Thema

UMR 8184

THéorie Économique, Modélisation et Applications

Thema Working Paper n°2014-26
Université de Cergy Pontoise, France

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

Seeun Jung

UNIVERSITÉ de Cergy-Pontoise

CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Novembre, 2014

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 the effort and quality of work as monitoring varies. The predictions of the theoretical model are tested in an original experimental setting in which levels of risk attitudes are measured and monitoring rates vary exogenously. Our results show that shirking decreases with risk aversion, being female, and monitoring. Moreover, monitoring is more effective at curtailing shirking behaviors with subjects who are less risk averse, although the size of the impact is small.

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 genuine support of the experimental design and to Maxim Frolov (Univ. Paris 1) for programming the experiments.

†Paris School of Economics, THEMA/ESSEC - seeun.jung@psemail.eu

1 Motivation

Firms frequently face the agency dilemma: although a firm's objective is to maximize its profits, its employees would like to maximize their 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 for opportunistic behaviors that might lower their costs.[1] These behaviors are not in the firm's interests because they may negatively affect workers' productivity and reduce firm profits. Therefore, employers attempt to curtail these behaviors with several tools, including monitoring. There have been debates regarding whether monitoring works as employers believe. Two different theories explain the different directions of the impact of monitoring. The crowding-out theory 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 (1971), Calvo and Wellisz (1978), Fama and Jensen (1983), Laffont and Martimort (2002), and Prendergast (1999)). A number of studies ((Rickman and Witt, 2007; Nagin et al., 2002; Kerkvliet and Sigmund, 1999; Bunn et al., 1992; Becker, 1968)) have investigated the rational cheater model and have observed variations in effort in response to monitoring. Nevertheless, an empirical validation of the rational cheater model is difficult to establish outside of

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

[2]For example, paying for blood donations would reduce the willingness to donate (Drago and Perlman (1989), Frey and Oberholzer-Gee (1997), Kreps (1997), Gneezy and Rustichini (2000),Bohnet et al. (2000), Frey and Jegen (2000), and Bnabou and Tirole (2003)).

an experimental setting. Cheating (shirking) is difficult to detect, and employers enforce 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 of the rational cheater model. Their field experiment 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. Through 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 are 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. In addition to the crowding-out effect, risk aversion may explain the heterogeneity observed by Nagin et al. (2002). 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)). Risk-averse workers are found to dislike the competitive and stressful work environment

[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, not successful. 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 when employees' shirking (cheating) behavior was caught was dismissal from employment.

that is also typically associated with higher compensation. When monitoring exists, would these 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, under any monitoring schemes, uncertainty is likely to play a role. 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 illustrate our point, we set up an experimental design in which individual risk aversion is measured. Because we build on Nagin et al. (2002) and exogenously alter the perceived monitoring rate and observe the variation in effort by the agents, we address some concerns regarding endogeneity.

In this paper, we use both performance-based payments and sanctions for bad outcomes as a penalty for 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 establishing 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 try 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).

Our experiment is original in its design because we observe the shirking behavior of individuals along with their level of risk aversion. Our results contribute to the literature by addressing the effectiveness of monitoring as a necessary condition for preventing risk-averse individuals from shirking and by investigating the heterogenous impact of monitoring on individuals with different risk attitudes. Thus, monitoring can be more

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

effective for less risk-averse workers. In addition, a better understanding of why risk-averse workers typically occupy low-status categories in labor markets (low wage, low productivity, etc.) will be discussed.

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 their 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 compare this figure to a given number k.[6] When the sum is equal to the given number k, the participant reports 'true'; otherwise, the participant reports 'false'. Because we put the alternative participants must compare to the answer that they calculate close to the correct answer (i.e., using a normal distribution with small variance), participants can learn that the given alternatives can be near the correct answer from the training session, which may tempt the participants to click 'yes' with guessing (i.e., shirking). Because we asked the participants to add all even numbers presented in a 15 digit-code, we only gave the even

[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. These experiments were held 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 p is, the easier the item 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, for example.

numbers as the alternative for each code.

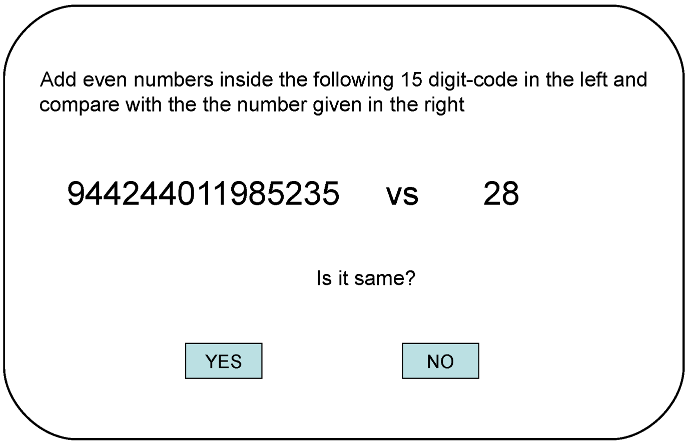


Figure 1: Example of Task Screen

Because these exercises require no particular skills, 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 salary is based on performance. Beginning with a training session, we gave the participants four sessions and assigned different monitoring rates. Each session lasted for 5 minutes.

Training Task: Participants are asked to decipher codes correctly in 5 minutes and are paid a flat rate of 5 euros. Nothing is reviewed.

Treatment Task 1: Participants are asked to decipher codes correctly in 5 minutes. They are also told that all answers will be checked and that they will be paid according to correct answers (10 cents per correct code) with penalties for wrong answers (a cents per wrong code).

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

Treatment Task 2: Participants are asked to decipher codes correctly in 5 minutes. They are told that they will be paid according to the number of answers (10 cents per code); there is a 60% chance that their answers will be reviewed, in which case a fine of 5 cents for each wrong answer will be imposed. 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 5 minutes. They are told that they will be paid according to the number of answers (10 cents per code); there is a 20% chance that their answers will be reviewed, in which case a fine of 5 cents per each wrong answer will be imposed. 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 10 cents per code deciphered with a flat wage of 5 euros for showing up and completing the training task. Then, we assigned 5 cent penalties for poor work quality (e.g., each code solved incorrectly) only when their 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 given the experimental setting. Assume that the number of codes, x, successfully deciphered by a participant depends on the effort supplied, E. The participant shirks whenever she considers a code deciphered by guessing. For instance, she makes a guess with the probability $p(< 1)$ whether the sum of even numbers in a 15-digit code corresponds to the given number. To simplify the model, we assume that if the participant applies enough effort, she obtains the answer right because counting is a reasonably easy task. We therefore assume a linear relationship between effort/shirking and performance. The performance function is defined as

$$x = E + pS + \epsilon$$

, 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. However, we define S as shirking behavior (i.e., the number of codes the subject answered by guessing). The subjects are offered a linear contract of the form $w = bx$. However, when an incorrect outcome is found, 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; we set up a linear cost function for each code on which a participant makes an effort ($c(E) = cE$) as a money equivalent. We consider shirking as shirking on quality. When an individual shirks, the quality of work decreases (i.e., the correction rate of solving codes is lower).

We have two possible states of wealth depending on the monitoring rate M, when 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.

