Quantitative Economics 15 (2024), 607–653 1759-7331/20240607

Incentive contracts when agents distort probabilities

Víctor González-Jiménez
Erasmus School of Economics, Erasmus University Rotterdam

I show that stochastic contracts generate powerful incentives when agents suffer from probability distortion. When implementing these contracts, the principal can target probability distortions in order to inflate the agent's perceived benefits of exerting high levels of effort. This novel source of motivation is absent in contracts traditionally regarded as optimal. A theoretical framework and an experiment demonstrate that stochastic contracts implemented with small probabilities, which expose the agent to a high degree of risk, generate higher performance than cost-equivalent contracts with lower or no risk exposure. I find that probability distortions that result from likelihood insensitivity—cognitive limitations that prevent the accurate evaluation of probabilities—account for this finding. The results highlight the limits of contracts traditionally regarded as optimal.

Keywords. Contracts, risk attitude, incentives, probability weighting, experiments.

JEL classification. C91, D81, D82, D86, D90.

1. Introduction

A fundamental result in contract theory is that the optimal contract emerges from a tradeoff between efficiency and insurance (Holmstrom (1979)). On the one hand, high effort levels in a task are elicited by a transfer schedule according to which the agent's compensation increases with his performance. On the other hand, the risk exposure implied by such a schedule should be moderate, such that it does not disincentivize the agent from accepting the contract.

This paper shows that such a tradeoff can disappear when the agent suffers from probability distortion, that is,m when the agent misperceives probabilities in a systematic way. Specifically, I demonstrate that contracts, which introduce additional risk into the agent's environment, can enhance his motivation to work without requiring the principal to offer higher average payments. Consequently, if probability distortion is a widespread phenomenon in decision-making, then contracts traditionally regarded as optimal need to be modified so as to impart more effective incentives.

Víctor González-Jiménez: victor.gonzalez@ese.eur.nl
I wish to thank Charles Noussair, Patricio Dalton, Wieland Müller, Aurelien Baillon, Peter Wakker, Robert Dur, Roberto Weber, Cole Williams, Geoffrey Castillo, and Stephanie Koornneef for valuable comments. I also thank participants in seminars at Erasmus University Rotterdam, Maastricht University, Tilburg University, University of Vienna, and Vienna University of Economics and Business, and in the Foundations of Utility and Risk Conference at the University of York. The experiments reported here were funded by the Center for Economic Research (CentER) at Tilburg University.

Available at http://qeconomics.org. https://doi.org/10.3982/QE2275

There is abundant empirical evidence that individuals distort probabilities; they overweight small probabilities and underweight medium to large probabilities (Tversky and Kahneman (1992), Abdellaoui (2000), Fehr-Duda and Epper (2012), L'Haridon and Vieider (2019)). With stochastic contracts, which involve lottery-like incentive schemes, the principal introduces additional risk into the agent's environment in order to *target* and exploit the probability distortions that motivate him to exert effort in the task. Since these incentives are absent in contracts traditionally regarded as optimal, because they generally seek to insure the agent against existing risk, stochastic contracts can generate greater output at no extra cost to the principal.

I consider a simple version of stochastic contracts in which the agent faces two possible outcomes: a monetary compensation that depends on performance or a performance-insensitive payment. Under such a contract, the agent faces the risk that his effort will not affect his compensation. Importantly, the principal can adjust this risk by determining the probability that the performance-contingent compensation will indeed be paid. Therefore, under full commitment, the agent's decision to exert effort not only depends on the compensation offered by the contract, but also on the perceived probability that his effort will affect his compensation.

To understand how stochastic contracts can outperform contracts traditionally regarded as optimal, consider a setting in which both are cost-equivalent for the principal. That is, for any given effort level, the agent's *expected* compensation under the stochastic contract is the same as that under a traditional contract. Accordingly, an expected value maximizer would be equally motivated under both incentive schemes. However, the stochastic contract creates greater motivation for the agent when he overweights the probability of obtaining the performance-contingent outcome. Such a distortion of probabilities erroneously inflates the perceived benefits of supplying higher levels of effort, thus motivating the agent to work harder. Therefore, under a stochastic contract, the principal can exploit the agent's probabilistic mistakes in order to achieve her objectives.

A simple theoretical framework serves two purposes. First, it pins down the conditions guaranteeing the main result of the paper, namely that stochastic contracts generate more motivation than contracts traditionally regarded as optimal. When the agent's probability weighting function attains a lower bound, representing the extent of probability overweighting needed to induce risk-seeking attitudes, the principal is better off offering the stochastic contract since it generates stronger incentives at similar average costs. Second, the theory provides a set of predictions that can be empirically tested using a laboratory experiment.

A laboratory experiment shows that stochastic contracts implemented with a small probability, that is, $p = 0.10$, yield higher performance in an effort-intensive task relative to a cost-equivalent piece-rate contract. In contrast, stochastic contracts implemented with larger probabilities, namely $p = 0.30$ or $p = 0.50$, yield no differences in performance relative to the cost-equivalent piece rate. The experiment also features an elicitation of the subjects' utility and probability weighting functions. I find that subjects display linear utilities and an average weighting function with a strong inverse-S shape.

Moreover, I demonstrate that this shape of the probability weighting function fully explains the higher performance generated by the stochastic contract with $p = 0.10$. Further analyses of the data show that probability distortions due to *likelihood insensitivity*, that is, the cognitive inability of individuals to evaluate probabilities accurately (Tversky and Wakker (1995), Wakker (2010)) explain the treatment effects.

While stochastic contracts are generally treated only as a theoretical possibility in the literature, their incentives can be put into practice using standard tools of personnel economics. For instance, contracts offering a bonus for achieving a production target expose the worker to different degrees of risk. When the production target is set high, the probability of obtaining the bonus is small and the agent is exposed to a high degree of risk. I show that when the worker overweights the probability that the bonus will be awarded, the principal elicits more motivation than if she were to use, say, a linear contract with no bonus. I provide a detailed explanation of this application and present others in the last section of the paper.

This paper contributes to several strands of the literature. Its results add to the literature on behavioral contract theory (see Koszegi (2014), for a review) by showing that when agents distort probabilities, stochastic contracts motivate the agent to a greater extent than contracts traditionally regarded as optimal. This is fundamentally at odds with the standard result from contract theory, according to which the principal faces a trade-off at the optimum between providing insurance and obtaining a high level of effort. Moreover, while the optimality of stochastic contracts has been shown in other theoretical settings, such as multitasking environments (Ederer, Holden, and Meyer (2018)), when agents exhibit aspiration levels (Haller (1985)), or when agents are loss averse (Herweg, Mueller, and Weinschenk (2010)), I am the first to demonstrate both theoretically and empirically that this type of contract is desirable for the principal when agents exhibit probability weighting.

To the best of my knowledge, only Spalt (2013) has studied optimal contract design under probability weighting. The main difference with that paper is that I include the agent's incentive compatibility constraint in the principal's program. That is, I study the incentives that result from offering risky contracts. His analysis does not and, as a result of the setting he chose, it ignores that constraint. Another difference is that I study the incentives of a general class of contracts that introduce risk into the agent's compensation. These incentives can be put into practice in multiple ways, which do not necessarily involve compensation plans with stock options.[1]

Second, the results also contribute to the literature on decision theory. To the best of my knowledge, I am the first to provide applications of probability weighting elicitation techniques in the context of incentives. Furthermore, the experimental results illustrate the importance of using parametrized probability weighting functions that separate likelihood insensitivity from optimism/pessimism. I use the various methods proposed by Wakker (2010) and Abdellaoui, Baillon, Placido, and Wakker (2011) to isolate

[1] Another important difference with Spalt (2013) is that I use analytical solutions rather than calibrations to show that stochastic contracts are more effective in motivating agents than more traditional contracting devices. Furthermore, I use an experiment to directly link the subjects' performance under stochastic contracts to their risk preferences, making it possible to cleanly establish whether or not the subjects' probability weighting functions drive the result that stochastic contracts can generate greater performance.

these two components of probability weighting and show that they contribute unequally to the effectiveness of stochastic contracts.

Third, the results add to the literature on incentives. Previous research has shown that replacing fixed payments with lotteries generates higher incentives. For example, allocating interest using lotteries enhances savings behavior as compared to using standard interest payments (Filiz-Ozbay, Guryan, Hyndman, Kearney, and Ozbay (2015), Cole, Iverson, and Tufano (2022), Jindapon, Sujarittanonta, and Viriyavipart (2022)). Payments by means of lotteries also improve healthy behaviors, such as medication adherence and diabetes monitoring, as compared to fixed payments (Volpp, John, Troxel, Norton, Fassbender, and Loewenstein (2008), Kimmel et al. (2012), Haisley, Volpp, Pellathy, and Loewenstein (2012), Sen et al. (2014)). Moreover, lottery-based incentives enhance compliance by reducing tax evasion (Naritomi (2019)) and encouraging law-abiding behavior (Fabbri, Barbieri, and Bigoni (2019)). This study provides a theoretical framework that unifies and generalizes the documented effectiveness of lottery payments. It shows that a principal can achieve a desirable action in a cost-effective way using stochastic contracts. Moreover, the experiment demonstrates that the motivational effect achieved by lotteries is due to the tendency of individuals to overweight small probabilities, in the spirit of Filiz-Ozbay et al. (2015). I take things one step further by showing that likelihood insensitivity—the cognitive component of probability weighting—can explain this amplification of incentives.

2. The model

Consider a principal who delegates a task to an agent. The agent's decision consists of choosing an effort level $e \in [0, \bar{e}]$ to carry out the task. Throughout, it is assumed that the agent experiences disutility from exerting effort, which is embodied in a strictly increasing and convex cost function, $c(e)$.

Assumption 1 (Cost of effort). *$c : [0, \bar{e}] \to [0, +\infty)$ is twice continuously differentiable with $c'(e) > 0$ and $c''(e) > 0$.*

I assume that the task's output, denoted by y, is stochastic. Therefore, y is modeled as a random variable that can take any possible value in the interval $[\underline{y}, \bar{y}]$. Both principal and agent know that production is distributed according to the cumulative density function $F(y|e)$ about which I make the following assumptions.

Assumption 2 (Distribution function). *The function $F : [\underline{y}, \bar{y}] \times [0, \bar{e}] \to [0, 1]$ admits a probability density $f(y|e)$, is twice continuously differentiable with respect to e, and exhibits $F_{ee}(y|e) > 0$.*

It is well known that convexity of the distribution function, $F_{ee}(y|e) > 0$, ensures the optimality of contracts derived from the first-order approach (Mirrlees (1999), Jewitt (1988)).

Importantly, this setting assumes that effort and production are linked according to the monotone likelihood ratio property.

Assumption 3 (Monotone likelihood ratio property). $\frac{\mathrm{d}}{\mathrm{d}y}(\frac{f_e(y|e)}{f(y|e)}) \geq 0$.

Assumption 3 implies that high performance realizations are more likely to be drawn from the distribution of performance when the agent's effort is high than when it is low.

Furthermore, the principal is an expected value maximizer and her objective function is given by

$$\Pi(t, e) = \int_{\underline{y}}^{\bar{y}} \big(S(y) - t(y)\big) f(y|e)\, \mathrm{d}y. \tag{1}$$

The function $S(y)$, which exhibits $S'(y) > 0$ and $S''(y) < 0$, captures the principal's benefits from production. Moreover, $t(y)$, represents the payment to the agent according to the contract; it is therefore the cost to the principal of incentivizing the agent. I assume that $t(y) \in [0, M_1)$ where $M_1 \in [1, +\infty)$; the agent's compensation exhibits an upper bound.

Throughout, I consider a setting in which the principal is interested in eliciting higher effort levels. Thus, the benefits from eliciting higher effort, that is, higher expected benefits from production, outweigh the implied costs, that is, higher expected transfers to the agent. Furthermore, the principal's desire to elicit higher effort and the fact that output and effort are positively linked (Assumption 3) imply that the contract must exhibit $\frac{\mathrm{d}t(y)}{\mathrm{d}y} \geq 0$. Therefore, it is assumed that the contract $t(y)$ increases monotonically in output.

I assume that $t(y)$ enters the agent's utility through the function u about which I make the following assumption.

Assumption 4 (Consumption utility). *Let* $M_2 \in (M_1, +\infty)$. *The function* $u : [0, M_2) \to [0, 1]$ *is twice continuously differentiable and exhibits* $u(0) = 0$, $u(M_2) = 1$, $u'(t) > 0$, *and* $u''(t) < 0$.

The assumptions that $u(0) = 0$ and $u(M_2) = 1$ fix the location and scale of utility. This property will be crucial in deriving the main results.

Consequently, the agent's expected utility (EU, henceforth) when offered a contract with transfer $t(y)$ is

$$\begin{aligned}\mathrm{EU}\big(t(y), e\big) &= \int_{\bar{y}}^{\underline{y}} u\big(t(y)\big)\, \mathrm{d}\big(1 - F(y|e)\big) \\ &= \int_{\underline{y}}^{\bar{y}} u\big(t(y)\big) f(y|e)\, \mathrm{d}y. \end{aligned} \tag{2}$$

Throughout, I relax the standard assumption that the agent's preferences conform to EU (equation (2)). Instead, his preferences are characterized by rank-dependent utility (Quiggin (1982), RDU henceforth) and, therefore, she can distort probabilities. That inaccurate perception of probabilities is modeled using a probability weighting function $w(p)$ about which I make the following assumptions.

Assumption 5 (Probability weighting functions). *Let* $p \in [0, 1]$. *The function* $w(p) : [0, 1] \to [0, 1]$ *is twice continuously differentiable with the following characteristics*:

- $w'(p) > 0$ *for all* $p \in [0, 1]$;
- $w(0) = 0$ *and* $w(1) = 1$;
- *There exists a* $\tilde{p} \in [0, 1]$ *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* $\tilde{p} < 1$;
- *If* $\tilde{p} \in (0, 1)$, *then there exists a* $\hat{p} \in (0, 1)$ *such that* $w(\hat{p}) = \hat{p}$.

By Assumption 5, $w(p)$ is an increasing function that maps the unit interval onto. It contains *at least* two fixed points: one at $p = 0$ and other at $p = 1$. Furthermore, the probability weighting function can exhibit three possible 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 generates an additional interior fixed point, $\hat{p} \in (0, 1)$.

The RDU agent distorts the probabilities associated with y as follows. He has a reference output level, which can be any $Y \in [\underline{y}, \bar{y}]$, and considers the *rank* or probability of obtaining a higher output level, that is, $1 - F(Y|e')$ for a given $e' \in [0, \bar{e}]$. That probability is perceived by this agent as $w(1 - F(Y|e'))$. Thus, for the RDU agent $p = 1 - F(Y|e')$ for any Y and a given e' and that p is transformed with w from Assumption 5. Accordingly, an outcome that is infinitesimally worse than Y generates a difference in perceived ranks captured by the expression $w'(1 - F(Y|e'))$.[2]

Therefore, the agent's rank-dependent utility when he works under a contract $t(y)$ is

$$\begin{aligned} \mathrm{RDU}\big(t(y), e\big) &= \int_{\bar{y}}^{\underline{y}} u\big(t(y)\big)\, \mathrm{d}\big(w\big(1 - F(y|e)\big)\big) \\ &= \int_{\underline{y}}^{\bar{y}} u\big(t(y)\big) w'\big(1 - F(y|e)\big) f(y|e)\, \mathrm{d}y, \end{aligned} \tag{3}$$

where the second equality is due to the continuity of w by Assumption 5. The expression after that equality shows the way in which the RDU agents evaluates $t(y)$. The agent weights the utility implied by a transfer specified in the contract, namely $u(t(Y))$ for some $Y \in [\underline{y}, \bar{y}]$, with the marginal contribution to the perceived rank of the outcome under consideration Y, that is, $w'(1 - F(Y|e))$. For related formulations of RDU in the context of continuous distributions, see Quiggin (1982) and Röell (1987).

Under RDU, the agent's risk attitudes are jointly determined by the curvatures of $u(t)$ and $w(p)$. The former is common to EU and RDU, while the latter is exclusive to RDU. In fact, note that when $w(p) = p$, RDU in equation (3) collapses to EU. The influence of $w(p)$ on risk attitude is informally known in the literature as *probabilistic risk attitude*,

[2]For illustrative purposes, suppose that there are three relevant output levels $y_1, y_2, y_3 \in [\underline{y}, \bar{y}]$ such that $y_3 > y_2 > y_1$ and consider a given $e' \in [0, \bar{e}]$. The rank of y_3 is 0 (as there is no better outcome than y_3), the rank of y_2 is $1 - F(y_2|e')$, and the rank of y_1 is $1 - F(y_1|e')$. Probability distortions are computed as the marginal contribution of the weighted probability associated with an outcome to its weighted rank. Accordingly, the perceived probability of y_3 is $w(1 - F(y_2|e'))$, that of y_2 is $w(1 - F(y_1|e')) - w(1 - F(y_2|e'))$, and that of y_1 is $1 - w(1 - F(y_1|e))$.

a term that is adopted here. I focus on the effects of this nontraditional source of risk attitude on optimal contracting.

To illustrate the importance of probabilistic risk attitude in optimal contracting, consider an agent with the preferences given in equation (3) who overweights the likelihood of obtaining high output levels. That error might make him less motivated to exert effort since he mistakenly believes that high output levels are more likely to be attained than they actually are. This probabilistic mistake makes him unwilling to exert high effort levels.

