Risk Attitudes and Shirking on the Quality of Work under Monitoring: Evidence from a Real-Effort Task Experiment*

Seeun JUNG†

Abstract

This paper studies the effects of risk attitudes on effort exerted under different monitoring schemes. Our design employs a theoretical model that relaxes the assumption that agents are risk neutral and investigates changes in workers' efforts and quality of work as monitoring varies. The predictions of the theoretical model are tested in an original experimental setting under which levels of risk attitudes are measured and monitoring rates vary exogenously. Our results show that shirking behavior decreases with risk aversion, being female, and as monitoring increases. Moreover, monitoring is more effective at curtailing shirking behaviors when subjects are less risk averse.

JEL Classification: C91 ; D61 ; D81 ; D86

Keywords: Shirking; Monitoring; Risk under Uncertainty; Effort; Work Quality

---

*This work was supported by the French National Research Agency through the program Investissements d'Avenir, ANR-10–LABX-93-01. The author is sincerely grateful to Kenneth Houngbedji (PSE) for his helpful support of the experimental design and to Maxim Frolov (Univ. Paris 1) for programming the experiments.

†THEMA/ESSEC - jung@essec.edu

1 Motivation

Firms frequently face the agency dilemma: although a firm's objective is to maximize its profits, its employees seek to maximize their own individual utility. On the firm's side, workers' efforts are needed to increase productivity and the firm's profits. However, effort is costly for workers. Because workers would like to minimize their costs to achieve higher utility, they look to engage in opportunistic behaviors to lower their costs.[1] This type of opportunism is not in firms' interests because such behavior may negatively affect workers' productivity and reduce firm profits. Therefore, employers attempt to curtail opportunistic behaviors with several tools, including monitoring. Whether monitoring works as employers believe is the subject of a long-running debate. Two different theories explain the different directions of the impact of monitoring. The crowding-out theory, which was developed in the psychology literature, suggests that monitoring may reduce overall intrinsic motivation and work effort. Deci (1971), Deci (1975), and Deci and Ryan (1999) argued that when workers feel that they are not trusted or are being controlled, they lose their motivation to work, and economic incentives such as monetary rewards or sanctions are not as effective as hoped.[2] Hence, according to this theory, monitoring would decrease workers' efforts.

Conversely, in the principal-agent problem, workers are rational cheaters who provide less than the optimal level of effort when the marginal benefit of doing so exceeds its cost. Therefore, monitoring motivates agents to raise their effort level to reduce the risk of a penalty if they are caught shirking (Alchian and Demsetz (1972), Calvo and Wellisz (1978), Fama and Jensen (1983), Laffont and Martimort (2002), and Prendergast (1999)). A number of studies (Cadsby et al. (2010), Rickman and Witt (2007), Nagin et al. (2002), Kerkvliet and Sigmund (1999), Bunn et al. (1992), Becker (1968)) have investigated cheating behaviors and observed different levels of effort in response to various environment

[1] Shirking behavior occurs when workers exert less effort than agreed to under their contracts with their employers.

[2] For example, under this theory, paying for blood donations would result in reduced willingness to give blood (Drago and Perlman (1989), Frey and Oberholzer-Gee (1997), Kreps (1997), Gneezy and Rustichini (2000), Bohnet et al. (2000), Frey and Jegen (2000), and Benabou and Tirole (2003)).

such as wage scheme and monitoring. However, an empirical validation of the rational cheater model is difficult to establish outside of an experimental setting. Because cheating (shirking) is difficult to detect, employers have implemented a wide range of schemes to discourage this behavior. Nagin et al. (2002) presented an experimental design that circumvents the empirical challenges listed above and provides empirical evidence in support of the rational cheater model. The field experiment developed in Nagin et al. (2002) in 16 call-center sites was designed to observe the relationship between monitoring and work motivation. The call-center operators were followed for weeks with different monitoring rates. Employing callbacks[3], the authors found that employees were acting as rational cheaters and shirked more as the monitoring rate decreased. More precisely, the number of 'bad calls' responded to the call-back rate. When the monitoring rate increased, the number of bad calls decreased. However, allowing for workers' heterogeneity in various dimensions, those who had 'positive attitudes' toward the firm did not respond to lower monitoring, which might be partly explained by the crowding-out theory. More recently, Dickinson and Villeval (2008) explained the complementarity between the crowding-out theory and the agent problem in a principal-agent experimental setting. Their findings were two-fold: (i) both principals and agents respond to extrinsic incentives, and (ii) intrinsic motivation is crowded out when monitoring exceeds a certain threshold.

In this paper, we follow the rational cheater model but relax the assumption of worker risk neutrality. Attitudes toward risk could explain various economic behaviors, such as job-sorting decisions (Pfeifer (2011); Bonin et al. (2007); Ekelund et al. (2005)) and wages (Pissarides (1974); Murphy et al. (1987); Moore (1995); Hartog et al. (2003); Pannenberg (2007)). In addition, it has been shown that individual risk aversion is often negatively correlated with productivity and wages (Gneezy et al. (2003); Grund and Sliwka (2006); Cornelissen et al. (2011); Dohmen and Falk (2011)). Moreover, risk-averse workers are found to dislike the competitive and stressful work environment that

[3]Callbacks involve monitoring in this setting. Callbacks could catch the 'bad calls' that the operators claimed to be successful but that were, in fact, unsuccessful. Of course, callbacks are costly, and it is impossible to monitor every call that is claimed to have been positive. Thus, 100% monitoring is not efficient. The penalty for shirking (cheating) was dismissal from employment.

is typically associated with higher compensation. With monitoring, would these risk-averse workers be more afraid of getting caught and take monitoring more harshly than others? Our interest lies at this juncture. Because full monitoring is not cost efficient, uncertainty is likely to play a role in any monitoring scheme. In this case, we should be able to identify different behaviors according to risk aversion.

The objective of this paper is to provide a unified theory of the rational cheater model in which the assumption that agents are risk neutral is relaxed. More specifically, we investigate whether the impact of monitoring depends on individual risk aversion. To elucidate our point, we construct an experimental design in which individual risk aversion is measured. We build on Nagin et al. (2002) and exogenously alter the perceived monitoring rate and observe the variation in effort by the agents.

In this paper, we use both performance-based payments and sanctions for bad outcomes to penalize shirking behaviors. Typically, the penalty for shirking is job dismissal in various economic settings. However, in real working conditions, it is difficult to detect shirking behavior and fire workers based on shirking. For this reason, firms can perform quality control as monitoring and make workers take responsibility for low-quality work by means of sanctions. We define the term 'shirking' as a decrease in work quality. Following Lazear (1995) and Lazear (2000), who argued that work quality decreases with performance-based pay as agents attempt to earn higher wages by producing higher quantities[4], we use performance-based payment to measure the direct impact of monitoring (control) on effort (quality) and shirking (increases in quantity).

The contributions of this paper are as follows. This paper is the first attempt to introduce risk aversion parameters into a utility model involving shirking and effort provision in work. Our experiment is original in its design because we observe the shirking behavior of individuals along with the levels of their risk aversion. Our results contribute to the literature by addressing the effectiveness of monitoring as a necessary condition for preventing workers from shirking and by investigating the heterogenous impact of

[4]For example, a typist who is paid by the number of words makes a greater number of errors.

monitoring on individuals with different risk attitudes. Thus, monitoring can be more effective for less risk-averse workers. In addition, this paper will help develop a better understanding of why risk-averse workers typically occupy low-status categories in labor markets (low wage, low productivity, etc.)

The remainder of the paper is organized as follows. The conceptual framework will be explained in Section 2. Section 3 will organize the experimental design and the relevant model and provide the simulation results. Finally, we will discuss the results of the experiment in Section 4 and conclude in Section 5.

2 Experiment

We conduct a controlled experiment in which subjects cannot select themselves into one group or another with respect to their risk aversion. Subjects are randomly allocated to a group and undertake the tasks assigned to their group. Therefore, the parameter of risk aversion is exogenous to the monitoring rate.[5]

The experiment involves solving tasks in a limited time period. Each task consists of counting even numbers in a 15-digit code to calculate the sum and comparing this figure to a given number, k.[6] When the sum is equal to k, the participant reports 'true'; otherwise, the participant reports 'false'. Because we make k close to the correct answer (i.e., using a normal distribution with small variance), participants can learn that the given numbers can be near the correct answer from the training session, which may tempt the participants to click 'yes' by guessing (i.e., shirking) instead of by calculating. Finally, we used only even numbers as the possible answer for each code because we asked

[5] All experiments were computerized using the REGATE software designed by Zeilliger (2000), and the program was set up by Maxim Frolov from the Centre d'Economie de la Sorbonne of the University of Paris 1. The experiments were conducted at the Laboratoire d'Economie Experimentale de Paris (L.E.E.P) of the Paris School of Economics.

[6] The code is generated by simulating n repeated Bernoulli draws of even and odd numbers with a probability, p, of obtaining an even number. The level of difficulty depends on the parameter, p, used to generate the code. The smaller that p is, the easier the item will be to calculate because there will be fewer even numbers to count in a code. The random number k is a code generated following the normal distribution with the mean of the sum.

the participants to add only the even numbers presented in a 15-digit code.

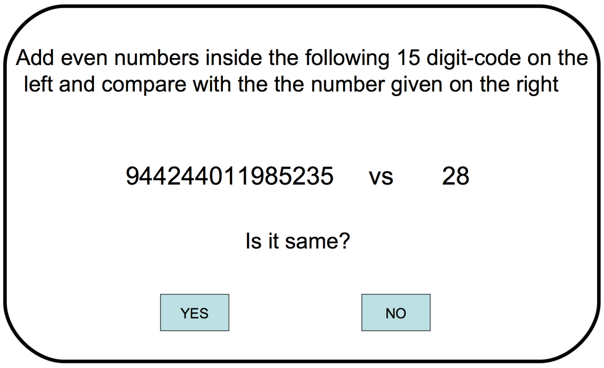


Figure 1: Example of Task Screen

Because these exercises require no particular skill, we believe that no participant was disadvantaged.[7] The participants were asked to solve codes with a computerized program and were paid according to their performance. Before the experiment, the instructions for the experiment were given to the participants. We asked them to imagine that they were working for a cryptography firm where the profit is based upon the number of codes solved correctly and that their pay is based on performance. Beginning with a training session, we gave the participants four sessions and assigned them different monitoring rates. Each session lasted five minutes.