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

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

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

T is the total time endowment that the subject can use in one task, and v_E and v_S are the times needed to provide effort and shirking, respectively. We assume that providing effort is more costly in terms of time constraints: $v_E > v_S$. Using the negative exponential utility function ($U(w) = -e^{-Rw}$) with risk-aversion parameter $R(> 0)$, we can solve for the optimal level of S^* via the first-order condition:[8]

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

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

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

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

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

$\partial S^*/\partial R$ is negative when $\ln(1 - M) - \ln(M) + \ln(bv_E - (b - c)v_S)) - \ln((b - c)v_S -$

[8] See Appendix1.

$v_E bp - a(1-p)) > 0$. 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.

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

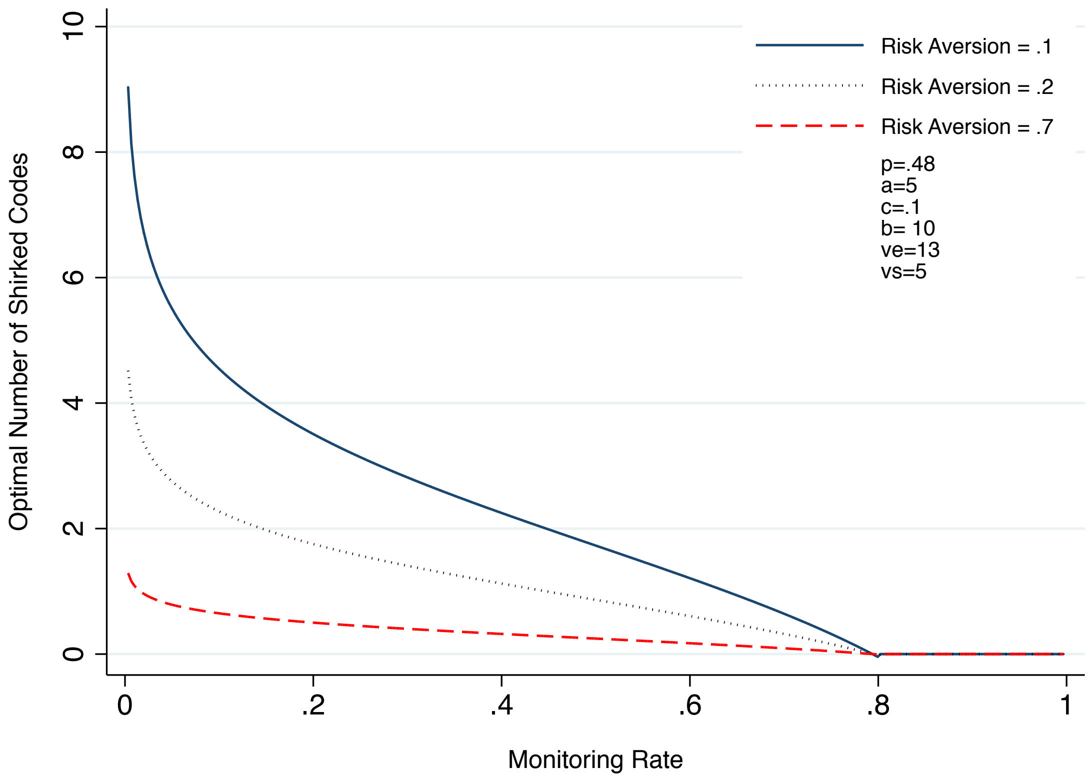


Figure 2: Optimal Shirking with Monitoring with Penalty

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.[9] This result 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 at a lower monitoring rate, less risk-averse

[9] $b = 10, a = 5, c = 0.1, p = 0.48, v_{s,i} = 5, v_{e,i} = 13, T = 300$.

subjects are more sensitive to the monitoring rate increase, whereas the slopes of more risk-averse subjects are rather flat and not much affected by the change in 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 below zero. We may infer that full monitoring is not necessary under this setting, particularly monitoring with a penalty. If enough monitoring is exerted that features a proper way of penalizing the 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. We would thus like to test whether (i) more risk-averse individuals shirk less ($\frac{\partial s^*}{\partial r} < 0$), (ii) shirking falls with monitoring ($\frac{\partial s^*}{\partial m} < 0$), and (iii) 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.[10]

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 was approximately 23 (the lab belongs to the University of Paris 1) and approximately half of the subjects were female and religious. In addition to socio-demographic information, we gathered data on individual risk aversion by letting them play an incentivized lottery game which was designed by Holt and Laury (2002). As Andersen et al. (2006) suggested, we employ a switching multiple price list design and therefore only allow the monotonic risk preference. As presented in Figure 5, 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. This would give the same expected gain for a risk neutral participant. Except the first choice

[10]We did another set of experiment without a penalty, whose results will be presented in the Appendix 3.

where there is no uncertainty of the gain, option B is riskier with possible higher gains. As the decision number increases, the probability of getting the higher amount of gain (€2 for Option A and €3.85 for Option B) reduces. In other words, the probability of getting a lower amount of gain increases. A rational participant would then switch to choose Option A instead of Option B at a certain decision number which varies over individual risk aversion. For the final gain, one row is randomly chosen, and then the lottery is played according to the choice made by the participant. In this paper, we use the number of Option A that the participant has chosen, as our measure of risk aversion. This measure is increasing as becoming more risk averse. Also, the gain from the lottery varies in between €0.10 and €3.85, which is similar to our gain for each task. The risk preference of this lottery should, therefore, be relevant for the one that we would use for 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
Monthly Expenditures (€)	107	30	425	500	2,200
Holt and Laury Type Incentivised Lottery Choice					
The number of Safe Options	107	0	0.61	1	1
Behavior-related questions					
The subject has used					
- Tobacco[†]	107	1	2.10	1	4
- Alcohol[†]	107	1	2.82	3	4
Attitudes toward 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 some money or a risk of losing money?"

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

We also asked some behavior-related questions, such as questions regarding smoking and drinking. Using these questions, we could 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 risk of losing money?"

Table 2 shows the correlation between the risk attitudes measure and related questions. In this paper, we use Holt and Laury type risk aversion, which is consistently correlated with the other questions: women and religious people are more risk averse, risk averse people smoke and drink less. In addition, the more educated, the less risk averse. At the end of the experiment, we asked subjects how they perceived the tasks they undertook: as 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 4 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 a strategy 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 20% of monitoring of their work (task3). Figure 5 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 a situation with more intensive monitoring.