In order to confine the analysis to probability distortion, and to limit the influence of other potential errors, it is assumed that the RDU individual is able to reduce compound lotteries. The following assumption is based on Segal (1990).

Assumption 6 (Reduction of compound lotteries). *Let* $Y = (p_1, y_1; \dots; p_n, y_n)$ *and* $X = (q_1, x_1; \dots; q_m, x_m)$ *be lotteries. Define their compound lottery as* $L_C = (r, Y; 1 - r, X)$ *and the reduced lottery as*

$$L_R = \big(p_1 r, y_1; \dots; p_n r, y_n; \dots; q_1(1-r), x_1; \dots; q_m(1-r), x_m\big).$$

The RDU decision maker knows that $\mathrm{RDU}(L_C, e) = \mathrm{RDU}(L_R, e)$ *for a given* e.

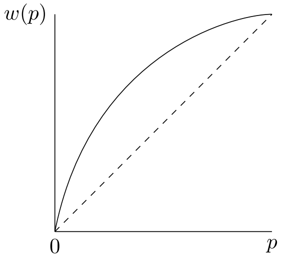
The ability to reduce compound lotteries is not incompatible with RDU. While violating Assumption 6 indeed implies RDU, as shown by Segal (1990), other errors can also lead to this nonexpected-utility representation, such as, for example, dynamic inconsistency, that is, not being able to perform backward induction, and/or inconsequentialism, that is, being affected by counterfactual risks (Machina (1989), Karni and Schmeidler (1991)). Empirical studies have found that consequentialism is often violated (Cubitt, Starmer, and Sugden (1998)). Thus, imposing Assumption 6 does not necessarily contradict RDU and is consistent with empirical evidence.[3]

Another theory of risk that incorporates probability weighting functions is Cumulative Prospect Theory (Tversky and Kahneman (1992), CPT henceforth), which is a more descriptive version of RDU. Like the RDU agent, an agent with CPT preferences exhibits probabilistic risk attitudes but also displays loss aversion. For the sake of brevity, the analysis of incentives produced by stochastic contracts for agents with CPT preferences is relegated to Supplemental Appendix C (Appendices B–G may be found in the Supplementary Material (Gonzalez-Jimenez (2024b))).

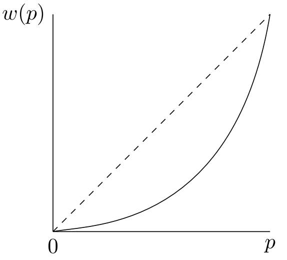
2.1 *Probabilistic risk attitudes and their decomposition*

This subsection can be omitted by readers already acquainted with the concepts of likelihood insensitivity, pessimism, and optimism toward risk (Tversky and Wakker (1995), Wakker (2010), Yaari (1987)). In order to exhaustively analyze the effect of probabilistic

[3]Notably, Assumption 6 can be weakened in order to obtain the main result of the paper. Alternatively, it can be assumed that the decision maker exhibits stronger probability overweighting for the compounded probability than for each probability separately. This weaker condition is likely to hold when the agent overweights smaller probabilities to a greater extent than larger ones.



(a) Example of optimism



(b) Example of pessimism

FIGURE 1. Motivational sources of probability distortion.

risk attitude on contracting, I follow Wakker (2010) by making a distinction between two components of probability weighting.

The first component captures *motivational* deviations from EU stemming from pessimist or optimist attitudes toward risk. This affects probability evaluations because the agent irrationally believes that unfavorable outcomes (in the case of pessimism), or favorable outcomes (in the case of optimism) realize more often than they actually do. Pessimism is represented by a convex weighting function while optimism is represented with a concave weighting function. Figure 1 presents graphical examples of each.

DEFINITION 1. Optimism (pessimism) is characterized by a probability weighting function $w(p)$ with the properties described in Assumption 5 and $\tilde{p} = 1$ ($\tilde{p} = 0$).

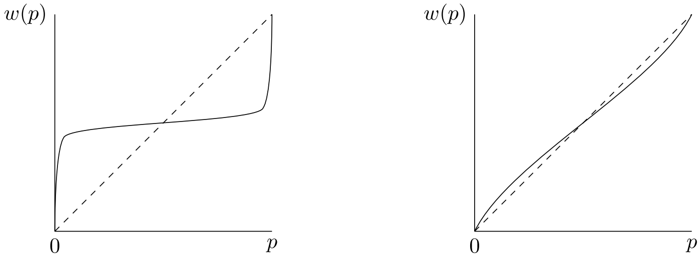
The following definition, due to Yaari (1987), makes it possible to compare degrees of optimism or pessimism.

DEFINITION 2. An agent i with weighting function $w_i(p)$ is more optimistic (pessimistic) than an agent j with weighting function $w_j(p)$ if $w_i(p) = \theta(w_j(p))$, where $\theta : [0, 1] \to [0, 1]$, is twice continuously differentiable with $\theta'(w_j) > 0$ and $\theta''(w_j) < 0$ ($\theta''(w_j) > 0$).

The second component of probability weighting is likelihood insensitivity (Tversky and Wakker (1995), Wakker (2010)), which captures the notion that individuals distort probabilities due to cognitive and perceptual limitations. As a result, they are insufficiently sensitive to changes in intermediate probabilities and overly sensitive to changes in extreme probabilities. An extreme characterization of likelihood sensitivity is a step-shaped probability weighting function that assigns $w(p) \approx 0.5$ to all interior probabilities $p \in (0, 1)$. At the other extreme is an EU agent who is fully sensitive to probabilities, that is, $w(p) = p$. Figure 2 presents graphical examples of likelihood insensitivity.[4]

DEFINITION 3. Likelihood insensitivity is characterized by a probability weighting function $w(p)$ with the properties described in Assumption 5 and $\tilde{p} = \hat{p} = 0.5$.

[4]These two phenomena are referred to as *curvature* and *elevation* in psychology (Gonzalez and Wu (1999)). I adhere to the jargon used in economics.



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

FIGURE 2. Cognitive sources of probability distortion.

The following definition, based on Tversky and Wakker (1995), makes it possible to compare degrees of likelihood insensitivity.

DEFINITION 4. An agent i with weighting function $w_i(p)$ is more likelihood insensitive than an agent j with weighting function $w_j(p)$ if $w(p)_i = \phi(w(p)_j)$, where $\phi : [0, 1] \to [0, 1]$ exhibits likelihood sensitivity in the sense of Definition 3.

Likelihood insensitivity and either optimism or pessimism can occur simultaneously. This coexistence generates probabilistic risk attitudes that can be represented by an inverse-S probability weighting function. However, the location of the interior fixed point, $\hat{p}$, crucially depends on the motivational factor. A likelihood-insensitive pessimist, exhibits a $w(p)$ with an interior fixed point located in the interval $\hat{p} \in (0, 0.5)$ and, therefore, he exhibits probabilistic risk aversion more often than probabilistic risk seeking. In contrast, a likelihood-insensitive optimist has a $w(p)$ with an interior fixed point in the interval $\hat{p} \in (0.5, 1)$ and, therefore, exhibits probabilistic risk seeking more often than probabilistic risk aversion.

2.2 *The optimality of stochastic contracts*

In the case of an RDU agent with probabilistic risk attitudes, the principal's problem is to implement a contract, which guarantees that the agent will accept the contract, and that it will incentivize him to exert a high level of effort. Formally,

$$\begin{aligned} \max_{\{t(y)\}} \quad & \int_{\underline{y}}^{\bar{y}} \big(S(y) - t(y)\big) f(y|e)\, \mathrm{d}y \\ \text{s.t.} \quad & \int_{\bar{y}}^{\underline{y}} u\big(t(y)\big)\, \mathrm{d}\big(w\big(1 - F(y|e)\big)\big) - c(e) \geq \bar{U}, \\ & \max_{\{e\}} \int_{\bar{y}}^{\underline{y}} u\big(t(y)\big)\, \mathrm{d}\big(w\big(1 - F(y|e)\big)\big) - c(e). \end{aligned} \tag{4}$$

The following lemma shows that the contract resulting from the first-order approach is optimal when the agent exhibits pessimism. The proofs of the main theoretical results are relegated to Appendix A.

Lemma 1. *Let Assumptions 1–6 hold. The candidate solution from the first-order approach, denoted by t^{fo}, is optimal if $w''(p) \geq 0$ for $p \in (0, 1)$.*

Offering the contract resulting from the first-order approach can be suboptimal when the shape of the weighting function generates probabilistic risk-seeking attitudes. There are two underlying reasons for this suboptimality. First, the contract offers too much insurance because the RDU agent might have a taste for risk, and that risk-seeking attitude makes the participation constraint to be slack at the optimum. Second, the contract might deliver insufficient incentives as the agent's misperception of probabilities may lead him to the erroneous belief that low effort suffices to obtain high production levels.

The following proposition formalizes the aforementioned reasoning by showing that an individual suffering from optimism or likelihood insensitivity can be demotivated by a deterministic contract $t(q)$. This includes, but is not limited to, the contract resulting from the first-order approach.

Proposition 1. *Let Assumptions 1–6 and either optimism or likelihood insensitivity hold. Consider any pair $e'', e' \in [0, \bar{e}]$ such that $e'' > e'$. The RDU agent working under $t(q)$ exhibits*:

(i) *a lower willingness to choose e'' over e' than if he were EU when $w'(1-F(y|e'')) < 1$;*

(ii) *a higher willingness to choose e'' over e' than if he were EU when $w'(1-F(y|e')) > 1$.*

An agent characterized by optimism or likelihood insensitivity is less motivated relative to the hypothetical situation in which he did not make any probabilistic mistakes when the increase in probability associated with choosing a higher level of effort occurs at intermediate and, in the case of the optimist, low output levels, that is, all y such that $w'(1 - F(y|\tilde{e})) < 1$ for any $\tilde{e} \in [0, \bar{e}]$. These output levels are erroneously perceived to be less likely to realize than they actually are, which leads this RDU agent to underestimate the profitability from exerting higher effort.

However, note that Proposition 1 also shows that probability distortions can be motivating. If the increase in probability associated with choosing a higher level of effort occurs at high output levels, that is, all y such that $w'(1 - F(y|\tilde{e})) > 1$, then an agent suffering from probability distortions due to optimism or likelihood insensitivity is more motivated than his hypothetical EU counterpart.

I next show that as optimism and likelihood insensitivity become more severe, the agent is more likely to be demotivated as a result of his erroneous perception of probabilities. Thus, the motivating effect of probability distortions is confined to an increasingly smaller output segment.

Corollary 1. *Let $e'', e' \in [0, \bar{e}]$ such that $e'' > e'$. Stronger optimism and likelihood insensitivity enlarge the interval of y in which $w'(1 - F(y|e')) < w'(1 - F(y|e'')) < 1$ holds, making Proposition 1(i) more likely.*

To remedy the demotivation caused by probabilistic risk-seeking attitudes and to account for the RDU agent's conceivable taste for risk, I now look at a contract that introduces additional risk into the agent's environment. Specifically, I consider stochastic contracts of the form $L := (r, T(y); 1-r, 0)$, where $T(y)$ represents a monetary compensation that depends on the output produced in the task and $r \in (0, 1)$ is a probability chosen by the principal. Throughout, I use r to refer to the probability included in the stochastic contract while p refers to the generic probability.

The timing of the stochastic contract L is as follows. The principal moves first by choosing r and $T(y)$, which are then communicated to the agent before he makes a decision. Next, the agent chooses e. Finally, a random device to which the principal has credibly committed determines whether or not the agent's compensation depends on output, y.

The demotivation formalized in Proposition 1(i) can be avoided using L because the principal can target the output segment in which the agent's motivation is the greatest. In other words, she can set r small enough so as to locate in the region where the weighting function exhibits $w' > 1$. In that region, the agent is motivated to exert a higher level of effort relative to the hypothetical scenario in which he did not make those mistakes, as shown by Proposition 1(ii). Such a tool was not available to the principal in the case of a deterministic contract $t(y)$. Therefore, she now has two channels by which to motivate the agent: (i) monetary rewards and (ii) changes in the likelihood that those rewards are realized.

In order to focus exclusively on the incentives generated by choosing different levels of probabilities r, I make the simplifying assumption that stochastic contracts offer, in expectation, the same monetary rewards as a traditional deterministic contract, $t(y)$.

Assumption 7 (Cost equivalence). *Stochastic and traditional contracts are, in expectation, cost-equivalent for the principal, that is,* $T(y) = \frac{t(y)}{r}$.

Thus, if the agent exhibits probabilistic risk-seeking attitudes, the principal can replace the suboptimal contract t^{fo} with the stochastic contract $L = (r, \frac{t^{\text{fo}}}{r}; 1-r, 0)$. In this way, she exploits the agent's risk preferences without incurring any additional average cost.

A consequence of Assumption 7 is that the payment included in the stochastic contract becomes larger as r becomes smaller. It is however assumed that the largest payment possible payment offered by the contract, which is obtained as r approaches zero, is not unbounded. Formally, I assume that $\lim_{r \to 0} \frac{M_1}{r} = M_2$. Consequently, the model has the restriction that for small enough r, then M_1 should be also sufficiently small so as to maintain the boundedness of $T(y)$.[5]

[5]In a previous version of the model, I did not encounter that restriction but the model relied on the normalization $u(1) = 1$, which is less general as it crucially depends on the scale used to define monetary outcomes.

The agent's rank-dependent expected utility when offered contract L is given by

$$\begin{aligned}\mathrm{RDU}(L, e) &= \int_{\bar{y}}^{\underline{y}} u\big(T(y)\big)\,\mathrm{d}\big(w\big(r\big(1 - F(y|e)\big)\big)\big) \\ &= \int_{\bar{y}}^{\underline{y}} u\big(T(y)\big)w(r)\,\mathrm{d}\big(w\big(1 - F(y|e)\big)\big) \\ &= \int_{\bar{y}}^{\bar{y}} u\big(T(y)\big)w(r)w'\big(1 - F(y|e)\big)f(y|e)\,\mathrm{d}y, \end{aligned} \tag{5}$$

where the second equality is due to Assumption 6 and the third equality is due to the continuity of w by Assumption 5.

The following proposition presents the condition ensuring the optimality of stochastic contracts.

Proposition 2. *Let Assumptions 1–7 hold. A contract L improves upon any contract $t(y)$ if and only if $w(r) > u(r)$.*

The optimality of stochastic contracts crucially depends on the functions u and w, the determinants of risk attitude under RDU. When the agent sufficiently overweights the probability chosen by the principal, such that the condition $w(r) > u(r)$ holds, the stochastic contract creates higher motivation. The expression $u(r)$ captures the probabilistic risk-attitudes equivalent of the risk attitude generated by the utility curvature. In other words, it captures the agent's risk attitudes emerging from utility curvature when these are brought to the probability space. This interpretation of $u(r)$ is possible because of the assumptions $u(0) = 0$ and $u(M_2) = 1$, which fix the location and scale of consumption utility and make the risk attitude resulting from the utility curvature comparable to the risk attitude embodied in the probability weighting function. Thus, the proposition states that if the r chosen by the principal ensures that the agent's probabilistic risk-seeking attitude is stronger than his risk-averse attitude resulting from the utility curvature, then the agent will be risk seeking, will have a taste for risky contracts, and will be more motivated under a stochastic contract implemented with probability r.

The intuition behind this result is illustrated by the following example.

Example 1 (Linear utility). Let $u(t(y)) = t(y)$. The condition in Proposition 2 now becomes $w(r) \geq r$. The stochastic contract L needs to be implemented with a probability that is overweighted by the agent. Thus, the stochastic contract is optimal as long as $\hat{r} > 0$, such that the agent exhibits overweighting of probabilities in a nonempty interval.

The example becomes more illuminating if we consider Prelec's (1998) weighting function $w(p) = \exp(-\beta(-\ln(p))^{\alpha})$. First, suppose that $\alpha = 1$, in which case the weighting function becomes $w(p) = p^{\beta}$. In this case, $\hat{p} = 1$ if $\beta < 1$ and the condition in Proposition 2, that is, $w(r) \geq r$, holds for any $r \in (0, 1)$. In words, when the weighting function is everywhere concave, which according to Definition 1 is equivalent to the agent exhibiting optimism, any interior probability ensures that the stochastic contract generates a higher level of effort.

Now suppose instead that $\beta = 1$. The weighting function becomes $w(p) = \exp(-(-\ln(p))^{\alpha})$, which has an inverse-S shape if $\alpha < 1$ and, therefore, is characterized by $\hat{p} = \frac{1}{\mathrm{e}}$. As a result, stochastic contracts will generate higher output if they target the region of probability overweighting by ensuring that $r \in (0, \frac{1}{\mathrm{e}})$ when $\alpha < 1$.

Example 2 in Appendix A provides further intuition by considering the more general case in which the agent has a CRRA utility function. Moreover, Supplemental Appendix B.1 shows the principal is better off implementing a stochastic contract with probability r under virtually the same condition as that in Proposition 2, that is, $w(r) > u(r)$, when production is assumed to be deterministic. Thus, the principal is better off introducing risk in a setting where the agent would not otherwise have to face any.

A less general but nevertheless important implication of Proposition 2 is that introducing risk in the agent's environment can be counterproductive when the agent's risk preference is characterized by EU. In that case, the principal would be better off incentivizing the agent with a standard contract, which highlights the importance of RDU preferences for the implementability of stochastic contracts.

