The dark side of bonuses*

Victor Gonzalez-Jimenez[a], Patricio S. Dalton[b], Charles N. Noussair[c]

[a]*Department of Economics, University of Vienna*
[b]*Department of Economics, Tilburg University*
[c]*Department of Economics, University of Arizona*

Abstract

To incentivize workers, firms often offer monetary bonuses for the achievement of production goals often chosen by the workers themselves. Such bonuses appeal to two types of motivation: an *extrinsic* motivation amounting to the money paid to achieve the goal, and an *intrinsic* motivation associated with the workers' desire to not fall short from their own goal. We develop a theoretical framework that predicts that if workers are sufficiently loss averse, offering a monetary bonus for achieving a self-chosen goal crowds out intrinsic motivation, it will make workers set more conservative goals. Instead, if no monetary bonuses are offered workers set more ambitious goals, which in turn improves performance. Results from a laboratory experiment corroborate this prediction. This paper highlights the limits of monetary bonuses as an effective incentive when workers may be loss averse.

JEL Codes: J41, D90, C91, D81
Keywords: Goal-setting, Contracts, Loss aversion, Bonuses, Experiment

* This paper has benefited from detailed insightful comments by Alexander Koch and Joaquin Gomez-Miñambres. We also thank Christine Eckert, Juha Tolvanen, Pol Campos-Mercade, participants at the 2019 Bogota Experimental Economics Conference, the 2019 M-BEES conference, and the 2019 Young Economists Meeting in Brno for their comments. The experiments reported in this paper were funded by the University of Vienna and the Economic Science Laboratory of the University of Arizona.

1. Introduction

Offering monetary bonuses for the achievement of milestones is a widespread practice that firms use to incentivize workers and company executives. According to Worldatwork (2018), close to 98% of publicly traded American companies use at least one compensation scheme that includes bonuses, and 73% of these companies report that these bonuses are triggered when an individual or organizational goal is reached. The theoretical rationale behind including bonuses in compensation packages is that the additional monetary incentive that the bonus creates boosts performance, as long as standard assumptions on the worker's preferences and cost of effort hold (Gibbons and Roberts, 2013).

There is also ample empirical evidence from psychology showing that setting a challenging but attainable goal can, on its own, lead to greater effort exertion, stronger attention, and higher endurance in physically and cognitively demanding tasks (Heath et al.,1999; Wu et al., 2008). This is presumably because the goal attains the status of a reference point, making the loss averse individual exert great effort to avoid experiencing the losses in utility that would result from falling short of the goal (Heath et al., 1999). Indeed, studies have shown that principals can take advantage of the motivating presence of non-binding goals in the design of workers' contracts (Gómez-Miñambres, 2012, Corgnet et al., 2015, 2018). If this intrinsic motivation to achieve a goal is sufficiently strong, monetary bonuses for goal achievement may not provide any additional incentive, and may even crowd out the motivation that the presence of a goal creates. If so, money spent on the bonuses would be a wasteful expenditure for an employer.

We consider an environment in which workers can set their own goals, rather than having them fully specified by an employer or another party. Allowing the worker to set own goals is particularly beneficial when the principal knows less about a worker's ability than the worker does. Self-chosen goals are also useful in settings where there is potential heterogeneity regarding workers' ability in the task, but it is not possible for the employer to impose different contracts on different workers. From an empirical standpoint, contracts with self-chosen goals have been shown to be more cost-effective than pure piece-rate contracts, both in the field (Groen et al., 2015; Brookins et al., 2017) and in the laboratory (Dalton et al., 2016a). Moreover, workers who set their own target feel that they have more control over the outcome, and that they are more involved with the decisions of the firm (Groen et al., 2012, 2015). Due to its many desirable features, firms are increasingly employing the practice of own-goal setting (Gallo 2011, Bourne et al., 2013, Groen et al., 2015, de Morree, 2018).[1]

We focus on the question of whether attaining a self-chosen goal should be rewarded with a monetary bonus or not. On one hand, offering monetary bonuses contingent on the achievement of production goals can incentivize accurate goal setting and motivate greater effort, particularly if more ambitious goals are accompanied by higher bonuses. Moreover, these bonuses can be designed in a way that incentivize workers to self-select into the goal, and corresponding bonus, that best fits their ability, solving the potential adverse selection problem (see Laffont and Tirole (1993)). On the other hand, as noted earlier, if the monetary bonus does not yield enough additional benefit to the principal than a goal with no monetary bonus, the employer

[1] Setting one's own goal is also common outside the workplace. Fitness apps allow the user to set her own target for minutes exercised or steps walked each day. Many banks offer a calculator in which the account holder sets a savings goal and the amount that has to be saved each month is calculated for them. Apps such as Goals-on-Track and Lifetick permit the user to set their goals in a variety of areas and to track their progress.

would be better off not including a monetary component. Including a bonus could also be counterproductive if the monetary payment crowds out the intrinsic motivation from setting a goal, as has been observed in other settings (Gneezy and Rustichini, 2000, Ariely et al., 2009a, Ariely et al., 2009b, Gneezy et al., 2011).

We develop a theoretical model, presented in Section 2, which shows that when workers have reference-dependent preferences and set their own goals, including monetary bonuses can actually lower performance. This is because a worker facing a potential monetary bonus will choose more conservative goals to increase her chance of obtaining the bonus, and as a consequence, exhibits lower performance. Workers with greater loss-aversion, that is, who are more sensitive to the losses associated with not attaining the bonus, set particularly lower goals and exhibit a larger decrease in performance.

We conduct a laboratory experiment to test the predictions of our model. In the experiment, described in Section 4, participants must complete a task that requires effort and attention. Participants are randomly assigned to one of four different contracts. The design of the experiment can be viewed from the perspective of a firm that is considering changes to a simple piece rate contract. The baseline treatment, called LOPR, is a low-powered piece rate contract. The second treatment is a higher-powered piece rate contract, and is called HIPR. Comparing the LOPR and HIPR conditions can indicate how much more (or less) performance one can get from increasing the piece rate. The third treatment, GOAL+BONUS, is a contract in which the low-powered piece rate is complemented with a self-chosen goal that yields a monetary bonus in the event that the goal is achieved. Comparing GOAL+BONUS to LOPR provides a measure of whether there is sufficient improvement in performance from adding a self-chosen GOAL to more than offset the bonuses that are paid. The final contract, GOAL, adds a self-chosen goal without any monetary bonus, to the low piece rate. This contract represents no extra expenditure on the part of the employer beyond that under LOPR.

Our model predicts that if individuals are sufficiently loss averse, GOAL would yield better performance than GOAL+BONUS. If this turns out to be the case, GOAL would dominate GOAL + BONUS because it would involve lower employer expenditure for higher output. Since our theoretical predictions depend on the extent of individuals' loss aversion, and because our model assumes risk-neutrality, we elicit the subjects' risk and loss aversion as well. We implement the parameter-free elicitation method developed by Abdellaoui et al. (2008). This elicitation method has the advantage that it allows the elicitation of the curvature of subjects' utility functions, as well as of their degree of loss-aversion, while accounting for the possibility that subjects might exhibit probability weighting (Tversky and Kahneman, 1992, Gonzalez and Wu, 1999, Abdellaoui, 2000).

The experimental data, reported in Section 5, confirm the main predictions of the model. On average, participants assigned to the GOAL treatment set more ambitious goals than those assigned to GOAL+BONUS and exhibit higher performance than subjects assigned to any of the other three contracts. Specifically, a non-paid goal contract leads to 11% more output, an increase in performance of 0.36 standard deviations, over a paid goal contract. Moreover, as predicted by the model, we find that the performance of loss averse subjects is especially greater when goals are not rewarded monetarily. Finally, we observe that performance under goals with no payment is significantly higher than performance under the piece-rate contract that offers the same monetary incentives without goals. We conclude that a GOAL contract with no monetary bonus is the cheapest way to improve performance among the contracts we study.

2. Theoretical Framework

2.1. Goals and effort under standard preferences

Consider a worker hired by a principal to produce output $y \in [0, \overline{y}]$ on a task. The agent's action consists of exerting an effort level $e \in \{e_L, e_H\}$. Exerting effort implies incurring in a cost $c(e)$, which is higher under high effort, e_H, than under low effort, e_L. For simplicity, we assume the following piece-wise function for $c(e)$:

Assumption 1. $c(e) = \begin{cases} c \text{ } if \text{ } e_H, \\ 0 \text{ } if \text{ } e_L. \end{cases}$

where $c > 0$. We consider a setting in which the agent's effort and factors beyond effort affect production. We thus model y as a random variable conditional on effort. Both parties, principal and agent, know that y is distributed according to the cumulative density function $F(y|e)$, which has a probability density function $f(y|e)$. To keep the problem tractable, we assume that the mean output produced is positive and bounded for any effort level, $0 < \mathbb{E}(y|e) < \infty$. Finally, we assume that higher effort boosts production in a manner exhibiting the monotone likelihood ratio property (MLRP):

Assumption 2. $\frac{\partial}{\partial y}\left(\frac{f(y|e_H)}{f(y|e_{L)}}\right) \forall \text{ } y \in [0, \overline{y}]$.

We are now in a position to describe the contracts available to the principal to incentivize the agent. We begin by analyzing a case in which the agent operates under a piece-rate contract, $w_p(y, a) := ay$, where $a > 0$ is a monetary amount that determines how much the agent is paid in exchange for some level of output y. As described in Section 4, this type of one-dimensional piece-rate contract is in effect in two treatments of our experiment, LOPR and HIPR. Under a piece-rate contract, the expected utility of the agent is:[2]

$$\mathbb{E}\left(U\left(e; w_p\right)\right) = \int_0^{\overline{y}} a \, y \, f(y|e)dy - c(e) \; = a\mathbb{E}(y|e) \; - c(e). \qquad (1)$$

When incentivized with $w_p(y, a)$ the agent exerts high effort as long as:

$$IC: \qquad a\left(\mathbb{E}(y|e_H) - \mathbb{E}(y|e_L)\right) \geq c. \qquad (2)$$

The incentive compatibility constraint in equation (2) shows that the agent is more likely to exert high effort, in that the constraint becomes more relaxed, when the piece rate a is higher and-or the cost of

[2] An assumption made in equation (1) is that monetary incentives enter into the worker's utility linearly. This assumption captures the notion that individuals facing small monetary amounts do not exhibit curvature in their utility functions (see Abdellaoui, 2000, Abdellaoui et al., 2008 and Wakker and Deneffe, 1996). As will be shown later, our data are consistent with this assumption.

exerting high effort, c, is lower.[3] Thus, there is a critical cost level below which higher effort is exerted, and above which low effort is put forth.

Now suppose that the agent can be incentivized with a goal contract, w_g, which in addition to offering a piece rate a, also offers a bonus $B(y, g)$. The bonus pays a monetary amount in the event that the agent attains or surpasses a goal, $g \in [0, \overline{y}]$. Specifically, the payoff of the agent is:

$$w_g(y, B(y, g)) := ay + B(y, g), \qquad (3)$$

where

$$B(y, g) = \begin{cases} 0 \; if \; y < g, \\ bg \; if \; y \geq g, \end{cases} \qquad (4)$$

and $b \geq 0$. That is, the agent receives a larger bonus for achieving more ambitious goals, and the bonus is not awarded if the goal is not attained. This type of contract is not novel. According to Chung et al. (2014), it is classified as a combination of linear commission and a bonus. Similar incentive schemes are used by firms.[4] We assume that the goal, g, is chosen by the agent. A key advantage of self-chosen goals as opposed to exogenously set goals is that the agent is more likely to know her cost value, c, and can set a goal that is tailored to this parameter. The incentives under the contract described above are clear: the agent will want to set a goal that is high enough so that she can earn a higher bonus, but not so high as to be unachievable.

The timing of the contract is as follows. First, the agent simultaneously decides g and e. Then, the level of output y is realized. Finally, at the end of the work period, the agent receives the benefits corresponding to y as well as the bonus if the goal is achieved.

When working under the self-chosen goal contract, the agent's expected utility is:

$$\begin{aligned} \mathbb{E}\left(U(e, g; w_g)\right) &= \int_0^{\overline{y}} ay \; f(y|e)dy + \int_g^{\overline{y}} bg \; f(y|e)dy - c(e) \\ &= a\mathbb{E}(y|e) + bg(1 - F(g|e)) - c(e) \end{aligned} \qquad (5)$$

In equation (5), the first term is the expected utility from the piece-rate payment and the second term the expected monetary benefit from reaching the goal. When incentivized with the self-chosen goal contract, the worker exerts high effort when the following inequality holds:[5]

$$IC: \quad a\big(\mathbb{E}(y|e_H) - \mathbb{E}(y|e_L)\big) + bg\big(F(g|e_L) - F(g|e_H)\big) \geq c. \qquad (6)$$

[3] We also assume that the individual rationality constraint $a\big(\mathbb{E}(y|e_H)\big) - c \geq 0$ holds.
[4] Larkin (2014) shows that firms use "accelerators" to motivate salesforce. That is after meeting a goal, the commission payment is further multiplied by a factor. In our case this factor is b. Oyer (1998) shows that firms use discrete bonuses when a quota is met. This can be interpreted as our discrete jump when the goal is met. Finally, Kaur et al. (2015) show how an incentive scheme with goals mitigates self-control problems.
[5] Here again, we assume that an individual rationality constraint ensuring that it is more profitable to exert high effort than to quit one's job holds. That is, we require that $a\big(\mathbb{E}(y|e_H)\big) + bg\big(F(g|e_H)\big) - c \geq 0$.

In what follows, we use the term *standard preferences* to describe an individual who derives no utility from achieving or failing to achieve the goal, other than from the monetary payment that it yields. Proposition 1 shows that for an agent with standard preferences high effort is more profitable under the self-chosen goals than under the piece-rate contract. The proposition also states that, under certain conditions on the probability density function, high effort choice is consistent with low goals, i.e. goals set below some threshold. All proofs are provided in Appendix A.