Table 3 shows the descriptive statistics of the variables measuring subjects' perfor-

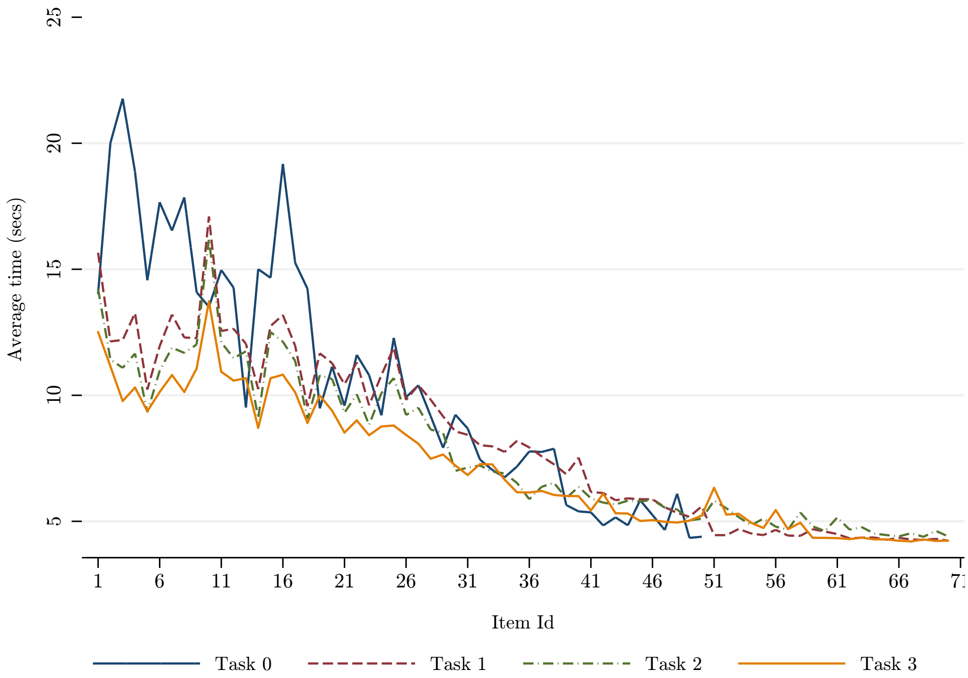


Figure 4: Average Time Spent on each Item Across Tasks

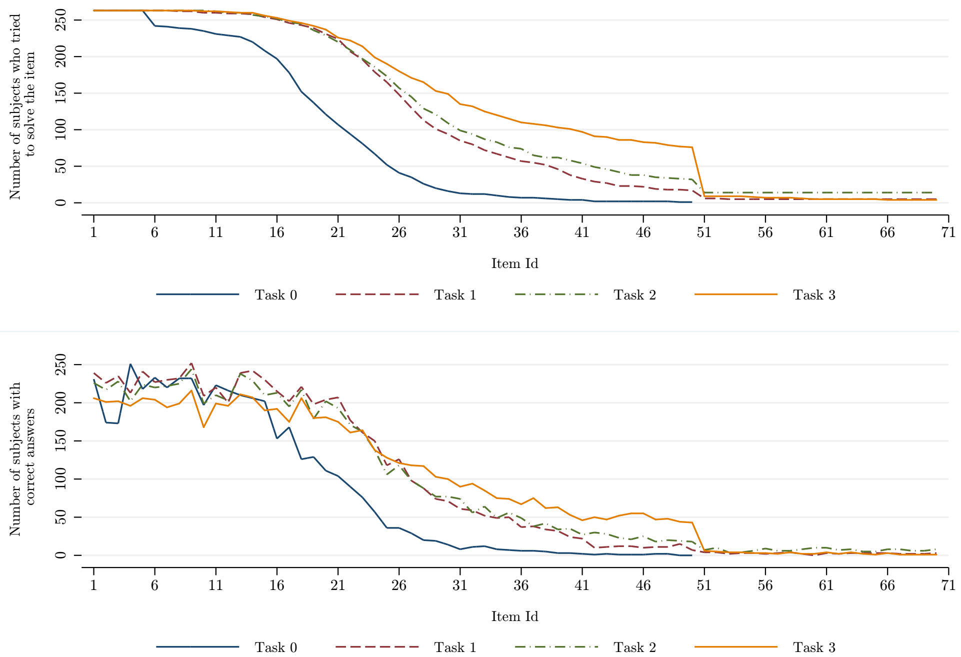


Figure 5: Performance by Item Across Tasks

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'.

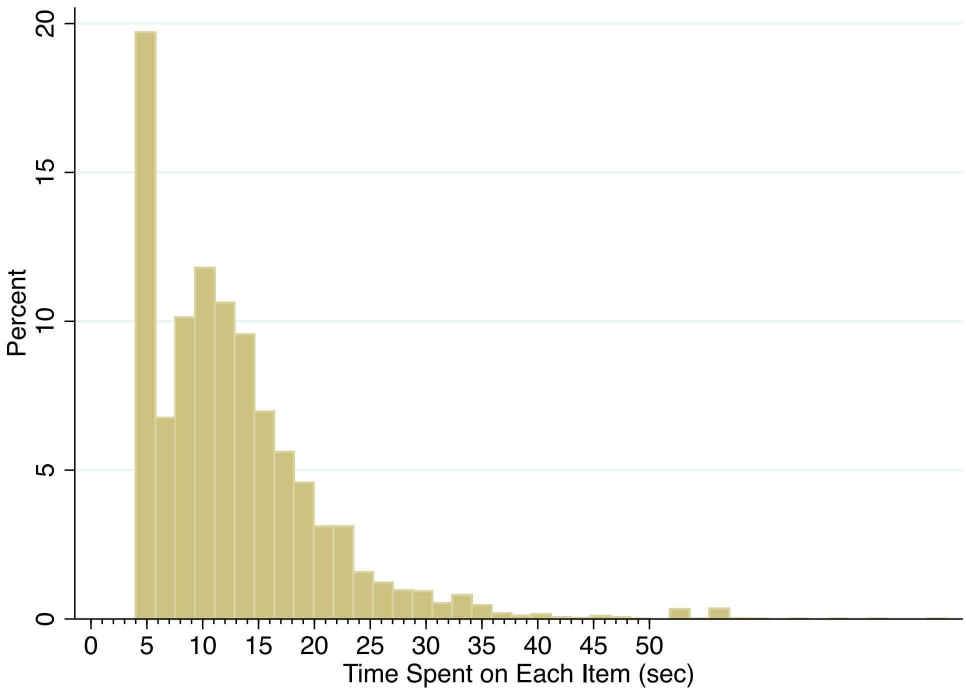


Figure 6: Distribution of Time Spent (sec.)

mance. In each task, we have information on the number of items reviewed by the subjects. 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[11]. Indeed, there is a jump in the frequencies of items solved within 4-5 seconds found in Figure 6. 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, depending on the code's difficulty. We expect that when the code seems more difficult with higher numbers to calculate, the probability of a correct answer and the time spent 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). Because we put the alternative participants compared to the answer that they calculated close to the correct answer (i.e., using a normal distribution with small variance), participants could learn that the given alternatives may be near the correct answer from the training session, which may have tempted participants to click 'yes' by guessing (i.e., shirking). If we examine the correlation coefficient between the shirking item and whether the answer is 'yes', it is positive and significant (R= 0.07 significant at the 1% level) in Table 5. In other words, when participants shirk, they tend to answer 'yes' more often.

Table 21 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 they become more careful on solving codes, but the correct rate 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 answer more correctly on each item.

Table 6 shows the subjects' performance across tasks. Monitoring rates in Tasks 1,

[11] 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^g.	Monitoring
d Shirking^b	-0.08*	-0.40*
Correct Rate^c	0.08*	0.37*
# of Clicks^d	-0.03*	-0.33*
# of Correct Answers^e	0.06*	-0.09*
# of Answered Shirked^f	-0.09*	-0.37*

[a] ⋆ 1% significance.
[b] The difference between the number of clicks and the number of correct answers: shirking proportion.
[c] % of the ratio: the number of correct answers/the number of clicks.
[d] The total number of clicks.
[e] The number of correct answers.
[f] The number of codes shirked (clicked within 5 seconds).
[g] The certainty equivalent rescaled, decided by 1,000.