Corollary 2. *Let Assumptions 1–4, 7, and $w(p) = p$ hold. Under those conditions, contract L cannot be optimal.*

The additional risk introduced by the stochastic contract reduces the EU agent's utility, which can lead to contract rejection. Furthermore, since Assumption 3 entails $-F_e(y|e) > 0$, the lower expected utility generated by the stochastic contract translates into a lower marginal expected benefit from exerting effort. Thus, under "standard risk aversion" the stochastic contract is more likely to be rejected and delivers lower motivation.[6]

Taken together, Proposition 2 and Corollary 2 imply that a stochastic contract is exploitative. It leads the RDU agent to deliver a higher level of effort than if he were an expected-utility decision maker. This is because the contract allows the principal to take advantage of the agent's misperception of probabilities to extract a higher level of effort.

I conclude the theoretical framework by analyzing the contribution of motivational and cognitive components of probability weighting to the optimality of stochastic contracts. The following proposition shows that optimism alone guarantees the optimality of stochastic contracts.

Proposition 3. *Let Assumptions 1–7 hold. Then there exists a level of optimism that guarantees Proposition 2.*

[6]The opposite result, namely that the stochastic contract generates higher motivation under EU, can be obtained under different theoretical assumptions. Specifically, if Assumption 7 is abandoned and if it is assumed that the agent perceives contract L as a fraction $rT(y)$ rather than as the payment $T(y)$ realized with probability r, then that result can hold. Under those conditions, an agent with sufficiently strong utility curvature exerts a higher level of effort under the stochastic contract. He will work harder under the stochastic contract implemented with small probabilities in order to compensate for the potentially lower earnings that he might otherwise receive.

To better understand Proposition 3, recall that optimism is equivalent to a concave probability weighting function, which implies that all interior probabilities are overweighted. Consequently, the agent exhibits probabilistic risk-seeking attitudes for all probabilities. Stronger optimism, in the sense of Definition 2, implies that interior probabilities are overweighted to a greater extent and that the agent exhibits stronger probabilistic risk-seeking. The proposition states that the risk-averse attitudes resulting from the curvature of the utility function can be offset by a sufficiently strong level of optimism. If an agent attains that degree of optimism, then he will be risk-seeking and more motivated under a stochastic contract.

I next show that likelihood insensitivity alone guarantees the optimality of stochastic contracts.

Proposition 4. *Let Assumptions 1–7 hold. Then there exists a level of likelihood insensitivity that guarantees Proposition 2 if* $u(r) < \tilde{p}$.

Likelihood insensitivity is equivalent to an inverse-S-shaped probability weighting function. Accordingly, the agent exhibits probabilistic risk-seeking attitudes in the interval $p \in (0, \hat{p})$. As the agent becomes more likelihood insensitive, in the sense of Definition 4, the probabilities in that interval are overweighted to an increasing extent and he exhibits stronger probabilistic risk-seeking for those probabilities. Proposition 4 states that the potential risk-averse attitudes emerging from the curvature of the utility function can be offset by a sufficiently strong degree of likelihood insensitivity. If the agent attains that degree of likelihood insensitivity, he will then be risk-seeking and more motivated under a stochastic contract with $r \in (0, \hat{p})$.

A number of extensions presented in the Supplemental Appendixes corroborate and generalize Proposition 2. Supplemental Appendix B.2 shows that a result similar to Proposition 2 is obtained in a setting where effort is binary. That framework is less realistic but nonetheless more widely used in the literature because the solutions from the first-order approach are optimal without requiring the restrictive assumption of a convex density function (Assumption 2). Supplemental Appendix C shows that the result resented in Proposition 2 holds under similar conditions when the agent exhibits reference-dependent preferences. The convexity of utility in the domain of losses, that is, all outcomes below the reference point, amplifies the agent's risk-seeking attitude and makes stochastic contracts more appealing to the agent.

3. Experimental method

3.1 *The general setup*

The experiment was conducted at Tilburg University's CentERLab (Gonzalez-Jimenez (2024a)). The participants were students at that university and were recruited by e-mail. The data consist of 15 sessions with a total of 172 subjects. On average, a session lasted approximately 80 minutes. Between 8 and 18 subjects took part in each session. The currency used in the experiment was euros. Z-Tree was used to conduct the experiment

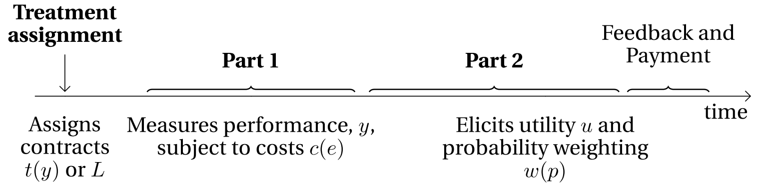


Figure 3. Timeline of the experimental design.

(Fischbacher (2007)). Subjects earned 15.83 euros on average. The instructions of the experiment are presented in Supplemental Appendix G.

Figure 3 provides a timeline of the experiment, which connects the parts of the experiment to the theoretical model presented in Section 2. It shows that the experiment consisted of two parts. Upon arrival, participants were informed that their earnings in either part 1 or part 2 of the experiment would become their final earnings and that choice would be made randomly at the conclusion of the experiment.[7]

In part 1, the subjects' motivation to perform a task that demanded their attention and effort was measured. In terms of the model, output (y) is measured when exerting effort is costly to a subject $(c(e))$. The task required subjects to sum five two-digit numbers multiple times, a real-effort paradigm used in previous research to measure effort exertion in work-related settings (see, for instance, Niederle and Vesterlund (2007), Charness, Cobo-Reyes, Lacomba, Lagos, and Pérez (2016)). The numbers to be summed were randomly chosen by the computer. On submitting an answer, a new set of numbers would appear on the subject's computer screen. Subjects went through 10 rounds of 4 minutes each in which they completed as many calculations as they could.

Real-effort tasks provide a more realistic account of field settings because they capture psychological aspects of working environments (Charness, Gneezy, and Henderson (2018)). These aspects are crucial to this paper. This advantage is offset, however, by the inability to observe the subjects' cost of effort (Charness, Gneezy, and Henderson (2018)). Therefore, it is not possible to know whether a subject exerted an optimal level of effort in the experiment. Previous research has found qualitatively similar results when real-effort tasks are compared to stated-effort paradigms, in which the cost of effort is exogenously determined by the experimenter (Brüggen and Strobel (2007), Charness et al. (2016), Dutcher, Salmon, and Saral (2015)). These findings suggest that subjects are equally capable of making optimal trade-offs in both types of tasks. Therefore, the considerable advantage of real-effort tasks appears to significantly outweigh its disadvantage, and thus justifies its use in the experiment.

In part 2, the subjects' task was to choose between two binary lotteries multiple times. As can be seen in Figure 3, this part of the experiment was designed to elicit the subject's consumption utility (u) and probability weighting functions (w). I used

[7]This randomization of payments may be a source of concern if subjects distort probabilities. However, as will be shown in Section 6, subjects on average exhibit $w(0.5) \approx 0.5$, so that the probability underlying this randomization of payments was more or less perceived accurately. Moreover, and as will explained below, isolation guarantees that this randomization of payments generates an accurate measurement of risk attitude and effort elicitation.

the two-step method developed by Abdellaoui (2000), a risk preference measurement tool that first measures utility and then probability weighting functions. 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.[8] Subjects were informed that one of their chosen lotteries would be selected randomly at the end of the experiment and that its realized outcome would determine their earnings in part 2 of the experiment.

At the conclusion of part 2, subjects were given feedback about their performance and were informed of their earnings in part 1 of the experiment. They were also informed about which lottery was chosen to determine their compensation in the second part of the experiment and of the lottery's outcome. In addition, subjects learned whether part 1 or part 2 counted toward their final earnings. Providing feedback at the end of the experiment minimized carry-over effects (Cox and Epstein (1989)).

3.2 *Treatments in part 1*

There were four treatments which differed according to the type of incentives offered to perform the real-effort task. Each subject was randomly assigned to one of them. This randomization attenuates the aforementioned disadvantage that the cost-of-effort function in real-effort tasks is unobservable by ensuring that potential deviations from the optimal effort level due to overconfidence, fatigue, or social desirability are on average similar across treatments. Therefore, when the treatments are compared, these potential deviations are averaged out. Thus, it is not essential that all subjects choose the optimal effort level, a strong requirement when using a real-effort task. Instead, it is only required that optimal effort is on average exerted in each treatment, and this assumption is further supported by the random allocation of subjects into the treatments.

A traditional linear contract of the form $t_d(y) := ay$, where $a > 0$ represents a monetary amount, served as the benchmark.[9] While linear contracts are only optimal under rather specific conditions (see, e.g., Holmstrom and Milgrom (1987)), they are commonly used by organizations and in the study of incentives (Chiappori and Salanié (2002), Bolton and Dewatripont (2005)). The treatment that used linear contracts is referred to as *Piecerate*. Subjects assigned to that treatment were given 0.25 euros for every correct summation ($a = 0.25$).

At this point, it is worth emphasizing that the assumption underlying part 1 is that performance in the real-effort task is stochastic. That is, subjects might fall short of or exceed their intended performance level due to external factors, such as distractions due to noise outside the lab, and/or internal factors, such as a run of easy numbers. The

[8] A drawback of this method is that it may violate incentive compatibility when subjects are aware of the chained nature of the questions they face. I overcame this disadvantage by randomly adding questions that were not used in the analysis of the data, and by randomizing the appearance of the lotteries in decision sets 7 to 11, which will be described in more detail in Section 3.3.

[9] A more general representation of these contracts would be $t = F + ay$ where $F \geq 0$ is a fixed payment that does not depend on the subject's performance. Since F does not generate any incentive for the agent to exert effort, the normalization $F = 0$ is assumed throughout.

noise inherent to the task is how the experiment captures Assumption 2.[10] Section 5 below will show that this assumption is validated by the data. Therefore, subjects assigned to Piecerate perceive the environment to be stochastic, such that their effort may translate into higher or lower production than intended. Importantly, the fact that performance in the task is stochastic does not make it unworthwhile to exert effort. This is because I also assume, and empirically corroborate, that greater effort pays off because it generates a higher likelihood that higher performance is achieved, which is in line with Assumption 3.

The other three treatments also involved monetary rewards that depend on individual performance in the task. However, subjects in those treatments faced the additional risk that performance in a particular round might not count toward their earnings. The magnitude of that risk was varied across treatments to simulate stochastic contracts that are implemented with different probabilities. Thus, treatments *LowPr*, *MePr*, and *HiPr* featured a low, medium, and high probability, respectively, that performance in a given round would count toward the subject's earnings. In LowPr, subjects faced a 10% chance that performance in a given round would count toward their earnings. This was implemented by randomly choosing only one round (out of ten) at the end of the experiment and paying according to the subject's performance in that round. Subjects were informed about this payment rule before they started working on the task. In MePr, three rounds were randomly chosen at the end of the experiment and only performance in those rounds was rewarded, while in HiPr five rounds were randomly chosen. Essentially, these three treatments can be understood as altering the payment terms of Piecerate to create stochastic contracts.

The chosen experimental representation of stochastic contracts assumes isolation between the rounds. That is, the subjects' decision to exert effort in any round is not affected by their decisions in other rounds. This property is strongly supported by the results of other experiments that used the *random incentive system*, which involves paying one task or one exercise in the experiment at random (see, for instance, Cubitt, Starmer, and Sugden (1998), Hey and Lee (2005), Lee (2008), Baltussen, Post, van den Assem, and Wakker (2012)).[11] Importantly, isolation guarantees the accurate measurement of risk preference when using the random incentive system (Baltussen et al. (2012), Hey and Lee (2005)). This includes utility and probability-weighting measurements (Bruhin, Fehr-Duda, and Epper (2010), Abdellaoui et al. (2011), L'Haridon and Vieider (2019), Bernheim and Sprenger (2020)). Therefore, under isolation, the chosen representation of stochastic contracts in the experiment generates the desired incentives resulting from

[10] Abstracting from that assumption, by treating those external and internal factors as negligible, does not necessarily invalidate the experimental design. Supplemental Appendix B.1 shows that in a setting in which output is deterministic, similar conditions imply that stochastic contracts are more motivating. Thus, the experiment remains valid regardless of the deterministic nature of output.

[11] A widespread misunderstanding regarding the random incentive system is that the independence axiom is a necessary condition in order to guarantee appropriate experimental measurement, that is, that subjects choose their level of effort as if they are paid for each decision and in the absence of income effects. While the independence axiom, along with some dynamic principles, *suffices* for proper experimental measurement, isolation, on its own, also guarantees it (Baltussen et al. (2012)).

probability distortion and produces the same incentives as in an experiment with a single round. In contrast, a randomization of payments across subjects has been shown to yield biased estimates of risk preference (Baltussen et al. (2012)). This led to my choice of randomizing across rounds, rather than across subjects.

There are conditions other than isolation that ensure proper measurement under the random lottery incentive system; however, they are not sufficient for the purposes of this experiment. Azrieli, Chambers, and Healy (2018) and Azrieli, Chambers, and Healy (2020) show that monotonicity guarantees proper measurement of effort when subjects are paid for only *one round*. Thus, monotonicity would not rule out income effects in MePr, HiPr, and Piecerate. Consequently, potential treatment effects would not be a direct consequence of incentives but might be due to income effects, or the lack thereof. Isolation, which according to Azrieli, Chambers, and Healy (2020) requires transitivity as well as monotonicity, is thus necessary for the experimental design to cleanly test incentives in the absence of income effects. Section 5 provides empirical evidence that treatment effects emerge even in the absence of income effects.

As in the theoretical framework, the monetary incentives offered in Piecerate, LowPr, MePr, and HiPr are calibrated so as to be equal in expectation (see Assumption 7). Accordingly, a subject assigned to LowPr received 2.50 euros for a correct calculation in the round that was chosen for compensation, which is tenfold the compensation in Piecerate for a correct calculation. This difference exactly corresponds to the difference in the probability of being compensated in a round between these two treatments. Similarly, subjects assigned to the MePr and HiPr treatments received 0.85 and 0.50 euros, respectively, for a correct calculation in the rounds that were chosen for compensation. An important implication of this calibration is that if subjects violated isolation and instead integrated their decisions in the various rounds by letting their decision in one round depend on those in others, then they would still exhibit, on average, similar performance across the treatments.

The probabilities in LowPr, MePr, and HiPr were based on a commonly reported finding in the literature on decision-making under uncertainty, according to which individuals distort probabilities using an inverse-S-shaped probability weighting function with an interior fixed point at approximately $p = 0.33$ (see Fehr-Duda and Epper (2012), Wakker (2010, p. 204)). If subjects indeed exhibit this behavior, then they should on average overweight the probability in the case of a round with a 10% chance of being chosen, underweight the probability in the case of a round with a 50% chance, and evaluate more or less accurately the probability in the case of a round with a 30% chance. The treatments were thus designed to generate differences in performance in the presence of overweighting or underweighting of probabilities.

3.3 *Elicitation of risk preference*

The second part of the experiment consisted of 11 decision sets. The first six decision sets are based on the first step in Abdellaoui's (2000) methodology, which in turn is based on Wakker and Deneffe (1996). These sets are meant to elicit a sequence of certainty equivalents $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ that make the subject indifferent between a lottery

Table 1. Example of Abdellaoui's (2000) algorithm.

Iteration	Left Panel			Right Panel		
#	Lotteries Available	Utility Interval	Choice	Lotteries Available	Probability Interval	Choice
1	$L = (0.66, 1; 0.33, 0.50)$ $R = (0.66, 3.70; 0.33, 0)$	[1, 6.40]	L	$L = (1, x_1)$ $R = (0.5, x_6; 0.5, 1)$	[0, 1]	L
2	$L = (0.66, 1; 0.33, 0.50)$ $R = (0.66, 5.05; 0.33, 0)$	[3.70, 6.40]	R	$L = (1, x_1)$ $R = (0.75, x_6; 0.25, 1)$	[0.50, 1]	L
3	$L = (0.66, 1; 0.33, 0.50)$ $R = (0.66, 4.38; 0.33, 0)$	[3.70, 5.05]	R	$L = (1, x_1)$ $R = (0.87, x_6; 0.13, 1)$	[0.75, 1]	R
4	$L = (0.66, 1; 0.33, 0.50)$ $R = (0.66, 4.04; 0.33, 0)$	[3.70, 4.38]	L	$L = (, x_1)$ $R = (0.81, x_6; 0.19, 1)$	[0.75, 0.87]	L
5	$L = (0.66, 1; 0.33, 0.50)$ $R = (0.66, 4.21; 0.33, 0)$	[4.04, 4.38]	L	$L = (1, x_1)$ $R = (0.85, x_6; 0.15, 1)$	[0.81, 0.87]	L
Result	$x_1 \in [4.21, 4.38]$			$p_1 \in [0.85, 0.87]$		

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

$L = (x_{j-1}, 2/3; 0.5, 1/3)$ and a lottery $R = (x_j, 2/3; 0, 1/3)$ for $j = \{1, \dots, 6\}$. Importantly, these lotteries were designed so that the elicited sequence of certainty equivalents guarantees equally spaced utility levels, that is, $u(x_j) - u(x_{j-1}) = u(x_{j-1}) - u(x_{j-2})$ for each $j = \{1, \dots, 6\}$. Thus, the subjects preference for monetary outcomes can be inferred by comparing $x_j - x_{j-1}$ to $u(x_j) - u(x_{j-1})$ for $j = \{1, \dots, 6\}$.