Proposition 1. *Let $\hat{c}_p$ and $\hat{c}_g$ be cost levels that make equations* (2) *and* (6)*, respectively, hold with equality. An agent with standard-preferences working under the goal contract:*

i. *exerts high effort more often than with a piece-rate contract, as $\hat{c}_g \geq \hat{c}_p$.*

ii. *There exists a unique threshold goal $\hat{g} \in (0, \overline{y})$ such that higher goals incentivize high effort if $g < \hat{g}$ and disincentivize high effort if $g > \hat{g}$.*

iii. *$\hat{g}$ does not depend on the bonus level, b.*

Proposition 1 has at least two key implications. The first one is that, under the self-chosen goal contract, high effort is profitable to the agent for a greater range of values of c as compared to the piece rate contract. This is mainly due to the inclusion of the bonus b.

The second implication is that the risk-neutral worker with standard preferences sets her goal lower than a threshold $\hat{g}$. This result formalizes the conventional wisdom that challenging but attainable goals should be set to generate motivation. This notion is typically attributed to non-standard preferences in the literature. In our setting, however, this result emerges from agents with standard preferences internalizing the probability of achieving their own goal. They are aware that too ambitious goals make obtaining the bonus improbable and are thus demotivating. By the same token, too easy goals yield a probable but small bonus, which is not motivating to them. Part (ii) of Proposition 1, is also at odds with the aforementioned discussion that the payoff maximizing agent should set high and challenging goals whenever possible with the self-chosen goal contract.

2.2. Goals and effort under reference-dependent preferences

We assume now that the worker has reference-dependent preferences. These preferences capture the notion that the agent not only derives utility from the monetary incentives offered by the goals contract, but also that the presence of a goal induces a psychological (dis) utility from (not) achieving it. Following Heath et al. (1999), Wu et al. (2008), and Markle et al. (2018), we assume that the goal acquires the status of a reference point, dividing the output space into gains, where the goal is attained or exceeded, and losses, where the goal is not attained. Hence, a production goal induces an intrinsic, non-monetary, psychological utility that satisfies the properties of Kahneman and Tversky's (1979) value function. We assume the following representation of this psychological utility:

Assumption 3. *The value function is the piecewise function* $v(y,g) = \begin{cases} \mu(y-g) & if\ y \geq g \\ -\mu\lambda(g-y) & if\ \ y < g \end{cases}$*, with* $\mu \geq 0$ *and* $\lambda > 1$.

The parameter $\lambda > 1$ reflects the agent's loss aversion: the psychological losses of failing short of a goal by some amount looms larger than the gains from surpassing the goal by the same amount. The parameter $\mu \geq 0$ represents the weight of the psychological component on the agent's overall utility. If $\mu = 0$, the agent's utility collapses to the case of standard preferences. Note that we do not include diminishing sensitivity. Not including this property is consistent with our assumption that individuals do not exhibit curvature of the utility function. Therefore, in our model, loss aversion is the agent's only source of risk preference (Wakker, 2010).

The expected utility of an agent with reference-dependent preferences, who is incentivized with a set-your-own goal contract is:

$$\mathbb{E}\left(U\left(w_g, e, g\right)\right) = \int_0^{\bar{y}} ay\, f(y|e)dy + \int_g^{\bar{y}} bg + \mu(y-g)\, f(y|e)dy - \int_0^g \lambda\mu(g-y)\, f(y|e)dy - c(e) \quad (7)$$

The first term of (7) is the expected monetary utility from the piece-rate payment. The second term is the expected utility from producing y above the goal. Note that this term includes both monetary (if $b > 0$) and psychological utility gains. The third term is the expected psychological disutility from producing y below the goal. It is informative to present equation (7) after solving the integrals, using integration by parts:

$$\mathbb{E}\left(U\left(w_g, e, g\right)\right) = (a+\mu)\mathbb{E}(y|e) - c(e) + bg\left(1 - F(g|e)\right) - \mu g - \mu(\lambda - 1)\int_0^g F(y|e)dy. \quad (8)$$

Using (8) we can establish that the worker with reference-dependent preferences chooses high effort as long as the following inequality holds:

$$IC: \quad (a+\mu)\left(\mathbb{E}(y|e_H) - \mathbb{E}(y|e_L)\right) + bg(F(g|e_L) - F(g|e_H)) + \mu(\lambda - 1)\int_0^g F(y|e_L) - F(y|e_H)dy \geq c. \quad (9)$$

Our next proposition states that a worker with reference-dependent preferences exerts high effort more often when incentivized with the self-chosen goal contract than under the piece-rate contract. It also shows that a challenging goal is more likely to lead to high effort if the worker is loss averse.

Proposition 2. *Let $\hat{c}_p$ and $\hat{c}_r$ be the cost levels that make equations* (2) *and* (9)*, respectively, hold with equality. An agent with reference-dependent preferences, $\mu > 0$, working under the goal contract:*

i. *exerts high effort more often than with a piece-rate contract, as $\hat{c}_r \geq \hat{c}_p$.*

ii. *There exists a unique threshold goal $\tilde{g} > \hat{g}$ such that higher goals incentivize high effort if $g < \tilde{g}$ and disincentivize high effort if $g > \tilde{g}$.*

iii. *$\tilde{g}$ increases in the agent's loss aversion, λ, and decreases in the bonus level, b.*

Equivalently, if the agent has reference-dependent preferences, the self-chosen goal contract makes exerting high effort profitable for a greater range of cost values compared to the piece-rate contract. This result is in line with Proposition 1. However, in contrast to the solution presented there, we find that higher goals, even when set above the threshold $\hat{g}$, can incentivize high effort; reference-dependence preferences make the set of motivating goals larger. Note that this motivational power of challenging goals is again not unbounded, since, to remain challenging, goals now need to be set below the threshold, $g < \tilde{g}$.

Moreover, according to Proposition 2 (iii), the larger the agent's degree of loss aversion, the larger the set of goals that are motivating, $g < \tilde{g}$. The intuition behind this result is that loss averse agents will exert high effort to minimize the probability of experiencing the psychological losses from falling short of a goal. This is evident from the third expression on the left-hand side of the inequality in eq. (9). A higher degree of loss aversion increases the disutility from missing the goal and further incentivizes the worker to choose high effort. This additional boost in motivation from stronger loss aversion is what makes the threshold $\tilde{g}$ larger.

Note that if the bonus feature of the self-chosen goal contract is what motivates the worker to exert high effort for more values of c as compared to the piece-rate, then the principal must weigh the benefits obtained from using this contract, namely receiving higher production levels, against the foregone payment in bonuses. However, it could also be that letting the worker set a goal, on its own, increases the worker's motivation apart from the effect of the bonus. If so, this practice provides benefit at no cost at all to the employer. This possibility makes the special case in which no bonus is offered, $b = 0$, of particular interest.

Under $b = 0$, the piece-rate and the self-chosen goal contract offer the same pecuniary incentives, and differ only in that the latter contract asks the worker to specify a personal goal which entails no monetary consequences. Indeed, an implication of Proposition 1 is that, under standard preferences, setting $b = 0$ means that the goal has no effect. However, the worker with reference-dependent preferences chooses high effort under the self-chosen goal contract with $b = 0$ as long as:

$$IC: \quad (a + \mu)\big(\mathbb{E}(y|e_H) - \mathbb{E}(y|e_L)\big) + \mu(\lambda - 1)\int_0^g F(y|e_L) - F(y|e_H)dy \geq c \qquad (10)$$

We show next that under reference-dependent preferences, the self-chosen goal contract yields higher motivation than a piece-rate contract with the same value of the piece rate a.

***Corollary 1**. Let $\hat{c}_p$ and $\hat{c}_r$ be the cost levels that make equations (2) and (10), respectively, hold with equality. An agent with reference-dependent preferences, $\mu > 0$, working under the goal contract with $b = 0$:*

i. *exerts high effort more often than with a piece-rate contract, as $\hat{c}_r \geq \hat{c}_p$.*

ii. *Higher goals always incentivize high effort.*

Corollary 1 states that the psychological utility from letting the agent set a goal is, on its own, able to motivate the agent to a greater extent than if she worked under a piece-rate contract offering the same monetary incentives. This greater motivation arises from letting the loss averse agent setting a goal, which divides the output space into gains and losses and induces her to work hard not to miss her goal.

It is important to note that, in contrast to Proposition 1 and Proposition 2, higher goals now always incentivize high effort. That is because not achieving a goal is now less costly, as the utility loss derived from failing to obtain a bonus is absent. Thus, when $b = 0$ challenging goals always generate more motivation since the loss averse worker will work hard to avoid incurring in the psychological losses from falling short from the goal without incurring in serious monetary costs from not making the goal.

To conclude this section, we compare the effort levels exerted by an agent who is offered a self-chosen goal contract with and without a monetary bonus. Proposition 3 states that more loss averse workers will set more ambitious goals and perform better under a self-chosen goal contract without a bonus (the GOAL treatment in our experiment) than under a similar contract with a bonus (GOAL+BONUS treatment).

Proposition 3. *There exists a threshold level of loss aversion $\hat{\lambda} > 1$ such that an agent with reference-dependent preferences, $\mu > 0$, working under the goal contract exerts high effort more often and sets more challenging goals when $b = 0$ than when $b > 0$ if $\lambda > \hat{\lambda}$.*

Proposition 3 is our main result. According to our analysis, offering a bonus for the achievement of a self-chosen goal backfires; it counteracts the motivational effects of letting the very loss averse worker set her own goal. It is thus better for the employer to offer a contract with $b = 0$, which generates higher output at a lower cost.

3. Hypotheses

The model yields a set of hypotheses that we test in our experiment. First, according to Proposition 3, the introduction of a bonus in a self-chosen goal contract leads to lower goals, lower effort, and, consequently, lower output, if subjects are sufficiently loss averse. Instead, Proposition 1 states that, under standard preferences, it is only when bonuses are included that self-chosen goals contract boost output. Because we expect a sizable fraction of participants to be loss averse, our first hypothesis is:

Hypothesis 1: *Goals and performance are greater under GOAL than under GOAL+BONUS.*

The second hypothesis concerns the interaction between the degree of loss aversion of individuals and the introduction of monetary bonuses. From Proposition 3 we hypothesize:

Hypothesis 2: *Goals and performance in the GOAL treatment exceed those in GOAL+BONUS by a greater difference, the more loss averse is the individual.*

From Proposition 2 we derive the third hypothesis, which states that a contract including a system of goal setting, regardless of whether it offers a bonus for the achievement of a self-chosen goal, outperforms a piece-rate contract:

Hypothesis 3: *Performance in GOAL and in GOAL+BONUS is greater than in LOPR.*

4. Experimental Procedures

4.1 General Procedures

The experiment was conducted at the University of Arizona's Economic Science Laboratory in May 2018. Participants were all students at the university and were recruited using an electronic system. The dataset consists of 12 sessions with a total of 161 subjects. On average, a session lasted approximately 70 minutes. Between 3 and 20 subjects took part in each session. The currency used in the experiment was US Dollars. We used Otree (Chen, et al., 2016) to implement and run the experiment. Subjects earned on average 20.7 US Dollars. The instructions of the experiment are given in Appendix D.

The experiment consisted of two parts: A and B. Upon arrival, participants were informed that their earnings from either Part A *or* Part B would be their earnings for the session, and that this would be decided by chance at the end of the session. Whether subjects faced Part A or Part B first was determined at random by the computer.

4.2. Treatment Structure: Comparison of the Contracts

In Part A, subjects performed a task that required their effort and attention. The task consisted of counting the number of zeros in a table of 100 randomly distributed zeros and ones. This task has been widely used by other researchers (e.g. Abeler et al., 2011, Gneezy et al., 2017, and Koch and Nafzinger, 2019). Subjects submitted their answers using the computer interface. Immediately after submission, a new table appeared on the computer screen and the subject was invited again to count the number of zeros in the new table.

Subjects had six rounds of five minutes each to complete as many tables correctly as they could. To get acquainted with the task, subjects also had a five-minute practice round where it was clear that their performance did not count toward their earnings. After each round ended, subjects were given feedback about the number of tables they solved correctly and their earnings for that round. If applicable, they were reminded of their goal for that round and were told whether that goal was achieved. Thus, aside from the practice round, subjects had 30 minutes to work on the task, and were given small intervals between rounds, so that they could assess their past performance. Since the time between goal setting and performance was almost immediate, we rule out by design any self-control problems, which have shown to affect goal-setting behavior.

There were four treatments, LOPR, HIPR, GOAL, and GOAL+BONUS. The treatments differed only with respect to the incentives offered to subjects. Each participant was randomly assigned to one of the four treatments. We ensured randomization in our design by having subjects in any given experimental session face the same chance of being assigned to any of the treatments. That is, within each session in the laboratory, different individuals were randomized into different treatments.[7] The incentives in effect in each treatment were the following.

- *LOPR*: Subjects were paid 0.20 dollars for each correctly solved table.
- *HIPR*: Subjects were paid 0.50 dollars for each correctly solved table.
- *GOAL*: Subjects were paid 0.20 dollars for each correctly solved table and were asked at the beginning of each round to set an individual goal regarding the number of tables that they aimed to solve in that round.
- *GOAL+BONUS:* Identical to the *GOAL* treatment, with the exception that subjects were offered a monetary bonus for reaching their goals. The bonus in dollars was equivalent to the goal set by the subject multiplied by a factor of 0.20.

LOPR can be viewed as a baseline condition to which different features that may improve performance are added. HIPR includes an increase in the piece rate, while GOAL adds a goal set by the worker. GOAL+BONUS adds the goal, as well as a monetary payment for reaching it, which is larger the more ambitious the goal.

4.3 Elicitation of Risk Attitudes

[7] This randomization implies that we expected to have 40 subjects in each treatment. An ex-ante power calculation shows that this number of subjects suffices to find differences between two treatments for a hypothesized effect size of 0.5 standard deviations with 0.8 power at the significance level used in the paper.