Table 5: Pairwise Correlation Matrix Item Level.

	Risk Attitudes^f	Difficulty^g	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 5 seconds.
[d] Total amount of time (sec.) spent to click.
[e] The code is correctly solved.
[f] The certainty equivalent rescaled, divided by -1,000.
[g] The difficulty of each item is the sum of all the numbers in the 15 digits.

Table 6: Subjects' Performance Across Tasks.

	Task 0[†]	Task 1	Task 2	Task 3
Task level				
Number of items reviewed	19.21	29.00	31.25	34.78
	(0.47)	(0.66)	(0.82)	(0.81)
Number of items correctly reviewed	16.98	24.16	24.33	24.84
	(0.43)	(0.46)	(0.46)	(0.47)
Number of items reviewed in less than 5s	0.34	3.17	6.92	12.85
	(0.14)	(0.47)	(0.94)	(1.14)
Payment received (€)	.	2.42	2.81	3.29
	(.)	(0.05)	(0.08)	(0.08)
Observations	263	263	263	263
Item level				
Time spent on an item (secs)	14.91	11.33	10.16	8.99
	(0.11)	(0.06)	(0.06)	(0.06)
Item correctly reviewed	0.88	0.83	0.78	0.71
	(0.00)	(0.00)	(0.00)	(0.00)
Item reviewed in less than 5s	0.02	0.11	0.22	0.37
	(0.00)	(0.00)	(0.00)	(0.01)
Observations	5053	7627	8220	9147

Standard errors are reported in parentheses.

[†] Task 0 is the practice session. Participants get paied 5 euros as fixed rate.

2, and 3 are 100%, 60%, and 20%, respectively. At the task level, the table shows that as the monitoring rate decreases, both the number of items reviewed and the number of shirked items rise. 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.

4.1 Analyses

We want to estimate $\frac{\partial y}{\partial M}$, $\frac{\partial y}{\partial R}$, and $\frac{\partial^2 y}{\partial M \partial R}$, where y is various proxies for shirking behavior, such as the number of shirking codes, the number of items reviewed, whether the code is shirked, and the time spent to solve each code. The following specification is estimated:

$$\mathbf{y_{it}} = \beta_0 + \beta_M \mathbf{M_t} + \beta_R \mathbf{R_i} + \beta_{M,R} \mathbf{R_t} \times \mathbf{R_i} + \beta_x \mathbf{x_i} + \eta_{it}$$

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

$$\mathbb{E}[\mathbf{M_t} \times \eta_{_{it}} \mid \mathbf{x}_{_{ik}}] = 0 \text{ and } \mathbb{E}[\mathbf{R_{it}} \times \eta_{_{it}} \mid \mathbf{x}_{_{it}}] \neq 0$$

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 9 is the specification at the item level. Our dependent variables are the proportion of the correct rate, the number of total clicks, the number of correct answers, the number of shirking codes, the gains, and the shirking proportion. In Table 7, the results present the impact of risk aversion on shirking behavior. With the interaction terms of risk aversion and task numbers with different monitoring rates, we allow for different slopes for risk aversion. Being more risk averse reduces shirking behavior (the number of items reviewed) but lowers gains[12]. As monitoring increases, shirking behaviors diminish significantly. When we look at the interaction terms, risk-averse individuals have a flatter downward curve with the monitoring rate, although the estimates are not significant (except for the number of items reviewed). 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, the estimates become more significant because we now have more observations. Allowing subject ability to vary at each task (random effect), we observe that risk-averse subjects spend more time on each item and succeed in finding the correct answer. However, again, with interactions, the size of marginal effects is reduced (the slopes are flatter) because the interactions have opposite signs. Risk-averse subjects shirk less from the beginning (20% monitoring rate) and do not respond as much as less risk-averse subjects, who shirk more and then modify their behavior more as the monitoring rate changes, as expected from our theoretical framework and simulation results.

More strictly, we run a fixed-effect model. The fixed effect can control for unobservable

[12] It is empirically true that risk-averse workers earn less.

Table 7: Task Level, Random Effect.

	(1) % Correct[b]	(2) # Clicks[c]	(3) # Correct[d]	(4) # Shirk[e]	(5) Gains	(6) d Shirk[f]
Risk Aversion	0.093	-1.629	2.345	-3.059	-0.010	-3.974
	(0.07)	(4.57)	(2.54)	(6.11)	(0.46)	(3.17)
Monitoring Rate 60%	0.112**	-6.902**	-0.862	-6.163	-0.181	-6.039**
	(0.05)	(3.27)	(1.46)	(4.97)	(0.38)	(2.46)
Monitoring Rate 100%	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%	-0.073	4.060	0.451	-0.610	0.173	3.609
	(0.07)	(5.08)	(2.27)	(7.71)	(0.59)	(3.81)
RA x Monitoring 100%	-0.114	9.026*	3.344	4.872	-0.030	5.683
	(0.07)	(5.08)	(2.27)	(7.71)	(0.59)	(3.81)
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
chi2	67.291	71.362	25.555	68.092	29.066	74.823
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 5 seconds: the number of shirking codes.
[f] The difference between the number of clicks and the number of correct answers: shirking proportion.
[g] Standard errors are clustered at the individual level.

Table 8: Task Level, Fixed Effect.

	(1) % Correct[b]	(2) # Clicks[c]	(3) # Correct[d]	(4) # Shirk[e]	(5) Gains	(6) d Shirk[f]
Monitoring Rate 60%	0.112**	-6.902**	-0.862	-6.163	-0.181	-6.039**
	(0.05)	(3.27)	(1.46)	(4.97)	(0.38)	(2.46)
Monitoring Rate 100%	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%	-0.073	4.060	0.451	-0.610	0.173	3.609
	(0.07)	(5.08)	(2.27)	(7.71)	(0.59)	(3.81)
RA x Monitoring 100%	-0.114	9.026*	3.344	4.872	-0.030	5.683
	(0.07)	(5.08)	(2.27)	(7.71)	(0.59)	(3.81)
Constant	0.764***	31.972***	22.486***	12.972***	2.538***	9.486***
	(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
chi2						
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 5 seconds: the number of shirking codes.
[f] The difference between the number of clicks and the number of correct answers: shirking proportion.
[g] Standard errors are clustered at the individual level.

Table 9: Item Level, Random Effect.

	(1) Shirk[c]	(2) Time[d]	(3) Correct[e]
Risk Aversion	-0.033	0.253	0.144**
	(0.10)	(1.75)	(0.06)
Monitoring Rate 60%	-0.132***	3.347***	0.142***
	(0.02)	(0.34)	(0.03)
Monitoring Rate 100%	-0.329***	5.753***	0.262***
	(0.02)	(0.35)	(0.03)
RA x Monitoring 60%	-0.037	-2.674***	-0.092**
	(0.04)	(0.54)	(0.04)
RA x Monitoring 100%	0.082**	-4.679***	-0.187***
	(0.04)	(0.54)	(0.04)
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
chi2	1473.742	2362.573	420.084
N	10455	10455	10455

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

Table 10: Item Level, Fixed Effect.