Training Task: Participants are asked to decipher codes correctly in five minutes and are paid a flat rate of five euros. Nothing is reviewed.

Treatment Task 1: Participants are asked to decipher codes correctly in five minutes. They are also told that all answers will be checked and that they will be paid according to correct answers (ten cents per correct code) with fines for wrong answers (five cents is deducted from their income per wrong code).

Treatment Task 2: Participants are asked to decipher codes correctly in five minutes. They are told that they will be paid according to the number of answers (ten cents

[7]It is possible that students who experience difficulty reading might be disadvantaged. We control for that by asking the participants to report whether they have been diagnosed with vision problems.

per code); there is a 60% chance that their answers will be reviewed, in which case a fine of five cents will be imposed for each wrong answer. Participants are told to play a lottery after the task to determine whether their answers will be reviewed.

Treatment Task 3: Participants are asked to decipher codes correctly in five minutes. They are told that they will be paid according to the number of answers (ten cents per code); there is a 20% chance that their answers will be reviewed, in which case a fine of five cents will be imposed for each wrong answer. Participants are told to play a lottery after the task to determine whether their answers will be reviewed.

In this experiment, we sampled 107 volunteers who were paid ten cents per code deciphered with a flat wage of five euros for showing up and completing the training task. Then, we imposed five-cent penalties for poor work quality (e.g., each code solved incorrectly) only when the work is monitored. Because we also wanted to control for heterogeneity in attitudes toward risk, the subjects took a test measuring individual risk aversion. To avoid learning bias and ordering effects, we randomized the order of tasks 1, 2, and 3 for each subject.

3 Model: Shirking Behavior with Monitoring

In this section, we derive a set of predictions based on the experimental setting. We assume that the number of codes, x, successfully deciphered by a participant depends on the effort supplied, E. Thus, a participant engages in shirking behavior whenever she deciphers a code by guessing instead of by calculating. When engaging in shirking behavior by guessing (and not calculating) whether the sum corresponds to the given number, the probability that the guessed answer is right is p, which is positive and less than 1 ($0 < p < 1$). To simplify the model, we assume that if the participant applies enough effort, she obtains the correct answer because addition is a reasonably easy task. We therefore assume a linear relationship between effort/shirking and performance. The performance function is defined as

$$x = E + pS + \epsilon, \tag{1}$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$. In our case, effort E corresponds to the number of codes for which the subject provided the full desired effort, and we define S as shirking behavior (i.e., the number of codes that the subject answered by guessing). The subjects are offered a linear contract of the form $w = bx$. Thus, based on the foregoing, when an incorrect outcome is discovered by monitoring, a penalty of a per code will be charged. We assume that supplying effort to solve the codes is costly; it takes longer to provide E than S. In addition, providing effort involves a disutility, and we construct a linear cost function for each code on which a participant makes an effort ($c(E) = cE$) as a money equivalent. We consider shirking to involve shirking on quality. When an individual shirks, the quality of her work decreases (i.e., the rate of solving codes correctly is lower).

We have two possible states of wealth that depend on the monitoring rate, M, and on how the subject allocates effort and determines shirking behavior (E and S).

- shirking is detected with probability M, and the subject receives a wage according to the real outcome with penalties on the wrong outcomes: $b(E + pS + \epsilon) - a(1 -$

$p)S - cE$. This utility is $U[b(E + pS + \epsilon) - a(1 - p)S - cE]$

- shirking is not detected with probability $1 - M$, and the subject receives the full wage: $b(E + S + \epsilon) - cE$. This utility is $U[b(E + S + \epsilon) - cE]$

The agent chooses the level of effort and shirking by maximizing her expected utility.

$$\max_{E,S} \ \mathbb{E}U[E, S|M]$$

$$\Rightarrow \max_{E,S} M\{\mathbb{E}U[b(E + pS + \epsilon) - a(1 - p)S - cE]\} + (1 - M)\{\mathbb{E}U[b(E + S + \epsilon) - cE]\}$$

$$s.t. \begin{cases} T \geq v_E E + v_S S \\ E \geq 0 \ and \ S \geq 0 \end{cases} \tag{2}$$

T is the total time that the subject can use in one task, and v_E and v_S are the time needed for full effort and shirking, respectively. In other words, providing effort is more costly in terms of time constraints. Using the negative exponential utility function with risk-aversion parameter, $R(> 0)$ $(U(w) = -e^{-Rw})$, we set the Lagrangian function as

$$\begin{aligned} \mathcal{L} \equiv & - M\mathbb{E}e^{-R\{b(E+pS+\epsilon)-a(1-p)S-cE\}} - (1 - M)\mathbb{E}e^{-R\{b(E+S+\epsilon)-cE\}} + \lambda(T - v_E E - v_S S) \\ \Rightarrow \mathcal{L} \equiv & - Me^{-R\{b(E+pS)-a(1-p)S-cE\}}\mathbb{E}e^{-Rb\epsilon} - (1 - M)e^{-R\{b(E+S)-cE\}}\mathbb{E}e^{-Rb\epsilon} \\ & + \lambda(T - v_E E - v_S S) \end{aligned} \tag{3}$$

Using properties of the mean of log-normal random variable: if $log\ x \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $\mathbb{E}(x) = e^{\mu_x^2 + \sigma_x^2/2})$, we can derive the following function:

$$\mathcal{L} \equiv -M\mathbb{E}e^{-R\{b(E+pS)-a(1-p)S-cE\}+\frac{R^2b^2\sigma^2}{2}} - (1-M)\mathbb{E}e^{-R\{b(E+S)-cE\}+\frac{R^2b^2\sigma^2}{2}} + \lambda(T - v_E E - V_S S) \tag{4}$$

Then, we can solve for the first-order conditions with respect to E, S, and λ in the case for positive E and S.

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial E} = 0 \Rightarrow & (b-c)RMe^{-R\{b(E+pS)-a(1-p)S-cE\}+\frac{R^2b^2\sigma^2}{2}} + (b-c)R(1-M)e^{-R\{b(E+S)-cE\}+\frac{R^2b^2\sigma^2}{2}} = \lambda v_E \\
\frac{\partial \mathcal{L}}{\partial S} = 0 \Rightarrow & RM\{bp-a(1-p)\}e^{-R\{b(E+pS)-a(1-p)S-cE\}+\frac{R^2b^2\sigma^2}{2}} + bR(1-M)e^{-R\{b(E+S)-cE\}+\frac{R^2b^2\sigma^2}{2}} = \lambda v_S \\
\frac{\partial \mathcal{L}}{\partial \lambda} = 0 \Rightarrow & T = v_e E + v_s S
\end{aligned} \tag{5}$$

Taking the ratio of the first two conditions yields

$$\begin{aligned}
\frac{(b-c)RMe^{-R\{b(E+pS)-a(1-p)S-cE\}+\frac{R^2b^2\sigma^2}{2}} + (b-c)R(1-M)e^{-R\{b(E+S)-cE\}+\frac{R^2b^2\sigma^2}{2}}}{RM\{bp-a(1-p)\}e^{-R\{b(E+pS)-a(1-p)S-cE\}+\frac{R^2b^2\sigma^2}{2}} + bR(1-M)e^{-R\{b(E+S)-cE\}+\frac{R^2b^2\sigma^2}{2}}} &= \frac{v_E}{v_S} \\
\Rightarrow e^{-R\{b(E+pS)-a(1-p)S-cE\}+\frac{R^2b^2\sigma^2}{2}} RM[(b-c)v_S - v_E\{bp-a(1-p)\}] & \\
= e^{-R\{b(E+S)-cE\}+\frac{R^2b^2\sigma^2}{2}} R(1-M)(bv_E-(b-c)v_S) & \\
\Rightarrow e^{SR(a+b)(1-p)} = \frac{1-M}{M}\frac{bv_E-(b-c)v_S}{(b-c)v_S - v_E\{bp-a(1-p)\}} &
\end{aligned} \tag{6}$$

We can solve this by taking natural logarithm. Then, finally we can obtain the optimal shirking as follows.

$$S = \frac{\ln(1-M) - \ln(M) + \ln(bv_E-(b-c)v_S) - \ln((b-c)v_S - v_E\{bp-a(1-p)\})}{R(a+b)(1-p)} \tag{7}$$

With respect to the case with negative S, applying the Kuhn-Tucker conditions indicates that optimal shirking is 0. Therefore, optimal shirking is as follows:

$$S^* = \begin{cases} \frac{\ln(1-M) - \ln(M) + \ln(bv_E-(b-c)v_S) - \ln((b-c)v_S - v_E\{bp-a(1-p)\})}{R(a+b)(1-p)} & \text{if } S > 0 \\ 0 & \text{if } S \leq 0 \end{cases} \tag{8}$$

Thus, the optimal shirking level falls with the monitoring rate:

$$\frac{\partial S^*}{\partial M} = \frac{-\frac{1}{1-M} - \frac{1}{M}}{R(a+b)(1-p)} < 0 \tag{9}$$

The relationship between risk aversion and optimal shirking is more complex, and more conditions must be applied to check the sign.

$$\frac{\partial S^*}{\partial R} = -\frac{\ln(1-M) - \ln(M) + \ln(bv_E - (b-c)v_S)) - \ln((b-c)v_S - v_E\{bp - a(1-p)\})}{R^2(a+b)(1-p)} \tag{10}$$

$\partial S^*/\partial R$ is negative when $\ln(1-M) - \ln(M) + \ln(bv_E - (b-c)v_S)) - \ln((b-c)v_S - v_E\{bp - a(1-p)\}) > 0$. However, the cross derivative of risk aversion and the monitoring rate on shirking behaviors becomes positive, which implies that shirking falls with the monitoring rate, but the slope is flatter for risk-averse agents. It is more straightforward to look at the simulated graph.

