb{R}$. It is evident that the right-hand side of that inequality is larger than one unless $p = 1$.

the probability space. Hence, under stronger likelihood insensitivity the result in Proposition 1 (ii) is more likely to hold, and the principal is better off offering $t_s(y)$ with small probabilities.

To summarize, the two parts in Proposition 1 yields opposing results regarding the effectiveness of the proposed contract. Part (i) formalizes the notion that introducing risk in the agent's compensation is detrimental to performance. Part (ii) presents the conditions under which, an RDU agent is better off with a contract that exposes him to a large amount of risk. Corollary 1 and Corollary 2 show that either optimism or likelihood insensitivity facilitate the effectiveness of stochastic contracts as shown by Part (ii) of Proposition 1.

An alternative analysis that not only incorporates the agent's incentive compatibility, as done throughout this section, but that also considers the solution to the full principal's program is presented in Appendix D. Such analysis confirms the main conclusion presented in this simple model: when contracting with an agent with RDU preferences, the principal is better off implementing stochastic contracts as long as they can target a probability that is sufficiently overweighted by the agent.

3. Experimental Method

3.1. The general setup

The experiment was conducted at Tilburg University's CentERLab. The participants were all students at that university and were recruited using an electronic system. The data consist 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 H.

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.[9] In the first part of the experiment subjects performed a task that demanded their effort and attention. The task consisted of summing five two-digit numbers multiple times. Each summation featured randomly drawn numbers by the computer, ensuring similar levels of difficulty between

[9]This randomization of payments could be a source of concern if subjects distort probabilities. However, as it will be shown in §6, subjects on average exhibit $w(p) \approx 0.5$, so the probability underlying this randomization of payments was not overweighted or underweighted. As it will explained later on, isolation also guarantees that this randomization of payments does not alter

subjects. When a participant knew the answer to the numbers that appeared in his screen, he could submit it using the computer interface. Immediately after submission, a new summation appeared on the computer screen and the participant was invited to solve the new summation. In total, subjects had 10 rounds of four minutes each to complete as many summations as they could.

In the second part of the experiment, the subjects' task was to choose between pairwise lotteries multiple times. 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.[10] Subjects were informed that only one lottery, chosen at random at the end of the experiment, was played and counted toward their earnings.

After the the second part of the experiment was over, subjects were given feedback about their performance and were informed about their earnings in the first part of the experiment. They were also 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.

3.2. Treatments

There were four treatments differing in the type of incentives given to subjects to perform the effort task. Subjects were randomly assigned to one of these treatments. The baseline treatment is *Piecerate*. Subjects assigned to that treatment were paid 0.25 Euros for every correctly solved summation. The other three treatments also offered monetary rewards that depended on individual performance on the task. However, they also included a risk that performance in a round did not count towards their earnings, and that risk was higher or lower depending on the treatment to which the subject was assigned. These treatments seek to represent stochastic contracts implemented with different probabilities.

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 (out of ten) 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 performance in those rounds was paid. This experimental representation of stochastic contracts assumes isolation, which, in

[10] A drawback of this method is that it can be violate incentive compatibility when subjects are aware of the chained nature of the questions that they face. I overcome this disadvantage by adding questions that are not used in the analysis of the data at random, and by randomizing the appearance of the lotteries corresponding to the decision sets 7 to 11, which will be described later on.

this setting, implies that the subjects' decision to exert effort in a round does not take into account decisions made or to be made in other rounds. A failure of isolation would yield similar average performance between the treatments.[11] [12]

As in the theoretical framework, the monetary incentives offered in the treatments Piecerate, LowPr, MePr and HiPr were calibrated such that subjects faced, on expectation, similar monetary incentives across the 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 probability difference that performance in a round is paid between these treatments. By the same token, 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.

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 $p = 0.33$ (See Wakker (2010) pp. 204 for an extensive list of references finding this pattern). If subjects in the experiment follow this empirical regularity, 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 treatments were designed to observe performance differences across the treatments as long as the incentives generated by probability distortions are strong enough.

3.3. Elicitation of risk preference

The second part of the experiment consisted of 11 decision sets. Decision sets 1 to 6 constitute the first step of Abdellaoui (2000)'s methodology, which is based on Wakker and Deneffe (1996). 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 lottery $L = (x_{j-1}, 2/3; 0.5, 1/3)$ and lottery

[11] Isolation is strongly supported by the literature of experimental economics in designs where the *random incentive system*, i.e. paying one round or one exercise at random at the end of the experiment, is implemented (See for instance Baltussen et al. (2012), Hey and Lee (2005) and Cubitt et al. (1998)).

[12] A common misunderstanding regarding the random incentive system is assuming that the independence axiom is a necessary condition to guarantee appropriate experimental measurement, which in the present setup implies that subjects making effort choices as if each decision was paid and in the absence of income effects. While the independence axiom, along with some dynamic principles, *suffices* to guarantee proper measurement, isolation also ensures proper experimental measurement under the random incentive system even when the independence axiom does not hold (Baltussen et al., 2012).

$R = (x_j, 2/3; 0, 1/3)$ for $j = \{1, ..., 6\}$. Indifference was found through bisection.[13]These two lotteries were designed so that the elicited sequence of outcomes yielded equally spaced utility levels for each subject, formally $u(x_j) - u(x_{j-1}) = u(x_{j-1}) - u(x_{j-2})$ for all for $j = \{1, ..., 6\}$. The starting point of the program, x_0, was set at $\frac{2}{5}$th of what a subject earned in the first part of the experiment. This is done to more accurately relate the subjects' risk preference, more specifically their utility curvature which can change with the magnitude of monetary incentives, to their behavior on the first part of the experiment. Subjects were not informed about this calibration.The left panel of Table 1 presents an example illustrating the bisection procedure used for Decision sets 1 to 6.

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 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}. These two lotteries 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 $j = \{2, ..., 6\}$. Again, indifference between these lotteries was found through bisection. The right panel of Table 1 presents an example illustrating the bisection procedure for these decision sets.

4. Hypotheses

The theoretical model generates a set of hypotheses that will be tested with the experiment. When formulating these hypotheses two additional assumptions about subjects' preference are made. First, subjects exhibit linear utility. This is consistent with experimental findings (Abdellaoui et al., 2008, 2007, Abdellaoui, 2000, Wakker and Deneffe, 1996) and with the fact that payments in laboratory experiments are modest. Second, it is assumed that if subjects have RDU preference, they exhibit a probability weighting function that conforms to the common finding of an inverse S-shape with a fixed-point located at $\hat{p} \approx 0.33$. Hence, Example 1 from Section §2.4 becomes essential to understand the following hypotheses.

[13]The bisection procedure works as follows: a subject was required to express his preference between two initial versions of lotteries L and R. After having made a choice, the outcome x_j of lottery R changed as a function of the subject's 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 and R. This process was repeated four times for each decision set.

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

	Left Panel			**Right Panel**		
iteration #	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=$(x_6, 0.50; 1, 0.5)$	$[0, 1]$	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=$(x_6, 0.75; 1, 0.25)$	$[.5, 1]$	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=$(x_6, 0.87; 1, 0.13)$	$[.75, 1]$	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=$(x_6, 0.81; 1, 0.19)$	$[.75, .87]$	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=$(x_6, 0.85; 1, 0.15)$	$[.81, .87]$	L
		$x_1 \in$[4.21, 4.38]			$p_1 \in$[0.85, 0.87]	

Note: This table illustrates the bisection method used to elicit utility and probability functions. The lotteries in this table are expressed in the form $(m, p; n, 1-p)$ where m and n 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.

The first hypothesis is based on Prediction 1 part (ii), which demonstrates that stochastic contracts implemented with small probabilities can outperform piece-rates as long as $w(p) > p$. Given the aforementioned assumption on the subjects' shape of the weighting function, the following hypothesis captures how performance is expected to compare across the treatments.

Hypothesis 1. *Subjects with RDU preference exhibit average performance levels across treatments that conform to the ranking:*

$$LowPr > MePr = Piecerate > HiPr.$$

Empirical support in favor of Hypothesis 1 does not conclusively validate the model. That is because factors other than probability weighting might generate these performance differences. Hence, if the model is correct, performance differences across treatments, if any, should be explained by the subjects' tendency to distort probabilities . This is captured in the following hypothesis.

Hypothesis 2. *(i) Subjects assigned to LowPr who overweight small probabilities exhibit higher performance as compared to subjects assigned to Piecerate.*

(ii) Subjects assigned to HiPr who underweight intermediate probabilities exhibit lower performance as compared to subjects assigned to Piecerate.

Finally, if Hypotheses 1 and 2 are corroborated by the data, we want to understand understand whether motivational or cognitive factors of probability distortion are behind these differences in perfromance. Corollary 1 predicts that optimism suffices to guarantee $w(p) > p$. Therefore, optimistic agents should display higher performance when exposed to more risk.

Hypothesis 3. *Optimistic subjects exhibit higher performance when assigned to treatments that expose them to more risk.*

Furthermore, Corollary 2 predicts that likelihood insensitivity, regardless of whether they also exhibit optimism or pessimism, suffices to guarantee probability overweighting $w(p) > p$ in the segment $p \in (0, \hat{p})$. Thus, likelihood insensitive subjects must exhibit treatment differences according to the ranking predicted by Hypothesis 1.

Hypothesis 4. *Likelihood insensitive subjects exhibit performance differences conforming to the ranking in Hypothesis 1.*

5. Results

5.1. Treatment effects

In this subsection I compare performance in the effort task across the treatments. Performance is 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 stochastic 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 ($t(84.454) = 2.361, p = 0.010$).[14] The effect size of this difference in performance is of 0.5 standard deviations which is significant at the 5% confidence level.[15]

In contrast, stochastic 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.[16] These

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

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

[16] 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.

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

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. Conjectures about this partial confirmation of Hypothesis 1 are provided at the end of the subsection.

Among the treatments representing stochastic contracts, 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 ($t(79.575) = 1.478, p = 0.0716$).[17] Therefore, statistical inference using pairwise testing suggests that LowPr generates highest average performance, while the other treatments produce similar performance.

I estimate regressions of performance on treatment dummies, dummies that capture different possible shapes of the utility function, and dummies that capture different possible shapes of the weighting function. These regressions have the purpose of establishing the robustness of the aforementioned treatment effects while controlling for average risk attitude. Thus, if the treatment effects found above are robust to the inclusion of these controls, the performance differences between the treatments are not an artifact of more risk seeking or less risk averse subjects assigned to some of the treatments.