	(1) Shirk[c]	(2) Time[d]	(3) Correct[e]
Monitoring Rate 60%	-0.132***	3.358***	0.140***
	(0.02)	(0.34)	(0.03)
Monitoring Rate 100%	-0.328***	5.739***	0.257***
	(0.02)	(0.35)	(0.03)
RA x Monitoring 60%	-0.036	-2.689***	-0.088**
	(0.04)	(0.54)	(0.04)
RA x Monitoring 100%	0.081**	-4.656***	-0.182***
	(0.04)	(0.54)	(0.04)
Item Difficulty	0.001***	0.158***	-0.003***
	(0.00)	(0.00)	(0.00)
Constant	0.315***	5.074***	0.831***
	(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
chi2			
N	10455	10455	10455

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

variables that might be correlated with individual risk aversion and ability and can correct for omitted variable bias. However, in the fixed-effect specification, we can only observe the marginal effect of monitoring changes and the interaction terms because the individual risk aversion level is absorbed into the individual fixed effect. Table 24 presents the results at the task level. Similar to the random-effect model, greater monitoring reduces shirking. In addition, we have different signs for the interaction terms, which means that risk-averse individuals respond less to monitoring. Table 25 shows more significant coefficients with a larger sample size. Intensive monitoring reduces shirking but at a lower rate for risk-averse subjects.

Now, we restrict samples only for women. Women are usually found to be more risk averse in general. Therefore it is interesting to see how risk aversion works in a specific group. Table 12, Table 11, Table 13, and Table 14 show the random-effect and the fixed-effect model at both task level and item level using only female participants, respectively. Now, the impact of risk aversion on monitoring change is stronger. Within women, the result again validates that risk averse individuals indeed respond less to the monitoring intensity and generally shirk less.

Overall, with our specification, we can observe that (i) risk-averse subjects shirk less, (ii) intensive monitoring reduces shirking, and (iii) risk-averse subjects respond less to the monitoring rate.[13]

[13] We observe similar results when we analyze shirking behaviors between genders. Because women are more risk averse in general, we consider women representative of risk-averse individuals. The results are presented at the task level in Table 26 and at the item level in Table 27. Women's behavior is similar to risk-averse individuals; they shirk less and respond less to changes in the monitoring rate.

Table 11: Task Level, Random Effect: Female Sample

	(1)	(2)	(3)	(4)	(5)	(6)
	% Correct[b]	# Clicks[c]	# Correct[d]	# Shirk[e]	Gains	d Shirk[f]
Risk Aversion	0.094	-2.856	1.266	-4.798	-0.420	-4.122*
	(0.06)	(3.83)	(2.51)	(3.79)	(0.49)	(2.42)
Monitoring Rate = 0.6	0.089*	-3.354	1.123	-4.175	-0.016	-4.477*
	(0.05)	(4.05)	(2.07)	(4.30)	(0.43)	(2.48)
Monitoring Rate = 1	0.171***	-7.666*	-0.744	-7.825*	-0.863**	-6.923***
	(0.05)	(4.05)	(2.07)	(4.30)	(0.43)	(2.48)
RA x Monitoring 60%	-0.082	7.122	2.282	2.965	0.067	4.839
	(0.07)	(6.08)	(3.10)	(6.46)	(0.64)	(3.73)
RA x Monitoring 100%	-0.171**	12.608**	5.320*	6.098	0.790	7.287*
	(0.07)	(6.08)	(3.10)	(6.46)	(0.64)	(3.73)
Age	-0.001	-0.522*	-0.427**	-0.415	-0.051*	-0.095
	(0.01)	(0.27)	(0.20)	(0.26)	(0.03)	(0.18)
Highest Diploma	0.050*	0.531	1.968*	-0.881	0.175	-1.437
	(0.03)	(1.47)	(1.09)	(1.37)	(0.16)	(0.95)
Constant	0.719***	35.750***	23.717***	18.405***	3.448***	12.033***
	(0.11)	(6.10)	(4.44)	(5.75)	(0.68)	(3.92)
r2_w	0.112	0.030	0.118	0.049	0.090	0.069
r2_b	0.069	0.081	0.113	0.107	0.056	0.084
r2_o	0.090	0.046	0.116	0.064	0.076	0.074
N	232	232	232	232	174	232

[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 5 seconds: the number of shirking codes.
[f] The difference between the number of clicks and the number of correct answers: shirking proportion.
[g] Standard errors are clustered at the individual level.

Table 12: Task Level, Fixed Effect: Female Sample

	(1)	(2)	(3)	(4)	(5)	(6)
	% Correct[b]	# Clicks[c]	# Correct[d]	# Shirk[e]	Gains	d Shirk[f]
Monitoring Rate = 0.6	0.089*	-3.354	1.123	-4.175	-0.016	-4.477*
	(0.05)	(4.05)	(2.07)	(4.30)	(0.43)	(2.48)
Monitoring Rate = 1	0.171***	-7.666*	-0.744	-7.825*	-0.863**	-6.923***
	(0.05)	(4.05)	(2.07)	(4.30)	(0.43)	(2.48)
RA x Monitoring 60%	-0.082	7.122	2.282	2.965	0.067	4.839
	(0.07)	(6.08)	(3.10)	(6.46)	(0.64)	(3.73)
RA x Monitoring 100%	-0.171**	12.608**	5.320*	6.098	0.790	7.287*
	(0.07)	(6.08)	(3.10)	(6.46)	(0.64)	(3.73)
Constant	0.824***	22.638***	17.569***	4.422***	2.260***	5.069***
	(0.01)	(0.80)	(0.41)	(0.85)	(0.10)	(0.49)
r2_w	0.112	0.030	0.118	0.049	0.090	0.069
r2_b	0.003	0.009	0.030	0.011	0.004	0.003
r2_o	0.039	0.020	0.074	0.028	0.043	0.030
N	232	232	232	232	174	232

[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 5 seconds: the number of shirking codes.
[f] The difference between the number of clicks and the number of correct answers: shirking proportion.
[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	-0.143*	-0.203	0.025
	(0.08)	(2.62)	(0.07)
Monitoring Rate = 0.6	-0.144***	0.548	0.065**
	(0.02)	(0.63)	(0.03)
Monitoring Rate = 1	-0.201***	1.888***	0.141***
	(0.02)	(0.63)	(0.03)
RA x Monitoring 60%	0.105***	-4.032***	0.015
	(0.03)	(0.93)	(0.05)
RA x Monitoring 100%	0.136***	-5.276***	-0.066
	(0.04)	(0.94)	(0.05)
Age	-0.012*	0.336	0.003
	(0.01)	(0.24)	(0.01)
Highest Diploma	-0.023	-1.833	0.068*
	(0.04)	(1.31)	(0.04)
Item Difficulty	-0.000	0.237***	-0.005***
	(0.00)	(0.01)	(0.00)
Constant	0.546***	3.572	0.752***
	(0.16)	(5.23)	(0.14)
r2_w	0.038	0.153	0.037
r2_b	0.126	0.161	0.136
r2_o	0.064	0.147	0.050
N	6804	6804	6804

[a] * 10%, ** 5%, and *** 1% significance.

[b] Standard errors are clustered at the individual level.

[c] The item is solved within 5 seconds: being shirked.