$$\frac{\partial S^{*2}}{\partial R \partial M} = \left(\frac{1}{1-M} + \frac{1}{M}\right)\frac{1}{R^2}\frac{1}{(a+b)(1-p)} > 0 \tag{11}$$

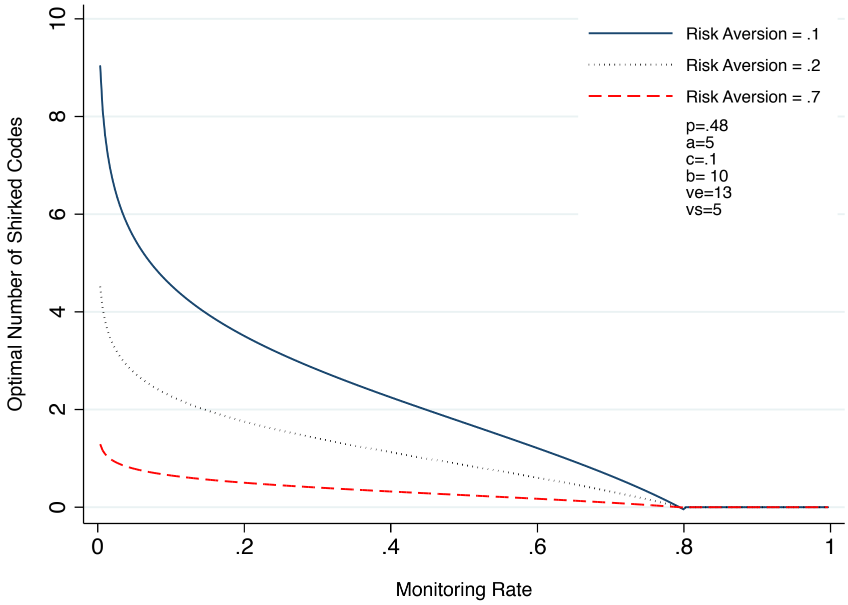


Figure 2: Optimal Shirking with Monitoring and Penalty

Therefore, we ran a simulation of our experimental environment. Figure 2 shows the simulation results for optimal shirking, which varies with the monitoring rate for different levels of risk aversion with the penalty. This experiment demonstrates that optimal shirking decreases as the subject is more risk averse. In addition, as the monitoring rate rises, optimal shirking decreases. It also shows that less risk-averse subjects are more sensitive to the monitoring rate increase at a lower monitoring rate, whereas the slopes of more risk-averse subjects are rather flat and not much affected by change in the monitoring rate. More risk-averse agents always have less incentive to shirk at any level of monitoring, whereas less risk-averse agents shirk significantly more when monitoring is low enough.

After the monitoring rate attains a certain level (80%), no one intends to shirk as optimal shirking falls towards zero. We may infer that full monitoring is not necessary under this setting, particularly when monitoring is associated with a penalty. If enough monitoring is exerted that penalizes shirking behavior, workers will not shirk. In this representation, the graph shows that the slope of the risk-averse agent is flatter than that of less risk-averse agents, beginning at the lower level of the shirking proportion. Thus, we seek to test for the following: (i) whether shirking falls with monitoring ($\frac{\partial S^*}{\partial M} < 0$), (ii) whether more risk-averse individuals shirk less ($\frac{\partial S^*}{\partial R} < 0$) under monitoring, and therefore (iii) whether more risk-averse individuals have a flatter shirking slope with monitoring ($\frac{\partial^2 S^*}{\partial R \partial M} > 0$). To test these three hypotheses, we construct an empirical model that we discuss in the following section.

4 Data

Table 1 presents the descriptive statistics of the samples we used for the experiment. Including the pilot session, we used 107 students. The average age of our participants was approximately 23 (the lab belongs to the University of Paris 1); approximately half of the subjects were female and half were religious. In addition to socio-demographic information, we gathered data on individual risk aversion by having the participants play an incentivized lottery game which was designed by Holt and Laury (2002). Following Andersen et al. (2006), we employ a switching multiple price list design, which therefore allows only monotonic risk preference. As presented in Figure 3, participants have ten choices between two options, A and B. Option A pays either €2 or €1.6 , whereas Option B pays either €3.85 or €0.10. Except with respect to the first choice (in which there is no uncertainty), option B is riskier and features higher and lower possible payments. As the decision number rises, the probability of receiving the higher amount (€2 for Option A and €3.85 for Option B) falls. In other words, the probability of receiving the lower amount increases. A rational participant would switch to Option A from Option B at a certain decision number, and that number varies with individual risk aversion. For example, a rational risk-neutral agent would switch his preferred option from B to A between decision numbers 5 and 6. For the final payment, one row is randomly chosen, and the lottery is then played according to the choice made by the participant. In this paper, we use the participant's numerical choice for Option A as our measure of risk aversion, which measure increases as the participant grows more risk averse. In addition, the payment from the lottery varies between €0.10 and €3.85, which is similar to the variance in payment for each task. The risk preference of this lottery should, therefore, be relevant to the preference structure that we use in our experimental setting.

Figure 3: Holt and Laury Type Lottery Game

Table 1: Descriptive Statistics of Subjects.

	Obs	Min	Mean	Median	Max
Age (years)	107	19	23	23	34
Woman	107	0	.54	1	1
Religious	107	0	0.48	0	1
Education level (highest achieved)					
- High School	107	0	.52	1	1
- Undergraduate	107	0	.46	0	1
- Graduate	107	0	.2	0	1
Holt and Laury Type Incentivized Lottery Choice					
The number of safe options	107	0	6.1	6	10
- Men	49	0	5.9	6	10
- Women	58	0	6.3	6	10
Behavior-related questions					
The subject has used					
- Tobacco[†]	107	1	2.10	1	4
- Alcohol[†]	107	1	2.82	3	4
Attitudes towards Task					
Winning lottery (=0) vs. Losing money (=1)[a]	107	0	0.39	0	1

[a] "Did you consider the task an opportunity to win money or a situation in which you risked losing money?"

[†] Indicates that the variable is discrete and takes the values of 1 for "No", 2 for "Rather not", 3 for "Sometimes" and 4 for "Yes".

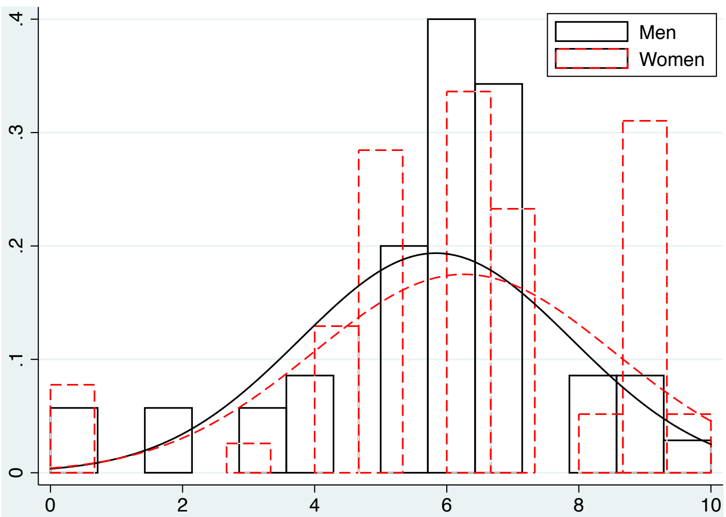


Figure 4: Risk Aversion Distribution by Gender

Figure 4 depicts the distribution of risk aversion (i.e., the number of safe options chosen) across genders. On average, women pick 0.4 safe options more than men do. The distribution of women has a slightly thicker tail, indicating greater variation in risk attitudes. We also asked certain behavior-related questions, such as questions involving smoking and drinking. These questions helped us validate the risk-aversion measure by significant correlation among various questions.

Table 2: Pairwise Correlation Matrix.

	Risk Attitudes
Being woman (=1)	0.12*
Smoking (=1)	-0.24*
Drinking Alcohol (=1)	-0.25*
Being Religious (=1)	0.22*
Education	-0.08*
Task[b]	0.11*

[a] * 1% significance.
[b] "Did you consider the task an opportunity to win money or a situation in which you risked losing money?"

Table 2 shows the correlation between the risk attitudes measured and related questions. In this paper, we use the Holt and Laury type of risk aversion, which is consistently correlated with the other questions: women and religious people are more risk averse, and risk-averse people smoke and drink less. In addition, the more educated a person is, the less risk averse. At the end of the experiment, we asked subjects whether they perceived the tasks they undertook as involving the possibility of winning money or losing money. Risk-averse subjects tended to perceive these tasks as losing money, which makes sense because risk-averse agents take the possibility of losing more seriously.

Figure 5 depicts the average time (in seconds) spent on each item. As the number of items reviewed increases, subjects tend to spend less time on each code, which might be due to the loss of concentration over time or to trying to go over more codes as time becomes more pressing. Compared with full monitoring (task1), subjects take less time for each code when there is only a 20% chance that there work will be monitored (task3).

Figure 6 shows the graph for the number of subjects on each item. It is clear that with 20% monitoring, more people reach higher item numbers compared with more intensive monitoring.