Indifference between lotteries was found through bisection. Thus, instead of asking a subject to directly report x_j, it was elicited through a series of binary choices, an approach that has been shown to minimize mistakes (Bostic, Herrnstein, and Luce (1990), Dixon and Mood (1948), Fischer, Carmon, Ariely, and Zauberman (1999)). The left panel of Table 1 presents an example of the bisection procedure for these decision sets.[12]

Decision sets 7 to 11 are based on the second step of Abdellaoui's (2000) methodology. They were designed to elicit the following sequence of probabilities:

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

These probabilities made subjects indifferent between a lottery $L = (x_6, w^{-1}(p_{j-1}); x_0, 1 - w^{-1}(p_{j-1}))$ and a degenerate lottery x_{j-1} for $j = \{2, \dots, 6\}$. Again, indifference between the lotteries was found through bisection. Note that these lotteries were designed

[12]The bisection algorithm starts with x_j being the midpoint of the considered interval $[x_{j-1}, x_{j-1} + \delta]$, where the value δ was set at one-fifth of what a subject earned in the first part of the experiment. This was done to more accurately relate the subjects' risk preference to their behavior in the first part of the experiment. Subjects were not informed about this calibration. After a subject expressed a preference, the outcome x_j of lottery R changed as a function of the subject's choice. If L was chosen, x_j was replaced by a more attractive outcome in the next iteration, namely the midpoint of the new interval $[\frac{x_{j-1}+\delta+x_{j-1}}{2}, \delta]$. On the other hand, if R was chosen, x_j was replaced by a less attractive outcome in the next iteration, namely the midpoint of the new interval $[x_{j-1}, \frac{x_{j-1}+\delta+x_{j-1}}{2}]$. The subject was then asked to choose again between lotteries L and R, a process that was repeated four times. The indifference point was assumed to be the midpoint of the remaining utility interval.

so that the elicited probabilities yield equally spaced probability weights, that is, $w(p_j) - w(p_{j-1}) = w(p_{j-1}) - w(p_{j-2})$ for each $p_{j-1} = \frac{j-1}{6}$ where $j = \{2, \dots, 6\}$. Probabilistic risk attitudes can be inferred from comparing $w(p_j) - w(p_{j-1})$ to $w^{-1}(p_j) - w^{-1}(p_{j-1})$. The right panel of Table 1 presents an example of the bisection procedure for these decision sets.

4. Hypotheses

The model generates a set of hypotheses that can be experimentally tested. To make more crisp predictions, I make two assumptions about the subjects' preferences. The first is linear utility, which is consistent with the moderate stakes used in the experiment and empirical evidence for linear utility functions in the case of small monetary rewards (Wakker and Deneffe (1996), Abdellaoui (2000), Epper, Fehr-Duda, and Bruhin (2011), Abdellaoui et al. (2011)).[13] The second assumption, which is also consistent with empirical evidence, is that subjects have an inverse-S-shaped probability weighting function with an interior fixed point at approximately $\hat{p} = 0.33$ (Wakker (2010), Fehr-Duda and Epper (2012)). Under these assumptions, the predictions of performance generated by Proposition 2 can be more easily understood using Example 1 in Section 2.2.

Since risk attitude is fully determined by the shape of the probability weighting function, a subject must be risk-seeking for $p \in (0, 0.33)$ and risk-averse for $p \in (0.33, 1)$. Therefore, a stochastic contract implemented with probability $r = 0.10$ should motivate subjects to a greater extent than the piece-rate contract. In contrast, a stochastic contract with $r = 0.30$ should yield the same level of motivation as the piece-rate contract, and one implemented with $r = 0.50$ should yield a lower level. This is formalized in the following hypothesis.

Hypothesis 1. *Subjects will exhibit average performance levels across treatments according to the following ranking*:

$$\text{LowPr} > \text{MePr} = \text{Piecerate} > \text{HiPr}.$$

Empirical support in favor of Hypothesis 1 would contradict Corollary 2, as well as the predictions of a model in which RDU agents are pessimistic, that is, they exhibit convex weighting functions. This would indicate that agents are risk averse and, as a result, Piecerate should generate the highest average performance and LowPr the lowest.

If the model is accurate, the performance differences predicted by Hypothesis 1 should be explained by the subjects' tendency to distort probabilities, as formalized in the following hypothesis.

Hypothesis 2. i. *Subjects assigned to LowPr who overweight small probabilities will exhibit higher average performance than subjects assigned to Piecerate.*

[13] This assumption is adopted for empirical purposes when stakes are moderate (Diecidue, Wakker, and Zeelenberg (2007), Homonoff (2018), L'Haridon and Vieider (2019)).

ii. *Subjects assigned to HiPr who underweight intermediate probabilities will exhibit lower average performance than subjects assigned to Piecerate.*

The validation of Hypothesis 2 would rule out explanations other than probability distortion, such as income effects, for potential treatment effects in performance.

Finally, I examine whether the motivational or cognitive factors of probability weighting lead to the predicted treatment effects in performance. The following hypotheses, based on Proposition 3 and Proposition 4, formalize the effect of these factors.

Since utility is expected to be linear, Proposition 3 can be interpreted as optimism toward risk (regardless of its degree), thus guaranteeing the higher performance of stochastic contracts. This intuition is also illustrated by Example 1 in the case that Prelec's (1998) function is assumed to exhibit $\alpha = 1$.

Hypothesis 3. *Optimism alone explains the higher average performance in LowPr relative to Piecerate.*

Furthermore, under linear utility, Proposition 4 can be interpreted as likelihood insensitivity (regardless of its degree) thus explaining the higher performance of stochastic contracts implemented with small probabilities, that is, probabilities such that $r < \frac{1}{e}$. This intuition is also illustrated by Example 1 in the case that Prelec's (1998) function is assumed to exhibit $\beta = 1$.

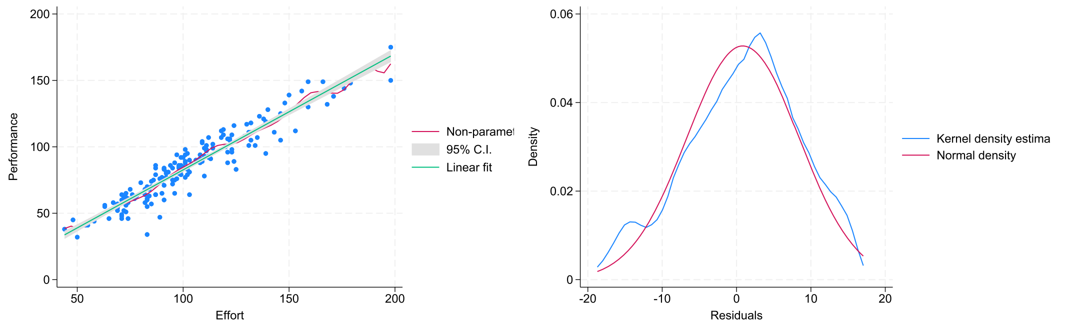
Hypothesis 4. *Likelihood insensitivity alone explains the higher average performance in LowPr relative to Piecerate.*

5. Results

5.1 *Effort and performance*

The initial analysis of the data is intended to determine whether performance in the experiment's real-effort task is stochastic. I define performance as the total number of correct calculations made by a subject, while effort is defined as the total number of calculations, whether correct or incorrect, attempted by a subject. Unless otherwise indicated, the analyses treat each subject as an independent observation. Statistical significance is assessed using t-tests and Wilcoxon Mann–Whitney tests (in the case of pairwise comparisons), and regressions. Standardized effect sizes complement the most relevant significant results.

Figure 4a presents a scatterplot of effort against performance from which two relevant insights emerge. First, different performance levels correspond to each level of effort, suggesting that performance is noisy, that is, subjects make mistakes of significant magnitude when performing the task, which is in line with Assumption 4. To further corroborate this conclusion, I examine the residuals from the nonparametric fit of performance against effort. Figure 4b presents the density of those residuals, which are unambiguously large. For example, between the 25th and 75th quartile of residuals performance increases by 0.34 standard deviations.



(a) Scatter performance and effort

(b) Density of residuals nonparametric fit

FIGURE 4. The relationship between effort and performance.

Second, the linear and nonparametric fits in Figure 4a show that despite the noise in performance, there is a positive relationship between effort and performance, confirming that higher effort leads to higher performance, as stated in Assumption 2.[14]

5.2 *Treatment effects*

To evaluate treatment effects, I focus on the variable Performance, which captures whether potential increments in motivation, as a result of using stochastic contracts, are also effective in boosting output. Importantly, all the main results hold when Performance is replaced by Effort.

Table 2 presents the descriptive statistics for performance by treatment. It can be seen that, as predicted by Hypothesis 1, the stochastic contract implemented with $r = 0.10$ generates higher average performance than the piece-rate contract. Subjects in LowPr carried out 20.56% more calculations on average than those in Piecerate ($t(86) = 2.366$, $p = 0.010$).[15] The effect size of this treatment difference is 0.50 standard deviations, which is significant at the 5% confidence level.[16] This is essentially the main empirical result of the paper, namely that subjects are more motivated under the stochastic contract implemented with a low probability than under a cost-equivalent piece-rate contract.

In contrast, the average performance under stochastic contracts implemented with higher probabilities is similar to that under a piece-rate contract. Subjects in the MePr treatment carried out 87.90 calculations correctly on average, and subjects in the HiPr

[14]The data also show that effort increases from one round to the next but at a decreasing rate. The OLS estimates in the regression $\text{Effort}_i = \beta_0 + \beta_1 \text{Round}_i + \beta_2 \text{Round}_i^2 + e_i$ are $\beta_1 = 0.865$, with standard error 0.052, and $\beta_2 = -0.056$, with standard error 0.004. Hence, subjects cannot maintain the effort rates they exerted in the previous rounds. This result supports Assumption 1.

[15]A Wilcoxon Mann–Whitney test also rejects the null hypothesis of no average difference between Piecerate and LowPr ($z = 2.634$, $p = 0.008$).

[16]The significance of the effect size was evaluated using a bootstrapped 95% confidence interval with 1000 replications.

Table 2. Descriptive statistics for performance by treatment.

Treatment	LowPr	MePr	HiPr	Piecerate	Total
Mean	98.116	87.900	83.750	81.378	87.686
Median	91.000	87.000	82.500	77.000	85.000
St.dev.	34.660	28.134	24.359	31.685	30.413
Observations	43	40	44	45	172

treatment carried out 83.75, neither of which is statistically different from the performance of the subjects in the Piecerate treatment.[17] These findings partially support Hypothesis 1, which correctly predicts that average performance will be similar in MePr and Piecerate but incorrectly predicts that HiPr will show lower performance than Piecerate. Possible explanations for this only partial confirmation of Hypothesis 1 are provided below.

Among the stochastic contract treatments, LowPr exhibited the highest average performance: 17% higher than HiPr ($t(85) = 2.241$, $p = 0.014$), and 11% higher than MePr ($t(81) = 1.467$, $p = 0.071$).[18] The effect sizes of these differences are 0.480 standard deviations and 0.322 standard deviations, respectively. Hence, statistical inference using pairwise testing suggests that LowPr generates the highest motivation.[19]

To complement these findings, I estimate regressions of performance on treatment dummies and dummies that capture the shape of each subject's utility and probability weighting functions. These regressions seek to establish the robustness of the aforementioned treatment effects when average risk attitude in the sample is controlled for. If the treatment effects are robust to the inclusion of these controls, it can then be concluded that the performance differences are not an artifact of subjects in some of the treatments being more risk-seeking or less risk averse than in others.

The regression controls were constructed using data from the second part of the experiment. The first set of controls classify a subject's utility as having either a linear,

[17]The t-tests of these comparisons are ($t(83) = 0.990$, $p = 0.321$) and ($t(87) = 0.395$, $p = 0.693$), respectively. Wilcoxon Mann–Whitney tests of these comparisons yield ($z = 1.321$, $p = 0.186$) and ($z = 0.895$, $p = 0.371$), respectively.

[18]Wilcoxon Mann–Whitney tests of these differences yield ($z = 1.966$, $p = 0.049$) and ($z = 1.035$, $p = 0.7$), respectively.

[19]Table 2 shows that there may be potential differences in the standard deviation of performance between the treatments. Nonetheless, I do not find a significant difference in standard deviations between LowPr and Piecerate, arguably the most relevant treatment comparison. Therefore, if the expected utility principal is risk averse, she would still prefer to implement LowPr over Piecerate, as that treatment generates higher performance and also a similar variability of performance. However, I do find that HiPr generates a lower standard deviation of performance relative to LowPr ($F(42, 43) = 2.024$, $p = 0.023$), MePr ($F(42, 44) = 1.691$, $p = 0.087$), and a similar standard deviation relative to Piecerate ($F(39, 43) = 1.334$, $p = 0.356$). I conjecture that this result is due to the documented tendency of individuals to perceive $p = 0.5$ more accurately than other probabilities (Abdellaoui (2000), Abdellaoui, Bleichrodt, and L'Haridon (2008)). Section 5.3 shows that this is indeed the case in this experiment. This more accurate probability perception leads to less heterogeneity in effort exertion. These results imply that when the principal is sufficiently risk averse, she might prefer to choose HiPr instead of LowPr, and thus she is willing to forgo higher average performance for achieving lower performance variability.

concave, convex, or mixed shape. These categories are captured by the dummy variables "Linear U.," "Concave U.," Convex U.," and "Mixed U.," respectively (further details are presented in Supplemental Appendix D).[20] The second set of controls classify a subject's probability weighting function as displaying overweighting of the probability $p = \frac{1}{6}$, overweighting of the probability $p = \frac{2}{3}$, and/or overweighting of the probability $p = \frac{1}{2}$. These dummy variables are called "Overw. 1/6," "Overw. 1/3," and "Overw. 1/2," respectively. Notably, these variables are all model-free. That is, they do not require that a subject exhibit RDU preferences to attain defined values. For example, under EU and risk aversion, Concave U. attains a value of one while all other dummy variables attain a value of zero.[21]

Columns (1)–(3) in Table 3 present the OLS regression estimates. The coefficient LowPr is positive and significant at the 5% significance level, corroborating the result that subjects assigned to that treatment display higher average performance than subjects in Piecerate (the benchmark of the regression). Similarly, the coefficient of LowPr is significantly larger than that of HiPr ($F(1, 163) = 5.75$, $p = 0.017$), as well as that of MePr ($F(1, 163) = 2.16$, $p = 0.071$). Thus, among the studied contracts, LowPr produces the highest performance, a result that is robust to unbalanced risk attitudes across the treatments.

The aforementioned analyses can be summarized as follows.

Result 1. Average performance across treatments displays the ranking:

$$\text{LowPr} > \text{MePr} = \text{Piecerate} = \text{HiPr}.$$

I perform several robustness checks that corroborate Result 1. First, I show that the result is not due to income effects. According to Azrieli, Chambers, and Healy (2020) and Azrieli, Chambers, and Healy (2018), LowPr generates higher performance, because in contrast to the other treatments, it fully circumvents income effects. Therefore, it keeps subjects motivated until the last round of the task whereas the other treatments do not. If this is the case, then there should not be performance differences between treatments in the first round, in which income effects are absent. Table 13 in Supplemental Appendix F shows that subjects in LowPr achieve higher average performance in the first round than subjects in Piecerate ($p = 0.025$), subjects in HiPr ($F(1, 171) = 5.97$, $p = 0.014$), and subjects in MePr ($F(1, 171) = 2.21$, $p = 0.066$). Hence, Result 1 remains valid regardless of income effects.[22]

[20] 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 of the values for Δ''_j are close to zero, concave utility if most are positive, convex utility if most are negative, and mixed utility otherwise.

[21] Note that in a two-outcome setup, like the one used here to elicit preferences, disappointment models and RDU models coincide. Therefore, these dummy variables can also capture behavior predicted by models other than RDU or EU. For example, under disappointment aversion, all overweighting dummies should take a value of zero.

[22] This robustness check also demonstrates that the documented treatment effects are not due to potential differences in learning to do the task across the treatments.

TABLE 3. Regression of performance on treatment.

	(1) Performance	(2) Performance	(3) Performance	(4) Performance	(5) Performance	(6) Performance
LowPr × Overw. 1/6				31.272		
				(12.652)		
LowPr × LS						23.267
						(11.524)
LowPr	16.739	16.558	19.368	18.874	16.001	6.144
	(7.090)	(7.508)	(7.654)	(9.039)	(7.532)	(8.192)
MePr × Overw. 1/6				16.811		
				(10.683)		
MePr × LS						8.911
						(9.243)
MePr	6.522	6.714	5.646	6.001	6.335	1.342
	(6.487)	(6.610)	(6.561)	(7.596)	(6.677)	(8.328)
HiPr × Overw. 1/6				−1.580		
				(8.334)		
HiPr × LS						−7.179
						(9.170)
HiPr	2.372	1.684	1.841	5.578	1.616	2.708
	(5.985)	(5.888)	(5.953)	(6.911)	(6.308)	(7.467)
Concave U.		14.359	13.895	12.819	15.067	14.309
		(9.401)	(9.373)	(9.283)	(9.529)	(9.099)
Convex U.		7.623	2.579	2.704	8.527	10.731
		(10.109)	(11.325)	(11.828)	(10.469)	(12.006)
Mixed U.		3.864	3.497	2.542	3.698	3.413
		(6.625)	(6.991)	(6.944)	(6.699)	(6.793)
Overw. 1/6			8.515	12.916		
			(6.554)	(11.753)		
Overw. 1/3			−20.128	−20.891		
			(12.178)	(12.822)		
Overw. 1/2			5.285	4.816		
			(13.129)	(13.360)		
US					0.904	0.723
					(5.183)	(5.204)
LS					2.924	−4.873
					(5.053)	(10.642)
Constant	81.378	79.819	79.214	78.290	78.497	81.417
	(4.726)	(5.025)	(5.373)	(6.070)	(5.242)	(5.923)
R^2	0.045	0.062	0.089	0.101	0.065	0.093
Observations	172	172	172	172	172	172

Note: This table presents OLS estimates of the model $\text{Performance}_i = \beta_0 + \beta_1 \text{LowPr} + \beta_2 \text{MePr} + \beta_3 \text{HiPr} + \text{Controls}'\Lambda + \varepsilon_i$, with $E(\varepsilon|\text{MePr, LowPr, HiPr, Controls}) = 0$. "Performance" is the number of calculations correctly solved by a subject in the first part of the experiment, "LowPr," "MePr," and "HiPr" are binary variables that indicate if a subject was assigned to the treatment offering stochastic contracts implemented with low, medium, or high probability, respectively. "Piecerate" is the benchmark of the regression. Robust standard errors are presented in parentheses.