The dummy variables included in the regressions reflected the shape of utility and weighting functions and were constructed on the basis of the subject's answers to the second part of the experiment. Utility functions were classified as having linear, concave, convex, or mixed shape. Details of this classification are provided in Appendix E.[18] Probability weighting functions were classified as displaying lower subadditivity (LS, from here onward) and/or upper subadditivity (US, from here onward). A weighting function with LS assigns

[17] 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.

[18] In short, a variable $\Delta_j^{''} := (x_j - x_{j-1}) - (x_{j-1} - x_{j-2})$ 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.

larger decision weights to low probabilities than to intermediate probabilities. A weighting function with US assigns larger decision weights to large probabilities than to intermediate probabilities.[19] In some specifications a different classification for weighting functions is used, such classification was based on 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.[20] Details of these two classifications are provided in Appendix F.

Table 3 presents the regression 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.

The first result is that the stochastic contract with small probability, exposing the agent to more risk, yields higher performance than the other contracts.

Result 1. *Average performance across treatments conforms to the ranking:*

$$LowPr > MePr = Piecerate = HiPr.$$

A possible explanation to Result 1 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 they can be a source of demotivation for subjects toward the last rounds of the experiment. That explanation also accommodates the result that LowPr generates higher performance than MePr and HiPr because, by paying less rounds, that treatment is more effective in minimizing income effects. If income effects are the reason behind the treatment effects, then we should not observe performance differences across treatments 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 show that in the first round subjects assigned to LowPr achieve1.6 higher average summations as compared to subjects in Piecerate ($p = 0.019$). Subjects in LowPr also exhibit higher average performance in the first round as compared to subjects in MePr ($F(1, 171) = 2.21, p = 0.069$ and subjects

[19] Specifically, a subject in the experiment exhibited LS when $w^{-1}\left(\frac{1}{6}\right) < w^{-1}\left(\frac{2}{6}\right) - w^{-1}\left(\frac{1}{6}\right)$. Also, a subject exhibited US when $1 - w^{-1}\left(\frac{5}{6}\right) < w^{-1}\left(\frac{5}{6}\right) - w^{-1}\left(\frac{4}{6}\right)$.

[20] A subject had a stronger possibility 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.

in HiPr ($F(1, 171) = 5.04, p = 0.026$). Hence, the aforementioned performance differences emerge in the absence of income effects.

To conclude, the data on performance in the effort task *partly* supports Hypothesis 1. Recall that Hypothesis 1 was structured around the common finding that individuals overweight probabilities smaller than $p = 0.33$ and underweight all probabilities thereafter. Instead, the analyses presented in this subsection suggest 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 subsection, it is shown that subjects indeed display an average weighting function with such a shape.

5.2. Probability weighting functions

In this subsection I analyze the data obtained in the second part of the experiment. These data feature the subjects' choices between pairwise lotteries designed to elicit the subjects' utility and probability weighting functions. As described in §3, decision sets 1 to 6 elicited the sequence $\{x_1, x_2, x_3, x_4, x_5, x_6\}$, which captures the subjects' preferences over the outcomes of the lotteries. Various analyses of these data demonstrate that the great majority of subjects exhibit linear utility functions, which is in line with the findings of Wakker and Deneffe (1996), Abdellaoui (2000), Abdellaoui et al. (2008), and Abdellaoui et al. (2011), and is consistent 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 stochastic contracts, I relegate the analysis of the shape of utility functions to Appendix E.

Decision sets 7 to 11 elicited 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 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 relevant factors behind the documented efficiency of stochastic contracts.[21] Alternative analyses of these data, including individual

[21] Comparisons across studies, see argument ii, 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.

analyses and non-parametric analyses, are presented in Appendix F.

Various and well-known proposals of probability weighting are assumed when performing the regressions. Specifically, I use the neo-additive probability weighting function (Chateauneuf et al., 2007), Tversky and Kahneman (1992)'s probability weighting function, Prelec (1998)'s two-parameter probability weighting function, and Goldstein and Einhorn (1987)'s log-odds probability weighting function.That various functionals of $w(p)$ are assumed ensures robustness, i.e. that the estimated results do not stem from the underlying assumptions of a particular functional form.

The resulting estimates are presented in Table 4 and Figure 3. Under all specifications it is found that subjects display an average probability weighting function with a strong inverse-S shape and, for those weighting functions with two parameters, less pessimism than previously documented. Detailed comparisons with respect to existing studies are provided in Appendix G. This shape of the weighting function along with the documented linearity of utility constitutes our second result.

Result 2. *Subjects exhibit linear utility and a probability weighting function with strong inverse S-shape and moderate pessimism.*

The coexistence of strong inverse-S shape and moderate pessimism properties produces a pattern a weighting function that strongly overweights small probabilities and moderately distorts intermediate probabilities. For example, using the estimates of Panel 1 in Table 4 it can be established that for the subjects in this experiment the probability $p = 0.10$ is on average perceived to be $w(0.10) = 0.25$, while the probabilities $p = 0.30$ and $p = 0.5$ are on average perceived to be $w(0.30) = 0.363$ and $w(0.50) = 0.477$, respectively. These patterns of probability distortion accommodate 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 shape of the probability weighting functions of subjects explains the treatment effects.

5.3. Overweighting of probabilities, likelihood insensitivity, and the treatment effect

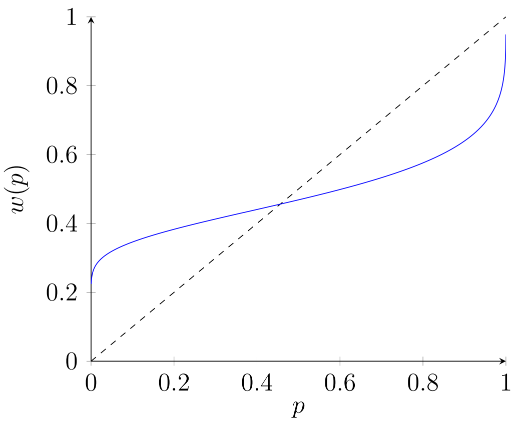
This subsection reconciles the results of the two parts of the experiment. First, I present empirical evidence supporting Hypothesis 2 by demonstrating that the higher average performance of subjects assigned to LowPr is caused by their tendency to overweight small probabilities. Second, I show that likelihood insensitivity, alone, explains the treatment effects documented in §5.1, validating Hypothesis 4.

To empirically verify the validity of Hypothesis 2, I extend the statistical models presented

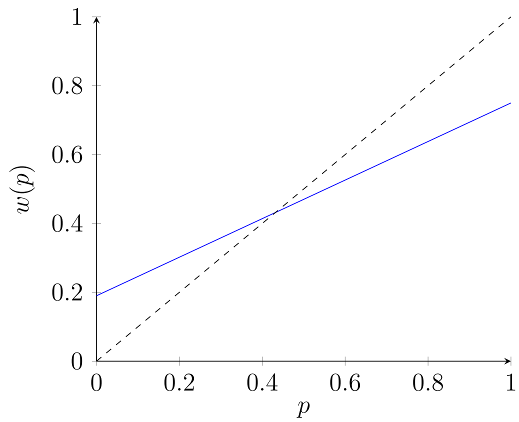
Table 4: Parametric estimates of the weighting function

Panel 1 $w(p) = c + sp$		
	$\hat{c}$	$\hat{s}$
	0.194***	0.566***
	(0.021)	(0.036)
Log-Likelihood		220.288
N		860
Panel 2 $w(p) = \frac{p^{\psi}}{\left(p^{\psi}+(1-p)^{\psi}\right)^{\frac{1}{\psi}}}$		
		$\hat{\psi}$
		0.598***
		(0.016)
Adj. R^2		0.838
N		860
Panel 3 $w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma}+(1-p)^{\gamma}}$		
	$\hat{\gamma}$	$\hat{\delta}$
	0.281***	0.921***
	(0.025)	(0.020)
Adj. R^2		0.863
N		860
Panel 4 $w(p) = \exp\left(-\beta(-\ln(p))^{\alpha}\right)$		
	$\hat{\alpha}$	$\hat{\beta}$
	0.284***	0.841***
	(0.025)	(0.015)
Adj. R^2		0.864
N		860

Note: This table presents estimates of the average probability weighting function of subjects when different parametric forms are assumed. Panel 1 presents the maximum likelihood estimates of the equation $w(p) = c + sp$ 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 $w(p) = \frac{p^{\psi}}{(p^{\psi}+(1-p)^{\psi})^{\frac{1}{\psi}}}$. The third panel presents the non-linear least squares estimates of the parametric form $\frac{\delta p^{\gamma}}{\delta p^{\gamma}+(1-p)^{\gamma}}$. The last panel presents the non-linear least squares estimates of the function $w(p) = \exp\left(-\beta(-\ln(p))^{\alpha}\right)$. 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.



(a) using Prelec (1998)



(b) using Chateauneuf et al. (2007)

Figure 3: Average probability weighting functions

in Table 3 by including interactions between the dummy variable indicating assignment to LowPr and variables that capture the shape of the probability weighting function. Because of the nature of the observed treatment effect, I focus on variables that indicate whether a subject exhibits a weighting function with overweighting of small probabilities. Specifically, I use the variables LS, Possibility, and $\text{Overweight}_{p=\frac{1}{6}}$. The first two variables were already defined in §5.1. The last variable takes a value of one if a subject overweights the probability $p = \frac{1}{6}$ and zero otherwise.[22]

Column (1) in Table 5 presents the OLS estimates when LS is used. 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. Columns (2) and (3) in Table 5 show that this conclusion is robust to using other variables that capture overweighting of small probabilities.[23] That overweighting

[22]These variables relate in the following way: a subject for whom LS takes a value of one, overweights the probability $p = \frac{1}{6}$ and might exhibit a possibility effect that is stronger than the certainty effect. Similarly, a subject for whom Possibility takes a value of one overweights $p = \frac{1}{6}$ and is likely to exhibit LS.

[23]Unlike the analyses in which LS and Possibility were used to capture probability overweighting, the coefficient associated to LowPr remains significant when $\text{Overweight}_{p=\frac{1}{6}}$ is used. This significance suggests that the treatment effect is not entirely captured by the mere tendency of subjects to overweight $p = \frac{1}{6}$, but instead by their tendency to overweight the probability $p = \frac{1}{6}$ *relative* to other elicited probabilities. For example, $p = \frac{1}{2}$ and $p = \frac{5}{6}$. This is already suggestive that overweighting of probabilities due to likelihood insensitivity, which entails a relative overweighting of small probabilities with respect to medium-sized

of probabilities explains the treatment effect in its totality is our third result.