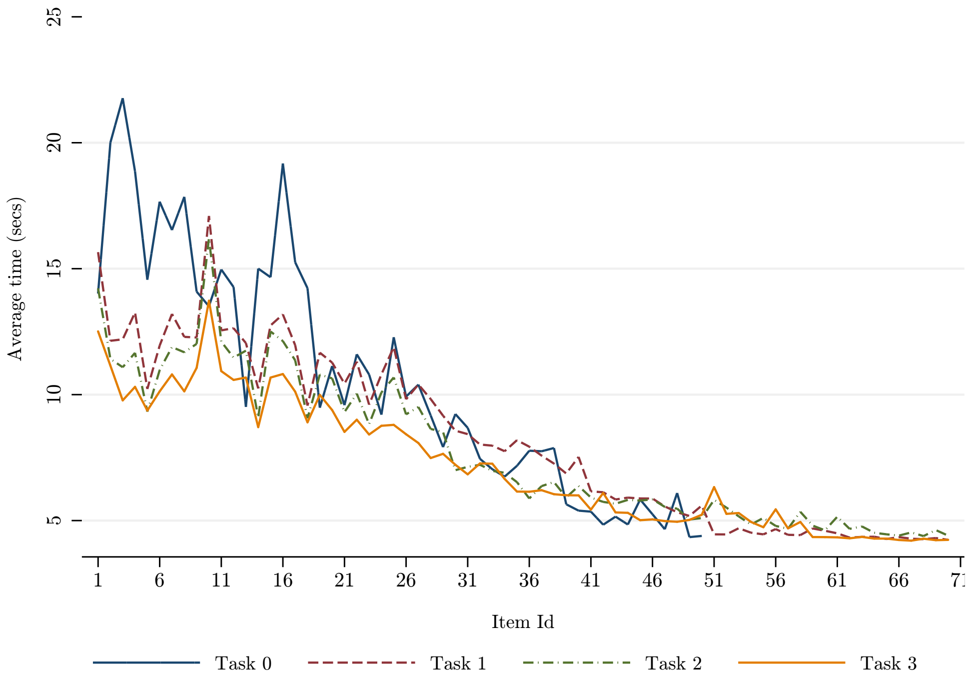


Figure 5: Average Time Spent on each Item Across Tasks

Table 3: Descriptive Statistics of Variables Measuring Subjects' Performance.

	Obs	Min	Mean	Median	Max
Number of items reviewed	428	6	25.42	23	50
Number of items reviewed in less than 5s.	428	0	5.34	0	50
Time spent on an item (secs)	13,359	4	12.85	11.36	76.61
Item is correctly reviewed (=1)	13,359	0	.81	1	1
Item is reviewed in less than 5s (shirking) (=1)	13,359	0	.17	0	1
Item difficulty[a]	13,359	4	30.06	30	74
Clicked yes (=1)[b]	13,359	0	0.4	0	1
Payment received at the end:					
- Task (€)	321	.45	2.4	2.1	5
- Experiment (€)	107	10	14.8	15	22

[a] The difficulty of each item is the sum of all the numbers in the 15 digits.
[b] The answer is clicked on 'Yes'.

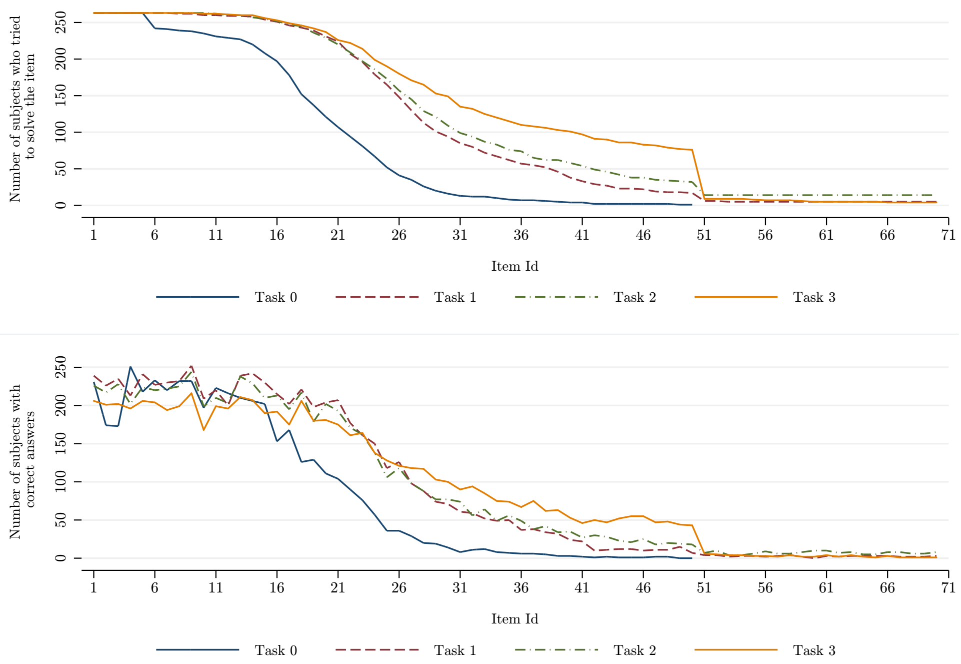


Figure 6: Performance by Item Across Tasks

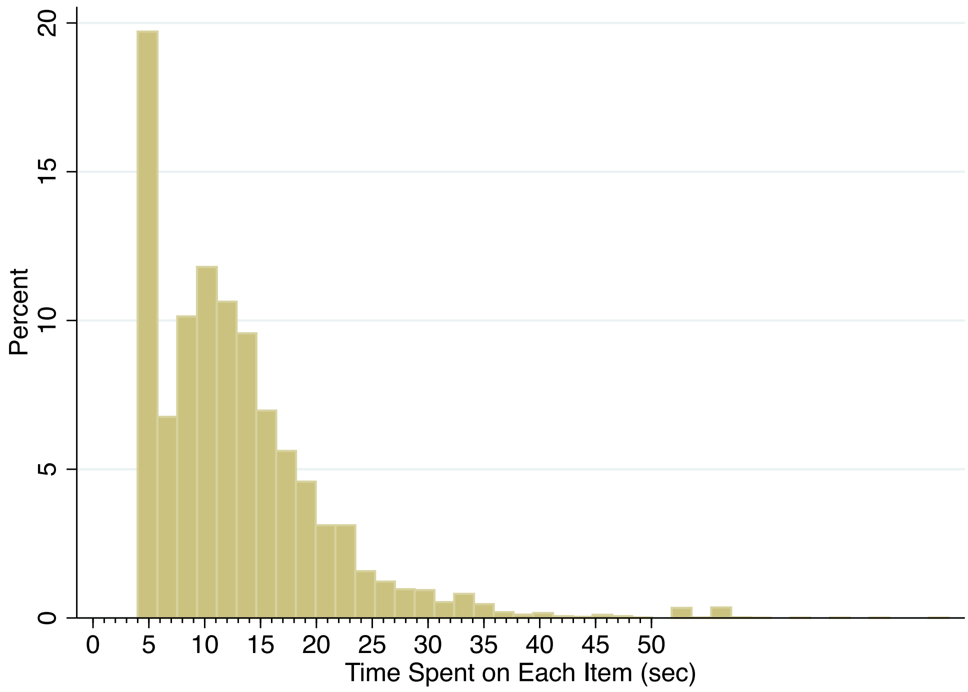


Figure 7: Distribution of Time Spent (sec.)

Table 3 shows the descriptive statistics of the variables measuring subjects' performance. We have information on the number of items reviewed by the subjects for each task. On average, subjects solved 25 items per task. If the time spent was under 5 seconds, we assume that the code was deciphered by shirking[8]. Indeed, there is a jump in the frequencies of items solved within 4-5 seconds found in Figure 7. In the experimental design, we set 4 seconds for the answer buttons to appear to allow participants to decide whether to shirk or to make an effort to solve the code, depending on its difficulty. We expect that when the code seems more difficult with higher numbers to calculate, the probability that the code is answered in five seconds (shirking) and the time spent determining that answer would increase. In addition, this 4-second rule could prevent the temptation to click continuously and decrease the incentive to shirk by peer effect (i.e., hearing that a neighbor is clicking continuously or shirking). Using a normal distribution with small variance, we made the given number close to the correct answer that participants would arrive at through their calculations; thus, participants might have learned in the training session that the given numbers may be close the correct answer, which may have tempted participants to click 'yes' by guessing (i.e., shirking). The correlation coefficient between the shirking item and whether the answer is 'yes' is positive and significant ($R = 0.07$ significant at the 1% level) in Table 5: thus, when participants shirk, they tend to answer 'yes' more often.

Table 4 is the correlation matrix of risk aversion and the monitoring rate with the performance outcomes at the task level. More risk-averse subjects have fewer items that we assume to be guessed by shirking and fewer items reviewed in each task. They also have more correct items. Monitoring works in the same way as risk aversion. As monitoring becomes more intensive, there is less shirking and fewer items solved as the participants become more careful when solving codes, and the rate of correct answers increases.

Table 5 is the correlation matrix at the item level. Similar to the task level, when agents are risk averse, they tend to shirk less and are more likely to answer correctly on

[8]This calculation is a proxy for shirking behavior. When we perform the calculation, it is difficult to solve the code within 6 seconds, but we nonetheless see codes deciphered within 5 seconds.

Table 4: Pairwise Correlation Matrix Task Level.

	Risk Attitudes	Monitoring
Correct Rate[b]	0.08*	0.37*
# of Clicks[c]	-0.03*	-0.33*
# of Correct Answers[d]	0.06*	-0.09*
# of Wrong Answers[e]	-0.08*	-0.40*
# of Answered Shirked[f]	-0.09*	-0.37*

[a] ⋆ 1% significance.
[b] % of the ratio: the number of correct answers/the number of clicks.
[c] The total number of clicks.
[d] The number of correct answers.
[e] The total number of wrong answers.
[f] The number of codes shirked (clicked within five seconds).

Table 5: Pairwise Correlation Matrix Item Level.

	Risk Attitudes	Difficulty[f]	Clicked 'Yes'	Shirked
Clicked 'Yes'[b]	-0.03*	-0.03*		
Shirked[c]	-0.07*	0.04*	0.07*	
Time Spent[d]	-0.003	0.28*	-0.13*	-0.56*
Correct[e]	0.04*	-0.10*	0.03*	-0.35*

[a] ⋆ 1% significance.
[b] The answer clicked is 'yes'.
[c] The item is solved within five seconds.
[d] Total amount of time (sec.) spent to click.
[e] The code is correctly solved.
[f] The difficulty of each item is the sum of all the numbers in the 15 digits.

each item.