In Part B of the experiment, the task was to choose between two binary lotteries in multiple trials. This part of the experiment was designed to elicit subjects' loss aversion and utility curvature. The lotteries yielded either only gains, or were mixed in the sense that either gains or losses were possible. We used the Abdellaoui et al. (2008) method, which has the advantage of eliciting risk and loss attitudes without making any assumptions about the decision model that subjects use to evaluate outcomes or probabilities.

Our implementation of Abdellaoui et al.'s (2008) method consisted of 10 decision sets. Each decision set was designed to elicit indifference between two initial lotteries through bisection. The algorithm was programmed so that the subject's choice between two initial lotteries determined the next choice problem that the subject faced. Specifically, in the next choice trial either the lottery chosen in the preceding trial was replaced by a less attractive alternative, or the one not chosen was replaced by a more attractive alternative, while the other choice remained the same. The subject was again invited to choose between the two available options. This process was repeated four times.

Decision sets 1 to 5 elicited subjects' utility curvature. In each decision set, the algorithm elicited the certainty equivalent x_j of a lottery of the form $Lottery_j = (H_j, 0.5; L_j, 0.5)$ with $j = 1,2,3,4,5$, $H_j \geq L_j \geq 0$.The values of H_j and L_j used in each decision set are shown in Table 1 below.

Table 1. High and Low Values Used in Lotteries to Measure Utility Curvature

Lottery	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
H_j	4	8	12	20	20
L_j	0	0	0	0	12

Panel A of Table 2 below details an example of the bisection algorithm used to find x_1, the certainty equivalent of L_1. Note that initially, option R is a degenerate lottery that pays the expected value of option L, which, in turn, is equal to L_1. The example shows that after having made a first choice, the subject faces a new problem whereby R, the option that was chosen before, becomes less attractive. In the remaining repetitions, the individual's preferred option is L, even though lottery R becomes more attractive. The certainty equivalent is eventually determined as x_2 =1.625, the midpoint between 1.75 and 1.5.

Table 2. Example of the Elicitation Procedure for Certainty Equivalents and Loss Aversion

	Panel A			**Panel B**		
Repetition	Lottery L	Lottery R	Choice	Lottery L	Lottery R	Choice
Initial lottery	(4,0.5;0,0.5)	2	**R**	(1.62,0.5; -1.62,0.5)	0	**R**
1	(4,0.5;0,0.5)	1	**L**	(1.62,0.5; -0.81,0.5)	0	**L**
2	(4,0.5;0,0.5)	1.5	**L**	(1.62,0.5; -1.20,0.5)	0	**L**
3	(4,0.5;0,0.5)	1.75	**L**	(1.62,0.5; -1.40,0.5)	0	**R**
Final Elicitation		x_1 **=1.625**		z_1 **=-1.40**		

This table presents an example of Abdellaoui et al.'s (2008) algorithm used to find certainty equivalents $\{x_1, x_2, x_3, x_4, x_5\}$ and the sequence of negative numbers $\{z_1 z_2, z_3, z_4, z_5\}$.The left panel presents how x_1 is elicited with the algorithm. The right panel shows how z_1 could be elicited.

Decision sets 6 to 10 elicited subjects' loss aversion. The program was designed to find the outcome $z_j < 0$ that made an individual indifferent between a sure outcome of zero and a mixed lottery of the form $(k_j, 0.5; z_j, 0.5)$, with $k_j > 0$ for $j = 1,2,3,4,5$. A loss averse subject would require low values of z_j to be indifferent, whereas a gain-seeking subject would require large values of z_j to be indifferent. The starting values of the program were set at the certainty equivalent of a decision set j, i.e. $k_j = x_i$, and its mirror image, that is $z_j = -x_j$.

Panel B of Table 2 presents an example of the bisection algorithm used to find these negative outcomes. Note that in this example the certainty equivalent elicited in Panel A, x_1 =1.625, is used as an outcome of the mixed lottery. Also, note that the mirror image of x_1, -1.625, is used initially as the other outcome of the mixed lottery. The elicited value in this example, z_1 =-1.40, was the value that made the subject indifferent between the lottery $(1.625_j, 0.5; -1.40, 0.5)$ and zero.

Once subjects finished both parts, they were reminded about their performance in each round of the real-effort task, as well as whether they achieved their goal in that round, if applicable. Also, subjects were informed about the lottery that was chosen for potential compensation for Part B and its realization. They were also informed about which part of the experiment (Part A or B) was chosen to become their final earnings. Finally, participants completed a questionnaire about their general willingness to take risks, as well as specific risks (health, job, and driving related). These questions were taken from Dohmen et al. (2011). The questionnaire also included a measure of self-efficacy. The questionnaire can be found in Appendix D.

5. Results

5.1 Goal Setting

Our first hypothesis stated that goals would be higher in GOAL than in GOAL+BONUS. Table 3 presents the descriptive statistics regarding the goals set by subjects throughout the experiment, by treatment. Figure 3a presents the distribution of goals in the two treatments. The table shows that participants assigned to GOAL set an average goal of 48.82 tables, while those in the GOAL+BONUS treatment averaged 37.48 tables. This represents a difference of 30.2%, or 0.57 standard deviations, between the two treatments. Figure 3a also shows that the distribution of goals in GOAL first-order stochastically dominates that in GOAL+BONUS.

Thus, we find that subjects assigned to GOAL set significantly higher goals than subjects assigned to GOAL+BONUS (t(53.069)= 2.61297, p=0.005).[8] This result supports the first part of Hypothesis 1, and we state it as our first result:

Result 1: Goals are higher under GOAL than under the GOAL+BONUS.

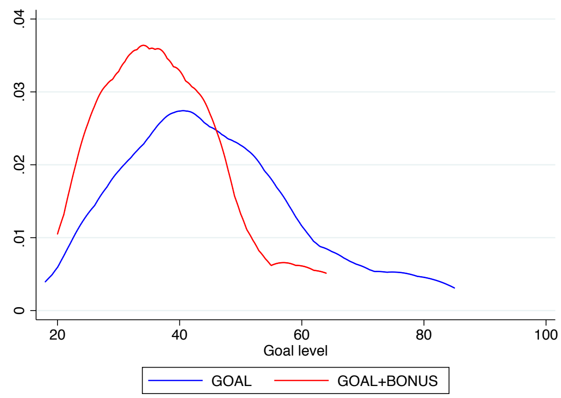
Table 3. Mean and Median Goals set, by Treatment

Treatment	N	Mean	Median	S.D.	25th perc.	75th perc.	Max.	Min.
GOAL +BONUS	39	37.487	38	10.308	29	43	64	20
GOAL	41	48.829	45	25.706	34	54	180	18
Total	**80**	**44.3**	**40**	**22.758**	**31**	**48.5**	**180**	**18**

[8] A Wilcoxon-Mann-Whitney test yields the same conclusion (U=2.842, p=0.004).

.

Figure 3a. Distribution of Goals by Treatment



In Appendix B, we show that the difference between the treatments becomes greater in later rounds, indicating that not only do subjects adjust their goals after being provided with feedback, but also that this adjustment induces a larger difference in goal setting between the treatments.

In Table 4, we report estimates from regressions of the goals set by individuals on the treatment to which they are assigned, controlling for loss aversion in some specifications. We use Poisson count regressions to account for the count nature of the goal data. The estimates confirm Hypothesis 1: subjects set higher average goals when goals are not monetarily rewarded with a bonus. Table 4 (col. 2 and 3) also shows that loss averse individuals set on average higher goals.

Table 4. Goals as Function of Treatment and Preference Parameters

	(1)	(2)	(3)
	Goal Level (all participants)	**Goal Level (all participants)**	**Goal Level (linear utility only)**
GOAL	0.264***	0.262***	0.378***
	(0.034)	(0.034)	(0.045)
Loss Averse		0.133***	0.241***
		(0.038)	(0.054)
Constant	3.624***	3.530***	3.426***
	(0.026)	(0.038)	(0.053)
Log-Likelihood	-475.870	-469.699	-280.819
N	80	80	44

Note: This table presents the estimates of Poisson count regressions of the statistical model $\text{Goal Level} = \beta_0 + \beta_1 \text{GOAL} + \Gamma' \text{Controls} + \varepsilon_i$ with $\varepsilon_i \sim \text{Poisson}(\omega)$. "Goal Level" equals the sum of a subject's goals over all six rounds of the real-effort task. Subjects were randomly assigned either to the "GOAL" or the "GOAL+BONUS" treatment. The latter is the benchmark condition of the regression. ``Loss Averse'' is a dummy variable that captures whether a subject is loss averse or not. A subject is classified as loss averse when at least four variables λ_j, where $\lambda_j \equiv x_j/z_j$, are larger than one. Model (3) presents estimates of a regression including only subjects classified as having linear utility. Standard errors presented in parenthesis. * indicates a p-value <0.1, ** indicates a p-value <0.05, ***indicates a p-value <0.0

5.2 Performance

Hypothesis 1 also asserts that performance in the GOAL treatment would be higher than in GOAL+BONUS. In addition, Hypothesis 3 states that GOAL and GOAL+BONUS lead to higher performance than LOPR, and Hypothesis 4 proposes that performance would be higher in HIPR than in LOPR. In this subsection, we examine how performance compares among treatments. Recall that we define performance in the experiment as the total number of tables an individual solves correctly.

Table 5 reports the descriptive statistics of performance by treatment. Figure 3b also shows the distribution of performance in GOAL and GOAL+BONUS. We find that paying a monetary bonus for achieving a goal does indeed backfire. Subjects assigned to GOAL solve a higher average number of tables, on average 47.58 tables, than subjects assigned to GOAL+BONUS, who solve 42.897 tables on average ($t = 1.485$, $p = 0.07$).[9] The size of this effect is 11.1%, or 0.36 standard deviations.[10] Figure 3b reveals that the distribution of performance in GOAL first order stochastically dominates that in GOAL+BONUS. According to our model, this difference implies widespread loss aversion among our participants. We consider, in Section 6, whether this is a correct view, using the data from our independent measure of loss aversion.

In addition, we observe that subjects assigned to GOAL display higher average performance than those assigned to either of the piece-rate treatments, LOPR ($t = 1.520$, $p= 0.066$) or HIPR ($t = 1.805$, $p=0.037$).[11] The effect size of the mean differences between the treatments is 0.401 standard deviations and 0.406 standard deviations, respectively. The fact that output is higher under GOAL than under HIPR shows that an employer setting up a GOAL contract rather than a higher piece rate can extract higher output from workers at lower cost. Overall, these results partially support Hypothesis 3.

Table 5. Performance by Treatment

Treatment	N	Mean	Median	S.D.	2^{5t}h	75th	Max	Min	Mean Cost
GOAL +BONUS	39	42.897	43	13.480	34	48	80	17	9.74
GOAL	41	47.585	47	14.747	39	56	86	17	9.517
HIPR	41	42.073	43	15.408	31	48.5	72	5	21.036
LOPR	39	42.179	41	15.022	27	55	72	16	8.436
Total	**160**	**44.3**	**43.5**	**14.734**	**34**	**54**	**86**	**5**	**12.259**

However, Hypothesis 3 is not fully supported in our data. We do not find significant differences in average performance between (a) GOAL+BONUS and LOPR ($t = -0.0477$, $p= 0.962$), (b) GOAL+BONUS and HIPR ($t = -0.316$, $p= 0.752$) or (c) LOPR and HIPR ($t = -0.316$, $p= 0.752$).[12] Raising the piece rate or adding a set-your-own-goal contract with a bonus did not improve performance over LOPR.

We also perform Poisson count regressions of individual performance on treatment dummies and relevant controls that confirm these patterns. Table 6 presents the regression estimates. The coefficient associated with GOAL is significant at the 5% level for all specifications indicating that subjects in GOAL exhibit higher average performance than subjects in GOAL+BONUS. Similarly, the coefficient of GOAL is

---

[9] A Wilcoxon-Mann-Whitney test generates the same conclusion ($U= 1.584$, $p=0.056$).

[10] The power of this test is 0.671. The design of the experiment aimed at a power of 0.8 with an effect size of 0.5 standard deviations. However, the sizeable variation in performance on the task, evident in larger-than anticipated standard deviations of performance, led to somewhat lower power.

[11] Wilcoxon-Mann-Whitney tests of these differences yield $U= 1.494$ ($p=0.074$) and $U= 1.512$ ($p=0.065$), respectively.

[12] These conclusions are also confirmed by Wilcoxon-Mann-Whitney tests. The U-statistics of these comparisons and their respective p-values are $U= 0.110$ ($p= 0.912$), $U= 0.048$ ($p= 0.961$), and $U= 0.034$ ($p= 0.973$), respectively.

significantly larger than the coefficient associated with LOPR ($\chi^2 = 13.28$, p= 0.001) and also larger than that for HIPR ($\chi^2 = 16.79$, p= 0.001).[13]

Figure 3b. Distribution of Performance in the Treatments with Goal Setting

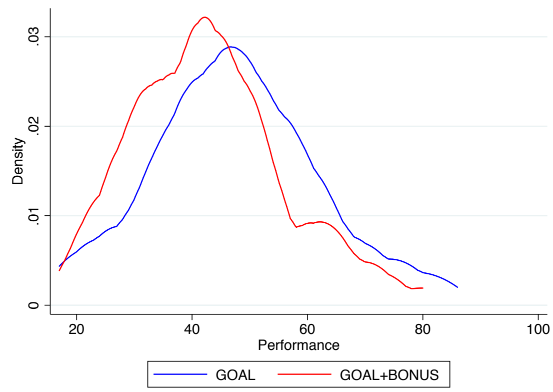


Table 6. Performance as Function of Treatment and Preference Parameters

	(1) **Performance (all participants)**	(2) **Performance (all participants)**	(3) **Performance (linear utility only)**
GOAL	0.104***	0.102***	0.176***
	(0.033)	(0.033)	(0.045)
HIPR	-0.019	-0.034	0.098**
	(0.034)	(0.034)	(0.038)
LOPR	-0.017	-0.020	-0.134***
	(0.035)	(0.035)	(0.048)
Loss Averse		0.135***	0.169***
		(0.028)	(0.038)
Constant	3.759***	3.664***	3.585***
	(0.024)	(0.032)	(0.043)
Log-Likelihood	-846.727	-834.826	-442.010
N	160	160	91

Note: This table presents the estimates of Poisson count regressions of the statistical model $\text{Performance}_i = \beta_0 + \beta_1\text{GOAL} + \beta_2\text{HIPR} + \beta_3\text{LOPR} + \beta_4\text{Loss Averse} + \varepsilon_i$ with $\varepsilon_i \sim \text{Poisson}(\omega)$. "Performance" is the total number of tables a subject solves correctly over the six rounds of the real-effort task. Subjects were assigned either to the GOAL, HIPR, LOPR, or GOAL+BONUS treatment. The GOAL+BONUS treatment is the benchmark category of the regression. ``Loss Averse'' is a dummy variable that indicates whether a subject is loss averse or not. A subject is classified as loss averse when at least four of her variables λ_j, where $\lambda_j \equiv x_j/z_j$, are greater than one. A subject is classified as having a linear utility function when for at least four Δ_{ij}s the null hypothesis that they are equal to zero was not rejected. Model (3) presents estimates of a regression including only those subjects classified as having linear utility. Standard errors are in parenthesis. * indicates a p-value <0.1, ** indicates a p-value <0.05, ***indicates a p-value <0.01

[13] We use the estimates of column (2) in Table 6 for statistical inference.