Second, Column (5) in Table 3 shows that the treatment effects are robust to using an alternative classification of probability weighting functions, according to whether a subject's weighting function exhibits lower subadditivity (hereafter, LS) and/or upper subadditivity (hereafter, US). A weighting function with LS assigns larger decision weights to low probabilities than to intermediate probabilities, while a weighting function with US assigns larger decision weights to large probabilities than to intermediate probabilities.[23] Note that this classification does not require RDU to attain values. For example, under EU and risk aversion, LS and US attain a value of zero.

Third, Table 14 in Supplemental Appendix F demonstrates that similar qualitative results are found when continuous variables capturing utility and probability weighting curvature are included in the regressions. Finally, Table 15 in Supplemental Appendix F shows that Result 1 is robust to abandoning the assumption of linearity between the dependent variable, that is, Performance, and variables representing treatment assignment.

Finally, it is worth mentioning that the data *partly* support Hypothesis 1. Recall that Hypothesis 1 was based on the commonly observed finding that individuals overweight probabilities smaller than $p = 0.33$ and underweight those larger than $p = 0.33$. In contrast, the analysis presented in this subsection suggests that subjects overweighted the probability $p = 0.10$ and evaluated the probabilities $p = 0.30$ and $p = 0.50$ fairly accurately. In the next subsection, I will show that subjects indeed display such an average weighting function.

5.3 *Utility and probability weighting functions*

This subsection analyzes the data obtained in part 2 of the experiment. Recall that part 2 was intended to elicit a sequence of certainty equivalents $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ for each subject. I use that data to examine each subject's preference over monetary outcomes.

I find that 75% of the subjects exhibit linear utility functions and that the average utility function exhibits very modest concavity. For instance, when a power utility function is assumed, that is, $u(x) = x^k$, the resulting nonlinear least squares estimate is $k = 0.995$. These results are in line with the findings of Wakker and Deneffe (1996) (in the case of small stakes), Bruhin, Fehr-Duda, and Epper (2010), Epper, Fehr-Duda, and Bruhin (2011), and Abdellaoui et al. (2011), and are consistent with the critique put forward by Rabin (2000). These results also support Assumption 4. Given these findings, and since the main focus of the analysis is probabilistic risk attitude, I relegate the full analysis of utility functions and their shape to Supplemental Appendix D.

The second part of the experiment also elicited the sequence of probabilities:

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

for each subject. These data are used to understand how the subjects evaluate probabilities. To that end, regressions are used to relate the elicited sequence of probabilities to the probability weights they map onto. The rationale for using regressions as the primary tool of analysis is that: (i) they provide a good indication of the average degree of probability weighting, (ii) the resulting estimates can be used to compare the degree of

[23] Specifically, subjects in the experiment exhibited LS when $w^{-1}(\frac{1}{6}) < w^{-1}(\frac{2}{6}) - w^{-1}(\frac{1}{6})$ and US when $1 - w^{-1}(\frac{5}{6}) < w^{-1}(\frac{5}{6}) - w^{-1}(\frac{4}{6})$.

probability weighting in this experiment to that reported in previous studies, and (iii) the resulting estimates can be used to construct indexes of likelihood insensitivity and optimism, which are, according to Proposition 3 and Proposition 4, relevant in explaining the efficiency of stochastic contracts. Alternative methods of analysis, including an analysis at the individual level and a nonparametric analysis, are presented in Supplemental Appendix E.

The regression analysis employs two parametric forms of weighting functions, namely, Prelec's (1998) two-parameter probability weighting function and the neo-additive probability weighting function (Chateauneuf, Eichberger, and Grant (2007)). The former is recommended for empirical purposes because it is more tractable than other alternatives and is better suited for fitting small and large probabilities (Wakker (2010)). The latter is used because of its clear interpretation of the parameters, which makes it better suited to separate likelihood insensitivity from optimism (Abdellaoui et al. (2011), Li, Müller, Wakker, and Wang (2018)). Their simultaneous usage thus provides a good overview of best fit and parsimony.[24]

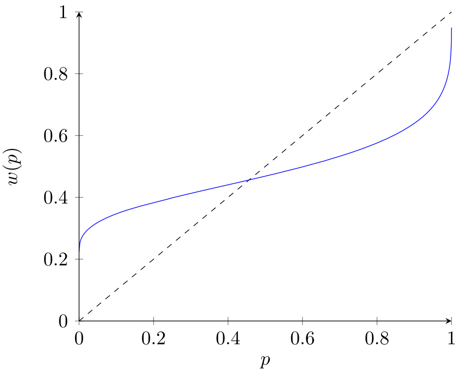
The coefficients of Prelec's (1998) weighting function are estimated by means of nonlinear least squares, which accounts for the nonlinear structure of the function. This estimation method is also used by Abdellaoui et al. (2011). The neo-additive weighting function is estimated using two approaches: an OLS regression and a truncated regression estimated using maximum likelihood. The former approach provides a goodness-of-fit measure that can be related to that of the nonlinear least squares but ignores potential discontinuities at zero and one. In contrast, the latter approach involves truncation at the end points, $w(0) = 0$ and $w(1) = 1$, in order to allow for the possibility that the probability weighting function is S-shaped. Furthermore, this method is favored in the literature (Wakker (2010), Abdellaoui et al. (2011), Li et al. (2018)).[25] The truncated regression is estimated using maximum likelihood in order to obtain a consistent variance-covariance matrix.

The estimation results are presented in Table 4 and Figure 5. According to the results, the average probability weighting function has a more pronounced inverse-S shape than in previous studies and shows less pessimism, which is the second main empirical result of the paper.[26] Moreover, both estimations of the neo-additive function yield quali-

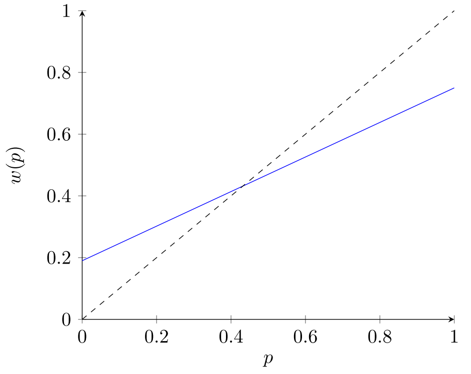
[24]Using these two functions simultaneously also provides the best compromise between consistency with the theory and correct hypothesis testing. The neo-additive function is not consistent with the theoretical framework because it violates the continuity included in Assumption 5. However, it does provide good measurements of likelihood insensitivity and optimism, which are key in testing Hypotheses 3 and 4. In contrast, Prelec's (1998) function satisfies continuity but, as will be explained below, is not able to separate likelihood insensitivity from optimism.

[25]This is because the structural equation to be estimated coincides with the theoretical one, which states not only that $w(p) = c + sp$ for $p \in (0, 1)$ but also that $w(0) = 0$ and $w(1) = 1$.

[26]The estimates in Table 4 indicate that when the Prelec (1998) function is assumed, subjects exhibit optimism ($\beta = 0.84 < 1$). In contrast, the neo-additive function indicates that subjects exhibit pessimism (e.g., the estimation with truncation yields $1 - c - s = 0.25 > c = 0.20$). This is paradoxical if one does not take into account that the α parameter of Prelec's (1998) weighting function also captures some degree of pessimism/optimism (Gonzalez and Wu (1999), Abdellaoui et al. (2011)). In other words, when using that weighting function, the parameters α and β are not as independent of each other as in the case of the



(a) using Prelec (1998)



(b) using Chateauneuf et al. (2007) with truncation

Figure 5. Average probability weighting functions.

tatively similar results, namely that subjects exhibit insensitivity and pessimism.[27,28] Finally, and in line with the discussion above, Prelec's (1998) function has better goodness-of-fit than the neo-additive, thus ex post rationalizing the notion that the former is preferable for data fitting.[29]

Result 2. Subjects exhibit, on average, a linear utility function and a probability weighting function with a pronounced inverse-S shape and moderate pessimism.

To the best of my knowledge, only L'Haridon and Vieider (2019) (for Nigeria) have found estimates of probability weighting comparable to those in Table 4. The findings of stronger insensitivity and less pessimism than in most previous studies can be explained by a number of factors. First, the method chosen to elicit subjects' preferences does not impose parametric assumptions on preferences, such as linear utility, which might bias parameter estimates. Note that under the assumption of linear utility, risk aversion pertaining to utility curvature can be erroneously attributed to probability weighting, which would lead to a finding of lower insensitivity. Second, probability weighting and utility

neo-additive function. This makes them less reliable as indexes of pessimism and likelihood insensitivity. Therefore, it may be that some pessimistic attitudes are being captured by the parameter α, leading to an underestimation of pessimism. This is confirmed by the fact that Goldstein and Einhorn's (1987) function, presented in Table 16 of Supplemental Appendix F, also indicates pessimism. That weighting function has been shown to better separate likelihood insensitivity from optimism than Prelec's (1998) function (Li et al. (2018)).

[27]These estimates do differ quantitatively. Specifically, the estimates presented in panel 2 imply an average probability weighting function with more insensitivity and optimism relative to that implied by the estimates presented in panel 1. Hence, ignoring truncation generates a bias that magnifies the findings summarized in Result 2.

[28]While the positive log-likelihood in panel 1 of Table 4 is unusual, it is a feasible value when the data exhibits little variance.

[29]Estimates of identical magnitude are obtained when Prelec's (1998) function is estimated using maximum likelihood. Therefore, the assumption of normally distributed errors, under which the maximum likelihood and nonlinear least squares methods are equivalent, seems to be plausible for these data.

Table 4. Parametric estimates of average probability weighting function.

Panel 1 $w(p) = c + sp$ *with truncation at* $w(0) = 0$ *and* $w(1) = 1$

	$\hat{c}$	$\hat{s}$
	0.194 (0.021)	0.566 (0.036)
Log-Likelihood		220.288
Observations		860

Panel 2 $w(p) = c + sp$ *without truncation*

	$\hat{c}$	$\hat{s}$
	0.258 (0.016)	0.448 (0.029)
Adj. R^2		0.307
Observations		860

Panel 3 $w(p) = \exp(-\beta(-\ln(p))^{\alpha})$

	$\hat{\alpha}$	$\hat{\beta}$
	0.285 (0.025)	0.842 (0.015)
Adj. R^2		0.865
Observations		860

Note: This table presents the subject's average probability weighting function when different parametric forms are assumed. Panel 1 presents the maximum likelihood estimates of the model $w(p) = c + sp$ with truncation at $w(0) = 0$ and $w(1) = 1$. Panel 2 presents the OLS estimates of the model $w(p) = c + sp$. Panel 3 presents the nonlinear least squares estimates of the model $w(p) = \exp(-\beta(-\ln(p))^{\alpha})$. Robust standard errors are presented in parentheses.

functions are immediately and fully identified in the experiment, which makes ancillary assumptions on error distributions for the *identification* of these functions unnecessary. Therefore, I avoid the possibility of obtaining biased estimates of probability weighting because the error distribution used for identification is misspecified. Finally, the fact that all subjects are compensated in the experiment can lead to differences in risk preference relative to studies that pay only some of the subjects (Baltussen et al. (2012)).

To further validate Result 2, I perform regressions using other functional forms of utility and probability weighting. Table 9 in Supplemental Appendix D shows that linearity of the utility function is also corroborated when an exponential utility function is assumed. Table 16 in Supplemental Appendix F shows that strong inverse-S shapes also emerge when Goldstein and Einhorn's (1987) log-odds probability weighting function and Tversky and Kahneman's (1992) probability weighting function are assumed.

The conjunction of a strong inverse-S shape and moderate pessimism generates a probability weighting function that strongly overweights small probabilities and moderately distorts medium-sized probabilities. For example, using the estimates of panel 1 in Table 4 it can be established that subjects perceived $p = 0.10$ to be on average equal to $w(0.10) = 0.25$, while $p = 0.30$ and $p = 0.5$ were on average perceived to be $w(0.30) = 0.363$ and $w(0.50) = 0.477$, respectively. These patterns of probability weighting can accommodate the result that LowPr generates higher output than Piecerate, and that HiPr, MePr, and Piecerate produce similar output.

5.4 *Overweighting of probabilities, likelihood insensitivity, and the treatment effect*

This subsection uses the data obtained in both parts of the experiment in order to test the remaining hypotheses. To evaluate the validity of Hypothesis 2, I extend the regression presented in column (4) of Table 3. Specifically, I add interactions between the variables for assignment to different treatments and the dummy variable that captures whether a subject overweights small probabilities. Small probability overweighting is captured using Overw. 1/6 in one specification and LS in the other. Using different variables to capture small probability overweighting ensures robustness.

Columns (4) and (6) in Table 3 present the OLS estimation results. Subjects in LowPr who overweight the probability $p = \frac{1}{6}$ display higher average performance than Piecerate subjects who do not. In contrast, subjects assigned to Piecerate who overweight the probability $p = \frac{1}{6}$ exhibit an average performance level that is statistically indistinguishable from that of subjects assigned to the same treatment but who do not overweight that probability. Similar conclusions are reached when LS is used.[30,31] The estimates presented in Table 17 in Supplemental Appendix F demonstrate further robustness for this result, by showing that higher probability overweighting, as measured by a continuous variable, generates higher performance in LowPr than in Piecerate.

While the previous approach is typically used to evaluate heterogeneous treatment effects in experiments (Young (2019)), it is based on assumptions that are not harmless. First, it assumes that utility curvature and probability weighting curvature affect performance independently. This is problematic because under RDU both curvatures jointly determine risk attitude, which is crucial for the effectiveness of stochastic contracts. A more complete analysis would account for the interaction between these two determinants of risk attitude. Second, the previous analysis assumed that the relationship between performance and the explanatory variables is linear, which also restricts the influence of risk preference on performance.

To address the first concern, the condition of Proposition 2, that is, $w(r) > u(r)$, is directly incorporated in a complementary analysis. Recall that that condition compares

[30]The main difference between these two regressions is that the coefficient associated with LowPr remains significant when Overw. 1/6 is used. This suggests that the treatment effect is *not entirely* captured by the tendency of subjects to overweight the probability $p = \frac{1}{6}$. In contrast, that effect can be fully explained by the subjects' tendency to overweight the probability $p = \frac{1}{6}$ *relative* to other probabilities and in the case of LS, overweighting $p = \frac{1}{6}$ relative to overweighting $p = \frac{1}{2}$. These findings already point to my final result, namely that overweighting of probabilities due to likelihood insensitivity, which entails overweighting of small probabilities relative to medium-sized and large ones, explains the treatment effect.

[31]Similar results are found when the variable "Possibility" is used to capture overweighting of probabilities. That variable takes a value of one if the subject's overweighting of small probabilities is stronger than his underweighting of large ones.

utility and probability weighting curvature to provide more precise predictions. Specifically, it states that subjects should be more motivated under a stochastic contract as long as the risk-seeking due to probability overweighting outweighs the risk aversion due to utility curvature at the probability chosen by the principal.

To evaluate whether subjects in the experiment meet that condition, I estimate the neo-additive probability weighting function for each subject, i:

$$w(p_{ij}) = c_i + s_i p_{ij} + e_i, \tag{6}$$

where $j = \{1, 2, 3, 4, 5\}$ denotes the elicited probability and e_i is a subject-specific error term. To allow for S-shaped functions, this regression is estimated with truncation at both end-points, that is, $w(0) = 0$ and $w(1) = 1$. I also estimate the following CRRA utility function for each subject using nonlinear least squares:

$$u(p_{ij}) = p_{ij}^{k_i} + v_i. \tag{7}$$

The resulting set of estimates $\{\hat{s}_i, \hat{c}_i, \hat{k}_i\}$ are used to construct a binary variable, denoted by $\mathbb{1}_{w(r)>u(r)}$, which takes the value of one if the subject fulfills the inequality $\hat{c}_i + \hat{s}_i r > r^{\hat{k}_i}$ and zero otherwise. This is done separately for the small probabilities of $r = 0.10$, $r = 0.20$, and $r = 0.05$, which should explain the treatment effect in performance, especially in the case of $r = 0.10$.

These binary variables were then included in OLS and nonparametric regressions that relate a subject's performance to the fulfillment of the condition in Proposition 2, that is, $\hat{c}_i + \hat{s}_i r > r^{\hat{k}_i}$, and to the assigned treatment. The nonparametric estimation seeks to address the second shortcoming mentioned above, that is, that performance and the explanatory variables might not relate linearly.