Table 6: Subjects' Performance Across Tasks.

	Task 0[†]	Task 1	Task 2	Task 3
Task level				
Number of items reviewed	18.91	25.77	30.35	35.55
	(7.42)	(7.32)	(11.93)	(13.45)
Number of items correctly reviewed	15.75	22.42	23.13	23.66
	(5.33)	(5.75)	(6.18)	(5.63)
Number of items reviewed in less than 5s	1.10	1.86	9.65	17.71
	(4.85)	(6.74)	(16.95)	(20.30)
Payment received (€)	.	2.32	2.59	2.69
	(.)	(1.03)	(1.18)	(1.23)
Observations	107	107	107	107
Item level				
Time spent on an item (secs)	18.54	12.80	11.45	9.86
	(10.33)	(5.97)	(6.33)	(6.40)
Item correctly reviewed	0.84	0.89	0.81	0.72
	(0.37)	(0.31)	(0.39)	(0.45)
Item reviewed in less than 5s	0.02	0.04	0.20	0.36
	(0.15)	(0.21)	(0.40)	(0.48)
Observations	2904	3108	3507	3840

Standard errors are reported in parentheses.

[†] Task 0 is the practice session. Participants get paid five euros as fixed rate.

Table 6 shows the subjects' performance across tasks. Monitoring rates in Tasks 1, 2, and 3 are 100%, 60%, and 20%, respectively. At the task level, the table shows that both the number of items reviewed and the number of shirked items rise as the monitoring rate decreases. At the item level, as the monitoring rate decreases, (i) the subjects spend less time on each item, (ii) the success rate for each item decreases, and (iii) the items are more likely to be reviewed by guessing. Thus, the intensity of monitoring indeed affects workers' behaviors. Now, we will investigate whether this impact can vary over risk aversion.

4.1 Analyses

We want to estimate $\frac{\partial S}{\partial M}, \frac{\partial S}{\partial R}$ when $M > 0$, and therefore the cross derivative $\frac{\partial^2 S}{\partial M \partial R}$. The following specification is estimated:

$$\mathbf{S_{it}} = \beta_0 + \beta_1 \mathbf{M_{60,t}} + \beta_2 \mathbf{M_{20,t}} + \beta_3 \mathbf{R_i} + \beta_4 \mathbf{M_{60,t}} \times \mathbf{R_i} + \beta_5 \mathbf{M_{20,t}} \times \mathbf{R_i} + \beta_6 \mathbf{X_i} + u_i + \eta_{it}$$

where S represents various proxies for shirking behavior, such as the number of codes shirked, the number of codes answered, whether the code is shirked, and the time spent to solve each code; and $M_{60,t}$ ($M_{20,t}$) is a dummy if the monitoring rate was 60% (20%) at treatment t. We have three treatments with different monitoring rates: $t = 1, 2, 3$. R_i is the Holt and Laury risk-aversion measure. To allow for different slopes for risk-averse individuals, we simply add interaction terms between the risk aversion measures, R_i and $M_{60,t}$ ($M_{20,t}$). We also control for basic socio-demographic variables X such as gender, age, and education. Because we use categorical dummy variables for monitoring rates, our reference points are (i) 100% for monitoring ($M_{100,t}$) and (ii) 100% x risk aversion for the interaction between monitoring and risk aversion ($M_{100,t} \times R_i$). We are interested in finding β_1 and β_2 for the impact of monitoring reductions on shirking behaviors, and β_4 and β_5 for capturing heterogeneity due to individual risk aversion. Because our reference point is perfect monitoring (100%), we expect that β_1 and β_2 are positive as monitoring falls, but that β_4 and β_5 are negative because risk-averse individuals won't increase their shirking behavior as much as others. Monitoring is completely exogenous in our experimental setting:

$$\mathbb{E}[\mathbf{M_t} \times \eta_{it} \mid \mathbf{x}_i] = 0$$

However, we might be concerned that the attitude toward risk is correlated with unobservables such as calculation ability or cognitive skills for computer work.

$$\mathbb{E}[\mathbf{R_i} \times \eta_{it} \mid \mathbf{x}_i] \neq 0$$

Table 7: Task Level, Random Effect.

	(1) % Correct[b]	(2) # Clicks[c]	(3) # Correct[d]	(4) # Shirk[e]	(5) Gains	(6) # Wrong[f]
Risk Aversion	-0.002	0.740	0.569**	0.181	-0.004	0.171
	(0.01)	(0.46)	(0.25)	(0.61)	(0.05)	(0.32)
Monitoring Rate 60%	-0.078*	6.237*	2.362	8.465*	0.182	3.875
	(0.05)	(3.27)	(1.46)	(4.97)	(0.38)	(2.46)
Monitoring Rate 20%	-0.190***	13.138***	3.224**	14.628***	0.363	9.914***
	(0.05)	(3.27)	(1.46)	(4.97)	(0.38)	(2.46)
RA x Monitoring 60%	0.004	-0.497	-0.289	-0.548	0.020	-0.207
	(0.01)	(0.51)	(0.23)	(0.77)	(0.06)	(0.38)
RA x Monitoring 20%	0.011	-0.903*	-0.334	-0.487	0.003	-0.568
	(0.01)	(0.51)	(0.23)	(0.77)	(0.06)	(0.38)
Subject is a woman	0.041*	-6.057***	-3.198***	-6.714***	-0.526***	-2.859***
	(0.02)	(1.52)	(0.94)	(1.82)	(0.13)	(0.99)
Age	0.001	-0.598**	-0.392**	-0.729**	-0.050*	-0.206
	(0.00)	(0.29)	(0.18)	(0.35)	(0.03)	(0.19)
Highest Diploma	0.033	1.287	1.747*	1.205	0.183	-0.460
	(0.02)	(1.67)	(1.04)	(1.99)	(0.15)	(1.09)
Constant	0.613***	48.178***	29.264***	33.565***	3.722***	18.914***
	(0.09)	(6.43)	(3.90)	(7.87)	(0.59)	(4.24)
r2_w	0.221	0.196	0.038	0.188	0.045	0.231
r2_b	0.069	0.165	0.144	0.160	0.158	0.103
r2_o	0.141	0.179	0.117	0.176	0.098	0.172
N	321	321	321	321	321	321

[a] * 10%, ** 5%, and *** 1% significance.
[b] The proportion of correct answers out of total clicks.
[c] The number of clicks.
[d] The number of correct answers.
[e] The number of codes clicked within five seconds, i.e., the number of codes shirked.
[f] The number of wrong answers.
[g] Standard errors are clustered at the individual level.

We will thus estimate the model first with a random effect and then a fixed effect.

Table 7 is the random-effect specification at the task level, and Table 8 is the specification at the item level. Our dependent variables include the correct solution rate (the total number of correct answers out of the total number of codes answered), the total number of codes answered (the total number of clicks), the number of correct answers, the number of codes shirked (the total number of codes that were answered in five seconds or less), the gains (payment), and the number of wrong answers. The results presented in Table 7 demonstrate the impact of risk aversion on shirking behaviors. Using the interaction terms of risk aversion and task numbers with different monitoring rates, we allow for different slopes for risk aversion: these are our cross-derivative estimates. In this table,

Table 8: Item Level, Random Effect.

	(1) Shirk[c]	(2) Time[d]	(3) Correct[e]
Risk Aversion	0.005	-0.443**	-0.004
	(0.01)	(0.17)	(0.01)
Monitoring Rate 60%	0.197***	-2.406***	-0.120***
	(0.02)	(0.35)	(0.03)
Monitoring Rate 20%	0.329***	-5.753***	-0.262***
	(0.02)	(0.35)	(0.03)
RA x Monitoring 60%	-0.012***	0.200***	0.009**
	(0.00)	(0.05)	(0.00)
RA x Monitoring 20%	-0.008**	0.468***	0.019***
	(0.00)	(0.05)	(0.00)
Subject is a woman	-0.150***	2.114***	0.061**
	(0.04)	(0.75)	(0.03)
Age	-0.017**	0.268*	-0.004
	(0.01)	(0.14)	(0.00)
Highest Diploma	0.030	-1.142	0.057**
	(0.05)	(0.82)	(0.03)
Item Difficulty	0.001***	0.158***	-0.003***
	(0.00)	(0.00)	(0.00)
Constant	0.726***	-0.053	0.726***
	(0.17)	(3.02)	(0.10)
r2_w	0.123	0.185	0.037
r2_b	0.212	0.104	0.141
r2_o	0.163	0.160	0.052
N	10,455	10,455	10,455

[a] * 10%, ** 5%, and *** 1% significance.
[b] Standard errors are clustered at the individual level.
[c] The item is solved within five seconds as a result of shirking.
[d] Time spent on the item (seconds).
[e] The answer is correct (=1).

Table 9: Task Level, Fixed Effect.

	(1) % Correct[b]	(2) # Clicks[c]	(3) # Correct[d]	(4) # Shirk[e]	(5) Gains	(6) # Wrong[f]
Monitoring Rate 60%	-0.078*	6.237*	2.362	8.465*	0.182	3.875
	(0.05)	(3.27)	(1.46)	(4.97)	(0.38)	(2.46)
Monitoring Rate 20%	-0.190***	13.138***	3.224**	14.628***	0.363	9.914***
	(0.05)	(3.27)	(1.46)	(4.97)	(0.38)	(2.46)
RA x Monitoring 60%	0.004	-0.497	-0.289	-0.548	0.020	-0.207
	(0.01)	(0.51)	(0.23)	(0.77)	(0.06)	(0.38)
RA x Monitoring 20%	0.011	-0.903*	-0.334	-0.487	0.003	-0.568
	(0.01)	(0.51)	(0.23)	(0.77)	(0.06)	(0.38)
Constant	0.884***	24.308***	21.290***	1.299	2.157***	3.019***
	(0.01)	(0.79)	(0.35)	(1.19)	(0.09)	(0.59)
r2_w	0.221	0.196	0.038	0.188	0.045	0.231
r2_b	0.004	0.000	0.009	0.008	0.002	0.004
r2_o	0.091	0.081	0.015	0.102	0.023	0.110
N	321	321	321	321	321	321

[a] * 10%, ** 5%, and *** 1% significance.
[b] The proportion of correct answers out of total clicks.
[c] The number of clicks.
[d] The number of right answers.
[e] The number of codes clicked within five seconds, i.e., the number of codes shirked.
[f] The number of wrong answers.
[g] Standard errors are clustered at the individual level.