We now consider how the form of the utility function affects these comparisons. When the regression is estimated including only subjects with linear utility, another significant difference among the treatments emerges.[14] Column (3) of Table 6 shows that subjects with linear utility assigned to GOAL+BONUS exhibit higher average performance than those in LOPR. For these risk-neutral individuals, a system of goal setting with a monetary bonus is effective in inducing better performance. Thus, when all assumptions of the model hold, including the requirement that workers are risk neutral for monetary payments, Hypothesis 3 is supported.[15] Nevertheless, as in the analyses reported previously, the estimates for risk-neutral participants only continue to show that subjects assigned to GOAL exhibit higher average performance than those assigned to any other treatment

Hypothesis 5 predicts piling-up (producing output strictly in excess of one's goal) in GOAL+BONUS, and to a lesser extent in GOAL as well. Table 7 shows that for subjects in GOAL, the average piling up is negative in the first round and becomes not significantly different from zero in subsequent rounds. Subjects in the treatment in which goals are not incentivized on average just meet their self-chosen goal. In contrast, for subjects in GOAL+BONUS, piling up is positive after round 1. The average piling up level over is significantly positive for subjects in GOAL+BONUS ($t = 4.934$, $p<0.001$) and not different from zero for participants in GOAL ($t = -0.322$, $p=0.749$) When comparing piling up between these treatments, we find a significant difference ($t = -1.62$, $p= 0.054$). These findings support Hypothesis 5 and are stated as result 3.

Result 3: Performance exceeds goals in GOAL+BONUS, but not in GOAL.

Table 7. Difference between Performance and Goals ($y - g$) by Round

	Round 1	**Round 2**	**Round 3**	**Round 4**	**Round 5**	**Round 6**	**Average**
GOAL+BONUS							
Mean	0.410	0.948	1	0.641	1.282	1.128	0.901
Median	0	1	1	1	1	1	1
S.D.	2.435	1.848	1.933	2.590	1.972	2.154	1.141
GOAL							
Mean	-1.463	-1.073	0.073	0.317	0.512	0.536	-0.207
Median	0	0	0	1	1	1	0.166
S.D.	5.532	4.880	4.797	4.687	4.675	4.135	4.123

6. Risk Attitudes and their Relationship to Crowding out

[14] Details of these classifications are given in Section 6 and Appendix B. To classify subjects according to their curvature we constructed variables $\Delta_{ij} \equiv x_{ij} - EV_j$, where x_{ij} is the certainty equivalent of subject i for lottery j, EV_j stands for the expected value of the lottery, and $j = \{1, \ldots, 5\}$ is an indicator of the lottery used. A subject was classified as having a linear utility function when for at least four Δ_{ij}s the null hypothesis that they are equal to zero was not rejected.

[15] The estimates presented in column (3) of Table 6 confirm that there is no empirical evidence in our data to support the notion that higher-powered piece rates yield higher output, even if only individuals who have linear utility are considered. This corroborates the results of the pairwise tests and the regression estimates using the full sample.

In this section, we describe the attitudes of participants toward risk and losses and test Hypothesis 2, which proposes that the crowding-out effect of the monetary bonus is more pronounced for more loss averse participants.

6.1. Risk Attitudes

To measure subjects' attitudes towards risky monetary payments, we use the data from part B of the experiment, where we elicit the certainty equivalents of five lotteries that offer positive payments, $\{x_1, x_2, x_3, x_4, x_5\}$, and the negative outcomes, $\{z_1, z_2, z_3, z_4, z_5\}$, that make subjects indifferent between receiving zero and a mixed lottery $(x_j, 0.5; z_j, 0.5)$ for each $j = \{1,2,3,4,5\}$. We classify participants according to the curvature of their utility function, and their sensitivity toward losses.

By our criterion, the majority of subjects in our sample have linear utility functions in the domain of gains. Specifically, 91 subjects are classified as having a linear utility function (proportion test against 0.5, p=0.048), while 65 have concave, and only five have convex, utility. Details of this classification are included in Appendix B. If a power utility function $u(x) = x^{\theta}$ is assumed, the pooled estimate overall participants is $\hat{\theta} = 0.95$, close to risk neutrality and slightly risk averse. In Appendix B, we show that similar conclusions are reached when other families of utility functions are assumed. These results confirm that our assumption that subjects have linear utility functions is reasonable.

In addition, we find that most subjects are loss averse. Specifically, 134 subjects are classified as loss averse and 27 as gain-seeking (more sensitive to gains than losses of the same magnitude). Appendix B presents further details of this classification. Importantly, participants are balanced across treatments with respect to both the mean level of loss aversion and the proportion of loss averse participants.[16]

Table 8 presents descriptive statistics for the loss aversion coefficients, λ_j, obtained by computing $\lambda_j = x_j / z_j$, for $j = 1, \ldots, 5$. Following Abdellaoui et al. (2008), we compute one loss aversion coefficient for each mixed lottery that we implement. We find that, on average, subjects exhibit loss aversion for every mixed lottery. The null hypothesis that the loss aversion coefficient is equal to one is rejected for each lottery. Moreover, we cannot reject the null that the five loss aversion coefficients are equal to each other (F(4,805)=0.26, sphericity-corrected p-value=0.613), corroborating the notion that sign-dependence, rather than the magnitude of the loss, determines loss aversion.

Aggregating the coefficients across lotteries and individuals, we observe that subjects exhibit an average loss aversion coefficient of 3.67 and a median coefficient of 2.14. This implies that, for our subjects, losses loomed on average 3.67 times larger than equally-sized gains. Previous studies that used the same definition of loss aversion found a median loss aversion parameter of similar magnitude to our 2.14. For instance, Tversky and Kahneman's (1992) median estimate was 2.25, Abdellaoui et al. (2007) reported an estimate

---

[16] The proportion of loss averse subjects is 0.717 in LOPR, 0.6923 in GOAL, 0.6923 in GOAL+BONUS and 0.804 in HIPR. We use a two-sample test of proportions and find that these proportions are not significantly different from each other (LOPR vs. GOAL (p=0.786), HIPR vs. GOAL (p= 0.230), GOAL+BONUS vs. GOAL (p= 0.985)). Mean loss aversion is on average 3.44 for GOAL, 3.94 for GOAL+BONUS, 3.55 for HIPR and 3.75 for LOPR. The distribution of loss aversion is not significantly different between pairs of treatments, according to t-tests (LOPR vs. GOAL (p= 0.7046), HIPR vs. GOAL (p= 0.889), GOAL+BONUS vs. GOAL (p= 0.564)).

of 2.54, Abdellaoui et al. (2016) observed 1.88, and Abdellaoui et al. (2008), using the same method to elicit loss aversion as we have employed here, obtained a median loss aversion parameter equal to 2.61. As in our study, Abdellaoui et al. (2008) found considerable heterogeneity, reflected in the size of the interquartile range (the difference between the 25th and 75th percentile).

Table 8. Loss Aversion Levels for Each of the Five Lotteries

$$\lambda_j = z_j/x_j$$

	λ_1	λ_2	λ_3	λ_4	λ_5	$\lambda_{average}$
Mean	3.459	3.676	3.712	3.548	3.945	3.668
Median	1.777	1.777	1.777	1.777	2.285	2.136
S.D.	4.798	4.716	4.699	4.166	4.566	3.752
25th perc.	0.516	1.066	1.066	1.230	1.066	1.208
75th perc,	3.2	3.2	3.2	5.333	5.333	4.48

6.2. Interaction of Loss Aversion with the Bonus Treatment

Next, we examine the role of loss aversion in explaining the treatment effects. We first evaluate the strength of the treatment effects among loss averse subjects. According to Hypothesis 1, sufficiently loss averse subjects would set higher goals and exhibit higher performance when assigned to GOAL than to GOAL+BONUS. Moreover, according to Hypothesis 2, more loss averse subjects would exhibit a larger positive difference in both goals and performance between GOAL than to GOAL+BONUS.

To perform a conclusive test of Hypotheses 1 and 2, we extend the OLS regression specifications presented in Tables 4 and 6 by adding an interaction dummy for whether a subject is both loss averse and is assigned the GOAL treatment. In this manner, we evaluate whether relatively loss averse subjects display greater positive differences in performance and the goals they set when assigned to GOAL rather than to GOAL+BONUS.

Table 9 presents the results of these regressions. Columns (1) and (3) show that the estimates of "GOAL", as well as those of the interaction between ``Loss averse'' and ``GOAL'', are positive and significant. Thus, Hypothesis 1 is supported by the estimates of the interaction term. Loss averse subjects assigned to GOAL exhibit higher performance and set higher goals compared to those assigned to GOAL+BONUS, and the treatment difference is larger than for individuals who are not loss averse (($\chi^2(1)$ =7.76, p=0.005) and ($\chi^2(1)$ =15.37, p<0.001), respectively).

To distinguish participants with a high degree of loss aversion from those who are only mildly loss averse, we use a median split. That is, we classify subjects as having a high degree of loss aversion if their average loss aversion coefficient is above the median (of 2.136). Subjects with an average loss aversion coefficient below the median but greater than 1 are classified as mildly loss averse. We use the same Poisson count regression specifications as in Tables 4 and 6 but we replace ``Loss Averse'' with dummies termed "High Loss Averse" and "Mild Loss Averse". Also, we interact the dummy variables for the different degrees of loss aversion with the GOAL treatment.

The estimates in Columns (2) and (4) of Table 9 show that subjects with a high degree of loss aversion set relatively higher goals when assigned to GOAL than to GOAL+BONUS ($\chi^2(1) = 22.30$, p= 0.001). As with goals, there is a difference in performance between the two treatments for subjects with high degrees of loss aversion ($\chi^2(1) = 4.89$, p= 0.027), with those in GOAL exhibiting better performance. Mildly loss averse subjects also exhibit a treatment difference, but the coefficient is smaller and less significant. Thus, the superior performance and higher goals under GOAL compared to GOAL+BONUS are primarily driven by highly loss averse individuals.

Table 9. Heterogeneity of Treatment Effects by Participant Loss Aversion Level

	(1)	**(2)**	**(3)**	**(4)**
	Performance	**Performance**	**Goal Level**	**Goal Level**
GOAL	0.093*	0.136**	0.148**	0.159*
	(0.056)	(0.064)	(0.067)	(0.083)
GOAL*Loss Averse	0.236***		0.351***	
	(0.043)		(0.054)	
GOAL*High Loss Averse		0.188***		0.327***
		(0.054)		(0.076)
GOAL* Mild Loss Averse		0.136**		0.151**
		(0.063)		(0.060)
Loss Averse	0.131***		0.045	
	(0.033)		(0.057)	
High Loss Averse		0.144***		0.074
		(0.043)		(0.079)
Mild Loss Averse		0.051		-0.015
		(0.045)		(0.078)
HIPR	-0.034	-0.013		
	(0.034)	(0.034)		
LOPR	-0.020	-0.014		
	(0.035)	(0.035)		
Constant	3.666***	3.663***	3.593***	3.602***
	(0.034)	(0.043)	(0.048)	(0.067)
Log-Likelihood	-834.808	-836.584	-467.624	-470.959
N	160	160	80	80

Note: This table presents the estimates of the Poisson regression of the specification $\text{Performance}_i = \beta_0 + \beta_1 \text{GOAL} * \text{Loss Averse} + \beta_2 \text{GOAL} + \beta_3 \text{LOPR} + \beta_4 \text{LOPR} + \beta_5 \text{Loss Averse} + \mathcal{E}_i$ with $\mathcal{E}_i \sim \text{Poisson}\ (\omega)$ as well as the Poisson regression of the specification $\text{Goal level}_i = \beta_0 + \beta_1 \text{GOAL} * \text{Loss Averse} + \beta_2 \text{GOAL} + \beta_3 \text{LOPR} + \beta_4 \text{LOPR} + \beta_5 \text{Loss Averse} + I$ with $\mathcal{E}_i \sim \text{Poisson}\ (\omega)$. "Performance" is the total number of correctly solved tables by a subject over all rounds and "Goal level" is the sum of the goals set by the subject over all rounds. Subjects were assigned either to the GOAL, HIPR, LOPR, or the GOAL+BONUS treatment. GOAL+BONUS is the benchmark category for the regression. ``Loss Averse'' is a dummy variable that captures whether a subject is loss averse or not. A subject is loss averse when at least four of her variables $\lambda_j$, where $\lambda_j \equiv x_j/z_j$, are greater than one, ``Mild Loss averse' equals 1 if a subject is loss averse and her average λ is lower than that of the median subject in the sample, and equals 0 otherwise. ``High Loss averse' equals 1 if a subject is loss averse and her average λ is greater than that of the median participant in the sample and 0 otherwise. See Appendix B for a detailed explanation of these measurements. Standard errors are in parenthesis. * indicates a p-value <0.1, ** indicates a p-value <0.05, ***indicates a p-value <0.01.