Table 5 presents the regression estimates, which show that fulfillment of the condition in Proposition 2 for small probabilities and being assigned to the stochastic contract with small probability leads to higher average performance, a result that is fully in line with the theoretical framework. As shown in columns (2),(4), and (6), this result is robust to dropping the assumption of linearity between the dependent and independent variables.

The overweighting of small probabilities as an explanation of the treatment effects presented in Result 1 constitutes the third empirical result of the paper.

Result 3. Subjects in LowPr who overweight small probabilities exhibit higher performance than subjects in Piecerate.

Finally, I investigate whether the higher performance of subjects in LowPr is due to optimism and/or likelihood insensitivity. To distinguish these factors, I follow the approach proposed by Wakker (2010) and Abdellaoui et al. (2011), which use the coefficients of the neo-additive weighting function, that is, $\hat{s}_i$ and $\hat{c}_i$, as indexes of insensitivity and optimism. Furthermore, the magnitude of these estimates can be used to classify subjects, which I now turn to.

Table 5. The influence of Proposition 1's condition on treatment effects.

	(1) Performance	(2) Performance	(3) Performance	(4) Performance	(5) Performance	(6) Performance
LowPr × $\mathbb{1}_{w(r)>u(r)}(r)$	22.428	22.428	21.361	21.361	21.934	21.934
	(9.835)	(9.834)	(9.739)	(9.739)	(9.635)	(9.635)
LowPr	14.736	7.191	13.669	8.373	12.410	6.700
	(8.500)	(8.897)	(8.389)	(9.119)	(9.181)	(9.215)
MePr × $\mathbb{1}_{w(r)>u(r)}(r)$	10.025	10.025	9.150	9.150	11.069	11.069
	(9.168)	(9.167)	(9.277)	(9.277)	(9.277)	(9.370)
MePr	9.278	1.733	8.019	2.722	6.604	0.894
	(8.719)	(9.106)	(8.394)	(9.124)	(8.906)	(8.941)
HiPr × $\mathbb{1}_{w(r)>u(r)}(r)$	0.262	0.262	0.393	0.393	−0.385	−0.385
	(8.289)	(8.289)	(8.122)	(8.122)	(8.122)	(8.465)
HiPr	10.853	3.308	8.979	3.682	10.205	4.495
	(8.196)	(8.607)	(8.190)	(8.936)	(8.374)	(8.411)
Piecerate	7.545	–	5.296	–	5.710	–
	(9.496)	–	(9.608)	–	(9.476)	–
Constant	78.192		79.259		78.840	
	(6.453)		(6.306)		(6.677)	
Value of r	0.10	0.10	0.20	0.20	0.05	0.05
Estimation Method	OLS	Nonparametric	OLS	Nonparametric	OLS	Nonparametric
R^2	0.061		0.056		0.062	
Cross-validation		944.124		948.980		943.025
Observations	172	172	172	172	172	172

Note: This table presents the estimates of the model $\text{Performance}_i = \beta_0 + \beta_1 \text{LowPr} * \mathbb{1}_{w(r)>u(r)}(r) + \beta_2 \text{LowPr} + \beta_3 \text{MePr} * \mathbb{1}_{w(r)>u(r)}(r) + \beta_4 \text{MePr} + \beta_5 \mathbb{1}_{w(r)>u(r)}(r) * \text{HiPr} + \beta_6 \text{HiPr} + \beta_7 \text{Piecerate} + \varepsilon_i$, with $E(\varepsilon_i | \text{MePr, LowPr, HiPr, Piecerate}, \mathbb{1}_{w(r)>u(r)}(r), \text{Controls}) = 0$. "Performance" is the number of calculations correctly solved by a subject in the first part of the experiment, "LowPr," "MePr," and "HiPr" are binary variables that indicate if a subject was assigned to a treatment offering a stochastic contract implemented with low, medium, or high probability, respectively. "Piecerate" is a binary variable indicating if a subject was assigned to the treatment offering a piece rate. "$\mathbb{1}_{w(r)>u(r)}(r)$" is a binary variable that takes a value of one if the subject overweights probability r and zero otherwise. Columns (1), (3), and (5) present OLS estimates. Columns (2), (4), and (6) present the conditional marginal of a cubic B-spline estimation. Robust standard errors are presented in parentheses.

The magnitude of the estimate $\hat{s}_i$ is a measure of a subject i's sensitivity to probabilities. If $\hat{s}_i < 1$, the subject is not fully responsive to changes in probabilities and is classified as likelihood insensitive and conversely, if $\hat{s}_i \geq 1$, the subject is classified as likelihood sensitive. It was found that 96 subjects in the sample are likelihood insensitive and 61 are likelihood sensitive.[32] Importantly, the degree of likelihood insensitivity is balanced across treatments. For example, there is no empirical evidence to reject the null hypothesis of no difference in likelihood insensitivity between LowPr and Piecerate $(t(86) = 0.746,\ p = 0.46)$.

The magnitude of $\hat{c}_i$ is a measure of subject i's optimism. If $\hat{c}_i > 0$, the subject assigns large weights to best-ranked outcomes and, therefore, exhibits optimism; conversely, if $\hat{c}_i < 0$ the subject does not exhibit optimism. It was found that 97 subjects in the sample display optimism while 75 subjects do not. Degrees of optimism are also balanced across

[32] Fifteen subjects were found to have $\hat{s}_i < 0$, which has no clear interpretation; they were left unclassified.

treatments. For example, there is no empirical evidence to reject the null hypothesis of no difference in optimism between LowPr and Piecerate ($t(86) = 1.03$, $p = 0.306$).

Interestingly, male subjects exhibit more optimism than female subjects, which is in line with the results of Sautmann (2013). However, this difference does not translate into higher performance of males in the aggregate nor at the treatment level. I conjecture that this is because likelihood insensitivity is similar between the genders and, as will be explained below, that factor is predominant in producing the treatment effects of Result 1.

Binary variables that capture these classifications (labeled "Optimism" and "Likelihood ins.") are added to the regression presented in column (3) of Table 3. Interactions between these variables and binary variables that capture assignment to treatment are included in some specifications. The coefficients associated with these interactions are an indication of the strength of the treatment effect among subjects who exhibit likelihood insensitivity and/or optimism.

The regression estimates are presented in Table 6. Columns (2) and (4) show that likelihood-insensitive subjects assigned to LowPr exhibit higher average performance than subjects in Piecerate. In contrast, the performance of subjects assigned to LowPr who were not classified as likelihood insensitive does not show any significant differences. These findings support Hypothesis 4. Furthermore, the estimates presented in columns (3) and (4) show that subjects displaying optimism who were assigned to LowPr achieve an average performance level that is statistically indistinguishable from that of subjects in Piecerate. Thus, optimism cannot explain the treatment effects.

The above analysis can be summarized as follows.

Result 4. Likelihood insensitivity explains the treatment effects presented in Result 1.

A set of additional analyses were performed in order to establish robustness. First, Table 15 in Supplemental Appendix F shows that Result 4 remains valid when nonparametric regressions are performed and, therefore, it is not an artifact of the underlying assumption of linearity in the relationship between performance and RDU components. Second, Table 18 in Supplemental Appendix F demonstrates that a similar conclusion is reached when the classification is performed using the functions proposed by Prelec (1998) and Goldstein and Einhorn (1987). Finally, Table 19 in Supplemental Appendix F shows that Result 4 is robust to using continuous measures of likelihood insensitivity and optimism, that is,m $1 - \hat{s}$ and $\frac{\hat{s}+\hat{c}}{2}$, respectively. There, I find that participants with higher likelihood insensitivity display a larger difference in performance between the treatments LowPr and Piecerate.

6. Applications and discussion

This paper demonstrated that stochastic contracts implemented with small probabilities can generate more motivation than contracts traditionally regarded as optimal. This is explained by the individuals' tendency to overweight small probabilities, which induces a preference for riskier contracts. Likelihood insensitivity, the cognitive component of probability weighting, was empirically found to be primarily responsible for the effectiveness of stochastic contracts.

TABLE 6. The influence of likelihood insensitivity and optimism on treatment effects.

	(1) Performance	(2) Performance	(3) Performance	(4) Performance
LowPr × Likelihood ins.			22.366	
			(9.443)	
LowPr × Optimism				8.075
				(10.446)
LowPr	16.059	16.062	8.340	11.852
	(7.216)	(7.581)	(8.500)	(11.790)
MePr × Likelihood ins.			13.600	
			(9.239)	
MePr × Optimism				−0.365
				(10.325)
MePr	6.372	6.514	2.232	2.537
	(6.483)	(6.628)	(7.495)	(9.136)
HiPr × Likelihood ins.			1.342	
			(6.922)	
HiPr × Optimism				−10.545
				(9.722)
HiPr	1.729	1.078	3.783	2.138
	(6.063)	(5.994)	(7.544)	(8.535)
Likelihood Ins.	6.275	5.883	2.333	6.198
	(5.242)	(5.325)	(10.324)	(5.310)
Optimism	−7.760	−7.529	−7.006	−10.368
	(5.381)	(5.407)	(5.420)	(10.116)
Concave		14.180	12.857	12.636
		(9.631)	(10.232)	(10.155)
Convex		5.834	3.593	5.725
		(11.758)	(12.470)	(11.182)
Mixed		3.201	3.603	3.363
		(6.645)	(6.500)	(6.559)
Constant	82.621	81.241	82.712	82.742
	(5.070)	(5.338)	(5.648)	(6.903)
R^2	0.058	0.074	0.085	0.078
Observations	172	172	172	172

Note: This table presents OLS estimates of the model $\text{Performance}_i = \beta_0 + \beta_1\text{LowPr} * \text{Likelihood ins.} + \beta_2\text{LowPr} * \text{Optimism} + \beta_3\text{LowPr} + \beta_4\text{MePr} + \beta_5\text{HiPr} + \beta_6\text{Likelihood ins.} + \beta_7\text{Optimism} + \text{Controls}'\Gamma + \varepsilon_i$, with $E(\varepsilon_i|\text{MePr, LowPr, HiPr, Piecerate, Optimism, Likelihood ins., Controls}) = 0$. "Performance" is the number of calculations correctly solved by a subject in the first part of the experiment, "LowPr," "MePr," and "HiPr" are binary variables that indicate if a subject was assigned to a treatment offering a stochastic contract 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 are presented in parentheses.

There are a number of ways in which these findings can be exploited as part of the tools offered by personnel economics:

- *Bonuses.* The principal can take advantage of the stochastic nature of performance by using a bonus contract that pays a sizable lump sum in the event that a performance target is attained. The findings show that the principal should set a high target, thus creating a small probability of achievement and a large degree of risk for the agent. Agents suffering from likelihood insensitivity will be more motivated under the bonus than under a cost-equivalent linear contract because they overweight the probability of achieving the bonus. This application and its relationship to Proposition 2 are further formalized in Appendix A.

- *Stock options.* A volatile firm can offer its CEOs compensation plans that include stock options whose future value is of course unknown when the contract is signed. This leads to two effects. First, as shown by Spalt (2013), the agent with likelihood insensitivity will accept such contracts despite the firm's high level of risk since he overweights the probability of obtaining a large profit by exercising the option. Second, the findings suggest that contracts with stock options can generate higher motivation than less risky performance-pay contracts because the agent erroneously inflates the contribution of his effort to the probability that the firm's stock will rise in the future.

A common property of these applications is that the incentives created by stochastic contracts are based on natural sources of uncertainty, that is, output realizations and future stock prices. Indexing the contract's outcomes to natural and uncertain events allows the principal to circumvent the problem of lack of credibility that might arise if she were to generate the risk using an artificial device, such as the spin of a roulette wheel or the throw of a dice. Ensuring that the principal has no influence over the realization of the uncertainty allows her to more credibly commit to the contract.

The present study has several limitations that might suggest avenues for future research. First, it assumes that the principal is fully informed about the agent's risk attitudes. Future research should relax that assumption. Specifically, the model presented in Section 2 can be extended to incorporate a stage of adverse selection. The principal's task in such a framework is to design a menu of contracts that will allow her to *screen* agents according to their risk preference *and* which will *motivate* the agents to exert high levels of effort in the delegated task. An experimental implementation of that model could test the notion that RDU decision makers self-select into the stochastic contract, provided they are properly disincentivized to choose a different contract, and that they exhibit higher motivation relative to EU decision makers selecting into a traditional contract.

Second, this paper considered a static setting. A more comprehensive investigation of stochastic contracts would examine its incentives in a setting of repeated interaction between principal and agent, which would allow for a more robust analysis of incentives. For example, it could shed light on whether and how probability weights are adjusted over time. Moreover, because stochastic contracts extract effort at the expense of the agent's welfare, it is unclear whether their motivating effect can be sustained in a dynamic setting. While the agent might remain unaware of his biases, thus making the exploitation sustainable over time, it is possible that the high effort costs incurred in the initial periods will encourage him to eventually reject this type of contract.

Appendix A: Proofs and examples

A.1 *Proofs*

Proof of Lemma 1. Using integration by parts, rewrite the agent's utility in equation (3) as

$$\mathrm{RDU}(t, e) = u\big(t(\underline{y})\big) - \int_{\bar{y}}^{\underline{y}} u'\big(t(y)\big)\frac{\mathrm{d}t(y)}{\mathrm{d}y} w\big(1 - F(y|e)\big)\,\mathrm{d}y - c(e). \tag{8}$$

Replace (3) with (8) in the participation constraint given in (4). Also, replace the incentive compatibility constraint in (4) with the first-order condition of equation (3) with respect to e. Those changes lead to the following program:

$$\begin{aligned}
\max_{\{t(y)\}} \quad & \int_{\underline{y}}^{\bar{y}} \big(S(y) - t(y)\big) f(y|e)\,\mathrm{d}y \\
\text{s.t.} \quad & u\big(t(\underline{y})\big) - \int_{\bar{y}}^{\underline{y}} u'\big(t(y)\big)\frac{\mathrm{d}t(y)}{\mathrm{d}y} w\big(1 - F(y|e)\big)\,\mathrm{d}y - c(e) \geq \bar{U}, \\
& \int_{\bar{y}}^{\underline{y}} u'\big(t(y)\big)\frac{\mathrm{d}t(y)}{\mathrm{d}y} w'\big(1 - F(y|e)\big)F_e(y|e)\,\mathrm{d}y - c'(e).
\end{aligned} \tag{9}$$

Denote by t^{fo} the solution to the program presented in equation (9).

The first-order approach, used to replace the agent's incentive compatibility constraint by the first-order condition of $\mathrm{RDU}(t, e)$ with respect to e, is necessary and sufficient if the following condition holds:

$$\int_{\bar{y}}^{\underline{y}} u'\big(t(y)\big)\frac{\mathrm{d}t(y)}{\mathrm{d}y}\Big(w'\big(1-F(y|e)\big)F_{ee}(y|e) - w''\big(1-F(y|e)\big)\big(F_e(y|e)\big)^2\Big)\,\mathrm{d}y - c''(e) < 0. \tag{10}$$

Since $c''(e) > 0$ (Assumption 1), $u' > 0$ (Assumption 4), $\frac{\mathrm{d}t(y)}{\mathrm{d}y} \geq 0$ (Assumption 3), the following condition suffices for the inequality in equation (10) to hold:

$$w'\big(1 - F(y|e)\big)F_{ee}(y|e) - w''\big(1 - F(y|e)\big)\big(F_e(y|e)\big)^2 < 0. \tag{11}$$

Because of $F_{ee}(y|e) > 0$ (Assumption 2) and $w'(1 - F(y|e)) > 0$ (Assumption 5), a probability weighting function with $w''(1 - F(y|e)) < 0$ cannot fulfill the sufficient condition presented in equation (11). Hence, for the optimality of t^{fo} it is required that $w''(1 - F(y|e)) > 0$. Letting $p = 1 - F(y|e)$, that condition can be written as $w''(p) > 0$. □

Proof of Proposition 1. *Preliminaries.* Consider an RDU agent suffering from optimism or likelihood insensitivity. Denote by $\tilde{y} \in [\underline{y}, \bar{y}]$ the output level satisfying $\tilde{p} = 1 - F(\tilde{y}|\tilde{e})$ for a given effort $\tilde{e} \in [0, \bar{e}]$. The existence and uniqueness of $\tilde{y}$ are guaranteed by Assumption 5. Accordingly, for the optimist $\tilde{y} = \underline{y}$ due to the assumption that $\tilde{p} = 1$ and for the likelihood insensitive agent $\tilde{y} \in (\underline{y}, \bar{y})$ due to the assumption that $\tilde{p} \in (0, 1)$.

Moreover, let $\hat{y} \in [\underline{y}, \bar{y}]$ be the *largest* output level satisfying $w'(1 - F(\hat{y}|\tilde{e})) = 1$ for a given effort $\tilde{e} \in [0, \bar{e}]$. The existence of $\hat{y}$ for optimists and likelihood insensitive agents is due to $\lim_{p\to 0} w'(p) > 1$ and $\lim_{p\to \tilde{p}} w'(p) < 1$ by Assumption 5.

Finally, let $p_h := 1 - F(y|e'')$ and $p_l := 1 - F(y|e')$ for any pair of efforts $e'', e' \in [0, \bar{e}]$ such that $e'' > e'$. Notice that $p_h > p_l$ due to Assumption 3.