Result 3. *Subjects in LowPr who overweight small probabilities exhibit higher performance with respect to subjects in Piecerate.*

Finally, I investigate the influence of likelihood insensitivity and optimism on performance to evaluate the validity of Hypotheses 3 and 4. To that end, subjects are first classified as likelihood insensitive and/or optimistic. Following Wakker (2010) and Abdellaoui et al. (2011), I estimate for each subject, i, the following neo-additive functional:

$$w(p_{ij}) = c_i + s_i p_{ij} + e_i, \tag{4}$$

where j indicates the elicited probability and e_i is an error term. The magnitude of the estimate $\hat{s}_i$ indicates subject's i sensitivity to probabilities. If $\hat{s}_i < 1$, the subject is not sufficiently responsive to changes in probabilities and is classified as likelihood insensitive. Instead, if $\hat{s}_i \geq 1$, the subject is sufficiently or too sensitive to changes in probabilities and is classified as likelihood sensitive. I find that 102 subjects are likelihood insensitive and 61 subjects are classified as likelihood sensitive.[24] Importantly, the degree of likelihood insensitivity is balanced across treatments.

In addition, the magnitude of $\hat{c}_i$ and that of the sum $\hat{c}_i + \hat{s}_i$ determine subject's i optimism. Whenever $\hat{c}_i > 0$ and $\hat{c}_i + \hat{s}_i \leq 1$, that 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 $\hat{c}_i - \hat{s}_i \leq 1$, that subject exhibits pessimism. I find that 80 subjects display optimism while 32 subjects display pessimism.[25] The degrees of optimism are also balanced across treatments.

Binary variables capturing the above classifications are added to the regressions presented in Table 3. Also, interactions between assignment to LowPr and the variables indicating whether a subject is likelihood insensitive and whether a subject exhibits optimism are included in the regressions. These interactions allow me to evaluate the strength of the treatment effect among likelihood insensitive subjects as well as the strength of the treatment effect among optimistic subjects.

The resulting estimates are presented in columns (1) and (2) of Table 6. I find empirical support for Hypothesis 4. Specifically, I find that likelihood insensitive subjects assigned to

probabilities, better explains the treatment effect.

[24] Nine subjects had a negative estimated parameter $\hat{s}_i < 0$ which has no clear interpretation and are thus left unclassified.

[25] 60 subjects are left unclassified since they display $\hat{c}_i + \hat{s}_i > 1$, which has no clear interpretation. In an alternative classification that ignores that restriction, the unclassified subjects become either optimistic or pessimistic. Under such relaxation, the relevant result that likelihood insensitivity explains the treatment effects, presented later on, also emerges.

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 variable	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 $\hat{s} + \hat{c} \leq 1$	$\hat{c} > 0$ and $\hat{s} + \hat{c} \leq 1$	$\hat{\beta} < 1$	$\hat{\delta} > 1$
Parametric family	Neo-additive	Neo-additive	Prelec (1998)	Goldstein and Einhorn (1987)

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.

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 result that likelihood insensitivity ensures the efficiency of stochastic contracts when they 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, on its own, is unable to explain the treatment effects documented in §5.1. This result invalidates Hypothesis 3.

For the sake of robustness, a similar exercise is performed using the probability weighting functions proposed by Prelec (1998) and Goldstein and Einhorn (1987). These functions also contain, each, two parameters. One of the parameters mainly influences likelihood insensitivity while the other parameter mainly influences optimism. On the basis of the magnitude of these parameters, I classify whether subjects are likelihood insensitive and/or optimistic. Columns (2) and (3) in Table 6 present the results of regressions when these alternative classifications are used. Altogether, the regression estimates corroborate the aforementioned results and, thus, the empirical validity of Hypothesis 4.

Result 4. *Likelihood insensitivity and not optimism explains the treatment effects.*

6. Applications and Discussion

This paper demonstrated that stochastic contracts motivate individuals who distort probabilities at a greater extent than more traditional contracts. To achieve this result, the principal is required to specify that the performance-contingent outcome realizes with a small probability. This implementation of the contract induces risk seeking attitudes in the RDU agent and, as a consequence, generates a preference for risky compensation schemes. I show that the agent's insensitivity to likelihoods, the cognitive component of probability weighting, explains this result.

While stochastic contracts are abstract and might be forbidden in countries where gambling is illegal, their incentives can be implemented using well-known tools of personnel economics. In the following, I discuss some ways to bring these incentives to practice.

- **Bonuses.** Consider a setting in which the agent's effort on the task and the resulting output relate stochastically. The principal can take advantage of this stochastic relationship by offering a contract that pays a bonus in the contingency that a threshold output level is attained. The findings of this paper show that the principal should set a high threshold output level, yielding a small probability of attainment, and the associated bonus should be large enough. The findings of this paper show that an agent with risk preferences characterized by RDU suffering from likelihood insensitivity is more motivated under this contract than under a cost-equivalent linear contract. Appendix B formalizes this application.

- **Entrepreneurship and autonomy.** The risk neutral principal can sell the agent a risky project taking place within the firm, making him residual claimant of that project. The RDU agent suffering from likelihood insensitivity will buy the project as he overweights the probability that it will deliver large revenue streams. Moreover, the results of his paper suggest that the agent will be more motivated when he is the residual claimant of the project, and is thus fully exposed to risk, than when the principal protects him from risk by offering a contract for his work on that project.

- **Stock options.** A volatile firm can offer CEOs compensation plans that include stock options. Naturally, when the contract is signed, the future stock price is unknown and this is the uncertainty the principal can take advantage of. First, as shown by Spalt (2013) the agent with RDU preferences will accept these contracts despite the firm being risky. These risk seeking attitudes emerge because the agent overweights the probability associated to obtain large gains from calling the option. Second, as the results of this paper suggest, when the agent's higher effort shifts the *perceived* distribution of future stock prices, this contract generates higher motivation than other standard performance-pay contracts. That is because the perceived contribution of the agent's effort to the probability of high future stock prices will be overweighted, inflating the perceived benefits of supplying high levels of effort under the proposed incentive scheme.

A common property among the aforementioned applications is that the incentives of stochastic contracts are brought to practice using natural sources of uncertainty: output realizations given effort, future stock prizes, and the success of projects. Indexing the outcomes of the contract to natural and uncertain events allows the principal to circumvent the problem of lack credibility that might arise if she were to generate the contract's uncertainty using an artificial device, e.g. a roulette or dice. Assuring that the principal has no influence over the realization of uncertainty, and by extension over the outcome to be paid, allows her to more credibly commit to the contract. Moreover, recent research suggests that individuals display more insensitivity toward ambiguity than toward risk (Baillon et al., 2018, Abdellaoui et al., 2011). Since likelihood insensitivity was found to be the main explanation for effectiveness of the contract, implementing the contract using natural sources of ambiguity can potentially enhance the gains from its usage.

The present study has several limitations that open avenues for future research. First, it is assumed through the paper that the principal is fully informed about the agent's risk attitudes. Future research could relax this assumption. Specifically, the model presented in §2 can be extended to incorporate a stage of adverse selection. The principal's task in that framework is not only to motivate agents but also to screen them according to their risk preference. In the simplest version of that model there are two types of agents EUT and RDU agents with likelihood insensitivity. The principal can screen and motivate these agents with a menu of contracts containing a deterministic contract, that targets agents with EUT preferences and concave utility, and a stochastic contract with a small probability, targeting agents with RDU preferences.

Second, this paper considered a static setting. A more comprehensive investigation of the probability contract could examine its incentives in a setting of repeated interaction between principal and agent. On the theoretical side, it has been shown that optimal contracts in this setting exhibit properties that depend on the agent's expectation over the contracting span. When expectations are distorted, due to probability weighting, it is unclear whether these properties are enhanced or fade away, and whether these conditions are more favorable for the implementation of stochastic contracts as compared to static settings. On the empirical side, an extension to a setup that admits repeated implementations of the contract allows for a more robust analysis of its incentives. That is because such setting could shed light on whether and how probability weights are adjusted with experience, and in turn how the incentives induced by stochastic contracts change.

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

Abdellaoui, M., Bleichrodt, H., and Paraschiv, C. (2007). Loss Aversion Under Prospect Theory: A

Parameter-Free Measurement. *Management Science*, 53(10):1659–1674.

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

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.

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.

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

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.

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

Corgnet, B., Gómez-Miñambres, J., and Hernán-Gonzalez, R. (2018). Goal setting in the principal-agent model: Weak incentives for strong performance. *Games and Economic Behavior*, 109:311–326.

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

Ederer, F., Holden, R., and Meyer, M. (2018). Gaming and strategic opacity in incentive provision. *RAND Journal of Economics*, 49(4):819–854.

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.

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

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

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.

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.

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

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

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

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.

Laffont, J. J. and Martimort, D. (2002). *The theory of incentives: The Principal-Agent Model*. Number January 2001. Princeton University Press.

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

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

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.

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

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

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

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.

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

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

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 and calculations

Proposition 1

Proof. Assumption 2 and the equivalence $A = \frac{a}{p}$ are used to rewrite equation (2) as follows,

$$RDU(t_s) = w(p)b\left(\frac{ay}{p}\right) - c\left(\frac{y}{\theta}\right). \tag{5}$$

The optimal output level y_R^{**} chosen by the agent with RDU preferences when working under t_s satisfies the following first order condition:

$$b'\left(\frac{ay_R^{**}}{p}\right)\frac{w(p)}{p}a - c'\left(\frac{y_R^{**}}{\theta}\right)\frac{1}{\theta} = 0. \tag{6}$$

That y_R^{**} is a maximum requires that the second order condition is negative. Formally,

$$b''\left(\frac{ay_R^{**}}{p}\right)\frac{w(p)}{p^2}a^2 - c''\left(\frac{y_R^{**}}{\theta}\right)\frac{1}{\theta^2} < 0. \tag{7}$$

If the principal sets $p = 1$, equation (5) becomes $U(t_s) = b(ay) - c\left(\frac{y}{\theta}\right)$, and the corresponding optimal output level, y^*, satisfies the following first-order condition:

$$b'(ay^*)\frac{w(p)}{p}a - c'\left(\frac{y^*}{\theta}\right)\frac{1}{\theta} = 0. \tag{8}$$

To investigate how y_R^{**} and y^* relate, implicitly differentiate (6) with respect to p and y_R^{**} which gives:

$$\frac{\mathrm{d}y_R^{**}}{\mathrm{d}p} = \frac{\left(\frac{w'(p)p - w(p)}{p^2}\right)ab'\left(\frac{ay_R^{**}}{p}\right) - \frac{w(p)a^2y}{p^3}b''\left(\frac{ay_R^{**}}{p}\right)}{c''\left(\frac{y_R^{**}}{\theta}\right)\frac{1}{\theta^2} - b''\left(\frac{ay_R^{**}}{p}\right)\frac{w(p)a^2}{p^2}}. \tag{9}$$

Notice that if y_R^{**} is a maximum, the denominator of (9) is positive due to (7).