Moreover, Table C.2 in Appendix C corroborates the pattern that subjects with both a linear utility for money and a high degree of loss aversion exhibit a stronger treatment effect. Overall, these findings are all supportive of Hypothesis 2. Additionally, in Table C.3 of Appendix C, we present evidence suggesting that there is no interaction between exhibiting a concave utility function for money and the presence of monetary bonuses in explaining goal levels. Thus, if concave utility functions have an effect in performance, this effect is not driven by the influence of concave utility on goal setting. We summarize the results of the section as follows.

Result 4: The superior performance of GOAL over GOAL+BONUS is magnified when workers are more loss averse

7. Conclusion

This paper shows that offering monetary bonuses for the achievement of self-chosen goals can significantly reduce performance. In a setting in which workers exhibit loss aversion, a monetary bonus for meeting a self-chosen goal can crowd out the motivational effect of the goal, since workers will prefer to set lower goals to increase the likelihood that the monetary bonus is achieved. This behavior is detrimental to performance because lower goals correlate with lower effort. Thus, a goal contract with no monetary payment can lead to better performance than one where achieving the goal comes with a monetary bonus. The results of the experiment demonstrate that this theoretical possibility can be observed in a real, incentivized task that requires cognitive effort.

The theoretical framework allows us to propose the mechanism behind this finding. The presence of loss aversion determines whether introducing bonuses is detrimental to performance. The experiment corroborates the theoretical proposition that loss averse individuals set especially lower goals and perform particularly worse when their self-chosen goal carries monetary consequences., In our sample, participants who exhibit a greater degree of loss aversion display a larger difference in performance between the GOAL and GOAL+BONUS treatments.

Insofar as our empirical findings generalize to less controlled, non-laboratory environments, this paper suggests that including monetary bonuses in a worker's compensation scheme does not necessarily guarantee better worker performance. Our finding in this regard is that it depends on worker preferences, with relatively high risk tolerance associated with an improvement in performance under the GOAL+BONUS contract relative to LOPR. Similarly, increasing the piece rate may also not improve performance, perhaps because income effects from having a greater piece rate that reduce effort may offset the substitution effect that increases effort. A scheme in which individuals set their own goals, when achieving the goal carries no monetary bonus, is the most effective incentive scheme that we have studied. From an employer' s point of view, given that there is already a piece rate in place, a GOAL scheme dominates the others we have studied, though obviously further research would be required in order to make stronger claims about the generality of our results.

In our view, this paper contributes to several strands of literature. It adds to the literature on incentives and contracting (Laffont and Tirole, 1993, Laffont and Martimort, 2002). We show that offering monetary bonuses can be counterproductive when they are linked to a production goal that is set by the workers. This result constitutes a proof of principle that in certain environments, offering additional monetary incentives could inhibit psychological motives that would otherwise stimulate effort.

Our study also adds to the recent and expanding area of behavioral contract theory (See Koszegi (2014) for a review). In contrast to the results of De Meza and Webb (2007) and Herweg et al. (2010), we show that in a setting of moral hazard the principal can obtain lower performance when he offers an incentive scheme that includes a performance bonus to loss averse agents. The results of this paper show that the principal should instead offer a contract that elicits a non-monetary self-chosen goal to take advantage of the worker's loss aversion.

Our paper also contributes to the emerging literature on goal setting in economics (Wu et al., 2008, Koch and Nafziger, 2011, 2019, 2020, Gómez-Miñambres, 2012, Corgnet et al., 2015, 2018, Kaur et al. 2015 Allen et al., 2017; Brookins et al., 2017, Markle et al., 2018). However, we depart from this existing literature in two ways. We do not assume dynamic inconsistency as in Hsiaw (2013,), Kaur et al. (2015), Hsiaw (2018) and Koch and Nafziger (2011, 2019). We also relax the assumption of performance being deterministic, as assumed by Wu et al. (2008), Corgnet et al. (2015, 2018), and Dalton et al. (2016a, 2016b). Including uncertainty about reaching a self-chosen goal yields the novel and perhaps counterintuitive prediction that adding monetary bonuses for the achievement of a goal can lead to lower performance. This paper also contributes to the goal-setting literature as it is, to our knowledge, the first to quantify loss aversion and utility curvature parameters elicited in an incentive compatible way to study its association with goal-setting, monetary incentives, and individual performance. Eliciting these preference parameters allows us to empirically validate the mechanisms of the model.

Additionally, our model shows that loss aversion on its own motivates higher effort. In the existing models of goal-setting, it is either diminishing sensitivity or the justification of goals through dynamic inconsistency that is necessary to make effort increase with goals. We show that when the worker faces uncertainty about reaching a goal, loss aversion alone motivates higher effort. This motivational aspect of loss aversion was highlighted in early work by Heath et al. (1999, p. 85): *"people who are below their goal by x units will perceive their current performance as a loss relative to their goal; thus they will work harder to increase their performance by a given increment than people who are above their goal by x units"*. Thus, our model provides one formalization of this notion of loss aversion as a motivating device.

Finally, we contribute to the literature that examines how extrinsic incentives crowd-out intrinsic motivation (see Gneezy et al., 2011 and Bowles and Polanía-Reyes, 2012 for reviews and Benabou and Tirole, 2003, for a theoretical framework). In this literature, crowding-out effects appear when monetary incentives inhibit individuals from signaling to others, or even themselves, a favorable attribute such as intelligence (Ariely, et al., 2009b), pro-sociality (Ariely et al., 2009a, Mellstrom and Johannesson, 2016) or norm-conforming behavior (Frey and Oberholzer-Gee, 1997, Gneezy and Rustichini, 2000). Our main contribution to this literature is to provide a completely different micro-foundation for the crowding out effect: individuals with reference dependence preferences exhibit more of a crowding out effect the more loss averse they are. That is because the intrinsic incentives of setting an ambitious goal, which stem from loss aversion, are offset by monetary incentives that lead one to set a more modest goal.

This paper opens avenues for future research. First, although laboratory experiments ensure internal validity, it would be interesting to replicate our experiment in a more natural environment with different pools of subjects. Second, future research could study the emergence and size of the crowding out effect under more general conditions. This may require giving-up on closed-form theoretical solutions and require the use of other methods to derive predictions. Finally, we restricted our study to the optimal choice of goals and output when monetary incentives are kept constant. It would be relevant to examine the optimal choice of monetary incentives (i.e. magnitude and possible non-linearity of the piece rate and eventual combination with bonus) in light of the crowding out effect that we have documented here.

References

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

Abdellaoui, M., Bleichrodt, H., and L'Haridon, O. (2008). A Tractable Method to Measure Utility and Loss Aversion Under Prospect Theory. *Journal of Risk and Uncertainty*, *36*(245), 1659-1674.

Abdellaoui, M., Bleichrodt, H., L'Haridon, O., and van Dolder, D. (2016). Measuring Loss Aversion under Ambiguity : A Method to Make Prospect Theory Completely Observable. *Journal of Risk and Uncertainty*, *52*, 1–20.

Abdellaoui, M., Bleichrodt, H., and Paraschiv, C. (2007). Loss Aversion Under Prospect Theory: A Parameter-Free Measurement. *Management Science*, *53*(10), 1659–1674.

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

Allen, E., Dechow P., Pope D. and Wu G. (2017). Reference-Dependent Preferences: Evidence from Marathon Runners", *Management Science*, 63, 1657-1672.

Ariely, B. D., Bracha, A., and Meier, S. (2009). Doing Good or Doing Well ? Image Motivation and Monetary Incentives in Behaving Prosocially. *The American Economic Review*, *99*(1), 544–555. (Ariely et al., 2009a).

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

Benabou, R., and Tirole, J. (2003). Intrinsic and Extrinsic Motivation. *Review of Economic Studies*, *70*(1), 489–520.

Bourne, M., Neely, A., Mills, J. and Platts, K. (2003). Implementing Performance Measurement Systems: a Literature review. International Journal of Business Performance Management, 5(1), 1-24.

Bowles, S., and Polanía-Reyes, S. (2012). Economic Incentives and Social Preferences: Substitutes or Complements? *Journal of Economic Literature*, *50*(2), 368–425.

Brookins, P., Goerg, S. J., and Kube, S. (2017). Self-chosen Goals, Incentives and effort. *Working Paper*.

Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree-An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97.

Chung, D. J., Steenburgh, T., and Sudhir, K. (2013). Do Bonuses enhance Sales Productivity? A Dynamic Structural Analysis of Bonus-based Compensation Plans. *Marketing Science*, *33*(2), 165-187.

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

Corgnet, B., Gómez-Miñambres, J., and Hernán-Gonzalez, R. (2018). Goal Setting in the Principal-Agent Model: Weak Incentives for Strong Performance. *Games and Economic Behavior*, *109*, 311–326.

Dalton, P. S., Gonzalez-Jimenez, V., and Noussair, C. N. (2016). Self-Chosen Goals: Incentives and Gender Differences, *CentER Discussion Paper Series* No. 2016-036. (Dalton et al., 2016a)

Dalton, P. S., Ghosal S., and Mani A. (2016). Poverty and Aspirations Failures, *Economic Journal*, *126(590)*, 165-188. (Dalton et al., 2016b)

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

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 522-550.

Frey, B. S., and Oberholzer-gee, F. (1997). The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding- Out. *The American Economic Review*, *87*(4), 746–755.

Gallo (2011, February 7) Making sure your employees succeed. Retrieved from https://hbr.org/2011/02/making-sure-your-employees-suc#comment-section

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

Gneezy, U., Goette, L., Sprenger, C., and Zimmermann, F. (2017). The limits of expectations-based reference dependence. *Journal of the European Economic Association*, *15*(4), 861–876.

Gneezy, U., Meier, S., and Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives*, *25*(4), 191–210.

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

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

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

Groen, B. A., Wouters, M. J., and Wilderom, C. P. (2012). Why do employees take more initiatives to improve their performance after co-developing performance measures? A field study. *Management Accounting Research*, *23*(2), 120-141.

Groen, B.A., and Wilderom, C.P. and Wouters M.J. (2015). High Job Performance through co-developing performance with employees. *Human resource management*, *56*(1), 111-132.

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

Herweg, F., Müller, D., and Weinschenk, P. (2010). Binary payment schemes: Moral hazard and loss aversion. *American Economic Review*, *100*(5), 2451-77.

Hsiaw, A. (2013). Goal-setting and self-control. *Journal of Economic Theory*, *148*(2), 601–626.

Hsiaw, A. (2018). Goal bracketing and self-control. *Games and Economic Behavior*, *111*(December), 100–121.

Kaur, S., Kremer, M., and Mullainathan, S. (2015). Self-control at work. *Journal of Political Economy*, *123*(6), 1227-1277.

Koch, A. K., and Nafziger, J. (2011). Self-regulation through Goal Setting. *Scandinavian Journal of Economics*, *113*(1), 212–227.

Koch, A. K., and Nafziger, J. (2016). Goals and bracketing under mental accounting. *Journal of Economic Theory*, *162*, 305–351.

Koch, A. K., and Nafziger, J. (2019). Motivational goal bracketing. *Journal of Economic Theory* (Forthcoming).

Koszegi, B. (2014). Behavioral contract theory. *Journal of Economic Literature*, *52*(4), 1075-1118.

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

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

Larkin, I. (2014). The cost of high-powered incentives: Employee gaming in enterprise software sales. *Journal of Labor Economics*, *32*(2), 199-227.

Markle, A., Wu, G., White, R., and Sackett, A. (2018). Goals as Reference Points in Marathon Running: A Novel Test of Reference Dependence. *Journal of Risk and Uncertainty*, *56*(1), 19-50.

Mellstrom, C., and Johannesson, M. (2016). Crowding out in Blood Donation: Was Titmuss Right? *Journal* *of the European Economic Association*, *6*(4), 845–863

Morree de, Pim (2018) The Science of Target Setting (and how most companies get it wrong). Retrieved from https://corporate-rebels.com/target-setting/

Oyer, P. (1998). Fiscal year ends and nonlinear incentive contracts: The effect on business seasonality. *The Quarterly Journal of Economics*, *113*(1), 149-185.

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

Wakker, P. (2010). Prospect theory: for risk and ambiguity. Cambridge university press.

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

Worldatwork. (2018). *2018 Incentive Pay Practices for Publicly-Traded Companies*.

Wu, G., Heath, C., and Larrick, R. (2008). A prospect theory model of goal behavior. *Mimeo*. Retrieved from http://faculty.chicagobooth.edu/george.wu/research/papers/wu heath larrick (prospect theory model of goal behavior).pdf

Appendix A. Proofs

Preliminaries

1. Stochastic orderings

Assumption 2 is equivalent to $\frac{f(y_2|e_H)}{f(y_2|e_L)} \geq \frac{f(y_1|e_H)}{f(y_1|e_L)}$ $\forall\, y_1, y_2 \in [0, \overline{y}]$ such that $y_2 \geq y_1$. Hence, for all y_1,

$$\int_{y_1}^{\overline{y}} \frac{f(y_2|e_H)}{f(y_1|e_H)} \mathrm{d}y_2 \geq \int_{y_1}^{\overline{y}} \frac{f(y_2|e_L)}{f(y_1|e_L)} \mathrm{d}y_2 \qquad (A1)$$

Then

$$\frac{1 - F(y_1|e_H)}{f(y_1|e_H)} \geq \frac{1 - F(y_1|e_L)}{f(y_1|e_L)} \Leftrightarrow \frac{1 - F(y|e_H)}{f(y|e_H)} \geq \frac{1 - F(y|e_L)}{f(y|e_L)} \qquad (A2)$$

Hence the monotone likelihood ratio property implies the *monotone hazard ratio* property, $\frac{f(y|e_L)}{1-F(y|e_L)} \geq \frac{f(y|e_H)}{1-F(y|e_H)}$.