Part (i). Consider all output levels in $y \in [\tilde{y}, \hat{y}]$. In that segment, the agent's weighting function exhibits $w'(1 - F(y|\tilde{e})) < 1$ for a given effort $\tilde{e} \in [0, \bar{e}]$. Integrating that inequality over the interval $[p_l, p_h]$ gives

$$\begin{aligned} &\int_{p_l}^{p_h} w'(s)\,\mathrm{d}s < \int_{p_l}^{p_h} \mathrm{d}s \\ \Leftrightarrow \quad & w\big(1 - F\big(y|e''\big)\big) - w\big(1 - F\big(y|e'\big)\big) < 1 - F\big(y|e''\big) - \big(1 - F\big(y|e'\big)\big). \end{aligned} \tag{12}$$

Letting e'' infinitesimally approach e' gives

$$-w'\big(1 - F(y|e)\big)F_e(y|e) > -F_e(y|e). \tag{13}$$

Multiply both sides of the inequality in equation (13) by $u(t(y))$ and integrate over the interval $y \in [\tilde{y}, \hat{y}]$ to obtain

$$-\int_{\tilde{y}}^{\hat{y}} u\big(t(y)\big)w'\big(1 - F(y|e)\big)F_e(y|e)\,\mathrm{d}y < -\int_{\tilde{y}}^{\hat{y}} u\big(t(y)\big)F_e(y|e)\,\mathrm{d}y. \tag{14}$$

Equation (14) shows that the agent's marginal utility of effort in $y \in [\tilde{y}, \hat{y}]$ is lower when he distorts probabilities relative to the hypothetical case in which he would have not distort them. Therefore, probability weighting due to optimism and likelihood insensitivity leads the agent to choose lower effort in the output segment in which the weighting function exhibits $w'(1 - F(y|\tilde{e})) < 1$ for any $\tilde{e} \in [0, \bar{e}]$.

Part (ii). The rationale presented in the first part of the proof is mirrored for all output levels $y \in [\hat{y}, \bar{y}]$. There, the agent's probability weighting function exhibits $w'(1 - F(y|\tilde{e})) > 1$ for a given $\tilde{e} \in [0, \bar{e}]$. □

Proof of Corollary 1. Let an agent i be more optimistic than an agent j. Denote by w_i and w_j their respective probability weighting functions. Let $\hat{y}_i$ be the *largest* output level satisfying $w_i'(1 - F(\hat{y}|\tilde{e})) = 1$ for a given effort $\tilde{e} \in [0, \bar{e}]$. The existence of $\hat{y}_i$ for optimists and likelihood insensitive agents is given by $\lim_{p\to 0} w_i'(p) > 1$ and $\lim_{p\to \tilde{p}} w_i'(p) < 1$ from Assumption 5. Similarly, let $\hat{y}_j$ be the *largest* output level satisfying $w_j'(1 - F(\hat{y}_j|\tilde{e}) = 1$ for a given effort $\tilde{e} \in [0, \bar{e}]$.

Suppose that $\hat{y}_i < \hat{y}_j$. Then, for any $y \in [\hat{y}_j, \bar{y}]$ and $e'', e' \in [0, \bar{e}]$ such that $e'' > e'$ it must be that

$$w_j'\big(1 - F\big(y|e''\big)\big) - w_j'\big(1 - F\big(y|e'\big)\big) > w_j'\big(1 - F\big(y|e''\big)\big) - w_j'\big(1 - F\big(y|e'\big)\big).$$

In words, agent j assigns more probability weight to the increase in probability associated with choosing a higher effort at high output levels than agent i. This weight assignment contradicts the initial assumption that i is more optimistic than j. Hence, it must be that $\hat{y}_i > \hat{y}_j$. Therefore, the segment $y \in [\underline{y}, \hat{y}_i]$, which according to Proposition 1(i) is where the RDU agent exhibits demotivation, is larger than the segment $y \in [\underline{y}, \hat{y}_j]$; agent i is demotivated over a larger output interval relative to j.

The proof for likelihood insensitive agents follows similar steps and differs only in that the output segment under consideration is restricted. Formally, the proof is performed in all $y \in [\tilde{y}_i, \hat{y}_i]$ where $\tilde{y}_i \in (\underline{y}, \bar{y})$ satisfies $1 - F(\tilde{y}_i|\tilde{e}) = \tilde{p}$ for a given effort level $\tilde{e} \in [0, \bar{e}]$. □

Proof of Proposition 2. First, let $w(r) > u(r)$. For the optimality of $L = (r, t(y); 1 - r, 0)$ it suffices that

$$\begin{aligned} &\mathrm{RDU}(L, e) > \mathrm{RDU}\big(t(y), e\big) \\ &\Leftrightarrow \quad w(r) \int_{\underline{y}}^{\bar{y}} u\big(T(y)\big) w'\big(1 - F(y|e)\big) f(y|e)\, \mathrm{d}y > \int_{\underline{y}}^{\bar{y}} u\big(t(y)\big) w'\big(1 - F(y|e)\big) f(y|e)\, \mathrm{d}y. \end{aligned} \tag{15}$$

Where the equivalence in the equation above follows from equations (3) and (5). The inequality in equation (15) holds if for any realization $\tilde{y} \in [\underline{y}, \bar{y}]$ it is true that

$$w(r) u\big(T(\tilde{y})\big) > u\big(t(\tilde{y})\big) \quad \Leftrightarrow \quad w(r) > \frac{u\big(t(\tilde{y})\big)}{u\big(T(\tilde{y})\big)}. \tag{16}$$

Using Assumption 7 and the concavity of u (Assumption 4), equation (16) can be rewritten as

$$w(r) > \frac{u\big(rT(\tilde{y})\big)}{u\big(T(\tilde{y})\big)} \quad \Rightarrow \quad w(r) > r \frac{u\big(T(\tilde{y})\big)}{u\big(T(\tilde{y})\big)} \quad \Leftrightarrow \quad w(r) > r. \tag{17}$$

The concavity of u, $u(0) = 0$, and $u(M_2) = 1$ (Asumption 4) imply $u(r) > r$. Hence, the last inequality in (17) is implied by $w(r) > u(r)$.

Since $\mathrm{RDU}(L, e) > R(t, e)$, offering L leaves the agent's participation and incentive compatibility constraint unchanged as compared to offering the generic contract $t(y)$; It provides the same utility and incentives (equation (17)) at lower costs to the principal (Assumption 7).

Next, let L be optimal. Consider a weighting function of the form $w(p) = \eta(u(p))$, where $\eta : [0, 1] \to [0, 1]$ exhibits the properties of Assumption 5. Notice that this weighting function can take any shape by virtue of Definition 2 and Definition 4. Specifically, η can be set to be sufficiently concave, convex, or likelihood insensitive to obtain a desired shape of w.

According to Lemma 1, a necessary condition for the optimality of L is $w''(p) < 0$. Under the assumed structure of $w(p)$, that condition becomes

$$-\frac{\eta''\big(u(p)\big)u'(p)}{\eta'\big(u(p)\big)} > \frac{u''(p)}{u'(p)}. \tag{18}$$

The inequality in equation (18) holds if η is strictly concave or, equivalently, when the following inequality is satisfied:

$$\frac{-\eta''(p)}{\eta'(p)} > 0. \tag{19}$$

Let $p_1, p_2 \in (0, 1)$ such that $p_2 \geq p_1$. Integrating both sides of equation (19) with respect to p over the interval $[p_1, p_2]$ gives

$$-\int_{p_1}^{p_2} \frac{\eta''(s)}{\eta'(s)} \, \mathrm{d}s > 0 \quad \Leftrightarrow \quad \ln\bigl(\eta'(p_2)\bigr) < \ln\bigl(\eta'(p_1)\bigr). \tag{20}$$

Integrating the last inequality in equation (20) with respect to $p_1 \in [0, p_2]$ gives

$$\begin{aligned} & \int_0^{p_2} \eta'(p_2) \, \mathrm{d}s < \int_0^{p_2} \eta'(s) \, \mathrm{d}s \\ \Leftrightarrow \quad & \eta'(p_2) p_2 < \eta(p_2) \\ \Leftrightarrow \quad & \frac{\eta'(p_2)}{\eta(p_2)} < \frac{1}{p_2}. \end{aligned} \tag{21}$$

Integrating the last inequality in equation (21) but this time with respect to $p_2 \in [q, 1]$ for arbitrary $q \in (0, 1)$ yields

$$\int_q^1 \frac{\eta'(s)}{\eta(s)} \, \mathrm{d}s < \int_q^1 \frac{1}{s} \, \mathrm{d}s \quad \Leftrightarrow \quad -\ln\bigl(\eta(q)\bigr) < -\ln(q) \quad \Leftrightarrow \quad \eta(q) > q. \tag{22}$$

Since $u(0) = 0$ and $u(M_2) = 1$ (Assumption 4), then $u(r) \in (0, 1)$. Let $q := u(r)$. Accordingly, the last inequality in equation (22) becomes $w(r) > u(r)$. □

Proof of Corollary 2. Let $w(p) = p$. In that case, the inequality in equation (10) becomes

$$\int_{\underline{y}}^{\bar{y}} u_t\bigl(t(y)\bigr) \frac{\mathrm{d}t(y)}{\mathrm{d}y} F_{ee}(y|e) f(y|e) \, \mathrm{d}y - c_{ee}(e) < 0. \tag{23}$$

Since $c_{ee}(e) > 0$ (Assumption 1), $u' > 0$ (Assumption 4), $\frac{\mathrm{d}t(y)}{\mathrm{d}y} \geq 0$ (Assumption 3), and $F_{ee}(y|e) > 0$ (Assumption 2), the condition in equation (23) holds and t^{fo} is optimal. Contract L cannot improve upon t^{fo}. □

Proof of Proposition 3. Denote by $o(p) : [0, 1] \to [0, 1]$ all probability weighting functions with the properties of Assumption 5 and $\tilde{p} = 1$. Let $o_i(p)$ and $o_j(p)$ be the weighting functions of optimistic decision-makers i and j, respectively.

First, I present an implication of Definition 2 that is useful for the proof. Let i be more optimistic than j. Accordingly, $o_i(p) = \theta(o_j(p))$ and

$$\frac{o_i''(p)}{o_i'(p)} = \frac{\theta''(p) o_j'(p)}{\theta'(p)} + \frac{o_j''(p)}{o_j'(p)}. \tag{24}$$

Due to the strict concavity of θ (Definition 2), equation (24) implies $\frac{o_i''(p)}{o_i'(p)} < \frac{o_j''(p)}{o_j'(p)}$.

Let $p_1, p_0 \in (0, 1)$ such that $p_1 > p_0$. Integrating $\frac{o_i''(p)}{o_i'(p)} < \frac{o_j''(p)}{o_j'(p)}$ with respect to p over $[p_1, p_2]$ gives

$$\int_{p_0}^{p_1} -\frac{o_i''(s)}{o_i'(s)}\,\mathrm{d}s > \int_{p_0}^{p_1} -\frac{o_j''(s)}{o_j'(s)}\,\mathrm{d}s \quad \Leftrightarrow \quad \frac{o_j'(p_1)}{o_j'(p_0)} > \frac{o_i'(p_1)}{o_i'(p_0)}. \tag{25}$$

Integrating the resulting inequality in equation (25) with respect to p_0 leads to

$$\int_0^{p_1} o_j'(p_1)o_i'(s)\,\mathrm{d}s > \int_0^{p_1} o_i'(p_1)o_j'(s)\,\mathrm{d}s \quad \Leftrightarrow \quad \frac{o_j'(p_1)}{o_j(p_1)} > \frac{o_i'(p_1)}{o_i(p_1)}. \tag{26}$$

Integrating again, but this time with respect to p_1 and over $[q, 1]$ for some $q \in (0, 1)$ gives

$$\int_q^1 \frac{o_j'(s)}{o_j(s)}\,\mathrm{d}s > \int_q^1 \frac{o_i'(s)}{o_i(s)}\,\mathrm{d}s \quad \Leftrightarrow \quad o_i(q) > o_j(q). \tag{27}$$

Thus, if i is more optimistic than j, his weighting function overweights q to a greater extent.

Second, it is shown that there is a degree of optimism that counteracts the risk aversion from u. Properties $u(0) = 0$ and $u(M_2) = 1$ (Assumption 4) imply $u(r) \in (0, 1)$ for given $r \in (0, 1)$. Moreover, $o_j(0) = 0$, $o_j(1) = 1$, and $o_j' > 0$ (Assumption 5) entail the existence of a q such that $o_j(q) = u(r)$ for any given function o_j and given r. Continuity of the weighting function (Assumption 5) ensures the existence of q such that $q = r$ for a sufficiently concave o_j. This equality can be found through adjustments in the concavity of o_j. Accordingly, when o_j is such that $o_j(r) < u(r)$ (resp., $o_j(q) > u(r)$), o_j can be set to be more (less) concave. Fix the concavity of o_j so that $o_j(r) = o_j(q) = u(r)$ holds. This is the degree of optimism counteracting the risk aversion from utility.

Finally, I show that stronger optimism than that implied by the fixed function o_j ensures the condition from Proposition 1. Any strictly concave weighting function θ ensures $\theta(o_j(r)) > o_j(r) = u(r)$. Letting $w(r) := \theta(o_j(r))$ and using Definition 2 gives $w(r) = \theta(o_j(r)) = o_i(q)$. Hence, for the chosen o_j guaranteeing $o_j(r) = o_j(q)$, the last inequality in equation (27) becomes $w(r) > u(r)$. □

Proof of Proposition 4. Let $l(p) : [0, 1] \to [0, 1]$ be all probability weighting functions with the properties of Assumption 5 and $\tilde{p} = 0.5$. Denote by $l_i(p)$ and $l_j(p)$ the probability weighting functions of the likelihood insensitive decision-makers i and j, respectively.

First, I present an implication of Definition 4 that is useful for the proof. Let i be more likelihood insensitive than j. Then $l_i(p) = \phi(l_j(p))$ and

$$\frac{l_i''(p)}{l_i'(p)} = \frac{\phi''(p)l_j'(p)}{\phi'(p)} + \frac{l_j''(p)}{l_j'(p)}. \tag{28}$$

Due to the concavity of l_i, l_j, and ϕ in $p \in (0, \tilde{p})$, equation (28) implies $\frac{l_i''(p)}{l_i'(p)} < \frac{l_j''(p)}{l_j'(p)}$ in $p \in (0, \tilde{p})$. Let $p_1, p_0 \in (0, \tilde{p})$ such that $p_1 > p_0$. Integrating $\frac{l_i''(p)}{l_i'(p)} < \frac{l_j''(p)}{l_j'(p)}$ with respect to

p over $[p_1, p_2]$ gives

$$\int_{p_0}^{p_1} -\frac{l_i''(s)}{l_i'(s)}\,\mathrm{d}s > \int_{p_0}^{p_1} -\frac{l_j''(s)}{l_j'(s)}\,\mathrm{d}s \quad \Leftrightarrow \quad \frac{l_j'(p_1)}{l_j'(p_0)} > \frac{l_i'(p_1)}{l_i'(p_0)}. \tag{29}$$

Integrating the resulting inequality from equation (29) with respect to p_0 leads to

$$\int_{0}^{p_1} l_j'(p_1)l_i'(s)\,\mathrm{d}s > \int_{0}^{p_1} l_i'(p_1)l_j'(s)\,\mathrm{d}s \quad \Leftrightarrow \quad \frac{l_j'(p_1)}{l_j(p_1)} > \frac{l_i'(p_1)}{l_i(p_1)}. \tag{30}$$

Integrating again, but this time with respect to p_1 over $[q, \tilde{p}]$ for some $q \in (0, \tilde{p})$ yields

$$\int_{q}^{\tilde{p}} \frac{l_j'(s)}{l_j(s)}\,\mathrm{d}s > \int_{q}^{\tilde{p}} \frac{l_i'(s)}{l_i(s)}\,\mathrm{d}s \quad \Leftrightarrow \quad l_i(q) > l_j(q). \tag{31}$$

Hence, if i is more likelihood insensitive than j, his weighting function overweights q to a greater extent.

Next, it is shown that there is a degree of likelihood insensitivity that counteracts the risk aversion from u. Properties $l_j(0) = 0$, $l_j(\tilde{p}) = \tilde{p}$, and $l_j' > 0$ (Assumption 5) along with $u(r) < \tilde{p}$ entail the existence of a $q \in (0, \tilde{p})$ such that $l_j(q) = u(r)$ for any given function l_j and given r. Continuity of the weighting function (Assumption 5) ensures the existence of q such that $q = r$ for a sufficiently likelihood insensitive function l_j. This equality can be found through adjustments in the likelihood insensitivity of l_j. Accordingly, when $l_j(q) < u(r)$ (resp., $l_j(q) > u(r)$), then l_j is set to be more (less) likelihood insensitive. Fix the likelihood insensitivity of l_j so that $l_j(q) = l_j(r) = u(r)$ holds. This is the degree of likelihood insensitivity that counteracts the risk aversion from utility.

Finally, I show that stronger insensitivity than that implied by the fixed function l_j ensures the condition from Proposition 1. Any likelihood insensitive weighting function ϕ ensures $\phi(l_j(r)) > l_j(r) = u(r)$. Letting $w(r) := \phi(l_j(r))$ and using Definition 2 gives $w(r) = \phi(l_j(r)) = l_i(q)$. Hence, for the chosen l_j ensuring $l_j(r) = l_j(q)$, the last inequality in equation (27) becomes $w(r) > u(r)$. □

A.2 *Examples*

Example 2 (CRRA utility). Let $u(T) = T^\gamma$ where $\gamma \le 1$. The condition in Proposition 2 becomes $w(r) \ge r^k$. To be more motivated under the stochastic contract, the agent's probabilistic risk seeking attitudes must exceed the agent's averse risk attitudes from utility curvature brought to the probability space, r^γ.