To prove part (i) of the Proposition, the sign of (9) is analyzed under the assumption $w(p) = p$. In that case, equation (9) collapses to:

$$\frac{\mathrm{d}y_R^{**}}{\mathrm{d}p} = \frac{-\frac{a^2y_R^{**}}{p^2}b''\left(\frac{ay_R^{**}}{p^2}\right)}{c''\left(\frac{y_R^{**}}{\theta}\right)\frac{1}{\theta^2} - b''\left(\frac{ay_R^{**}}{p}\right)\frac{a^2}{p}}. \tag{10}$$

The restrictions $a > 0$, $y_R^{**} > 0$, and $p \in [0,1]$ along with equation (10) imply that, at the optimum, $\frac{\mathrm{d}y_R^{**}}{\mathrm{d}p} > 0$ if $b''\left(\frac{a^2y_R^{**}}{p^2}\right) < 0$. In such case, $y^* \geq y_R^{**}$ and the principal is better off giving t_s with $p = 1$, which is equivalent to offering contract t_d. Instead, if $b''\left(\frac{a^2y_R^{**}}{p^2}\right) > 0$, then $\frac{\mathrm{d}y_R^{**}}{\mathrm{d}p} < 0$. In that case the principal offers t_s with $p \to 0^+$, exposing the agent to a large degree of risk.

To prove part (ii), consider the general case in which $w(p)$ acquires the properties of Assumption 4. A necessary condition for equation (9) to be negative, at the optimum, is:

$$w'(p) \leq \frac{w(p)}{p}\left(1 - \rho\left(\frac{ay_R^{**}}{p}\right)\right), \tag{11}$$

where $\rho(x) := -\frac{xb''(x)}{b'(x)}$, the coefficient of relative risk aversion. The requirements for $w(p)$ to satisfy (11) are found using Grönwall's lemma. First, I find the solution to:

$$v'(p) = \frac{v(p)}{p}\left(1 - \rho\left(\frac{ay_R^{**}}{p}\right)\right), \tag{12}$$

where $v(p)$ is a weighting function that exhibits the same properties as $w(p)$. The solution to that ordinary differential equation is given by:

$$\int \frac{v'(p)}{v(p)}\mathrm{d}p = \int \frac{1}{p}\left(1 - \rho\left(\frac{ay_R^{**}}{p}\right)\right)\mathrm{d}p \Leftrightarrow v(p) = \exp\left(-\int_p^1 \frac{1 - \rho\left(\frac{ay_R^{**}}{\mu}\right)}{\mu}\mathrm{d}\mu\right). \tag{13}$$

Where the fact that $v(1) = 1$, implies that the integration constant or initial condition of the differential equation is $c = 0$. Second, the way in which $w(p)$ and $v(p)$ relate is investigated computing the following derivative of their ratio:

$$\frac{\mathrm{d}\left(\frac{w(p)}{v(p)}\right)}{\mathrm{d}p} = \frac{(v(p)w'(p) - w(p)v'(p))}{v(p)^2} = \frac{v(p)\left(w'(p) - \frac{w(p)}{p}\left(1 - \rho\left(\frac{ay_R^{**}}{p}\right)\right)\right)}{v(p)^2}, \tag{14}$$

where the second equality results from replacing $v'(p)$ using equation (13). Using the above equation together with (11) it can be established that $\frac{\mathrm{d}\left(\frac{w(p)}{v(p)}\right)}{\mathrm{d}p} \leq 0$. Hence, the minimum of $\frac{w(p)}{v(p)}$ must be attained at $p = 1$ and it must be that for any $p \in (0, 1]$:

$$\frac{w(p)}{v(p)} \geq \frac{w(1)}{v(1)} = 1. \tag{15}$$

Using the above inequality it can be established that the solution to (11) is bounded by (12), in the following way:

$$w(p) \geq p \exp\left(\int_p^1 \frac{\rho\left(\frac{ay_R^{**}}{\mu}\right)}{\mu}\mathrm{d}\mu\right). \tag{16}$$

Hence, long as $w(p)$ satisfies the inequality in equation (16), then $\frac{\mathrm{d}y_R^{**}}{\mathrm{d}p} < 0$, and the principal must offer t_s with $p \to 0^+$. Alternatively, when the inequality in (16) cannot hold, then the principal must offer t_s with $p = 1$, which is equivalent to the piece-rate contract.

■

Corollary 1

Proof. Let $o(p) : [0, 1] \to [0, 1]$ be a probability weighting function with $\hat{p} = 1$, i.e. optimism in the sense of Definition 1. Under these conditions $o(p) > p$ holds for all $p \in (0, 1)$. Thus, optimism generates probability overweighting, a necessary, condition for Proposition 1 part (ii). However, the sufficient condition for Proposition 1 part (ii) to hold under optimism is

$$o(p) \geq p \exp\left(\int_p^1 \frac{r\left(\frac{ay_R^{**}}{\mu}\right)}{\mu} \mathrm{d}\mu\right). \tag{17}$$

By Definition 2, stronger optimism implies $o(p)$ becoming more concave, or, equivalently, that the expression $\left|\frac{o''(p)}{o'(p)}\right|$ becomes larger. Extreme levels of optimism entail $\lim_{p\to 0^-}\left|\frac{o''(p)}{o'(p)}\right| = \infty$, while the smallest possible degree of optimism implies $\lim_{p\to 0^-}\left|\frac{o''(p)}{o'(p)}\right| = \epsilon$ for arbitrarily small $\epsilon > 0$. These two extreme characterizations of optimism, together with the assumptions that $o(p)$ is $\mathcal{C}^2$ and $\exp\left(\int_p^1 \frac{r\left(\frac{ay_R^{**}}{\mu}\right)}{\mu}\mathrm{d}\mu\right) < B$ for some $B < \infty$, imply that there exists a weighting function with degree of concavity such that (17) holds with equality as $p \to 0^+$. Denote that threshold probability weighting function by $\hat{o}(p)$. Any probability weighting function such that $\left|\frac{o''(p)}{o'(p)}\right| \geq \left|\frac{\hat{o}''(p)}{\hat{o}'(p)}\right|$, that is exhibiting more or equal optimism than the threshold weighting function $\hat{o}(p)$ ensures (17). In such case, the principal chooses $p \to 0^+$ because that (17) holds implies $\frac{\mathrm{d}y_R^{**}}{\mathrm{d}p} < 0$. Hence, stronger optimism in the sense of Definition 2 implies that $\left|\frac{o''(p)}{o'(p)}\right|$ increases so that the concavity of $o(p)$ approaches or surpasses that of $\hat{o}(p)$ as $p \to 0^+$.

■

Corollary 2

Proof. Let $l(p) : [0, 1] \to [0, 1]$ be a probability weighting function with likelihood insensitivity as defined in Definition 3. Under these conditions $l(p) > p$ holds in the probability interval $p \in (0, 0.5)$. Thus, a probability weighting function with likelihood insensitivity $l(p)$ exhibits probability overweighting in some non-empty probability interval, a necessary condition for Proposition 1 part (ii) to hold. The sufficient condition for Proposition 1 part (ii) states that a weighting function with likelihood insensitivity must satisfy:

$$l(p) \geq \exp\left(\int_p^1 \frac{r\left(\frac{ay_R^{**}}{\mu}\right)}{\mu} \mathrm{d}\mu\right). \tag{18}$$

By Definition 4, stronger likelihood insensitivity imply $\lim_{p\to 1^-} l'(p)$ and $\lim_{p\to 0^+} l'(p)$ becoming larger. At the limit extreme likelihood insensitivity implies $\lim_{p\to 0^+} l'(p) = \infty$, while the mildest form of likelihood insensitivity is $\lim_{p\to 0^+} l'(p) = \epsilon$ for ϵ arbitrarily close to one. These results, together with the assumptions that $l(p)$ is $\mathcal{C}^2$ and $\exp\left(-\int_p^1 \frac{1-r\left(\frac{ay_R^{**}}{\mu}\right)}{\mu}\mathrm{d}\mu\right) < B$ for some $B < \infty$, imply that there exists a weighting function with a degree of likelihood insensitivity such that (18) holds with equality as $p \to 0^+$. Denote that weighting function by $\hat{l}(p)$. Any weighting function such that $\lim_{p\to 0^+} \hat{l}(p) > \lim_{p\to 0^+} l(p)$, that is exhibiting the more or equal likelihood insensitivity than $\hat{l}(p)$, also ensures (18). In such case, the principal chooses $p \to 0^+$ because that (18) holds implies that $\frac{\mathrm{d}y_R^{**}}{\mathrm{d}p} < 0$. Hence, stronger likelihood insensitivity means that $\lim_{p\to 0^+} l'(p)$ increases so that the weights given to small probabilities surpasses or approaches that of $\lim_{p\to 0^+} \hat{l}'(p)$.

■

Appendix B. Stochastic output and applications

In this Appendix I show that the main result of the theoretical framework applies when the relationship between effort, e, and output y, is assumed to be stochastic. Also, taking advantage of the introduced framework I formally introduce one of the applications of the contract discussed in the last section of the main body of the paper.

Let $y \in [0, \bar{y}]$ be a stochastic variable distributed according to the probability density function $g(y|e)$ which admits a cumulative density function $G(y|e)$. To make things simple assume that the agent's action consists on exerting a high effort level or a low effort level $e = \{e_L, e_H\}$. In this setting, only high effort is costly, that is $c(e) = c$ if $e = e_H$ and $c(e) = 0$ if $e = e_L$. Finally, as it is standard in the literature, I assume that output and effort relate according to the monotone likelihood ratio property: $\frac{\partial}{\partial y}\left(\frac{g(y|e_H)}{g(y|e_L)}\right) > 0$.

The following proposition shows that the agent is more motivated under $t_s(y)$ when implemented with small probabilities.