The monotone hazard ratio property in $(A2)$ can be written as:

$$\frac{d}{dy}\Big(-ln\big(1 - F(y|e_H)\big)\Big) \leq \frac{d}{dy}\Big(-ln\big(1 - F(y|e_L)\big)\Big). \qquad (A3)$$

Using the fundamental theorem of calculus rewrite $(A3)$ as:

$$F(y|e_H) = 1 - \exp\left(-\int_0^y \frac{1 - F(t|e_H)}{f(t|e_H)} \mathrm{d}t\right) \leq 1 - \exp\left(-\int_0^y \frac{1 - F(t|e_L)}{f(t|e_L)} \mathrm{d}t\right) = F(y|e_L). \quad (A4)$$

Which is equivalent to first order stochastic dominance. Thus, the monotone likelihood ratio property from Assumption 2 implies *first order stochastic dominance* (FOSD), $F(y|e_H) \geq F(y|e_L)$.

To summarize, MLRP ⇒ Monotone hazard ratio ⇒ FOSD

2. Convexity ordering and monotone likelihood ratio property

The distribution $F(y|e_H)$ is more convex than $F(y|e_L)$ if $F(y|e_H)F(y|e_L)^{-1}$ is a convex function in [0,1]. The function $F(y|e_H)F(y|e_L)^{-1}$ is convex if and only if $\frac{F(y|e_L)^{-1} f(y|e_L)}{F(y|e_L)^{-1} f(y|e_H)}$ is increasing in y. To see how compute the derivative of $F(y|e_H)F(y|e_L)^{-1}$ with respect to y:

$$\frac{F(y|e_L)^{-1}f(y|e_H)}{F(y|e_L)^{-1}f(y|e_L)} + F(y|e_H)F(y|e_L)^{-1} > 0, \qquad (A5)$$

Convexity, that the second derivative is strictly positive, is guaranteed if and only if $\frac{F(y|e_H)^{-1}f(y|e_L)}{F(y|e_H)^{-1}f(y|e_H)}$ increases in y. Since, $F(y|e_H)^{-1}$ is an increasing function, then it is sufficient and necessary that the ratio $\frac{f(y|e_H)}{f(y|e_L)}$ increases in y. That is equivalent to the monotone likelihood ratio property (Assumption 2).

To summarize, MLRP $\Rightarrow$ $F(y|e_H)$ is more convex than $F(y|e_L)$

3. Properties of the monotone likelihood ratio property.

Given $F(g|e_L) - F(g|e_H) \geq 0$, then $\lim_{y\to\underline{y}^+} f(y|e_H) = 0$ and $\lim_{y\to\underline{y}^+} f(y|e_H) = \epsilon$ for arbitrarily small $\epsilon > 0$. Hence, the monotone likelihood ratio property implies that $\lim_{y\to\underline{y}^+} \frac{f(y|e_H)}{f(y|e_L)} = 0$. Also, by the FOSD implies then $\lim_{y\to\underline{y}^+} f(y|e_H) = \epsilon$ for arbitrarily small $\epsilon > 0$ and $\lim_{y\to\underline{y}^+} f(y|e_H) = 0$. Hence, the monotone likelihood ratio implies $\lim_{g\to\overline{y}^-} \frac{f(g|e_H)}{f(g|e_L)} = \infty$. These two results along with the monotone likelihood ratio property, $\frac{\partial}{\partial y}\left(\frac{f(y|e_H)}{f(y|e_L)}\right) > 0$ for all y, entail that at the low end of the output space $f(y|e_H) < f(y|e_L)$ up to a single point where $f(y|e_H) = f(y|e_L)$, and after which $f(y|e_H) > f(y|e_L)$.

To summarize, MLRP$\Rightarrow$ $f(y|e_L)$ and $f(y|e_H)$ cross only once in the support y and $f(y|e_H) \geq f(y|e_L)$ at high values of y.

Proposition 1

Proof. To prove part (i) of the Proposition, compare the incentive compatibility constraints in (2) and (6) to establish for which ranges of costs c they are satisfied. Let $\hat{c}_p > 0$ be a cost level satisfying:

$$a\big(\mathbb{E}(y|e_H) - \mathbb{E}(y|e_L)\big) = \hat{c}_p. \quad (A6)$$

Thus, for any $c \leq \hat{c}_p$ equation (2) holds. Next, let $\hat{c}_g$ be a cost level that satisfies (2):

$$a\big(\mathbb{E}(y|e_H) - \mathbb{E}(y|e_L)\big) + bg\big(F(g|e_L) - F(g|e_H)\big) = \hat{c}_g\,. \quad (A7)$$

Hence, equation (6) is satisfied for cost levels such that $c \leq \hat{c}_g$. The left-hand sides of equations $(A6)$ and $(A7)$ show that $\hat{c}_g \geq \hat{c}_p$ if $bg\big(F(g|e_L) - F(g|e_H)\big) \geq 0$. Since Assumption 2 implies FOSD, $F(g|e_L) - F(g|e_H) \geq 0$, the requirement $\hat{c}_g \geq \hat{c}_p$ amounts to having $b > 0$. Therefore, for positive bonus levels, $b > 0$, contract w_g elicits e_H for higher costs of effort c as compared to contract w_p.

To prove part (ii), take the derivative of equation (6) with respect to g:

$$\frac{\partial IC}{\partial g} = b\big(F(g|e_L) - F(g|e_H)\big) + bg\big(f(g|e_L) - f(g|e_H)\big). \qquad (A8)$$

The first expression in $(A8)$ is non-negative for all $g \in [0, \overline{y}]$ since $F(g|e_L) - F(g|e_H) \geq 0$ is implied by Assumption 2. However, the second expression is not necessarily positive. To see why, suppose instead that $f(g|e_L) > f(g|e_H)$ for all $g \in [0, \overline{y}]$. Since, $\int_0^{\overline{y}} f(g|e_L)\ dg = 1$ then, if $f(g|e_L) > f(g|e_H)$, it must be that $\int_0^{\overline{y}} f(g|e_H)\ dy < 1$, a contradiction. Therefore, $f(g|e_H) \geq f(g|e_L)$ must be true some interval in $g \in (0, \overline{y}]$, and equation $(A9)$ can be negative, implying that higher goals g can disincentivize e_H as long as $f(g|e_H) > f(g|e_L)$ to an extent that causes equation (A8) to be negative.

Let $\hat{g} \in [0, \overline{y}]$ be threshold goal making equation $(A8)$ equal to zero. That threshold can be written as:

$$\hat{g} = \frac{F(\hat{g}|e_L) - F(\hat{g}|e_H)}{f(\hat{g}|e_H) - f(\hat{g}|e_L)} \qquad (A9)$$

Due to FOSD, implied by Assumption 2, the numerator of the right-hand side of $(A9)$ is positive and the existence of $\hat{g}$ requires $f(\hat{g}|e_H) > f(\hat{g}|e_L)$. That is because if $f(\hat{g}|e_L) > f(\hat{g}|e_H)$, then by $(A9)$ the threshold goal is $\hat{g} < 0$, contradicting the initial assumption that $g \in [0, \overline{y}]$. As shown in part 3 of the proof preliminaries, the monotone likelihood ratio property (Assumption 2) guarantees $f(g|e_H) > f(g|e_L)$ at the highest end of the output interval. Therefore, $\hat{g}$ exists at high output levels.

To investigate the uniqueness of $\hat{g}$, take the second derivative of (6) with respect to g to obtain:

$$\frac{\partial^2 IC}{\partial g^2} = 2b\big(f(g|e_L) - f(g|e_H)\big) + bg\big(f'(g|e_L) - f'(g|e_H)\big). \qquad (A10)$$

The threshold goal $\hat{g}$ is unique if $\frac{\partial^2 IC}{\partial g^2} < 0$. Equation $(A10)$ together with $f(g|e_H) > f(g|e_L)$, the condition guaranteeing the existence of $\hat{g}$, imply that a sufficient condition for uniqueness is $f'(g|e_L) < f'(g|e_H)$. In other words, the function $F(g|e_H)$ must be more convex than $F(g|e_L)$. That condition is implied by Assumption 2 as shown in part 2 of the preliminaries.

Hence, if $g > \hat{g}$ equation $(A8)$ is negative and $\frac{\partial IC}{\partial g} < 0$. In this case setting goals beyond the threshold, $g > \hat{g}$, disincentivize choosing high effort e_H. Alternatively, if $g < \hat{g}$ equation $(A8)$ is positive and $\frac{\partial IC}{\partial g} > 0$, higher goals incentivize high effort.

Part (iii) of the Proposition results from b not showing in $(A9)$. ■

Proposition 2.

Proof. To prove part (i) we compare the incentive compatibility constraints presented in (2) and (9). Let $\hat{c}_r > 0$ be a cost level satisfying (2) with equality:

$$(a+\mu)\big(\mathbb{E}(y|e_H) - \mathbb{E}(y|e_L)\big) + bg\big(F(g|e_L) - F(g|e_H)\big) + \mu(\lambda - 1)\int_0^g F(y|e_L) - F(y|e_H)\mathrm{d}y = \hat{c}_r. \qquad (A11)$$

Thus, equation (9) holds for any cost level $\hat{c}_r \geq c$.

The left-hand side of equations $(A6)$ and $(A11)$ show that $\hat{c}_r \geq \hat{c}_p$ if $bg\big(F(g|e_L) - F(g|e_H)\big) \geq 0$ and $\mu(\lambda - 1)\int_0^g F(y|e_L) - F(y|e_H)dy \geq 0$. Assumption 2 implies the FOSD, $F(g|e_L) \geq F(g|e_H)$, so $bg\big(F(g|e_L) - F(g|e_H)\big) \geq 0$ if $b > 0$ and $\mu(\lambda - 1)\int_0^g F(y|e_L) - F(y|e_H)dy \geq 0$ if $\mu > 0$ and $\lambda > 1$. The restrictions $b > 0, \mu > 0, \lambda > 1$ correspond to the assumptions on the variables of the model. Hence, $\hat{c}_r \geq \hat{c}_p$ and the contract w_g elicits e_H at higher cost levels as compared to contract w_p.

To prove part (ii) first differentiate (9) with respect to g:

$$\frac{\partial IC}{\partial g} = \big(b + \mu(\lambda - 1)\big)(F(g|e_L) - F(g|e_H)) + bg(f(g|e_L) - f(g|e_H)). \qquad (A12)$$

When $\mu = 0$ equation $(A12)$ becomes identical to $(A6)$, which in the proof of Proposition 1 was established to be negative if $g > \hat{g}$, where $\hat{g}$ is the threshold goal given by $(A9)$. Let $\tilde{g} \in [0, \overline{y}\,]$ be the threshold goal that makes $(A12)$ equal to zero, that goal can be written as:

$$\tilde{g} = \left(1 + \frac{\mu}{b}(\lambda - 1)\right)\frac{F(\tilde{g}|e_L) - F(\tilde{g}|e_H)}{f(\tilde{g}|e_H) - f(\tilde{g}|e_L)}. \qquad (A13)$$

Assumption 2 implies $F(\tilde{g}|e_L) - F(\tilde{g}|e_H) \geq 0$, making the numerator of the right-hand side of $(A13)$ positive. Thus, that $\tilde{g}$ exists requires $f(\tilde{g}|e_H) > f(\tilde{g}|e_L)$, a plausible restriction at high goal levels due to Assumption 2, as show in part 3 of the preliminaries. Therefore, $\tilde{g}$ exists at high goal levels.

We investigate the uniqueness of $\tilde{g}$ by computing the second derivative of (9) with respect to g, which gives:

$$\frac{\partial^2 IC}{\partial g^2} = \big(2b + \mu(\lambda - 1)\big)(f(g|e_L) - f(g|e_H)) + bg(f'(g|e_L) - f'(g|e_H)). \qquad (A14)$$

Equation $(A14)$ together with $f(\tilde{g}|e_H) > f(\tilde{g}|e_L)$, the condition for the existence of $\tilde{g}$, imply that a necessary condition for $\frac{\partial^2 IC}{\partial g^2} < 0$, which amounts to $\tilde{g}$ being unique is $f'(g|e_H) > f'(g|e_L)$. That is, the function $F(g|e_H)$ must exhibit more convexity than function $F(g|e_L)$. This condition is implied by Assumption 2 as it was shown in part 2 of the preliminaries.

Therefore, if $g > \tilde{g}$ equation $(A12)$ is negative and $\frac{\partial IC}{\partial g} < 0$. In this case setting goals beyond the threshold, $g > \tilde{g}$, disincentivize choosing high effort e_H. Alternatively, if $g < \tilde{g}$ equation $(A12)$ is positive and $\frac{\partial IC}{\partial g} > 0$, higher goals incentivize high effort.

Next, we prove that $\tilde{g} > \hat{g}$. Suppose instead that $\hat{g} \geq \tilde{g}$. Using $(A9)$ and $(A13)$, we rewrite the assumed inequality as:

$$\frac{F(\hat{g}|e_L) - F(\hat{g}|e_H)}{f(\hat{g}|e_H) - f(\hat{g}|e_L)} \geq \left(1 + \frac{\mu}{b}(\lambda - 1)\right)\frac{F(\tilde{g}|e_L) - F(\tilde{g}|e_H)}{f(\tilde{g}|e_H) - f(\tilde{g}|e_L)}. \qquad (A15)$$

For the special case in which $\hat{g} = \tilde{g}$ equation $(A15)$ cannot hold unless $\frac{\mu}{b}(\lambda - 1) = 0$ contradicting the assumptions on these parameters of the model $\mu > 0, \lambda > 1,$ and $b > 0$. Furthermore, for the more general case $\hat{g} > \tilde{g}$, a sufficiently large μ and/or λ can be chosen to contradict inequality $(A15)$ Hence, it must be that $\tilde{g} > \hat{g}$.