As in Example 1, consider Prelec's (1998) weighting function. Suppose that $\alpha = 1$. The condition for Proposition 2 becomes $r^\beta \ge r^\gamma \Leftrightarrow \beta \le \gamma$. In words, the agent needs to be sufficiently optimistic for stochastic contracts to be more motivating. Now consider $\beta = 1$. The stochastic contract is optimal if and only if $\exp(-(-\ln(r))^\alpha) > r^\gamma$. A sufficient condition for that inequality to hold is that the agent is likelihood insensitive to an extent such that the concavity implied by α in the interval $r \in (0, \frac{1}{\mathrm{e}})$ is larger than that implied by γ. In that case, the contract must be implemented with $r < \frac{1}{\mathrm{e}}$.

Lump-sum bonus versus flat wage Consider a principal who is deciding to switch from a performance-insensitive wage paying $G > 0$ to a lump-sum bonus contract, t_B. Specifically, the bonus contract is given by

$$t_B := \begin{cases} B & \text{if } y \geq \hat{y}, \\ 0 & \text{if } y < \hat{y}, \end{cases}$$

where $\hat{y} \in [\underline{y}, \bar{y}]$. The bonus contract can introduce considerable risk in the agent's environment when $\hat{y}$ is set high. That these two payment modalities are cost-equivalent (Assumption 7) implies

$$B\big(1 - F(\hat{y}|e)\big) = G. \tag{32}$$

The bonus contract leads to higher utility, and thus encompasses stronger incentives, as long as

$$\begin{aligned} &\int_{\underline{y}}^{\bar{y}} u(G)w'\big(1 - F(y|e)\big)f(y|e)\,\mathrm{d}y - c(e) < \int_{\hat{y}}^{\bar{y}} u(B)w'\big(1 - F(y|e)\big)f(y|e)\,\mathrm{d}y - c(e) \\ &\quad \Leftrightarrow \quad u(F) < u(G)w\big(1 - F(\hat{y}|e)\big). \end{aligned}$$

Using equation (32), rewrite the last inequality in the above equation as

$$\begin{aligned} & u(B)w\big(1 - F(\hat{y}|e)\big) > u\big(B\big(1 - F(\hat{y}|e)\big)\big) \\ & \quad \Rightarrow \quad w\big(1 - F(\hat{y}|e)\big) > 1 - F(\hat{y}|e), \end{aligned} \tag{33}$$

where the implication in the previous equation is due to the concavity of u. Letting $r := 1 - F(\hat{y}|e)$, the last inequality in equation (33) becomes $w(r) > r$, which is implied by $w(r) > u(r)$ the condition from Proposition 2.

Lump-sum bonus versus piece-rate Consider a principal who is deciding to switch from a linear piece-rate paying $t_P = ay$ to a lump-sum bonus contract with the same properties as the one described in the previous example. That these two payment modalities comply with Assumption 7 implies that

$$B\big(1 - F(\hat{y}|e)\big) = a\mathbb{E}(y|e). \tag{34}$$

The bonus contract leads to higher utility than the piece-rate contract if

$$\int_{\underline{y}}^{\bar{y}} u(ay)w'\big(1 - F(y|e)\big)f(y|e)\,\mathrm{d}y < \int_{\hat{y}}^{\bar{y}} u(B)w'\big(1 - F(y|e)\big)f(y|e)\,\mathrm{d}y.$$

For each $\tilde{y} \in [\underline{y}, \bar{y}]$, the above equation implies that

$$u(B)w\big(1 - F(\hat{y}|e)\big) > u(a\tilde{y}). \tag{35}$$

Using the equality in equation (34), one obtains

$$\begin{aligned} u(B)w\big(1 - F(\hat{y}|e)\big) > u(a\tilde{y}) \quad & \Leftrightarrow \quad u(B)w\big(1 - F(\hat{y}|e)\big) > u\big(B\big(1 - F(\hat{y}|e)\big)\big) \\ & \Rightarrow \quad w\big(1 - F(\hat{y}|e)\big) > 1 - F(\hat{y}|e), \end{aligned} \tag{36}$$

where the last implication is due to the concavity of u. Letting $r := 1 - F(\hat{y}|e)$, the last inequality in equation (36) becomes $w(r) > r$, a condition implied by the sufficient condition from Proposition 2.

References

Abdellaoui, Mohammed (2000), "Parameter-free elicitation of utility and probability weighting functions." *Management Science*, 46 (11), 1497–1512. [608, 622, 624, 625, 626, 629]

Abdellaoui, Mohammed, Aurélien Baillon, Laetitia Placido, and Peter P. Wakker (2011), "The rich domain of uncertainty: Source functions and their experimental implementation." *American Economic Review*, 101 (2), 695–723. [609, 623, 626, 632, 633, 637]

Abdellaoui, Mohammed, Han Bleichrodt, and Olivier L'Haridon (2008), "A tractable method to measure utility and loss aversion under prospect theory." *Journal of Risk and Uncertainty*, 36, 245. [629]

Azrieli, Yaron, Christopher P. Chambers, and Paul J. Healy (2018), "Incentives in experiments: A theoretical analysis." *Journal of Political Economy*, 126 (4), 1472–1503. [624, 630]

Azrieli, Yaron, Christopher P. Chambers, and Paul J. Healy (2020), "Incentives in experiments with objective lotteries." *Experimental Economics*, 23, 1–29. [624, 630]

Baltussen, Guido, G. Thierry Post, Martijn J. van den Assem, and Peter P. Wakker (2012), "Random incentive systems in a dynamic choice experiment." *Experimental Economics*, 15, 418–443. [623, 624, 635]

Bernheim, Douglas B. and Charles Sprenger (2020), "On the empirical validity of cumulative prospect theory: Experimental evidence of rank-independent probability weighting." *Econometrica*, 88 (4), 1363–1409. [623]

Bolton, Patrick and Mathias Dewatripont (2005), *Contract Theory*. MIT press. [622]

Bostic, Raphael, Richard J. Herrnstein, and Duncan R. Luce (1990), "The effect on the preference-reversal phenomenon of using choice indifferences." *Journal of Economic Behavior & Organization*, 13 (2), 193–212. [625]

Brüggen, Alexander and Martin Strobel (2007), "Real effort versus chosen effort in experiments." *Economics Letters*, 96 (2), 232–236. [621]

Bruhin, Adrian, Helga Fehr-Duda, and Thomas Epper (2010), "Risk and rationality: Uncovering heterogeneity in probability distortion." *Econometrica*, 78 (4), 1375–1412. [623, 632]

Charness, Gary, Ramón Cobo-Reyes, Juan A. Lacomba, Francisco Lagos, and José María Pérez (2016), "Social comparisons in wage delegation: Experimental evidence." *Experimental Economics*, 19, 433–459. [621]

Charness, Gary, Uri Gneezy, and Austin Henderson (2018), "Experimental methods: Measuring effort in economics experiments." *Journal of Economic Behavior & Organization*, 149, 74–87. [621]

Chateauneuf, Alain, Jürgen Eichberger, and Simon Grant (2007), "Choice under uncertainty with the best and worst in mind: Neo-additive capacities." *Journal of Economic Theory*, 137 (1), 538–567. [633]

Chiappori, Pierre-Andre and Bernard Salanié (2002), "Testing contract theory: A survey of some recent work." Available at SSRN 318780. [622]

Cole, Shawn, Benjamin Iverson, and Peter Tufano (2022), "Can gambling increase savings? Empirical evidence on prize-linked savings accounts." *Management Science*, 68 (5), 3282–3308. [610]

Cox, James C. and Seth Epstein (1989), "Preference reversals without the independence axiom." *American Economic Review*, 408–426. [622]

Cubitt, Robin, Chris Starmer, and Robert Sugden (1998), "Dynamic choice and the common ratio effect: An experimental investigation." *The Economic Journal*, 108 (450), 1362–1380. [613, 623]

Diecidue, Enrico, Peter P. Wakker, and Marcel Zeelenberg (2007), "Eliciting decision weights by adapting de Finetti's betting-odds method to prospect theory." *Journal of Risk and Uncertainty*, 34 (3), 179–199. [626]

Dixon, Wilfrid J. and Alexander M. Mood (1948), "A method for obtaining and analyzing sensitivity data." *Journal of the American Statistical Association*, 43 (241), 109–126. [625]

Dutcher, Glenn, Tim Salmon, and Krista J. Saral (2015), "Is 'real' effort more real?" Working Paper available at SSRN 2701793. [621]

Ederer, Florian, Richard Holden, and Margaret Meyer (2018), "Gaming and strategic opacity in incentive provision." *The Rand Journal of Economics*, 49 (4), 819–854. [609]

Epper, Thomas, Helga Fehr-Duda, and Adrian Bruhin (2011), "Viewing the future through a warped lens: Why uncertainty generates hyperbolic discounting." *Journal of Risk and Uncertainty*, 43 (3), 169–203. [626, 632]

Fabbri, Marco, Paolo Nicola Barbieri, and Maria Bigoni (2019), "Ride your luck! A field experiment on lottery-based incentives for compliance." *Management Science*, 65 (9), 4336–4348. [610]

Fehr-Duda, Helga and Thomas Epper (2012), "Probability and risk: Foundations and economic implications of probability-dependent risk preferences." *Annual Review of Economics*, 4 (1), 567–593. [608, 624, 626]

Filiz-Ozbay, Emel, Jonathan Guryan, Kyle Hyndman, Melissa Kearney, and Erkut Y. Ozbay (2015), "Do lottery payments induce savings behavior? Evidence from the lab." *Journal of Public Economics*, 126, 1–24. [610]

Fischbacher, Urs (2007), “Z-tree: Zurich toolbox for ready-made economic experiments.” *Experimental Economics*, 10, 171–178. [621]

Fischer, Gregory W., Ziv Carmon, Dan Ariely, and Gal Zauberman (1999), “Goal-based construction of preferences: Task goals and the prominence effect.” *Management Science*, 45 (8), 1057–1075. [625]

Goldstein, William M. and Hillel J. Einhorn (1987), “Expression theory and the preference reversal phenomena.” *Psychological Review*, 94 (2), 236. [634, 635, 639]

Gonzalez, Richard and George Wu (1999), “On the shape of the probability weighting function.” *Cognitive Psychology*, 166, 129–166. [614, 633]

Gonzalez-Jimenez, Victor (2024a), “Replication package for ‘Incentive contracts when agents distort probabilities’.” *Quantitative Economics*, data posited at Zenodo https://doi.org/10.5281/zenodo.10911176. [620]

González-Jiménez, Víctor (2024b), “Supplement to ‘Incentive contracts when agents distort probabilities’.” *Quantitative Economics Supplemental Material*, 15, https://doi.org/10.3982/QE2275. [613]

Haisley, Emily, Kevin G. Volpp, Thomas Pellathy, and George Loewenstein (2012), “The impact of alternative incentive schemes on completion of health risk assessments.” *American Journal of Health Promotion*, 26 (3), 184–188. [610]

Haller, Hans (1985), “The principal-agent problem with a satisficing agent.” *Journal of Economic Behavior & Organization*, 6 (4), 359–379. [609]

Herweg, Fabian, Daniel Mueller, and Philipp Weinschenk (2010), “Binary payment schemes: Moral hazard and loss aversion.” *American Economic Review*, 100 (5), 2451–2477. [609]

Hey, John D. and Jinkwon Lee (2005), “Do subjects separate (or are they sophisticated)?” *Experimental Economics*, 8, 233–265. [623]

Holmstrom, Bengt (1979), “Moral hazard and observability.” *Bell Journal of Economics*, 10 (1), 74–91. [607]

Holmstrom, Bengt and Paul Milgrom (1987), “Aggregation and linearity in the provision of intertemporal incentives.” *Econometrica*, 303–328. [622]

Homonoff, Tatiana A. (2018), “Can small incentives have large effects? The impact of taxes versus bonuses on disposable bag use.” *American Economic Journal: Economic Policy*, 10 (4), 177–210. [626]

Jewitt, Ian (1988), “Justifying the first-order approach to principal-agent problems.” *Econometrica*, 1177–1190. [610]

Jindapon, Paan, Pacharasut Sujarittanonta, and Ajalavat Viriyavipart (2022), “Prize-linked savings games: Theory and experiment.” *Games and Economic Behavior*, 133, 202–229. [610]

Karni, Edi and David Schmeidler (1991), "Atemporal dynamic consistency and expected utility theory." *Journal of Economic Theory*, 54 (2), 401–408. [613]

Kimmel, Stephen E., Andrea B. Troxel, George Loewenstein, Colleen M. Brensinger, Jane Jaskowiak, Jalpa A. Doshi, Mitchell Laskin, and Kevin G. Volpp (2012), "Randomized trial of lottery-based incentives to improve warfarin adherence." *American Heart Journal*, 164 (2), 268–274. [610]

Koszegi, Botond (2014), "Behavioral contract theory." *Journal of Economic Literature*, 52, 1075–1118. [609]

Lee, Jinkwon (2008), "The effect of the background risk in a simple chance improving decision model." *Journal of Risk and Uncertainty*, 36 (1), 19–41. [623]

L'Haridon, Olivier and Ferdinand M. Vieider (2019), "All over the map: A worldwide comparison of risk preferences." *Quantitative Economics*, 10 (1), 185–215. [608, 623, 626, 634]

Li, Zhihua, Julia Müller, Peter P. Wakker, and Tong V. Wang (2018), "The rich domain of ambiguity explored." *Management Science*, 64 (7), 3227–3240. [633, 634]

Machina, Mark J. (1989), "Dynamic consistency and non-expected utility models of choice under uncertainty." *Journal of Economic Literature*, 27 (4), 1622–1668. [613]

Mirrlees, James A. (1999), "The theory of moral hazard and unobservable behaviour: Part I." *The Review of Economic Studies*, 66 (1), 3–21. [610]

Naritomi, Joana (2019), "Consumers as tax auditors." *American Economic Review*, 109 (9), 3031–3072. [610]

Niederle, Muriel and Lise Vesterlund (2007), "Do women shy away from competition? Do men compete too much?" *The Quarterly Journal of Economics*, 122 (3), 1067–1101. [621]

Prelec, Drazen (1998), "The probability weighting function." *Econometrica*, 66 (3), 497–527. [618, 627, 633, 634, 639, 647]

Quiggin, John (1982), "A theory of anticipated utility." *Journal of Economic Behavior & Organization*, 3 (4), 323–343. [611, 612]

Rabin, Matthew (2000), "Risk aversion and expected-utility theory: A calibration theorem." *Econometrica*, 68 (5), 1281–1292. [632]

Röell, Ailsa (1987), "Risk aversion in Quiggin and Yaari's rank-order model of choice under uncertainty." *The Economic Journal*, 97 (Supplement), 143–159. [612]

Sautmann, Anja (2013), "Contracts for agents with biased beliefs: Some theory and an experiment." *American Economic Journal: Microeconomics*, 5 (3), 124–156. [639]

Segal, Uzi (1990), "Two-stage lotteries without the reduction axiom." *Econometrica*, 349–377. [613]

Sen, Aditi P., Taylor B. Sewell, E. Brooks Riley, Beth Stearman, Scarlett L. Bellamy, Michelle F. Hu, Yuanyuan Tao, Jingsan Zhu, James D. Park, George Loewenstein et al.

(2014), “Financial incentives for home-based health monitoring: A randomized controlled trial.” *Journal of General Internal Medicine*, 29, 770–777. [610]

Spalt, Oliver (2013), “Probability weighting and employee stock options.” *Journal of Financial and Quantitative Analysis*, 48 (4), 1085–1118. [609, 641]

Tversky, Amos and Daniel Kahneman (1992), “Advances in prospect theory: Cumulative representation of uncertainty.” *Journal of Risk and Uncertainty*, 5 (4), 297–323. [608, 613, 635]

Tversky, Amos and Peter P. Wakker (1995), “Risk attitudes and decision weights.” *Econometrica*, 63 (6), 1255–1280. [609, 613, 614, 615]

Volpp, Kevin G., Leslie K. John, Andrea B. Troxel, Laurie Norton, Jennifer Fassbender, and George Loewenstein (2008), “Financial incentive-based approaches for weight loss: A randomized trial.” *JAMA*, 300 (22), 2631–2637. [610]

Wakker, Peter P. (2010), *Prospect Theory for Risk and Ambiguity*. Cambridge University Press. [609, 613, 614, 624, 626, 633, 637]

Wakker, Peter P. and Daniel Deneffe (1996), “Eliciting von Neumann–Morgenstern utilities when probabilities are distorted or unknown.” *Management Science*, 42 (8), 1131–1150. [624, 626, 632]

Yaari, Menahem E. (1987), “The dual theory of choice under risk.” *Econometrica*, 55 (1), 95–115. [613, 614]

Young, Alwyn (2019), “Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results.” *The Quarterly Journal of Economics*, 134 (2), 557–598. [636]

Co-editor Garance Genicot handled this manuscript.

Manuscript received 24 October, 2022; final version accepted 10 May, 2024; available online 20 May, 2024.

The replication package for this paper is available at https://doi.org/10.5281/zenodo.10911176. The Journal checked the data and codes included in the package for their ability to reproduce the results in the paper and approved online appendices.