Proposition 2. *Under Assumptions 2, 3, 4, and stochastic y, an agent with RDU preferences exerts high effort more often under the stochastic contract implemented with small probabilities if $\frac{w(p)}{p} > 1$.*

Proof. The RDU agent is motivated under $t_d(y)$ if the following incentive compatibility condition holds:

$$\int_0^{\bar{y}} b(ay)\mathrm{d}w\left(1 - G(y|e_H)\right) - \int_0^{\bar{y}} b(ay)\mathrm{d}w\left(1 - G(y|e_L)\right) \geq c. \tag{19}$$

Instead, the RDU agent is motivated under $t_s(y)$ when the following inequality holds:

$$w(p)\left(\int_0^{\bar{y}} b\left(\frac{ay}{p}\right)\mathrm{d}w\left(1 - G(y|e_H)\right) - \int_0^{\bar{y}} b\left(\frac{ay}{p}\right)\mathrm{d}w\left(1 - G(y|e_L)\right)\right) \geq c. \tag{20}$$

Denoting the non-additive expectation as $\tilde{\mathbb{E}}\left(y|e\right) := \int_0^{\bar{y}} y\mathrm{d}\left(1 - G(y|e)\right)$, the stochastic contract elicits e_H more often as long as:

$$w(p)\left(\tilde{\mathbb{E}}\left(b\left(\frac{ay}{p}\right)\middle| e_H\right) - \tilde{\mathbb{E}}\left(b\left(\frac{ay}{p}\right)\middle| e_L\right)\right) > \tilde{\mathbb{E}}\left(b\left(ay\right)\middle| e_H\right) - \tilde{\mathbb{E}}\left(b\left(ay\right)\middle| e_L\right). \tag{21}$$

Rearranging and multiplying by $\frac{1}{p}$ gives:

$$\frac{w(p)}{p} > \frac{1}{p}\left(\frac{\tilde{\mathbb{E}}\left(b\left(ay\right)\middle| e_H\right) - \tilde{\mathbb{E}}\left(b\left(ay\right)\middle| e_L\right)}{\tilde{\mathbb{E}}\left(b\left(\frac{ay}{p}\right)\middle| e_H\right) - \tilde{\mathbb{E}}\left(b\left(\frac{ay}{p}\right)\middle| e_L\right)}\right). \tag{22}$$

The derivative of left-hand side of (22) with respect to p is

$$\frac{\partial}{\partial p}\left(\frac{w(p)}{p}\right)=\frac{w'(p)p-w(p)}{p^2}, \tag{23}$$

From the proof of Proposition 1 and from Example 1, it can be established that $\frac{\partial}{\partial p}\left(\frac{w(p)}{p}\right) > 0$ if $w(p) < p$, i.e. under probability underweighting. However, at $p = 1$, the largest probability that can be set by the principal, (22) cannot hold since $w(1) > 1$, an impossibility. Hence, under probability underweighting the stochastic contract does not elicit higher effort.

Let $\frac{\partial}{\partial p}\left(\frac{w(p)}{p}\right) < 0$ which, according to Proposition 1, holds under $w(p) > p$. In such case p should be set as small as possible to make $\frac{w(p)}{p}$ as large as possible. For (22) to hold, it suffices to show that the right-hand side of (22) shrinks as p decreases. Use the integration by parts, to rewrite the denominator of (22) as:

$$\left(\tilde{\mathbb{E}}\left(b\left(\frac{ay}{p}\right)\middle|\, e_H\right)\right)-\tilde{\mathbb{E}}\left(b\left(\frac{ay}{p}\right)\middle|\, e_L\right)=\frac{a}{p}\int_0^{\bar{y}} b'\left(\frac{ay}{p}\right)\left(w\Big(1-F(y|e_H)\Big)-w\Big(1-F(y|e_L)\Big)\right). \tag{24}$$

The derivative of $p\left(\tilde{\mathbb{E}}\left(b\left(\frac{ay}{p}\right)\middle|\, e_H\right)-\tilde{\mathbb{E}}\left(b\left(\frac{ay}{p}\right)\middle|\, e_L\right)\right)$ with respect to p is:

$$-\frac{a^2}{p^2}\int_0^{\bar{y}} yb''\left(\frac{ay}{p}\right)\left(w\Big(1-F(y|e_H)\Big)-w\Big(1-F(y|e_L)\Big)\right), \tag{25}$$

which is positive under $b''(\cdot) < 0$ since $w\Big(1-F(y|e_H)\Big)-w\Big(1-F(y|e_L)\Big) > 0$ due to the monotone likelihood ratio property. Hence, as p shrinks (22) becomes less stringent since $\frac{w(p)}{p}$ increases while the right-hand side of (22) decreases and its largest value is one when $p = 1$. Hence, under probability overweighting $t_s(y)$ elicits high effort more often than $t_d(y)$.

Finally, to relate this result to that given in Proposition 1 part (ii) notice that the derivative of $\left(\tilde{\mathbb{E}}\left(b\left(\frac{ay}{p}\right)\middle|\, e_H\right)-\tilde{\mathbb{E}}\left(b\left(\frac{ay}{p}\right)\middle|\, e_L\right)\right)$ with respect to p is:

$$\begin{aligned}-\frac{a}{p^2}\int_0^{\bar{y}} b'\left(\frac{ay}{p}\right)&\left(w\Big(1-F(y|e_H)\Big)-w\Big(1-F(y|e_L)\Big)\right)-\\ &\frac{a^2}{p^3}\int_0^{\bar{y}} yb''\left(\frac{ay}{p}\right)\left(w\Big(1-F(y|e_H)\Big)-w\Big(1-F(y|e_L)\Big)\right).\end{aligned} \tag{26}$$

A sufficient condition for (26) to be strictly positive is:

$$1 < -\frac{\int_0^{\bar{y}} \frac{ay}{p} b''\left(\frac{ay}{p}\right)\left(w\Big(1-F(y|e_H)\Big) - w\Big(1-F(y|e_L)\Big)\right)}{\int_0^{\bar{y}} b'\left(\frac{ay}{p}\right)\left(w\Big(1-F(y|e_H)\Big) - w\Big(1-F(y|e_L)\Big)\right)}. \tag{27}$$

Thus, a necessary, but not sufficient, condition for equation (22) to hold when multiplied on both sides by p is given by (27), a condition that resembles the lower bound from the condition of Proposition 1 part (ii). ■

Proposition 2 generalizes the result that the agent is better off introducing additional risk in the agent's environment by implementing the stochastic contract with a small probability. Thus, the result from Proposition 1 part (ii) is not due to the considered deterministic relationship between output and effort. A condition for this result to holds is that the agent overweights probabilities in some non-empty interval. Using this framework I formalize the bonus application discussed in the last section of the paper next.

Lump-sum bonus vs. linear contract

As an example of the incentives of the stochastic contract consider a lump-sum bonus against a linear piece-rate. The bonus contract pays $B > 0$ in the contingency that output level surpasses some threshold $\hat{y}$. Instead, the piece-rate pays ay for any $y \in [0, \bar{y}]$. Assume, as it is done in the theoretical framework of the paper, that these two contracts are cost-equivalent to the principal:

$$B\left(1 - G(\hat{y}|e_H)\right) = a\mathbb{E}\left(y|e_H\right). \tag{28}$$

Let b be strictly concave and $w((1 - G(\hat{y}|e_H) > (1 - G(\hat{y}|e_H))$, i.e. probability overweighting, then using (28) and Jensen's inequality:

$$b(B) = b\left(\frac{\mathbb{E}\left(ay|e_H\right)}{(1 - G(\hat{y}|e_H))}\right) > \frac{\mathbb{E}\left(b(ay)|e_H\right)}{w\left(1 - G(\hat{y}|e_H)\right)}. \tag{29}$$

Denote the non-additive expectation as $\tilde{\mathbb{E}}\left(y|e\right) := \int_0^{\bar{y}} y \mathrm{d}\left(1 - G(y|e)\right)$. Assume that $\tilde{\mathbb{E}}\left(b(ay)|e_H\right) < \mathbb{E}\left(b(ay)|e_H\right)$. Thus, despite $w((1 - G(\hat{y}|e_H) > (1 - G(\hat{y}|e_H))$ the agent also exhibits pessimism. This can is possible is $w(p)$ exhibits an inverse-S shape with pessimism. Under those assumptions (29) can be rewritten as:

$$b(B) > \frac{\tilde{\mathbb{E}}\left(b(ay)|e_H\right)}{w\left(1 - G(\hat{y}|e_H)\right)} > \frac{\tilde{\mathbb{E}}\left(b(ay)|e_H\right) - \tilde{\mathbb{E}}\left(b(ay)|e_L\right)}{w\left(1 - G(\hat{y}|e_H)\right)}. \tag{30}$$

Finally, let $w\left(1-G(\hat{y}|e_H)\right)-w\left(1-G(\hat{y}|e_L)\right)<\delta$ for arbitrarily small $\delta>0$. Indeed, such a condition can be due to likelihood insensitivity (inverse-S $w(p)$). Then (30), becomes:

$$b(B) \overset{\delta}{>} \frac{\tilde{\mathbb{E}}\left(b(ay)|e_H\right)-\tilde{\mathbb{E}}\left(b(ay)|e_L\right)}{w\left(1-G(\hat{y}|e_H)\right)-w\left(1-G(\hat{y}|e_L)\right)}. \tag{31}$$

The above equation can be rewritten as:

$$b(B)\left(w\left(1-G(\hat{y}|e_H)\right)-w\left(1-G(\hat{y}|e_L)\right)\right) \overset{\delta}{>} \tilde{\mathbb{E}}\left(b(ay)|e_H\right)-\tilde{\mathbb{E}}\left(b(ay)|e_L\right), \tag{32}$$

the condition for the lump-sum bonus to elicit high effort more often than the linear piece-rate. Hence, under likelihood insensitivity (inverse-S) with pessimism, the decision maker is more motivated under the lump-sum bonus contract.

Appendix C: Agents with CPT preferences

In this Appendix, I analyze the incentives generated by stochastic contracts when agents have risk preferences characterized by CPT. I find that under mild additional conditions, the result stated in Proposition 1 part (ii) holds, and stochastic contracts can generate higher output than piece-rate contracts. This finding is not surprising since CPT incorporates probability distortions in the same way as RDU. Therefore, when contracting with an agent with CPT preferences, the principal can generate higher output with stochastic contracts when the agent sufficiently overweights small probabilities.

Agents with CPT preferences evaluate the possible outcomes in the stochastic contract relative to a reference point $r \geq 0$. Outcomes below the reference point are coined *losses* and outcomes above it are *gains*. In the original formulation of CPT, r represents the status quo, or the monetary amount that the agent owns and is thus exogenous to the principal's choice. In the following, I adopt that assumption.

The novelty of CPT with respect to RDU is that the agent can exhibit different risk preferences between the domain of gains and the domain of losses. This is, in part, because possible outcomes are evaluated with a value function with the following properties:

Assumption 5. *$V(t_s(y), r)$ is the piecewise function,*

$$V(t_s, r) = \begin{cases} b(\frac{ay}{p} - r) & , \textit{ if } \frac{ay}{p} \geq r, \\ -\lambda b(r - \frac{ay}{p}) & , \textit{ if } \frac{ay}{p} < r. \end{cases}$$

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

In words, the value function is an increasing function that is concave in the domain of gains and convex in the domain of losses, generating risk averse and risk seeking attitudes in the agent, respectively. Additionally, the worker is loss averse, i.e. for him losses loom larger than equally-sized gains. This property is captured by $\lambda > 1$ which enters the value function for the domain of losses.