[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%	-0.145***	0.595	0.068**
	(0.02)	(0.63)	(0.03)
Monitoring Rate 100%	-0.200***	1.907***	0.141***
	(0.02)	(0.63)	(0.03)
RA x Monitoring 60%	0.106***	-4.084***	0.013
	(0.03)	(0.93)	(0.05)
RA x Monitoring 100%	0.134***	-5.290***	-0.066
	(0.04)	(0.94)	(0.05)
Item Difficulty	-0.000	0.237***	-0.005***
	(0.00)	(0.01)	(0.00)
Constant	0.143***	8.086***	0.929***
	(0.01)	(0.25)	(0.01)
r2_w	0.038	0.153	0.037
r2_b	0.003	0.144	0.038
r2_o	0.029	0.138	0.037
N	6804	6804	6804

[a] * 10%, ** 5%, and *** 1% significance.

[b] Standard errors are clustered at the individual level.

[c] The item is solved within 5 seconds: being shirked.

[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) d Shirk[f]	(7) # Shirk[b]	(8) # Clicks[c]	(9) # Correct[d]	(10) # Correct[e]	(11) Gains	(12) d Shirk[f]
Monitoring Rate 60%	-6.347**	-3.469*	0.041	0.064**	-0.197	-3.510**	-6.347*	-3.469*	0.041	0.064**	-0.197	-3.510**
	(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 100%	-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%	-0.343	-1.789	-1.162	0.006	0.223	-0.628	-0.343	-1.789	-1.162	0.006	0.223	-0.628
	(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 100%	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	-9.069***	-6.451**	-2.732*	0.057*	-0.608**	-3.719**						
	(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	33.216***	47.021***	29.929***	0.656***	3.755***	17.093***	12.972***	3 1.972***	22.486***	0.764***	2.538***	9.486***
	(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	0.080	0.077	0.025	0.092	0.010	0.109
chi2	75.320	73.672	22.255	68.768	30.016	79.235						
N	321.000	321.000	321.000	321.000	321.000	321.000	321.000	32 ¿ 1.000	321.000	321.000	321.000	321.000

[a] * 10%, ** 5%, and *** 1% significance.
[b] The number of codes clicked within 5 seconds: the number of shirking codes
[c] The number of clicks.
[d] The number of right answers.
[e] The proportion of correct answers out of total clicks.
[f] The difference between the number of clicks and the number of correct answers: shirking proportion.
[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%	-0.134***	1.642***	0.068***	-0.135***	1.620***	0.069***
	(0.01)	(0.15)	(0.01)	(0.01)	(0.17)	(0.01)
Monitoring Rate 100%	-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%	-0.040***	0.193	0.037**	-0.040***	0.079	0.038**
	(0.01)	(0.22)	(0.02)	(0.01)	(0.24)	(0.02)
Woman x Monitoring 100%	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	-0.178***	2.223***	0.061**			
	(0.04)	(0.77)	(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	0.712***	0.455	0.812***	0.347***	9.865***	0.728***
	(0.16)	(3.00)	(0.10)	(0.00)	(0.08)	(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 5 seconds: being shirked.
[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 explains that risk-averse subjects switch from shirking to not shirking at a lower monitoring rate compared with less risk-averse subjects. In our setting, subjects decide whether to shirk or to make an effort to solve given codes across tasks with different monitoring rates. We derive a relevant theoretical model. The simulation results show that risk aversion is negatively correlated with shirking in general, and the monitoring rate is similarly negatively correlated. When more intensive monitoring is exerted, individuals shirk less, but the size of any change differs by risk aversion. Because more risk-averse agents shirk less at any level of monitoring, they respond less to the monitoring change compared with less risk-averse agents.

The experiment utilizes a series of codes to decipher. The objective of the task is to solve codes carefully to obtain a correct answer under piece-rate payment schemes. With an uncertain probability, subjects can be paid either only for correct answers with/without a sanction for wrong answers or with more luck for the number of items they attempt to solve regardless of whether they are correct. This setting corresponds to a real work environment in which the firm cannot monitor every piece produced by its employees. In this setting, we observe that risk-averse subjects behave differently than risk-seeking subjects: they shirk less under uncertainty. In addition, monitoring works as we expected from 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 that of less risk-averse subjects, which is expected from 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 effectiveness of monitoring as a necessary condition for 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 bet on the outcome of not being caught when shirking but continuously provide more effort to avoid mistakes, they produce less. However, the quality of their work is higher, 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 productivity or better work quality. This subject is useful for further research to be undertaken on the firms' side.

References

Abdellaoui, M., A. Driouchi, and O. L'Haridon (2011). Risk aversion elicitation: reconciling tractability and bias minimization. *Theory and Decision 71*, 63–80.

Abeler, J., A. Falk, L. Goette, and D. Huffman (2009, January). Reference points and effort provision. IZA Discussion Papers 3939, Institute for the Study of Labor (IZA).

Alchian, A. A. and H. Demsetz (1971, May). Production, information costs and economic organizations. UCLA Economics Working Papers 10A, UCLA Department of Economics.

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.

Arrondel, L. and A. Masson (2010, June). French savers in the economic crisis : What has changed ? variation of preferences or simple adaptation to a new environment. mimeo, Paris-Jourdan Sciences Economiques.

Arrondel, L. and A. Masson (2011). La crise a-t-elle rendu l'pargnant plus prudent ? *Retraite et socit 60*(1), 111–135.

Basov, S. and X. Yin (2010). Optimal screening by risk-averse principals. *The B.E. Journal of Theoretical Economics 10*(1), 8.

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

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). Crosssectional earnings risk and occupational sorting: The role of risk attitudes. *Labour Economics 14*(6), 926–937.

Bruggen, A. and M. Strobel (2007, August). Real effort versus chosen effort in experiments. *Economics Letters 96*(2), 232–236.

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.

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

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.

Chouikhi, O. and S. V. Ramani (2004, 03). Risk aversion and the efficiency wage contract. *LABOUR 18*(1), 53–73.

Clementi, F. (2011, March). The role of the experiment in the evolution of labor legislation: A perspective on constitutional law. In *The Role of Experiments for the Advancement of Effective Labour Legislation*. SSRN.

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.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. Wagner (2009). Individual risk attitudes: Measurement, determinants and behavioral consequences. Research Memoranda 007, Maastricht : ROA, Research Centre for Education and the Labour Market.

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.

Falk, A. and J. J. Heckman (2009, October). Lab experiments are a major source of knowledge in the social sciences. IZA Discussion Papers 4540, Institute for the Study of Labor (IZA).

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

Fochmann, M., J. Weimann, K. Blaufus, J. Hundsdoerfer, and D. Kiesewetter (2010).

Grosswage illusion in a real effort experiment. FEMM Working Papers 100009, Otto-von-Guericke University Magdeburg, Faculty of Economics and Management.

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. and J. A. List (2006, 09). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica 74*(5), 1365–1384.

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.

Golan, L., C. Parlour, and U. Rajan (2007, November). Racing to the bottom: Competition and quality. GSIA Working Papers 2008-E33, Carnegie Mellon University, Tepper School of Business.

Goldstone, R. L. and C. Chin (1993). Dishonesty in Self-Report of copies made - moral relativity and the copy machine. *Basic and Applied Social Psychology 14*(1), 19–32.

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.

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

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

Jitsophon, S. and T. Mori (2013). The hidden costs of control in the field. *Working Paper, Osaka University*.

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.

Mazar, N., O. Amir, and D. Ariely (2008, November). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research 45*(6), 633–644.

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.