Finally, to prove part (iii) of the Proposition compute the derivative of $(A13)$ with respect to λ:

$$\frac{\partial \tilde{g}}{\partial \lambda} = \frac{\mu}{b}\left(\frac{F(\tilde{g}|e_L) - F(\tilde{g}|e_H)}{f(\tilde{g}|e_H) - f(\tilde{g}|e_L)}\right). \qquad (A16)$$

Equation $(A16)$ shows that higher λ increases the threshold $\tilde{g}$ and enlarges the set $g < \tilde{g}$, in which high effort is consistent with higher goals for the agent with reference-dependent preferences. Furthermore, the derivative of $(A17)$ with respect to b yields:

$$\frac{\partial \tilde{g}}{\partial b} = -\frac{\mu(\lambda - 1)}{b^2}\left(\frac{F(\tilde{g}|e_L) - F(\tilde{g}|e_H)}{f(\tilde{g}|e_H) - f(\tilde{g}|e_L)}\right). \qquad (A17)$$

Hence, higher bonuses reduce $g < \tilde{g}$ the set in which high goals lead to high effort for the agent with reference-dependent preferences. ■

***Corollary 1*.**

Proof. To prove part (i) suppose that $\hat{c}_r > \hat{c}_p$. Using equation $(A11)$ with $b = 0$ and equation $(A6)$ the assumed inequality is equal to:

$$\mu(\lambda - 1)\int_0^g F(y|e_L) - F(y|e_H)\mathrm{d}y < 0. \quad (A18)$$

The above equation contradicts the assumptions on the parameters of the model that $\mu > 0$ and $\lambda > 1$ or the FOSD, an implication of Assumption 2 $F(y|e_L) \geq F(y|e_H)$. Hence, it must be that $\hat{c}_r \geq \hat{c}_p$ and the set of costs $\hat{c}_r \geq c$ is larger than the set of costs $\hat{c}_p \geq c$. Therefore, w_g with $b = 0$ elicits high effort e_H at higher costs than w_p when the agent has reference-dependent preferences.

To prove part ii) compute the derivate of (10) with respect to g:

$$\frac{\partial IC}{\partial g} = \mu(\lambda - 1)\big(F(g|e_L) - F(g|e_H)\big). \qquad (A19)$$

Equation $(A19)$ is always positive due to $\mu > 0, \lambda > 1$, and FOSD, an implication of Assumption 2. Hence, in the absence of bonuses, higher g ***always*** incentivize e_H. ■

Proposition 3.

Proof. First we show that the agent chooses lower g if w_g is offered with $b > 0$. First, notice that when $b = 0$, equation (9) collapses to:

$$IC: \quad (a + \mu)\big(\mathbb{E}(y|e_H) - \mathbb{E}(y|e_L)\big) + \mu(\lambda - 1)\int_0^g F(y|e_L) - F(y|e_H)\mathrm{d}y \geq c. \qquad (A20)$$

The derivative of $(A20)$ with respect to g gives:

$$\frac{\partial IC}{\partial g} = \mu(\lambda - 1)\big(F(g|e_L) - F(g|e_H)\big). \qquad (A21)$$

The assumptions on the parameters of the model $\mu > 0$ and $\lambda > 1$, and FOSD, which is implied by Assumption 2, entail that $(A21)$ is positive for all g. Second, the proof of Proposition 2 established that when $b > 0$, the incentive compatibility constraint in equation (9) can be negative if $g > \tilde{g}$. Thus, under $b > 0$ the agent avoids setting higher goals in the segment $g > \tilde{g}$, while under $b = 0$ those goals yield the highest motivation.

Let g_b be the goal when the contract with $b > 0$ is offered and let g_{nb} be the goal level when $b = 0$ is offered. We established above that $g_{nb} > g_b$ if $g_{nb} \geq \tilde{g}$. An implication of this goal setting difference is:

$$\mu(\lambda - 1)\int_0^{g_{nb}}\big(F(y|e_L) - F(y|e_H)\big)\mathrm{d}y - \mu(\lambda - 1)\int_0^{g_b} F(y|e_L) - F(y|e_H)\mathrm{d}y > 0. \qquad (A22)$$

In words, psychological utility is larger under $b = 0$ as compared to a situation in which $b > 0$ if $g_{nb} \geq \tilde{g}$. To compare the incentive compatible constraints when the agent works w_g with $b = 0$ as compared to a situation in which w_g is given with $b > 0$, subtract (9) from (10) obtaining:

$$\mu(\lambda - 1)\int_{g_b}^{g_{nb}}\big(F(y|e_L) - F(y|e_H)\big)\mathrm{d}y - bg_b\big(F(g_b|e_L) - F(g_b|e_H)\big). \qquad (A23)$$

The first expression in the above equation is positive if $g_{nb} \geq \tilde{g}$ due to $(A22)$. The second expression in $(A23)$ is negative by FOSD. Thus, $(A23)$ can become positive as long as $g_{nb} \geq \tilde{g}$. That is, w_g with $b = 0$ can elicit e_H at higher cost levels c as compared to w_g with $b > 0$ when goals are sufficiently challenging.

We turn to investigate how higher λ influences the agent's motivation when working without $b = 0$ as compared to a situation in which the bonus was given $b > 0$. To that end, derive $(A23)$ with respect to λ to obtain:

$$\mu \int_{g_b}^{g_{nb}} \big(F(y|e_L) - F(y|e_H)\big)dy + \mu(\lambda - 1)\left(\big(F(g_{nb}|e_L) - F(g_{nb}|e_H)\big)\frac{\partial g_{nb}}{\partial \lambda} - \big(F(g_b|e_L) - F(g_b|e_H)\big)\frac{\partial g_b}{\partial \lambda}\right) - b\frac{\partial g_b}{\partial \lambda}\Big(F(g_b|e_L) - F(g_b|e_H) - g_b\big(f(g_b|e_L) - f(g_b|e_H)\big)\Big) \qquad (A24)$$

To evaluate the sign of $\frac{\partial g_b}{\partial \lambda}$, take the implicit derivative of $(A14)$ with respect to b and λ:

$$\frac{\partial g_b}{\partial \lambda} = -\frac{\mu\big(F(g|e_L) - F(g|e_H)\big)}{\Big(\big(2b + \mu(\lambda - 1)\big)\big(f(g|e_L) - f(g|e_H)\big) + bg\big(f'(g|e_L) - f'(g|e_H)\big)\Big)}. \qquad (A25)$$

From the above equation it is evident that $\frac{\partial g_b}{\partial \lambda} > 0$. From $(A25)$ we also establish that $\frac{\partial g_{nb}}{\partial \lambda} > \frac{\partial g_b}{\partial \lambda}$. Thus, all expressions in $(A26)$ are positive if $g_{nb} > \tilde{g}$ due to the FOSD, the fact that $\frac{\partial g_{nb}}{\partial \lambda} > \frac{\partial g_b}{\partial \lambda}$ and that $g_{nb} > g_b$ if $g_{nb} \geq \tilde{g}$, and because $\tilde{g} > \hat{g}$. Hence, larger loss aversion, λ, unequivocally boosts motivation when the goal contract is given with $b = 0$ is given instead of $b > 0$.

Next, we prove the existence of a threshold level of loss aversion $\hat{\lambda} > 1$ that makes $(A23)$ hold with equality. Note that as $\lim_{\lambda \to 1^+} \mu(\lambda - 1) = 0$ making equation $(A23)$ negative as $\lambda \to 1^+$. From the previous result we can establish that as λ increases the expression in equation $(A23)$ becomes less negative, up to a level of loss aversion $\hat{\lambda} > 1$ that makes that equation hold with equality. Since λ is unbounded above, the existence of such $\hat{\lambda}$ is guaranteed. Hence, for any $\lambda > \hat{\lambda}$, the contract with $b = 0$ yields higher motivation than $b > 0$. Notice that $(A23)$ being positive due to $\lambda > \hat{\lambda}$ necessarily implies that $g_{nb} > \tilde{g}$ to make the first expression in $(A23)$ sufficiently large.

Finally, we show how offering a contract with larger b alters the comparison represented by equation $(A23)$. The derivative of $(A23)$ with respect to b gives:

$$-\mu(\lambda - 1)\big(F(g_b|e_L) - F(g_b|e_H)\big)\frac{\partial g_b}{\partial b} - g_b\big(F(g_b|e_L) - F(g_b|e_H)\big) - \frac{\partial g_b}{\partial b} b\Big(\big(F(g_b|e_L) - F(g_b|e_H)\big) + g_b\big(f(g_b|e_L) - f(g_b|e_H)\big)\Big). \qquad (A26)$$

To evaluate the sign of $\frac{\partial g_b}{\partial b}$, implicitly derive $(A12)$ with respect to b and g to get:

$$\frac{\partial g_b}{\partial b} = -\frac{\big(F(g|e_L) - F(g|e_H)\big) + g\big(f(g|e_L) - f(g|e_H)\big)}{\big(2b + \mu(\lambda - 1)\big)\big(f(g|e_L) - f(g|e_H)\big) + bg\big(f'(g|e_L) - f'(g|e_H)\big)}. \qquad (A27)$$

The denominator of $(A27)$ is equal to $(A14)$ which is negative due to Assumption 2, as shown in the second part of the proof preliminaries. Thus, $\frac{\partial g_b}{\partial b} < 0$ if $g_b > \hat{g}$, since for that segment the numerator of $(A25)$ is negative. Instead, $\frac{\partial g_b}{\partial b} > 0$ if $g_b < \hat{g}$.

We are now in a position to analyze the sign of $(A26)$. If $g_b > \hat{g}$, the first expression in that equation is positive, while the second and third expressions are negative. Thus, that $(A26)$ is positive requires:

$$\frac{\mu(\lambda - 1)}{b} + 1 > -\frac{\frac{g_b}{b}}{\frac{\partial g_b}{\partial b}} - \frac{g_b\big(f(g_b|e_L) - f(g_b|e_H)\big)}{\big(F(g_b|e_L) - F(g_b|e_H)\big)}. \qquad (A28)$$

Note that larger levels of loss aversion λ, make the inequality in $(A28)$ less stringent. Corroborating our previous result. Hence, contracts with larger bonuses $b > 0$ can be less motivating if $g_b > \hat{g}$ and the more loss averse the agent is. For completeness note that if $g_b < \hat{g}$, then all expressions in $(A28)$ are negative and, in that case, it is better to offer the contract with $b > 0$. ■

Appendix B. Details of the Risk Attitude Classifications

In this appendix, we report some additional analysis of the data from part B of the experiment, in which utility curvature and loss-aversion are measured. These elicited data consist of a vector of positive certainty equivalents $\{x_1, x_2, x_3, x_4, x_5\}$ for five different lotteries in the domain of gains, and the vector of offsetting loss outcomes $\{z_1, z_2, z_3, z_4, z_5\}$ for each subject. These values can be analyzed to understand (1) the risk attitudes of subjects when outcomes are restricted to the gain domain, (2) whether risk attitudes have a sign-dependent component as proposed in prospect theory, and (3) how loss averse a participant is.

B.1. Risk attitudes in the domain of gains

We begin by studying the elicited sequence $\{x_1, x_2, x_3, x_4, x_5\}$ which informs us about the risk attitudes of subjects in the domain of gains. We classify each subject according to their risk attitude. To that end we compute the difference $\Delta_{ij} \equiv x_{ij} - EV_j$, where the index j indicates the lottery number, and EV_j stands for the expected value of that lottery, that is, $EV_j = 0.5H_j + 0.5L_j$. The sign of Δ_{ij} is a non-parametric measure of the risk attitude of subject i with respect to lottery j. If the subject exhibits $\Delta_{ij} > 0$, the lowest price at which she is willing to sell the lottery is larger than its expected value, denoting a risk-seeking attitude. If $\Delta_{ij} < 0$, the subject has a risk averse attitude toward that lottery. Also, whenever $\Delta_{ij} = 0$, the subject is risk neutral.

We perform a classification of subjects based on the statistical significance of their elicited Δ_{ij}. We compute confidence intervals around zero to determine whether a Δ_{ij} is statistically relevant given the overall variation in the data. Specifically, we calculate the standard deviation of $\sum_i \Delta_{ij}$ for each lottery j, and multiply it by the factors 0.64 and -0.64. A significantly positive Δ_{ij} indicates that subject i is risk seeking with respect to lottery j, while a significantly negative Δ_{ij} indicates risk aversion. Under the assumption that the data follow a normal distribution, approximately 50% of the data must lie within this confidence interval. Furthermore, to account for response error, we classify a subject to have a risk averse attitude when at least four of her Δ_{ij}s are negative. A subject is risk seeking when at least four of her Δ_{ij}s are positive, and a subject has linear utility when at least four Δ_{ij}s are not different from zero. This is also the approach followed by Abdellaoui (2000) and Abdellaoui et al. (2008).

Table C.1 shows that, by this criterion, the majority of subjects, 57%, are classified as having linear utility. A proportion test rejects the hypothesis that the fraction is 0.5, p=0.048, indicating that a significant majority has linear utility. 40% of participants are classified as having concave utility and only five individuals (3%) as having convex utility.

The second analysis we conduct on these data assumes that the utility function of subjects follows a particular functional form. We fit the certainty equivalents to these functionals to examine the subjects' risk attitudes. Specifically, we assume that the utility of subjects is of the power utility form, $x_{ij} = {EV_j}^{\alpha}$, which belongs to the CRRA family, or of the exponential utility form, $x_{ij} = 1 - \exp(-\rho EV_j)$, which belongs to the CARA family. We estimate the parameters of these functionals, using the non-linear least squares method, for the pooled data of all participants.

Table B.1 Classification of individuals' risk attitudes towards lotteries with gains

Lottery/Classification	Concave Utility	Convex Utility	Linear Utility
Lottery 1	60	24	77
Lottery 2	80	15	66
Lottery 3	98	16	47
Lottery 4	77	9	75
Lottery 5	100	11	50
Total num. subjects	**65**	**5**	**91**

Note: This table presents the classification of individuals according to their risk attitudes in the domain of gains. Each row presents the number of subjects classified as having concave, convex or linear utility, which is equivalent to saying that they are risk averse, risk seeking, or risk neutral, respectively. A subject i is classified as having concave utility for lottery j whenever the difference $\Delta_{ij} \equiv x_{ij} - EV_j < 0$. A subject i is classified as having convex utility for lottery j whenever the difference $\Delta_{ij} > 0$. A subject i is classified as having linear utility for lottery j whenever the difference Δ_{ij} is not significantly different from zero. The last row presents the number of subjects classified as having concave, convex, and linear utility over all lotteries. A subject is classified to have concave utility if she displays risk averse attitudes toward at least four lotteries. A subject is classified to have convex utility if she displays risk seeking attitudes toward at least four lotteries. Having risk neutral attitudes for at least four lotteries classify a subject as having linear utility.