The CPT agent also transforms the probabilities associated to the outcomes of the contract t_s using a probability weighting function. However, transformations of probability can be different for gains and losses. Let $w(p)$ be the probability weighting function used to transform probabilities in the domain of gains. This weighting function exhibits the properties from Assumption 4. Moreover, let $z(p)$ be the probability weighting function used to transform probabilities in the domain of losses. To simplify matters, it is assumed that $w(p)$ and $z(p)$

relate through the duality $z(p) = 1 - w(1-p)$. Hence, the weighting function for losses adopts the same properties as that for gains, and only differs in that probability transformations are applied to loss ranks or a ranking of outcomes from least-desirable to most-desirable.

All in all, the utility of the agent with CPT preferences when offered $t_s(y)$ is equal to:

$$CPT(t_s(y), t) = \begin{cases} w(p)v\left(\frac{ay}{p} - r\right) - c(e) & \text{, if } \frac{ay}{p} \geq r \geq 0, \\ -z(p)\lambda v\left(r - \frac{ay}{p}\right) - c(e) & \text{, if } r > \frac{ay}{p} > 0. \end{cases} \tag{33}$$

Consider next the case in which the agent with CPT preferences works under t_d. While this contract does not contain risk, which disregards probability weighting, the assumption that the agent still makes decisions relative to a reference point is kept. Maintaining this assumption is consistent with abundant evidence showing that in settings of deterministic choice, individuals exhibit reference-dependent preferences Kahneman et al. (1991).

When offered $t_d(y)$, the agent with riskless prospect theory preferences (Kahneman et al., 1991) exhibits the following preference:

$$CPT(t_d(y)) = \begin{cases} b\left(\frac{ay}{p} - r\right) - c(e) & \text{, if } \frac{ay}{p} \geq r \geq 0, \\ -\lambda b\left(r - \frac{ay}{p}\right) - c(e) & \text{, if } r > \frac{ay}{p} > 0. \end{cases} \tag{34}$$

We are in a position to compare the two contracts with respect to the output that they deliver. Proposition 3 provides the conditions under which the principal is better off exposing the agent to large amounts of risk with the stochastic contract.

Proposition 3. *Under Assumptions 1, 2, 4, and 5, an agent with CPT preferences exerts higher effort with a stochastic contract implemented with small probabilities if either:*

(i) $w(p) \geq p \exp\left(ay \int_p^1 \frac{\mathcal{A}\left(\frac{ay}{\mu} - r\right)}{\mu^2} d\mu\right)$ *and* $\frac{ay_C^{**}}{p} \geq r$, *or*

(ii) $z(p) \geq p \exp\left(ay \int_p^1 \frac{\mathcal{A}^l\left(\frac{ay}{\mu} - r\right)}{\mu^2} d\mu\right)$ *and* $\frac{ay_C^{**}}{p} < r$.

Where $\mathcal{A}\left(\frac{ay}{\mu} - r\right) := \frac{-b''\left(\frac{ay}{\mu} - r\right)}{b'\left(\frac{ay}{\mu} - r\right)}$ *and* $\mathcal{A}^l := \frac{-b''\left(r - \frac{ay}{\mu}\right)}{b'\left(r - \frac{ay}{\mu}\right)}$.

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

$$\frac{a}{p} w(p) b'\Big(\frac{ay_C^{**}}{p} - r\Big) - c'\left(\frac{y_C^{**}}{\theta}\right)\frac{1}{\theta} = 0, \text{ if } \frac{ay}{p} \geq r, \tag{35}$$

$$\frac{a}{p}\left(1 - w(1-p)\right)\lambda b'\Big(r - \frac{ay_C^{**}}{p}\Big) - c'\left(\frac{y_C^{**}}{\theta}\right)\frac{1}{\theta} = 0, \text{ if } \frac{ay}{p} < r. \tag{36}$$

That y_C^{**} is a maximum requires that the second order condition is negative. Formally,

$$\frac{a^2}{p^2}w(p)b''\Big(\frac{ay_C^{**}}{p}-r\Big)-c''\left(\frac{y_C^{**}}{\theta}\right)\frac{1}{\theta^2}<0,\text{ if }\frac{ay}{p}\geq r, \tag{37}$$

and

$$\frac{a^2}{p^2}\left(1-w(1-p)\,b''\Big(r-\frac{ay_C^{**}}{p}\Big)-c''\left(\frac{y_C^{**}}{\theta}\right)\frac{1}{\theta^2},\text{ if }\frac{ay}{p}<r. \tag{38}$$

Let us first analyze the case in which the agent is in the domain of gains. To investigate whether the principal must set $p<1$, using t_s differentiate implicitly equation (35) with respect to y_C^{**} and p to obtain:

$$\frac{\mathrm{d}y_C^{**}}{\mathrm{d}p}=\frac{\left(\frac{w'(p)p-w(p)}{p^2}\right)ab'\left(\frac{ay_C^{**}}{p}-r\right)-\frac{w(p)a^2y}{p^3}b''\left(\frac{ay_C^{**}}{p}-r\right)}{c''\left(\frac{y_C^{**}}{\theta}\right)\frac{1}{\theta^2}-b''\left(\frac{ay_C^{**}}{p}-r\right)\frac{w(p)a^2}{p^2}}. \tag{39}$$

That $\frac{\mathrm{d}y_C^{**}}{\mathrm{d}p}>0$ implies that the principal is better off setting $p=1$ and obtaining y_C^{**} that satisfies

$$ab'\Big(ay_C^{**}-r\Big)-c'\left(\frac{y_C^{**}}{\theta}\right)\frac{1}{\theta}=0. \tag{40}$$

Instead, $\frac{\mathrm{d}y_C^{**}}{\mathrm{d}p}\leq 0$ implies that the principal must set the smallest possible. From (39) it can be established that a necessary condition to obtain $\frac{\mathrm{d}y_C^{**}}{\mathrm{d}p}<0$, and thus that the principal has incentives to choose $p<1$, is:

$$w'(p)\leq\frac{w(p)}{p}\left(-\frac{ay}{p}\mathcal{A}\left(\frac{ay}{p}-r\right)+1\right). \tag{41}$$

Where $\mathcal{A}\left(\frac{ay}{p}-r\right)=-\frac{b''\left(\frac{ay}{\mu}-r\right)}{b'\left(\frac{ay}{\mu}-r\right)}$. I proceed as in the proof of Proposition 1. Let $v(p)$ be a weighting function with the properties of $w(p)$. First, consider the following ordinary differential equation,

$$v'(p)=\frac{v(p)}{p}\left(-\frac{ay}{p}\mathcal{A}\left(\frac{ay}{p}-r\right)+1\right), \tag{42}$$

which is solved by

$$v(p)=\exp\left(-ay\int_p^1\frac{1-\mathcal{A}\left(\frac{ay}{\mu}-r\right)}{\mu^2}\mathrm{d}\mu\right). \tag{43}$$

Second, to study how $w(p)$ and $v(p)$ relate compute the following derivative of their ratio:

$$\frac{\mathrm{d}\left(\frac{w(p)}{v(p)}\right)}{\mathrm{d}p} = \frac{(v(p)w'(p) - w(p)v'(p))}{v(p)^2} = \frac{v(p)\left(w'(p) - \frac{w(p)}{p}\left(-\frac{ay}{p}\mathcal{A}\left(\frac{ay}{p} - r\right) + 1\right)\right)}{v(p)^2}, \tag{44}$$

where the second equality results from replacing $v'(p)$ using (43). Using the above equation together with (41) it can be established that $\frac{\mathrm{d}\left(\frac{w(p)}{v(p)}\right)}{\mathrm{d}p} \leq 0$. Hence, the minimum of $\frac{w(p)}{v(p)}$ must be attained at $p = 1$ and it must be that $\frac{w(p)}{v(p)} \geq \frac{w(1)}{v(1)} = 1$ for any $p \in (0, 1]$.

Therefore, the solution to (41) is bounded by (43), in the following way:

$$w(p) \geq p \exp\left(ay \int_p^1 \frac{\mathcal{A}\left(\frac{ay}{\mu} - r\right)}{\mu^2} \mathrm{d}\mu\right). \tag{45}$$

As long as (45) holds in the domain of gains, $y \geq \frac{ay}{p}$, the principal is better setting smallest possible probability $p \to 0^+$. Consider now the domain of losses. To analyze the influence between y_C^{**} and p, implicitly differentiate (36) with respect to those variables:

$$\frac{\mathrm{d}y_C^{**}}{\mathrm{d}p} = \frac{\left(\frac{\frac{w'(1-p)}{p} - (1-w(1-p))}{p^2}\right)\frac{a}{p}b'\left(r - \frac{ay_C^{**}}{p}\right) - \frac{(1-w(1-p))a^2y}{p^3}\lambda b''\left(r - \frac{ay_C^{**}}{p}\right)}{c''\left(\frac{y_C^{**}}{\theta}\right)\frac{1}{\theta^2} + \lambda b''\left(r - \frac{ay_C^{**}}{p}\right)\frac{(1-w(1-p))a^2}{p^2}}. \tag{46}$$

From (46) a sufficient condition for $\frac{\mathrm{d}y_C^{**}}{\mathrm{d}p} \leq 0$ is:

$$w'(1-p) \leq \frac{(1 - w(1-p))}{p}\left(1 - \frac{ay_C^{**}}{p}\mathcal{A}^l\left(r - \frac{ay_C^{**}}{p}\right)\right), \tag{47}$$

where $\mathcal{A}^l\left(r - \frac{ay_C^{**}}{p}\right) := -\frac{b''\left(r - \frac{ay_C^{**}}{p}\right)}{b'\left(r - \frac{ay_C^{**}}{p}\right)}$. Following a similar procedure than the one performed in the analysis of the domain of gains, and recognizing the duality $z(p) = 1 - w(1-p)$ the solution to the above differential inequality is:

$$z(p) \geq p \exp\left(ay \int_p^1 \frac{\mathcal{A}^l\left(\frac{ay}{\mu} - r\right)}{\mu^2} \mathrm{d}\mu\right). \tag{48}$$

If equation (48) holds in the domain of losses, $y < \frac{ay}{p}$, then $\frac{\mathrm{d}y_C^{**}}{\mathrm{d}p} \leq 0$ and the principal is better setting smallest possible probability $p \to 0^+$.