Preston, J. and D. M. Wegner (2007). The eureka error: Inadvertent plagiarism by misattributions of effort. *Journal of Personality and Social Psychology 92*(4), 575 – 584.

Rick, S., G. Loewenstein, J. R. Monterosso, D. D. Langleben, N. Mazar, O. Amir, and D. Ariely (2008, November). Commentaries and rejoinder to the dishonesty of honest people. *Journal of Marketing Research 45*(6), 645–653.

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

Schwieren, C. and D. Weichselbaumer (2010, June). Does competition enhance performance or cheating? a laboratory experiment. *Journal of Economic Psychology 31*(3), 241–253.

van Dijk, F., J. Sonnemans, and F. van Winden (2001, February). Incentive systems in a real effort experiment. *European Economic Review 45*(2), 187–214.

Appendix 1: Solving for optimal shirking

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

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

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

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

T is the total time endowment that the subject can use in one task, and v_E and v_S are the time needed to provide 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 ($U(w) = -e^{-Rw}$), we set the Lagrangian function as

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

$$\Rightarrow \mathcal{L} \equiv -Me^{-R\{b(E+pS)-a(1-p)S-cE\}}\mathbb{E}e^{-Rb\epsilon}-(1-M)e^{-R\{b(E+S)-cE\}}\mathbb{E}e^{-Rb\epsilon}-\lambda(T-v_E E-V_S S)$$

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

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

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

$$\frac{\partial \mathcal{L}}{\partial E} = 0 \Rightarrow (b-c)RMe^{-R\{b(E+pS)-a(1-p)S-cE\}+\frac{R^2b^2\sigma^2}{2}}+(b-c)R(1-M)e^{-R\{b(E+S)-cE\}+\frac{R^2b^2\sigma^2}{2}} = \lambda v_E$$

$$\frac{\partial \mathcal{L}}{\partial S} = 0 \Rightarrow RM\{bp-a(1-p)\}e^{-R\{b(E+pS)-a(1-p)S-cE\}+\frac{R^2b^2\sigma^2}{2}}+bR(1-M)e^{-R\{b(E+S)-cE\}+\frac{R^2b^2\sigma^2}{2}} = \lambda v_S$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0 \Rightarrow T = v_e E + v_s S$$

Taking the ratio of the first two conditions yields

$$\frac{(b-c)RMe^{-R\{b(E+pS)-a(1-p)S-cE\}+\frac{R^2b^2\sigma^2}{2}} + (b-c)R(1-M)e^{-R\{b(E+S)-cE\}+\frac{R^2b^2\sigma^2}{2}}}{RM\{bp-a(1-p)\}e^{-R\{b(E+pS)-a(1-p)S-cE\}+\frac{R^2b^2\sigma^2}{2}} + bR(1-M)e^{-R\{b(E+S)-cE\}+\frac{R^2b^2\sigma^2}{2}}} = \frac{v_E}{v_S}$$

$$\Rightarrow e^{-R\{b(E+pS)-a(1-p)S-cE\}+\frac{R^2b^2\sigma^2}{2}} RM[(b-c)v_S - v_E\{bp-a(1-p)\}]$$

$$= e^{-R\{b(E+S)-cE\}+\frac{R^2b^2\sigma^2}{2}} R(1-M)(bv_E - (b-c)v_S)$$

$$\Rightarrow e^{SR(a+b)(1-p)} = \frac{1-M}{M} \frac{bv_E - (b-c)v_S}{(b-c)v_S - v_E\{bp-a(1-p)\}}$$

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

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

In the case with the negative S, by applying Kuhn-Tucker condition, optimal shirking is 0. Therefore, optimal shirking is as follows:

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

Appendix 2

Instructions for the experiment[14]

Slide 1.

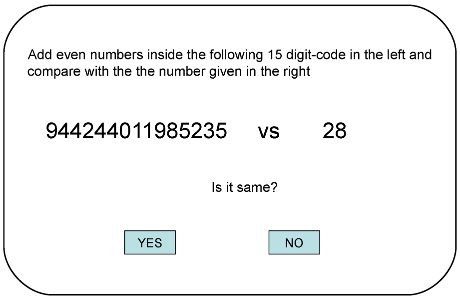
Good morning. Thank you for participating in this experiment. Please read these instructions carefully and, should you have any questions, raise your hand to call the administrator. Communication between participants is forbidden. Please turn off cellular phones.

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 an effort task during 4 identical rounds of the same experiment. Each round lasts for 5 minutes. During each round, the time remaining is displayed in the corner of the screen (in seconds). A euro payoff will be delivered at the end of the experiment. The payment is connected to performance at the task 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.

[14]The experiment was conducted in French.

Slide 4.

Please raise your hand if you have any questions. If you are ready, please start the first training session by clicking the button below.

Slide 5. Task 0: Training Session

Please decipher the codes correctly in 5 minutes. You will be paid a flat rate of €5.

Slide 6. Treatment Task 1

Please decipher the codes correctly in 5 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 the Task 1 is €[Task1].

Slide 8. Treatment Task 2

Please decipher the codes correctly in 5 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 5 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.

(with 60% chance) All your answers will be checked. You will be paid according to your correct answers (10 cents per correct code) with penalties for wrong answers (5 cents per wrong code).

(with 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 the Task 2 is €[Task2].

Slide 11. Treatment Task 3

Please decipher codes correctly in 5 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 (10 cents per correct code) with penalties for wrong answers (a 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 the Task 3 is €[Task3].

Slide 14. Payoff for the experiment

You have earned 5 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? [Bachelor's/Undergraduate/Graduate]

Monthly Expenditures €[]

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

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

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

Slide 16. Lottery Game

Here, you will play a lottery game which will add up your final gain. 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 gain will be added to your final gain.

Slide 17. Lottery Game

You choice Option[A/B]. And your gain 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 come to the administrator to collect your wage.