Table 10: Item Level, Fixed Effect.

	(1) Shirk[c]	(2) Time[d]	(3) Correct[e]
Monitoring Rate 60%	0.196***	-2.381***	-0.117***
	(0.02)	(0.35)	(0.03)
Monitoring Rate 20%	0.328***	-5.739***	-0.257***
	(0.02)	(0.35)	(0.03)
RA x Monitoring 60%	-0.012***	0.197***	0.009**
	(0.00)	(0.05)	(0.00)
RA x Monitoring 20%	-0.008**	0.466***	0.018***
	(0.00)	(0.05)	(0.00)
Item Difficulty	0.001***	0.158***	-0.003***
	(0.00)	(0.00)	(0.00)
Constant	0.036***	7.977***	0.977***
	(0.01)	(0.14)	(0.01)
r2_w	0.123	0.185	0.037
r2_b	0.135	0.008	0.015
r2_o	0.101	0.115	0.034
N	10,455	10,455	10,455

[a] * 10%, ** 5%, and *** 1% significance.
[b] Standard errors are clustered at the individual level.
[c] The item is solved within five seconds as a result of shirking.
[d] Time spent on the item (seconds).
[e] The answer is correct (=1).

Column (1) represents the proxy for effort provision, whereas Columns (2), (4), and (6) represent the proxies for shirking behavior. Columns (3) and (5) are mixed because both shirking and effort can increase the number of correct answers and also payment. As monitoring falls, shirking behaviors increase significantly, which is shown by the negative coefficients of monitoring rates in Column (1) and positive coefficients in Column (2), (4), and (6). The interaction terms demonstrate that risk-averse individuals have a flatter upward curve with the monitoring rate (i.e., the opposite sign of the coefficients of the monitoring rates: positive coefficients in Column (1) and negative coefficients in Column (2), (4), and (6)), although these estimates are not significant (except for the number of items reviewed in Column (2)). We have only a small sample size (321), so capturing any small marginal change due to risk aversion might be difficult.

At the item level (Table 8), the estimates become more significant because we now have more observations. The dependent variables are whether the item is shirked (solved in 5 seconds or less), the time spent on each item, and whether the item is solved correctly. Column (1) represents the proxy for shirking behavior on each item, whereas Columns (2) and (3) represent the proxies for effort provision. Allowing the subject's ability to vary at each task (random effect) shows that shirking behaviors dramatically increase as the monitoring rate falls: the coefficients of monitoring rates are positive (negative) in Column (1) (Columns (2) and (3)). However, the size of marginal effects is again reduced with interactions (the slopes are flatter) because the interactions have the opposite signs. Risk-averse subjects do not respond as strongly as less risk-averse subjects who modify their behavior and seek opportunistic behaviors more rapidly as the monitoring rate changes, which is as expected from our theoretical framework and simulation results.

More strictly, we run a fixed-effect model. The fixed effects can control for unobservable variables that might be correlated with individual risk aversion and ability and can correct for omitted variable bias. However, in the fixed-effects specification, we can only observe the marginal effect of monitoring changes and the interaction terms because the individual level of risk aversion is absorbed into the individual fixed effects. Table 9 presents the results at the task level. Similar to the random-effect model, higher monitoring levels reduces shirking. In addition, we have different signs for the interaction terms, which means that risk-averse individuals have less of a response to monitoring. Table 10 shows more significant coefficients with a larger sample size. Mild monitoring increases shirking but at a lower rate for risk-averse subjects.

Next, we restrict the sample to women. Women are typically found to be more risk averse, and in our sample, women's risk attitudes are more varied than men's, which would allow us to better capture a clear picture of the impact of risk aversion on behaviors. Therefore it is worthwhile to demonstrate how risk aversion works in a specific group. Table 11, Table 12, Table 13, and Table 14 show the random-effect and the fixed-effect models at both the task level and the item level using only female participants. Now, the impact of risk aversion on monitoring change is stronger. Within the group of women, the result confirms that risk-averse individuals indeed respond less to monitoring intensity and generally shirk less.

Overall, with our specification, we observe that (i) reduced monitoring yields increased shirking, (ii) risk-averse subjects shirk less with monitoring, and (iii) (therefore) risk-averse subjects respond less to the monitoring rate. We observe similar results when we analyze shirking behaviors between the genders. Because women are more risk averse in general, we consider women as representative of risk-averse individuals. The results are presented at the task level in Table 15 and at the item level in Table 16. Women's behavior is similar to that of risk-averse individuals; they shirk less and respond less to changes in the monitoring rate, and these behavior patterns are driven purely by the differences in risk attitudes between genders, as in our setting. The only difference across treatments is the level of monitoring (risk).[9]

[9] One might argue that this result ensues because women are more afraid of being caught and feel more ashamed when they get caught. Nevertheless, these features also characterize risk-averse individuals.

Table 11: Task Level, Random Effect: Female Sample

	(1) % Correct[b]	(2) # Clicks[c]	(3) # Correct[d]	(4) # Shirk[e]	(5) Gains	(6) # Wrong[f]
Risk Aversion	-0.008	0.971**	0.661**	0.120	0.037	0.310
	(0.01)	(0.49)	(0.27)	(0.58)	(0.05)	(0.34)
Monitoring Rate 60%	-0.082	4.313	1.867	3.651	0.847**	2.446
	(0.06)	(4.11)	(1.84)	(5.16)	(0.43)	(2.91)
Monitoring Rate 20%	-0.232***	18.714***	6.766***	15.107***	0.863**	11.947***
	(0.06)	(4.11)	(1.84)	(5.16)	(0.43)	(2.91)
RA x Monitoring 60%	0.009	-0.549	-0.304	-0.313	-0.072	-0.245
	(0.01)	(0.62)	(0.28)	(0.78)	(0.06)	(0.44)
RA x Monitoring 20%	0.022**	-2.009***	-0.908***	-1.075	-0.079	-1.102**
	(0.01)	(0.62)	(0.28)	(0.78)	(0.06)	(0.44)
Age	-0.002	-0.618*	-0.492**	-0.549	-0.051*	-0.126
	(0.01)	(0.32)	(0.21)	(0.34)	(0.03)	(0.22)
Highest Diploma	0.056*	0.458	2.203*	-1.177	0.175	-1.745
	(0.03)	(1.70)	(1.13)	(1.83)	(0.16)	(1.18)
Constant	0.889***	30.455***	24.109***	14.204*	2.585***	6.346
	(0.13)	(7.18)	(4.63)	(7.89)	(0.68)	(4.99)
r2_w	0.224	0.231	0.120	0.196	0.090	0.244
r2_b	0.072	0.081	0.118	0.106	0.056	0.084
r2_o	0.142	0.168	0.119	0.163	0.076	0.180
chi2	36.565	38.429	22.514	33.618	14.288	41.177
N	174	174	174	174	174	174

[a] * 10%, ** 5%, and *** 1% significance.
[b] The proportion of correct answers out of total clicks.
[c] The number of clicks.
[d] The number of correct answers.
[e] The number of codes clicked within five seconds, i.e., the number of codes shirked.
[f] The number of wrong answers.
[g] Standard errors are clustered at the individual level.

Table 12: Task Level, Fixed Effect: Female Sample

	(1) % Correct[b]	(2) # Clicks[c]	(3) # Correct[d]	(4) # Shirk[e]	(5) Gains	(6) # Wrong[f]
Monitoring Rate 60%	-0.082	4.313	1.867	3.651	0.847**	2.446
	(0.06)	(4.11)	(1.84)	(5.16)	(0.43)	(2.91)
Monitoring Rate 20%	-0.232***	18.714***	6.766***	15.107***	0.863**	11.947***
	(0.06)	(4.11)	(1.84)	(5.16)	(0.43)	(2.91)
RA x Monitoring 60%	0.009	-0.549	-0.304	-0.313	-0.072	-0.245
	(0.01)	(0.62)	(0.28)	(0.78)	(0.06)	(0.44)
RA x Monitoring 20%	0.022**	-2.009***	-0.908***	-1.075	-0.079	-1.102**
	(0.01)	(0.62)	(0.28)	(0.78)	(0.06)	(0.44)
Constant	0.888***	22.862***	20.155***	0.414	1.891***	2.707***
	(0.01)	(0.99)	(0.44)	(1.25)	(0.10)	(0.70)
r2_w	0.224	0.231	0.120	0.196	0.090	0.244
r2_b	0.001	0.002	0.017	0.012	0.004	0.003
r2_o	0.089	0.090	0.006	0.128	0.049	0.134
N	174	174	174	174	174	174

[a] * 10%, ** 5%, and *** 1% significance.
[b] The proportion of correct answers out of total clicks.
[c] The number of clicks.
[d] The number of correct answers.
[e] The number of codes clicked within five seconds, i.e., the number of codes shirked.
[f] The number of wrong answers.
[g] Standard errors are clustered at the individual level.