■

As with RDU preferences, the principal can generate more output when she offers the agent the stochastic contract with a small probability. To achieve this result, the agent needs to sufficiently overweight probabilities so as to outweigh the potential risk averse attitudes stemming from the value function. As in Proposition 1 part (ii), the lower-bound from Proposition 2, $\exp\left(ay\int_p^1 \frac{\mathcal{A}\left(\frac{ay}{\mu}-r\right)}{\mu^2}\mathrm{d}\mu\right)$ reflects the agent's risk attitudes stemming from the value function which are translated to the probability space. When probabilistic risk attitudes are such that this lower-bound is attained, implying that the agent exhibits risk seeking at small probabilities, the principal should fully expose the agent to risk.

To conclude this appendix, I comment on the role of the additional components of risk attitudes when agents have CPT preference. The first component is loss aversion, λ. Note that the measure of loss aversion does not enter the condition guaranteeing Proposition 2. Thus, that the agent exhibits more or less loss aversion is immaterial to the effectiveness of the stochastic contract. After all these contracts introduce and adjust probabilistic risk and do not alter how losses are evaluated with respect to gains.

The second component is diminishing sensitivity, i.e. that the value function is concave for gains and convex for losses. Notice that $\mathcal{A}\left(\frac{ay}{\mu}-r\right)$ included in the lower bound $\exp\left(ay\int_p^1 \frac{\mathcal{A}\left(\frac{ay}{\mu}-r\right)}{\mu^2}\mathrm{d}\mu\right)$ changes depending on whether the agent is in the domain of gains or in the domain of losses. In the domain of losses this expression becomes negative, while in the domain of gains is positive. That makes the lower-bound less stringent in the domain of losses than in the domain of gains. Intuitively, the curvature of the value function in the domain of losses motivates risk seeking and eases that the agent becomes more motivated with a contract that introduces risk. Instead, in the domain of gains, the agent exhibits concave value function which goes against risk seeking.

Appendix D. The principal's problem

The purpose of this Appendix is to complement the theoretical model presented in §2 by providing a solution to the full principal problem. In that problem not only is the revenue of the principal maximized using the incentive compatibility constraint, but also including the participation constraint. The results from this Appendix show that Proposition 1 (ii) holds when the participation constraint is taken into account.

Assume that the agent's risk preferences are characterized by RDU. That is, the agent distorts cumulative probabilities using the weighting function $w(p)$ described by Assumption 4. To simplify matters, also assume that the agent's utility belong to the CRRA family:

Assumption 6. *Let $b(t) = t^{\rho}$ where $\rho \in \mathbb{R}$.*

This assumption must be taken with a grain of salt. It states that Proposition 1 (ii) can hold, as shown in Example 2. Therefore, the present exercise is an investigation of whether by including the participation constraint, the principal is further restricted in her actions.

Finally, I consider a setting in which the principal be risk-neutral, and her decision consists on choosing the probability $p \in (0, 1]$ included in $t_s(y)$ such that makes the agent accept the contract and that best incentivizes the agent to exert as much effort as possible. Notice that, due to the equivalence $A = \frac{a}{p}$, that the principal chooses $p = 1$ is equivalent to the piece-rate contract $t_d(y)$. In other words, the principal problem amounts to choose among contracts that are cost-equivalent but that differ on the amount of risk that will be faced by the agent.

All in all, the principal's program is:

$$\begin{aligned} \min_{p \in (0,1]} \quad & Apy \\ \text{s.t.} \quad & w(p)\,(Ay)^{\rho} - c\left(\frac{y}{\theta}\right) \geq \bar{U}, \\ & A\rho w'(p)\,(Ay)^{\rho-1} - c'\left(\frac{y}{\theta}\right)\frac{1}{\theta}. \end{aligned} \tag{49}$$

The solution to the principal's problem is presented in Proposition 4. The main message is that the solution to the principal's problem is identical to that presented in Proposition 1 (ii). That is, the contract $t_s(y)$ applied small probabilities should be implemented when the agent's weighting function sufficiently overweights small probabilities.

Proposition 4. *Under Assumptions 1, 2, 4, and 6, the solution to (49) consists on implementing $t_s(y)$ with small probabilities if $w(p) \geq p^{\rho}$ and implementing $t_d(y)$ otherwise.*

Proof. Recognizing the equivalence $A = \frac{a}{p}$, and denoting by ν_1 and ν_2 the Lagrangian multipliers of the participation and incentive compatibility constraints, respectively, the Lagrangian of the principal program can be written as:

$$\mathcal{L} = ay - \nu_1 \left(w(p) \left(\frac{ay}{p} \right)^{\rho} - c\left(\frac{y}{\theta} \right) - \bar{U} \right) - \nu_2 \left(\frac{arw(p)}{p} \left(\frac{ay}{p} \right)^{\rho-1} - c'\left(\frac{y}{\theta} \right) \frac{1}{\theta} \right). \tag{50}$$

The first-order condition of the Lagrangian in (50) with respect to p and taking into account that changes in p might generate changes in y is:

$$\begin{aligned} -a\frac{\mathrm{d}y}{\mathrm{d}p} - \nu_1 \left(w'(p) \left(\frac{ay}{p} \right)^{\rho} - \frac{a\rho w(p)}{p} \left(\frac{\mathrm{d}y}{\mathrm{d}p} - \frac{y}{p} \right) \left(\frac{ay}{p} \right)^{\rho-1} \right) \\ - \nu_2 \left(\frac{a\rho w'(p)}{p} \left(\frac{ay}{p} \right)^{\rho-1} - \frac{a\rho w(p)}{p^2} \left(\frac{ay}{p} \right)^{\rho-1} - \frac{a^2\rho(\rho-1)w(p)}{p^2} \left(\frac{\mathrm{d}y}{\mathrm{d}p} - \frac{y}{p} \right) \left(\frac{ay}{p} \right)^{\rho-2} \right) = 0. \end{aligned} \tag{51}$$

After some manipulations, equation (51) can be rewritten as:

$$(\nu_1 y + \nu_2 \rho) \left(w'(p)y + \rho w(p) \left(\frac{\mathrm{d}y}{\mathrm{d}p} - \frac{y}{p} \right) \right) = \nu_2 \rho \left(w(p) \frac{\mathrm{d}y}{\mathrm{d}p} \right) + \frac{a\frac{\mathrm{d}y}{\mathrm{d}p}}{\frac{a^2}{p^2} \left(\frac{ay}{p} \right)^{\rho-2}}. \tag{52}$$

To analyze the optimal value of p determined by in (52) first, assume that $\frac{\mathrm{d}y}{\mathrm{d}p} = 0$, in that case, equation (52) gives:

$$(\nu_1 y + \nu_2 \rho) \left(w'(p) - \frac{w(p)\rho}{p} \right) = 0. \tag{53}$$

Hence, under $\nu_1 > 0$ and $\nu_2 > 0$ equation (53) holds if

$$w'(p) = \frac{\rho w(p)}{p}. \tag{54}$$

The solution to the ordinary differential equation in (54) is

$$w(p) = p^{\rho}. \tag{55}$$

Hence, $\frac{\mathrm{d}y}{\mathrm{d}p} = 0$ is achieved for agents with weighting functions complying with $w(p) = p^{\rho}$ and in such case the principal is indifferent between offering t_s with sufficiently small p or t_d.

Consider now the more interesting case in which $\frac{\mathrm{d}y}{\mathrm{d}p} \leq 0$. In such case, the right-hand side of (52) is negative, and that, for the equality to be maintained under $\nu_1 > 0$ and $\nu_2 > 0$, it is necessary that:

$$w'(p) \leq \frac{\rho w(p)}{p}. \tag{56}$$

.

From Grönwall's lemma, applied in the proof of Proposition 1, it can be established that (56) holds for any weighting function such that $w(p) \geq p^{\rho}$. Hence, the principal must choose $t_s(y)$ with $p \to 0^+$ if $w(p) \geq p^{\rho}$ because in that case $\frac{\mathrm{d}y}{\mathrm{d}p} < 0$. Notice that this condition is identical to the one presented in Example 2. Instead, when $w(p) < p^{\rho}$ the principal is better off offering $t_d(y)$.

Finally it is analyzed whether the aforementioned solution to the principal's program is valid. To do so, I investigate the shape of the Lagrangian in Equation (50). The second-order condition of this Lagrangian is:

$$\frac{a\rho y}{p^2}\left(\frac{ay}{p}\right)^{\rho-1}\left(\nu_1 + \frac{\nu_2 \rho}{y}\right)\left(w'(p) - \frac{\rho w(p)}{p}\right) - \left(\frac{ay}{p}\right)^{\rho}\left(\nu_1 + \frac{\nu_2 \rho}{y}\right)\left(w''(p) - \frac{p\rho w'(p) - rw(p)}{p^2}\right) \tag{57}$$

Equation (57) becomes positive if $w''(p) < 0$. Hence the candidate solution, $p \to 0$ if $w(p) \geq p^{\rho}$, is valid as long as $p \in (0, \tilde{p})$. This is consistent with the fact that $\frac{\mathrm{d}y}{\mathrm{d}p} \leq 0$ under $w(p) \geq p^{\rho}$. However, if $w''(p) > 0$, the second order condition in (57) becomes negative, implying that the objective function attains a minimum value at one of the extremes, $p = \tilde{p}$ or $p = 1$. Hence, according to the Lagrangian method there exists multiple solutions to the principal's problem.

Note that according to Assumption 4, $w(p)$ might display $\hat{p} \neq \tilde{p}$, i.e. the inflection point is different than the fixed point. Consider first $\hat{p} > \tilde{p} > 0$. In such case, for $p \in (0, \tilde{p})$, the candidate solution can be implemented. Instead, for $p \in [\tilde{p}, 1]$ the solution is either at $p = \tilde{p}$ or at $p = 1$. Since at $p = \tilde{p}$ probabilities are overweighted, the incentive compatibility and participation constraints of the program presented become larger than at $p = 1$, which implies that at $p = \tilde{p}$ the Lagrangian attains a lower value. Thus, according to the Lagrangian method the principal chooses $p = \tilde{p}$ if $p \in (\tilde{p}, 1)$ and $\hat{p} > \tilde{p}$. However, since $\hat{p} > \tilde{p}$ and if $w(p) \geq p^{\rho}$ holds, implying $\frac{\mathrm{d}y}{\mathrm{d}p} \leq 0$, the principal is better off choosing sufficiently small p. So the candidate solution is corroborated for this case.

Let now $0 < \hat{p} \leq \tilde{p}$. As before, for the interval $p \in (0, \tilde{p})$, p as small as possible is implemented if $w(p) \geq p^{\rho}$ holds. Moreover, for the interval $p \in [\tilde{p}, 1]$, the solution is $p = 1$ since $p = \tilde{p}$ yields $\frac{w(p)}{p} < 1$ which leads to lower values of the incentive compatibility and participations constraints than those implied by $p = 1$. This solution is however invalid if $w(p) \geq p^{\rho}$ holds, since in that case, the principal is better off setting small p. In this case the Lagrangian solution is contradicted by the candidate solution.