Table B.2 presents the estimates of the parameters. We find that the estimate $\hat{\alpha}$ is significantly less than one, though only modestly so. This implies a utility function for a representative agent that has slightly risk averse attitudes (F(1,160)=127.651, p<0.001). Similarly, we find that the estimate $\hat{\rho}$ is significantly greater than zero, though still fairly small in magnitude. This also reveals a small degree of risk aversion on the part of the representative individual (F(1,160)= 3.4e+05, p<0.001). Thus, we find that subjects in aggregate display slight risk aversion, when functional forms with CARA and CRRA are estimated. This is in line with the statistical analysis of the data at the individual level, which found that the majority of participants were risk-neutral, but that there was also a considerable share of subjects more appropriately classified as risk averse.

Table B.2 Estimates of parameters of the utility function of the representative participant

Parametric form	**Coefficient**	**St. Error**	R^2
$x_{ij} = EV_j^{\alpha}$	α =.9480115	.0048	0.946
$x_{ij} = \dfrac{1 - \exp(-\rho EV_j)}{\rho}$	ρ =. 01854	.0016	0.942

B.2 Loss aversion

Next, we analyze the sequence of negative outcomes $\{z_1, z_2, z_3, z_4, z_5\}$ that made the subjects indifferent between receiving zero for sure and a lottery consisting of $(z_j, 0.5; x_j, 0.5)$, where x_j was an elicited certainty equivalent of one of the lotteries containing only positive outcomes. The analysis of this data reveals the degree of a subject's loss aversion. The measure of loss aversion is the coefficient $\lambda_j \equiv x_j/z_j$. When the λ_i coefficient takes the value of one, the subject is indifferent between accepting and declining a lottery that consists of a gain and a loss of equal magnitude, each occurring with probability 0.5, and thus the subject exhibits no loss aversion or gain seeking. If the coefficient λ_i takes on a value larger than one, it indicates the presence of loss aversion.

We first analyze the loss aversion coefficients at the individual level. A subject is classified as loss averse when at least four (out of five) of her loss aversion coefficients satisfy $\lambda_i > 1$. A subject is classified as gain-seeking when at least four (out of five) of her loss aversion coefficients have $\lambda_i < 1$. Finally, a subject is classified as having mixed attitudes toward losses if she cannot be classified as either loss averse or gain seeking. Table B.3 shows that the large majority of subjects is loss averse. Specifically, 72% of participants are classified as loss averse and 14% as gain-seeking.

Table B.3 Individuals classified as loss averse for each x_j

Classification	**Loss Averse**	**Gain-Seeking**	**Mixed**
Lottery 6	106	55	-
Lottery 7	125	36	-
Lottery 8	128	33	-
Lottery 9	128	33	-
Lottery 10	124	37	-
Total	**117**	**23**	**21**
Average	**134**	**27**	**0**

We also perform a second analysis of these data featuring the average of the loss aversion coefficient that an individual exhibits over all five lotteries. The last row in Table B.3 shows that according to this analysis, 134 subjects, or 85% of participants, are loss averse, and the remaining 27 subjects are gain-seeking. Thus, both analyses conclude that the great majority of the subjects is loss averse.

Appendix C. Additional Tables and Analyses

Table C.1 Descriptive statistics of goals set by round in the GOAL+BONUS and GOAL treatments

	round 1	**round 2**	**round 3**	**round 4**	**round 5**	**round 6**
GOAL+BONUS						
mean	6.076	5.948	6.179	6.205	6.461	6.615
median	6	5	6	6	6.5	6
S.D.	2.056	1.986	2.186	2.142	1.889	2.034
GOAL						
mean	8.658	8.024	7.902	8	7.829	8.414
median	7	7	7	7	7	8
S.D.	4.908	4.345	4.353	4.399	4.371	4.549

Table C.2 Effect of treatment and loss aversion on performance and goals for subjects with linear utility only

	(1) **Performance**	(2) **Performance**	(3) **Goal Level**	(4) **Goal Level**
GOAL* Loss Averse	0.346^{***}		0.555^{***}	
	(0.058)		(0.073)	
GOAL* High Loss Averse		0.231^{***}		0.543^{***}
		(0.077)		(0.107)
GOAL*Mild Loss Averse		0.048		0.176^{*}
		(0.093)		(0.097)
Loss Averse	0.170^{***}		0.134^{*}	
	(0.045)		(0.077)	
High Loss Averse		0.220^{***}		0.114
		(0.056)		(0.109)
Mild Loss Averse		0.048		-0.038
		(0.093)		(0.111)
GOAL	0.181^{**}	0.309^{**}	0.218^{**}	0.211
	(0.082)	(0.099)	(0.095)	(0.1306)
HIPR	0.098^{***}	0.145^{***}		
	(0.044)	(0.044)		
LOPR	-0.134^{**}	-0.129^{***}		
	(0.048)	(0.048)		
Constant	3.584^{***}	3.588^{***}	3.505^{***}	3.564^{***}
	(0.464)	(0.056)	(0.065)	(0.097)
Log-Likelihood	-442.008	-437.376	-278.999	-282.066
N	91	91	44	44

Note: This table presents the estimates of the Poisson regression of the specification $\text{Performance}_i = \beta_0 + \beta_1 \text{GOAL} * \text{Loss Averse} + \beta_2 \text{GOAL} + \beta_3 \text{LOPR} + \beta_4 \text{LOPR} + \beta_5 \text{Loss Averse} + \mathcal{E}_i$ with $\mathcal{E}_i \sim \text{Poisson}(\omega)$, as well as the Poisson regression of the statistical model $\text{Goal level}_i = \beta_0 + \beta_1 \text{GOAL} * \text{Loss Averse} + \beta_2 \text{GOAL} + \beta_3 \text{LOPR} + \beta_4 \text{LOPR} + \beta_5 \text{Loss Averse} + \mathcal{E}_i$ with $\mathcal{E}_i \sim \text{Poisson}(\omega)$. "Performance" is the total number of tables a subject solves correctly over all rounds and "Goal setting" is the sum of the goals set by the subject over all rounds. Subjects were assigned either to the GOAL, HIPR, LOPR, or GOAL+BONUS treatment. GOAL+BONUS is the benchmark category for the regression. ``Loss Averse'' is a dummy variable that captures whether a subject is loss averse or not. A subject is loss averse when at least four variables $\lambda_j$, where $\lambda_j \equiv x_j/z_j$, are greater than one, ``Loss averse Mild' equals 1 if a subject is loss averse and her average λ is lower than the median subject in the sample, and 0 otherwise. ``Loss averse High' equals 1 if a subject is loss averse and her average λ is lower than the median subject in the sample and 0 otherwise. All models include only subjects classified as having linear utility. Standard errors presented in parenthesis. * indicates a p-value <0.1, ** indicates a p-value <0.05, ***indicates a p-value <0.01.

Table D.3 Association of concave utility, loss aversion and treatment, with goals and performance

	(1)	(2)	(4)	(5)
	Performance	**Performance**	**Goal Level**	**Goal Level**
GOAL*Concave Utility	-0.024	-0.021	0.014	0.015
	(0.021)	(0.021)	(0.027)	(0.026)
Concave Utility	-0.019*	-0.012	-0.013	-0.01
	(0.011)	(0.011)	(0.020)	(0.072)
GOAL	0.073*	0.075*	0.280***	0.280***
	(0.042)	(0.042)	(0.046)	(0.046)
LOPR	-0.016	-0.019		
	(0.034)	(0.034)		
HIPR	-0.020	-0.033		
	(0.034)	(0.034)		
Loss Averse		0.123***		0.139***
		(0.028)		(0.039)
Constant	3.735***	3.657***	3.61***	3.552***
	(0.027)	(0.033)	(0.035)	(0.043)
Log-Likelihood	-841.956	-832.509	-475.665	-469.368
N	160	160	80	80

Note: This table presents the estimates of the Poisson regression of the specification $\text{Performance}_\text{i} = \beta_0 + \beta_1\text{GOAL} * \text{Concave Utility} + \beta_2\text{GOAL} + \beta_3\text{LOPR} + \beta_4\text{LOPR} + \beta_5\text{Loss Averse} + \mathcal{E}_\text{i}$ with $\mathcal{E}_\text{i} \sim \text{Poisson}\ (\omega)$, as well as the Poisson regression of the statistical model $\text{Goal level}_\text{i} = \beta_0 + \beta_1\text{GOAL} * \text{Concave Utility} + \beta_2\text{GOAL} + \beta_3\text{LOPR} + \beta_4\text{LOPR} + \beta_5\text{Loss Averse} + \mathcal{E}_\text{i}$ with $\mathcal{E}_\text{i} \sim \text{Poisson}\ (\omega)$. "Performance" is the number of tables a subject correctly solves over all rounds and "Goal setting" the sum of the goals set by the subject over all rounds. Subjects were assigned either to the GOAL,"HIPR, LOPR, or GOAL+BONUS treatment. The last is the benchmark category for the regression. ``Loss Averse'' is a dummy variable that indicates whether a subject is loss averse or not. A subject is loss averse when at least four of her $\lambda_j$, where $\lambda_j \equiv x_j/z_j$, are greater than one. ``Concave Utility" is a dummy variable that equals 1 if a subject exhibits concave utility and zero otherwise. A subject exhibits concave utility when at least four of her variables Δ_j, where $\Delta_\text{j} \equiv x_j - EV_j$, are less than zero. Standard errors are in parenthesis. * indicates a p-value <0.1, ** indicates a p-value <0.05, ***indicates a p-value <0.01.

Appendix D. Experimental Instructions

D.1 Welcome

This is an experiment in the economics of decision-making. The instructions are simple and if you follow them and make good decisions, you might earn a considerable amount of money, which will be paid to you at the end of the experiment. The amount of payment you receive depends on your decisions, your effort, and partly on luck. The currency used throughout the experiment is Dollars.

Once the experiment has started, no one is allowed to talk to anybody other than the experimenter. Anyone who violates this rule will lose his or her right to participate in this experiment. If you have further questions when reading these instructions please do not hesitate to raise your hand and formulate the question to the experimenter.

The experiment consists of two parts. A **counting task** part and a **decision-making** part. Both tasks will count towards your final earnings.

D.2 Part A: Counting task

This part of the experiment consists of a sequence of **6 rounds** of **5 minutes** each. In each round you need to complete as many tasks as possible. A task is completed when you count the correct number of zeros in a table that contains 100 zeros and ones.

As soon as you know the correct number of zeros contained in a table, you have to input your answer using the keyboard. Once you have entered the number, click "Next". Immediately afterwards a new table will be displayed and, again, you have to count the number of zeros in this new table. This procedure is repeated until the time is up. A timer is displayed in the upper part of your screen. After each round is over you receive feedback about your performance in that round.

Counting Tips: Of course you can count the zeros in any way you want. Speaking from experience, however, it is helpful to always count two zeros at once and multiply the resulting number by two at the end. In addition, you miscount less frequently if you track the number you are currently counting with the mouse cursor.

you will see an example when you press "Next".

(example is displayed)

Payments

(LOPR treatment)

For each correct task that you complete you receive 0.20 Dollars, but if your answer was wrong you receive nothing.

The formula we use to calculate your earnings is Earnings = # correct tasks * 0.20 Dollars.

(HIPR treatment)

For each correct task that you complete you receive 0.50 Dollars, but if your answer was wrong you receive nothing.

The formula we use to calculate your earnings is Earnings = # correct tasks * 0.50 Dollars.

(GOAL treatment)

For each correct task that you complete you receive 0.20 Dollars, but if your answer was wrong you receive nothing.

Goal: At the beginning of each round you will be asked to set a goal in terms of the number of tables you aim at solving correctly in that specific round. Provide this target at this best ability, we would like to see how accurate is your prediction

The formula we use to calculate your earnings is Earnings = # correct tasks * 0.20 Dollars

(GOAL+BONUS treatment)

For each correct task that you complete you receive 0.20 Dollars, but if your answer was wrong you receive nothing.

Goal: At the beginning of each round you will be asked to set a goal in terms of the number of tables you aim at solving correctly in that specific round. If you achieve your goal or you surpass it, you will be paid an additional bonus. The bonus is larger the large the goal is set. If your goal is high and you achieve it or surpass you will be given a high monetary reward. But, if your goal is low and you accomplish it or surpass it you will be given a low monetary reward. Also be aware that if you set a very high goal and you cannot accomplish it you will get no bonus.

The formula we use to calculate your earnings is Earnings = # correct tasks * 0.20 Dollars + goal * 20 cents.

Are you ready to start now? As soon as you press "Next" the task will start.

D.3 Part B: decision-making part

the following, you will face a series of decisions. Your task in each decision is to choose among two possible alternatives. Your earnings on this part of the experiment depend on your choices.

You will be faced with 10 decision sets. Each decision set contains several choices. In each of decision you need to choose between the option R, that delivers a sure amount of money, and the option L that results in one of two different monetary amounts.

Note that in each decision set you need to choose between L and R multiple times. But, you need to be careful since the offered amounts of money could change from one decision to the next.

you will see an example when you press "Next"

Payments

At the end of this part of the experiment **one** randomly chosen decision will be played and paid. Hence, a random number chosen by the program will be drawn to determine which decision counts towards your earnings.

This means that each choice that you make might be chosen and paid.
If it is clear what you have to do in this part of the experiment, press "Next" to start.

D.4 Exit Questionnaire.

This is the last part of the experiment. Please answer the following questions at your best ability.

Enter the computer (Letter plus digit) you are at.
Enter your age.
What is the program you are studying?
Enter your gender.
Male Female
What is your nationality?