Table 13: Item Level, Random Effect: Female Sample

	(1) Shirk[c]	(2) Time[d]	(3) Correct[e]
Risk Aversion	-0.001	-0.506***	-0.003
	(0.01)	(0.19)	(0.01)
Monitoring Rate 60%	0.046*	-1.480***	-0.075**
	(0.03)	(0.51)	(0.04)
Monitoring Rate 20%	0.313***	-8.486***	-0.247***
	(0.03)	(0.49)	(0.04)
RA x Monitoring 60%	-0.001	0.148**	0.008
	(0.00)	(0.07)	(0.01)
RA x Monitoring 20%	-0.016***	0.971***	0.019***
	(0.00)	(0.07)	(0.01)
Age	-0.013	0.282	-0.005
	(0.01)	(0.18)	(0.01)
Highest Diploma	-0.033	-0.616	0.070**
	(0.05)	(0.94)	(0.03)
Item Difficulty	-0.000	0.218***	-0.003***
	(0.00)	(0.01)	(0.00)
Constant	0.386**	4.957	1.032***
	(0.19)	(3.77)	(0.14)
r2_w	0.115	0.268	0.039
r2_b	0.193	0.134	0.140
r2_o	0.143	0.233	0.053
chi2	683.978	1907.203	216.624
N	5,236	5,236	5,236

[a] * 10%, ** 5%, and *** 1% significance.
[b] Standard errors are clustered at the individual level.
[c] The item is solved within five seconds as a result of shirking.
[d] Time spent on the item (seconds).
[e] The answer is correct (=1).

Table 14: Item Level, Fixed Effect: Female Sample

	(1) Shirk[c]	(2) Time[d]	(3) Correct[e]
Monitoring Rate 60%	0.044	-1.460***	-0.070*
	(0.03)	(0.52)	(0.04)
Monitoring Rate 20%	0.311***	-8.482***	-0.242***
	(0.03)	(0.49)	(0.04)
RA x Monitoring 60%	-0.001	0.146*	0.007
	(0.00)	(0.07)	(0.01)
RA x Monitoring 20%	-0.016***	0.970***	0.018***
	(0.00)	(0.07)	(0.01)
Item Difficulty	-0.000	0.218***	-0.003***
	(0.00)	(0.01)	(0.00)
Constant	0.040***	6.915***	0.988***
	(0.01)	(0.21)	(0.02)
r2_w	0.115	0.268	0.039
r2_b	0.163	0.020	0.070
r2_o	0.116	0.198	0.042
N	5236	5236	5236

[a] * 10%, ** 5%, and *** 1% significance.
[b] Standard errors are clustered at the individual level.
[c] The item is solved within five seconds as a result of shirking.
[d] Time spent on the item (seconds).
[e] The answer is correct (=1).

Table 15: Gender Test: Task Level.

	(1) # Shirk[b]	(2) # Clicks[c]	(3) # Correct[d]	(4) % Correct[e]	(5) Gains	(6) # Wrong[f]	(7) # Shirk[b]	(8) # Clicks[c]	(9) # Correct[d]	(10) # Correct[e]	(11) Gains	(12) # Wrong[f]
Monitoring Rate 60%	9.224***	6.000***	1.367	-0.085***	0.199	4.633***	9.224***	6.000***	1.367	-0.085***	0.199	4.633***
	(2.46)	(1.63)	(0.73)	(0.02)	(0.19)	(1.22)	(2.46)	(1.63)	(0.73)	(0.02)	(0.19)	(1.22)
Monitoring Rate 20%	15.571***	9.469***	1.327	-0.149***	0.396*	8.143***	15.571***	9.469***	1.327	-0.149***	0.396*	8.143***
	(2.46)	(1.63)	(0.73)	(0.02)	(0.19)	(1.22)	(2.46)	(1.63)	(0.73)	(0.02)	(0.19)	(1.22)
Woman x Monitoring 60%	-7.535*	-5.121*	-1.402	0.058	0.196	-3.719*	-7.535*	-5.121*	-1.402	0.058	0.196	-3.719*
	(3.34)	(2.22)	(1.00)	(0.03)	(0.26)	(1.66)	(3.34)	(2.22)	(1.00)	(0.03)	(0.26)	(1.66)
Woman x Monitoring 20%	-7.192*	-3.331	-0.240	0.051	-0.027	-3.091	-7.192*	-3.331	-0.240	0.051	-0.027	-3.091
	(3.34)	(2.22)	(1.00)	(0.03)	(0.26)	(1.66)	(3.34)	(2.22)	(1.00)	(0.03)	(0.26)	(1.66)
Subject is a woman	-1.877	-3.119	-2.492*	0.006	-0.581**	-0.627						
	(2.64)	(1.98)	(1.11)	(0.03)	(0.20)	(1.37)						
Age	-0.747*	-0.569	-0.353	0.001	-0.050	-0.216						
	(0.35)	(0.29)	(0.18)	(0.00)	(0.03)	(0.19)						
Highest Diploma	1.327	1.083	1.478	0.031	0.180	-0.395						
	(1.96)	(1.65)	(1.03)	(0.02)	(0.15)	(1.07)						
Constant	17.644*	37.552***	28.602***	0.805***	3.359***	8.950*	1.299	24.308***	21.290***	0.884***	2.157***	3.019***
	(7.20)	(6.01)	(3.74)	(0.09)	(0.53)	(3.91)	(1.18)	(0.78)	(0.35)	(0.01)	(0.09)	(0.59)
r2_w	0.210	0.205	0.037	0.227	0.048	0.243	0.210	0.205	0.037	0.227	0.048	0.243
r2_b	0.159	0.160	0.121	0.066	0.158	0.102	0.118	0.128	0.087	0.038	0.127	0.079
r2_o	0.187	0.180	0.100	0.142	0.100	0.178	0.166	0.144	0.039	0.127	0.014	0.166
N	321	321	321	321	321	321	321	321	321	321	321	321

[a] * 10%, ** 5%, and *** 1% significance.
[b] The number of codes clicked within five seconds: the number of codes shirked
[c] The number of clicks.
[d] The number of right answers.
[e] The proportion of correct answers out of total clicks.
[f] The number of wrong answers.
[g] Standard errors are clustered at the individual level.

Table 16: Gender Test: Item Level.

	(1) Shirking[c]	(2) Time[d]	(3) Correct[e]	(4) Shirking	(5) Time	(6) Correct
Monitoring Rate 60%	0.212***	-1.796***	-0.101***	0.211***	-1.772***	-0.098***
	(0.01)	(0.16)	(0.01)	(0.01)	(0.18)	(0.01)
Monitoring Rate 20%	0.346***	-3.439***	-0.169***	0.346***	-3.392***	-0.167***
	(0.01)	(0.16)	(0.01)	(0.01)	(0.17)	(0.01)
Woman x Monitoring 60%	-0.174***	1.234***	0.077***	-0.174***	1.176***	0.076***
	(0.02)	(0.23)	(0.02)	(0.02)	(0.25)	(0.02)
Woman x Monitoring 20%	-0.134***	1.041***	0.040**	-0.134***	1.097***	0.038**
	(0.01)	(0.23)	(0.02)	(0.01)	(0.25)	(0.02)
Subject is a woman	-0.044	1.183	0.022			
	(0.04)	(0.78)	(0.03)			
Age	-0.016**	0.237	-0.004			
	(0.01)	(0.15)	(0.00)			
Highest Diploma	0.029	-0.949	0.054**			
	(0.05)	(0.83)	(0.03)			
Item Difficulty	0.001***	0.158***	-0.003***			
	(0.00)	(0.00)	(0.00)			
Constant	0.365**	3.894	0.981***	0.068***	12.707***	0.876***
	(0.16)	(3.00)	(0.10)	(0.01)	(0.09)	(0.01)
R^2				0.132	0.053	0.027
Observations	10455	10455	10455	10455	10455	10455

[a] * 10%, ** 5%, and *** 1% significance.
[b] Standard errors are clustered at the individual level.
[c] The item is solved within five seconds as a result of shirking.
[d] Time spent on the item (seconds).
[e] The answer is correct (=1).

5 Concluding Remarks

This paper has investigated shirking with risk aversion under different monitoring schemes. A conceptual model demonstrates that risk-averse subjects shirk less with monitoring. In our setting, subjects decide whether to shirk or whether to make an effort to make simple calculations involving given numerical codes across tasks with different monitoring rates. The simulation results show that the monitoring rate is negatively correlated with shirking and that risk aversion is negatively correlated with shirking under monitoring. When there is less intensive monitoring, individuals shirk more, but the size of any change differs according to risk aversion. Because more risk-averse agents shirk less at any level of monitoring, they respond less to the monitoring changes compared with less risk-averse agents.

The experiment utilizes a series of codes to decipher and calculate. The objective of the task is to solve the calculation presented with the codes carefully to obtain a correct answer under piece-rate payment schemes. With an uncertain probability (monitored), subjects can be paid either only for correct answers with a sanction for wrong answers, or they can be paid for the number of items they attempt to solve regardless of whether they are correct with more luck (not monitored). This setting corresponds to a real work environment in which a firm cannot monitor all the production activities of its employees. In this setting, risk-averse subjects are shown to behave differently than risk-seeking subjects and shirk less under uncertainty. In addition, monitoring works as we expected according to the rational cheater model: it reduces shirking behavior. Examining the slope of the impact of monitoring on shirking reveals that the slope for risk-averse subjects is flatter than the slope for less risk-averse subjects, which is expected under the theoretical framework. Less risk-averse subjects shirk more at lower monitoring rates, and they modify their behaviors sharply as monitoring increases.

We therefore validate the impact of monitoring as an effective method of preventing shirking. In addition, we suggest that risk-averse agents may earn less under a piece-rate contract due to their lower productivity. Because they do not risk shirking behaviors that can yield more production at a faster rate but instead continuously attempt to avoid mistakes, they produce less. However, the quality of their work is higher (e.g. the correct rate), which might help us understand why risk-averse workers (e.g., female workers) occupy low statuses in labor markets. This leaves open the question of whether firms should search for greater quantity of productivity or better work quality.

References

Alchian, A. A. and H. Demsetz (1972, December). Production , Information Costs, and Economic Organization. *American Economic Review 62*(5), 777–95.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutstrom (2006). Elicitation using multiple price list formats. *Experimental Economics 9(4)*, 383–405.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy 76*, 169.

Benabou, R. and J. Tirole (2003). Intrinsic and extrinsic motivation. *Review of Economic Studies 70*(3), 489–520.

Bohnet, I., B. S. Frey, and S. Huck (2000). More order with less law: On contract enforcement, trust, and crowding. *Institute for Empirical Research in Economics - University of Zurich, Working Paper 52*(052).