■

Proposition 4 shows that considering the participation constraint and the objective function of the principal does not affect the implementation of incentives presented in Proposition 1 part (ii). This is because of the nature of the considered problem. In this framework the principal is, on expectation, not affecting the compensation of the agent by exposing him to more or less risk. This is due to the equivalence $A = \frac{a}{p}$. Hence, that the expected value maximizer accepts t_d necessarily implies that t_d will be accepted. Furthermore, an agent who is incentivized by t_s to exert higher effort than in t_d in the sense of Proposition 1 part (ii), must also be willing to accept the contract. That is because his distortion of probabilities enhances the utility that he will derive from the contract. Therefore, the participation constraint becomes superfluous to the problem.

I conclude this Appendix by summarizing its main results. I find that the principal's optimal choice is identical to the case in which the incentive constraint was considered alone. This is because the incentives that are analyzed keep constant the agent's monetary reward, on expectation. Also, because that the agent is willing to work harder under the stochastic contract, implies that his distortion of probabilities guarantees the participation constraint.

Appendix E: Utility functions

This appendix investigates the properties of the elicited utility functions. Decision sets 1 to 6 of the second part of the experiment are designed to elicit the sequence of outcomes $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ for each subject. This elicited sequence has the relevant property that it ensures equally-spaced utility values, i.e. $u(x_j) - u(x_{j-1}) = u(x_{j-1}) - u(x_{j-2})$, allowing me to characterize a subject's preference over monetary outcomes by mapping each utility value, $u(x_j)$to the subject's stated preference x_j.

I focus on two properties of the utility function: the sign of the slope and the curvature. To that end, I construct two variables, the first variable is $\Delta'_i := x_j - x_{j-1}$, for $j = 1, ..., 6$ and the second is $\Delta''_j := \Delta'_j - \Delta'_{j-1}$ 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 to smaller 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 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 an increasing and concave utility function.

The first analysis focuses on classifications at the individual level. I classify subjects 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.[26]

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 7 presents the classification of subjects according to the curvature of their utility function. The data suggest that the majority of subjects exhibit linear utility functions.

[26] More stringent confidence intervals were also used for the analysis. These confidence intervals were also constructed using the standard deviation of a Δ''_j which was multiplied by different factors, such as 1 and -1, 1.64 and -1.64, and 2 and -2. The qualitative results of these analyzes are not different from the main result presented here 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.

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). 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 ($p<0.001$) 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), their 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 curvature. Therefore, one of the advantages of the experimental design, that it elicits the utility function and the probability weighting function of subjects using monetary stakes that reflect the monetary incentives in the first part of the experiment, is also the reason that diminishing sensitivity is not be observed.

Table 7 also presents the results of the aforementioned analysis when it is assumed that subjects have CPT preferences with a reference point equal to the monetary equivalent of a subject's beliefs about his performance in the first part of the experiment. Monetary outcomes above this reference point are considered gains and outcomes below the reference point are considered 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 x_j is averaged for all subject. Table 8 presents the descriptive statistics of the resulting outcomes. I find that the average outcome x_j is increasing with j, implying that on average subjects exhibit a taste for larger monetary outcomes. Moreover, the column displaying the average values of the variable Δ'_j shows that as j increases, increments of x_j become larger. Thus, while on average subjects exhibit linear utility, this tendency ceases as monetary outcomes in the lotteries become larger. In fact, for large values of x_j the average utility function displays concavity. This result is also found by Abdellaoui (2000).

The last analysis of the data consists on fitting well-known parametric families of utility functions. 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 9 the regression estimates when non-linear least squares is used to fit the data to the

Table 7: 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

assumed utility function. For 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 parameter attains a value of 0.995. This finding is consistent with the large proportion of subjects that were classified as having a linear utility function in the individual analysis and the modest increments that the averaged outcomes x_j exhibit as j increases presented in Table 8.

These analyses are also performed under the assumption that subjects have CPT preferences with a reference point equal to the monetary equivalent of the subject's belief in the first part of the experiment. According to Table 8, 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 outcome becomes larger in the domain of gains and lower in the domain of losses. The latter finding implies that in the domain of gains the average utility function tends to concavity, while in the domain of losses the function it tends to convexity. Furthermore, the data suggest that diminishing sensitivity manifests at different degrees across the domains, with subjects exhibiting more in the domain of gains. 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 8: 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 also 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 $\phi = 0.992$ when a power utility function is assumed. A similar result is 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 finding is robust to the assumption that subjects have CPT preferences and the reference point is assumed to be their belief. This is not a surprising finding given the magnitude of the stakes used to elicit the subject's risk preferences. Furthermore, the conclusion that the utility function is linear implies that probability risk attitudes fully determine the risk attitudes. Implying that performance differences across treatments must be explained by probability distortions.

Table 9: 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 F: 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 classifies each subject according to the shape of the elicited probability weighting function and is based on Bleichrodt and Pinto (2000). There were five possible shapes of the probability weighting function. A subject could display a weighting function with either lower subadditivity (LS), upper subadditivity (US) or with both properties. These three properties result from comparing the behavior of the probability weighting function at extreme probabilities to the behavior of the same function at intermediate probabilities. Moreover, a subject could display a concave or a convex probability weighting function.

To classify a subject into one of these five categories, I created the variable $\partial_{j-1}^{j} := \frac{w(p_j)-w(p_{j-1})}{w^{-1}(p_j)-w^{-1}(p_{j-1})}$, which captures the average slope of the probability weighting function between probabilities j and $j-1$. I also created the variable $\bigtriangledown_{j-1}^{j} \equiv \partial_{j-1}^{j} - \partial_{j-2}^{j-1}$, which represents the change of the average slope of the weighting function between successive probabilities.

To understand the subjects' behavior at extreme and intermediate probabilities I focus on the sign of the variables $\bigtriangledown_{0.16}^{0.33}$ and $\bigtriangledown_{0.83}^{1}$. If a subject exhibits $\bigtriangledown_{0.16}^{0.33} < 0$, 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 $\bigtriangledown_{0.83}^{1} > 0$, 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. The resulting dummy variables LS and US or Both were used in the main body of the paper to investigate the effect of these properties of the weighting function on the treatment effects.

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

Table 10 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. These proportions are however 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 exhibit pessimism. 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 10: 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 $\bigtriangledown^{j}_{j-1}$. 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.

For the sake of robustness, I perform an alternative classification of LS and US also proposed by Bleichrodt and Pinto (2000). In comparison to the above classification, weights given to extreme probabilities are contrasted to the corresponding objective probability. 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. The reason is that assigning large weights to extreme probabilities does does not guarantee that the weights assigned to medium-ranged probabilities are small.

The results of the alternative classification are presented in Table 11. 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).

Table 11: 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.

The last considered classification, evaluates the strength of the possibility effect relative to the certainty effect. A subject exhibits a weighting function with 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 12 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 close to 32 % of subjects.

Table 12: 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 a reference point equal to the monetary equivalent of their beliefs about performance in the first part of the experiment. All previous analyses are also performed under the assumption that the monetary equivalent of a subject's belief in the real-effort task is the reference point.[27] The results of these analyses are also presented in Table 10, Table 11, and Table 12. 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 less 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.

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

Appendix G: Additional analyses

Descriptions and comparisons with previous studies

Panel 1 in Table 4 presents the estimates of a truncated regression of the neo-additive functional, $w(p) = c + sp$.[28] 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. Furthermore, $\hat{c}$ and $\hat{s}$ are larger and smaller, respectively, than the estimates reported in Abdellaoui et al. (2011), suggesting that subjects in my experiment exhibit 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). Their proposal relates probabilities and their associated weights according to the following non-linear function: $w(p) = \frac{p^{\psi}}{\left(p^{\psi}+(1-p)^{\psi}\right)^{\frac{1}{\psi}}}$. The second panel of Table 4 shows that the non-linear least squares method generates an 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 probability distortion.

A crucial disadvantage of Tversky and Kahneman's (1992) weighting function is that likelihood insensitivity and optimism/pessimism influence ψ, so their effect on probabilistic risk attitudes cannot be identified. To overcome such disadvantage, I also use the log-odds weighting function proposed by Goldstein and Einhorn (1987), $w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma}+(1-p)^{\gamma}}$, which can, up to some extent, separate these two components. The estimates of a 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 $\hat{\delta}$ an strikingly small degree of pessimism. 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 optimism than previously documented.

Lastly, I also estimate a regression assuming Prelec (1998)'s probability weighting function with two parameters, $w(p) = \exp\left(-\beta(-\ln(p))^{\alpha}\right)$. 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,

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

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 on average display optimism. 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, these subjects display an average probability weighting function with a stronger inverse-S shape and more optimism as compared to previous studies.

Including an expectations-based reference point

For the sake of robustness, I perform the aforementioned estimations accounting for the possibility that subjects have CPT preferences. In such analysis I assumed the subject's reference point to be the monetary equivalent of each subject's belief in the first part of the experiment. Lottery outcomes above this reference point belong to the domain of gains, while lottery outcomes below this reference point belong to the domain of losses. I perform separate regressions for each domain. The results are presented in Table 13. I find that for all considered functional forms of probability weighting and for both domains, subjects display weighting functions with inverse-S shapes and more optimism than previously found. As a consequence, the results presented in this section are robust to the assumption that subjects' preferences can be represented by CPT preferences.

Table 13: Parametric estimates of the weighting function with reference point

	(1)	(2)	(3)
Panel 1: Neo-additive (truncated)			
$w(p) = c + sp$			
$\hat{c}$	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)			
$w(p) = \frac{p^{\psi}}{\left(p^{\psi}+(1-p)^{\psi}\right)^{\frac{1}{\psi}}}$			
$\hat{\psi}$	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)			
$w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma}+(1-p)^{\gamma}}$			
$\hat{g}$	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)			
$w(p) = \exp\left(-\beta(-\ln(p))^{\alpha}\right)$			
$\hat{\alpha}$	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 $w(p) = \frac{p^{\psi}}{(p^{\psi}+(1-p)^{\psi})^{\frac{1}{\psi}}}$. The third panel presents the non-linear least squares estimates of the parametric form $\frac{\delta p^{\gamma}}{\delta p^{\gamma}+(1-p)^{\gamma}}$. The last panel presents the non-linear least squares estimates of the function $w(p) = \exp\left(-\beta(-\ln(p))^{\alpha}\right)$. 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.

Appendix H: 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 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 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 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 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"