Bonin, H., T. Dohmen, A. Falk, D. Huffman, and U. Sunde (2007, December). Cross-sectional earnings risk and occupational sorting: The role of risk attitudes. *Labour Economics 14*(6), 926–937.

Bunn, D. N., S. B. Caudill, and D. M. Gropper (1992). Crime in the classroom: An economic analysis of undergraduate student cheating behavior. *The Journal of Economic Education 23*(3), pp. 197–207.

Cadsby, C. B., F. Song, and F. Tapon (2010). Are you paying your employees to cheat? an experimental investigation. *The B.E. Journal of Economic Analysis & Policy 10*(1), 35.

Calvo, G. A. and S. Wellisz (1978, October). Supervision, loss of control, and the optimum size of the firm. *Journal of Political Economy 86*(5), 943–52.

Cornelissen, T., J. S. Heywood, and U. Jirjahn (2011, April). Performance pay, risk attitudes and job satisfaction. *Labour Economics 18*(2), 229–239.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology 18(1)*, 105–115.

Deci, E. L. (1975). *Intrinsic Motivation.* New-York: Plenum Press.

Deci, Edward L.; Koestner, R. and R. M. Ryan (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin 125(3)*, 627–68.

Dickinson, D. and M.-C. Villeval (2008). Does monitoring decrease work effort? the complementarity between agency and crowding-out theories. *Games and Economic Behavrior 63*(1), 56–76.

Dohmen, T. and A. Falk (2011, April). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review 101*(2), 556–90.

Drago, R. and R. Perlman (1989). *Microeconomic Issues in Labour Economics: New Approaches.*, Chapter Supervision and High Wages as Competing Incentives: A Basis for Labour Segmentation Theory, pp. 41–61. New-York: Harvester Wheatsheaf.

Ekelund, J., E. Johansson, M.-R. Jarvelin, and D. Lichtermann (2005, October). Self-employment and risk aversion–evidence from psychological test data. *Labour Economics 12*(5), 649–659.

Fama, E. F. and M. C. Jensen (1983, June). Separation of ownership and control. *Journal of Law and Economics 26*(2), 301–25.

Frey, B. S. and R. Jegen (2000). Motivation crowding theory: A survey of empirical evidence. *Institute for Empirical Research in Economics - University of Zurich, Working Paper 49*(049).

Frey, B. S. and F. Oberholzer-Gee (1997, September). The cost of price incentives: An empirical analysis of motivation crowding-out. *American Economic Review 87*(4), 746–55.

Gneezy, U., M. Niederle, and A. Rustichini (2003, August). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics 118*(3), 1049–1074.

Gneezy, U. and A. Rustichini (2000, January). A fine is a price. *The Journal of Legal Studies 29*(1), 1–17.

Grund, C. and D. Sliwka (2006, March). Performance pay and risk aversion. Discussion Papers 101, SFB/TR 15 Governance and the Efficiency of Economic Systems, Free University of Berlin, Humboldt University of Berlin, University of Bonn, University of Mannheim, University of Munich.

Hartog, J., E. Plug, L. Diaz-Serrano, and J. Vieira (2003, July). Risk compensation in wages - a replication. *Empirical Economics 28*(3), 639–647.

Holt, C. A. and S. K. Laury (2002, December). Risk aversion and incentive effects. *American Economic Review 92*(5), 1644–1655.

Kerkvliet, J. and C. L. Sigmund (1999). Can we control cheating in the classroom? *Journal of Economic Education 30*(4), 331–343.

Kreps, D. M. (1997, May). Intrinsic motivation and extrinsic incentives. *American Economic Review 87*(2), 359–64.

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

Lazear, E. (1995). *Personnel Economics*. MIT Press.

Lazear, E. (2000). Performance pay and productivity. *American Economic Review 90(5)*, 1346–1361.

Moore, M. J. (1995, January). Unions, employment risks, and market provision of employment risk differentials. *Journal of Risk and Uncertainty 10*(1), 57–70.

Murphy, K. M., R. H. Topel, K. Lang, and J. S. Leonard (1987). Unemployment, risk, and earnings: Testing for equalizing wage differences in the labor market. In K. Lang and J. S. Leonard (Eds.), *Unemployment and the Structure of Labor Markets*, pp. 103–140. New York: Basil Blackwell.

Nagin, D. S., J. B. Rebitzer, S. Sanders, and L. J. Taylor (2002). Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment. *American Economic Review 92*(4), 850–873.

Pannenberg, M. (2007). Risk aversion and reservation wages. IZA Discussion Papers 2806, Institute for the Study of Labor (IZA).

Pfeifer, C. (2011). Risk aversion and sorting into public sector employment. Technical Report vol. 12(1), pages 85-99, 02, German Economic Review, Verein fr Socialpolitik.

Pissarides, C. A. (1974, Nov.-Dec.). Risk, job search, and income distribution. *Journal of Political Economy 82*(6), 1255–67.

Prendergast, C. (1999, March). The provision of incentives in firms. *Journal of Economic Literature 37*(1), 7–63.

Rickman, N. and R. Witt (2007). The determinants of employee crime in the uk. *Economica 74*(293), 161–175.

Appendix: Instructions for the experiment[10]

Slide 1.

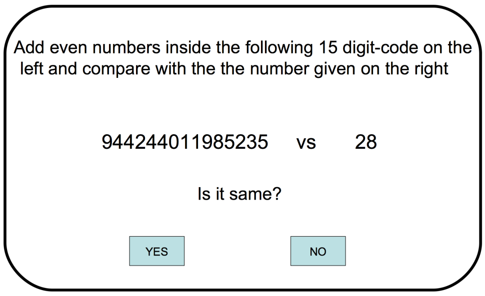
Good morning. Thank you for participating in this experiment. Please read these instructions carefully and, should you have any questions, raise your hand to call the administrator. Communication between participants is forbidden. Please turn off cellular phones.

Slide 2.

You are invited to solve codes. It is important for us to get the right answers for quality control. You will be required to perform effort tasks during four identical rounds of the same experiment. Each round lasts for five minutes. During each round, the time remaining is displayed in the corner of the screen (in seconds). A payoff in euros will be delivered at the end of the experiment. The payment depends on performance of tasks according to rules known by everyone.

Slide 3.

Here is an example of the task screen. For each code, you will have to compare the sum of even numbers in the 15-digit code on the left with the given answer on the right. If



you click your answer, the next code will appear in the screen. At the end of each round, the computer will display the wage you have earned during the round.

Slide 4.

Please raise your hand if you have any questions. If you are ready, please start the first

[10] The experiment was conducted in French.

training session by clicking the button below.

Slide 5. Task 0: Training Session

Please decipher the codes correctly in five minutes. You will be paid a flat rate of five €.

Slide 6. Treatment Task 1

Please decipher the codes correctly in five minutes. All answers will be checked, and you will be paid according to your correct answers (b cents per correct code) with penalties for wrong answers (a cents per wrong code).

Slide 7. Results on Task 1

You have solved [nitem] for Task 1.

The number of correct answers is [nitemc].

The number of wrong answers is [nitmew].

Your payment for Task 1 is €[Task1].

Slide 8. Treatment Task 2

Please decipher the codes correctly in five minutes.

You will be paid according to the number of answers (10 cents per correct code). There is a 60% chance that your answers will be reviewed, and a fine of five cents for each wrong answer will be imposed.

At the end of this round, you will play a lottery to determine whether your answers are reviewed.

Slide 9. Lottery 60/40

Please click the button below to play the lottery 60/40.

(60% chance) All your answers will be checked. You will be paid according to your correct answers (ten cents per correct code) with penalties for wrong answers (five cents per wrong code).

(40% chance) You will be paid according to the number of clicks.

Slide 10. Results on Task 2

You have solved [nitem] for the Task 2.

The number of correct answers is [nitemc].

The number of wrong answers is [nitmew].

Your payment for Task 2 is €[Task2].

Slide 11. Treatment Task 3

Please decipher codes correctly in five minutes.

You will be paid according to the number of answers (b cents per correct code). There is a 20% chance that your answers will be reviewed, and a fine of a cents per each wrong answer will be imposed.

At the end of this round, you will play a lottery to determine whether your answers are reviewed.

Slide 12. Lottery 20/80

Please click the button below to play the lottery 20/80.

(with 20% chance) All your answers will be checked. You will be paid according to your correct answers (ten cents per correct code) with penalties for wrong answers (five cents per wrong code).

(with 80% chance) You will be paid according to the number of clicks.

Slide 13. Results on Task 3

You have solved [nitem] for the Task 3.

The number of correct answer is [nitemc].

The number of wrong answer is [nitmew].

Your payment for Task 3 is €[Task3].

Slide 14. Payoff for the experiment

You have earned five euros for Task 0.

You have earned €[gains1] for Task 1.

You have earned €[gains2] for Task 2.

You have earned €[gains3] for Task 3.

All the tasks are over. Please fill out these questionnaires regarding your socio-demographic information.

Slide 15. Questionnaires

You were born in [year].

Your gender is [man/woman].

Religion [None/Christian/Muslim/Jewish/Buddhist/Other]

What is the highest level of education have you have achieved? [High School/Undergraduate/Graduate]

Monthly Expenditures €[]

Do you smoke? [No/RatherNot/Sometimes/Yes]

Do you drink? [No/RatherNot/Sometimes/Yes]

Concerning this experiment, did you consider the task an opportunity to win some money or a situation with a risk of losing money? [win/lose]

Slide 16. Lottery Game

Here, you will play a lottery game that will add up your final payment. Please choose what you prefer between Option A and Option B. One row will be randomly chosen, and the lottery option of your choice will be played. The payment will be added to your final payment.

Slide 17. Lottery Game

You choice Option[A/B]. And your payment is €[gainlottery]. Your total wage is €[totalgain] for the experiment.

Slide 18. At the end of the experiment

Thank you for participating in this experiment. Please see the administrator to collect your wage.