Appendix 3: Results of Experiment, No Penalty Setting

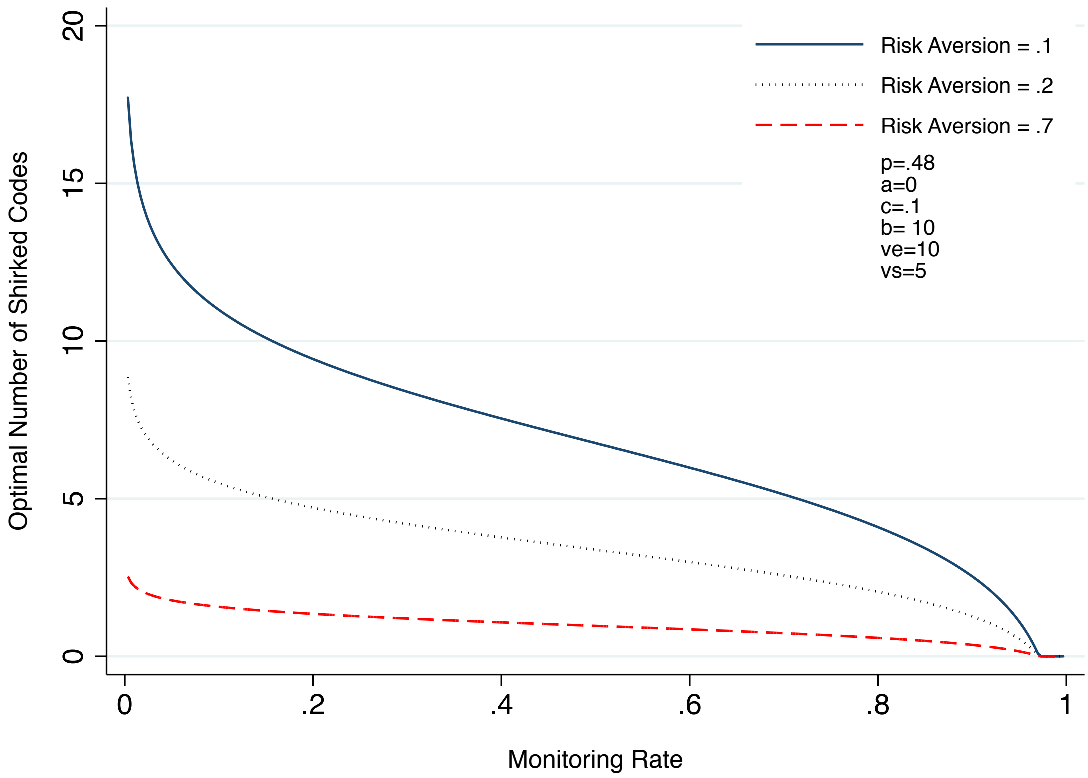


Figure 7: Optimal Shirking with Monitoring without Penalty

Figure 7 depicts the simulation result when there is no penalty on the wrong answer. Still we observe that risk averse individuals shirk less, and also respond less to the monitoring variation. Table 17 presents the descriptive statistics of the samples we used for the experiment. In this experiment, we ask participants the following question:

How much are you willing to pay for a lottery ticket with a 50% chance of winning €1,000?

From this question, we can define the amount of willingness to pay as a certainty equivalent, CE, which then satisfies

$$U[CE] = 0.5U[1,000] + 0.5U[0].$$

We use this certainty equivalent to control for risk attitudes. We rescale the certainty equivalent as $(1000 - CE)/1000$ to fashion a variable that is increasing with risk aversion.

Table 17: Descriptive Statistics of Subjects.

	Obs	Min	Mean	Median	Max
Age (years)	263	18	24	23	63
Woman	263	0	.47	0	1
Religion					
- Christian	263	0	.27	0	1
- Muslim	263	0	.19	0	1
- Jewish	263	0	.034	0	1
- Other	263	0	.061	0	1
- None	263	0	.44	0	1
Education level (highest achieved)					
- High School	263	0	.61	1	1
- Undergraduate	263	0	.38	0	1
- Graduate	263	0	.011	0	1
Monthly Expenditures (€)	263	0	467	300	3,500

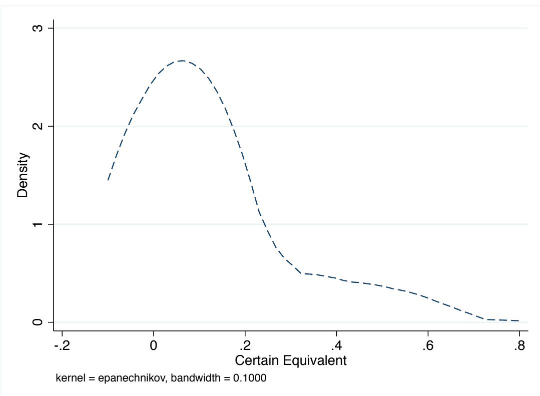


Figure 8: Distribution of Certainty Equivalent.

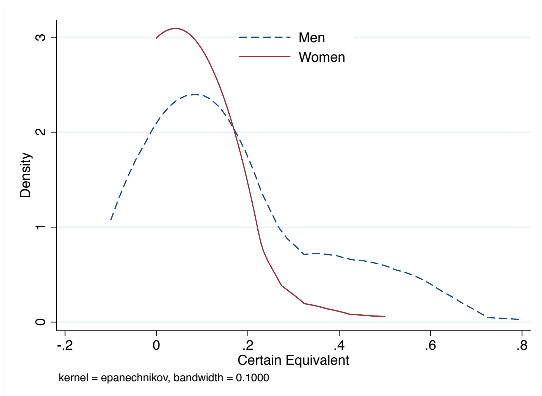


Figure 9: Distribution of Certainty Equivalent by Gender.

Table 18: Descriptive Statistics of Variables Related to Risk Aversion.

	Obs	Min	Mean	Median	Max
Lottery {1000€, 0.5, 0€}[a]					
Certainty Equivalent (CE)	263	0	105	50	700
$CE >$ €500: Risk Seeking	9	505	601.7	600	700
$CE =$ €500: Risk Neutral	36	500	500	500	500
$CE <$ €500: Risk Averse	744	0	79.99	50	499
Related Behaviors					
The subject has used					
- Tobacco[†]	263	1	1.9	1	4
- Alcohol[†]	263	1	2.5	3	4
Attitudes toward Task					
Winning lottery (=0) vs. Losing money (=1)[b]	263	0	0.37	0	1

[a] The subject is asked to give the maximum price she or he is willing to pay for a lottery ticket that offers a best outcome of € 1000 with a probability of 50% or a worse outcome of € 0.

[b] "Did you consider the task an opportunity to win some money or a risk of losing money?"

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

Table 19: Pairwise Correlation Matrix.

	Risk Attitudes
Being woman (=1)	0.37*
Smoking (=1)	-0.15*
Drinking Alcohol (=1)	-0.13*
Being Religious (=1)	0.05*
Education	-0.01*
Task[b]	0.10*

[a] ⋆ 1% significance.

[b] "Did you consider the task an opportunity to win money or a risk of losing money?"

Table 20: Descriptive Statistics of Variables Measuring Subjects' Performance.

	Obs	Min	Mean	Median	Max
Number of items reviewed	1,052	5	29	26	70
Number of items reviewed in less than 5s.	1,052	0	13	6	70
Time spent on an item (secs)	24,994	4	10.1	9.1	77.2
Item is correctly reviewed (=1)	24,994	0	.77	1	1
Item is reviewed in less than 5s (shirking) (=1)	24,994	0	.24	0	1
Item difficulty[a]	24,994	4	29.9	30	74
Clicked yes (=1)[b]	24,994	0	0.4	0	1
Payment received at the end:					
- Task (€)	789	.7	2.8	2.5	7
- Experiment (€)	263	7.5	14	13	24

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

Table 21: Pairwise Correlation Matrix Task Level.

	Risk Attitudes[h].	Monitoring
d Shirking[b]	-0.14*	-0.26*
Correction Rate [c]	0.10*	0.27*
# of Clicks[d]	-0.17*	-0.19*
# of Correct Answers[e]	-0.13*	-0.04
# of Answered Shirked[f]	-0.16*	-0.23*
% Shirking[g]	-0.15*	-0.26*

[a] ⋆ 1% significance.
[b] The difference between the number of clicks and the number of correct answers: shirking proportion.
[c] % of the ratio: the number of correct answers/the number of clicks.
[d] The number of clicks.
[e] The number of correct answers.
[f] The number of codes shirked (clicked within 5 seconds).
[g] The proportion of shirking codes out of the total number of clicks.
[h] The certainty equivalent rescaled, decided by 1,000.

Table 22: Pairwise Correlation Matrix Item Level.

	Risk Attitudes[f]	Difficulty[g]	Clicked 'Yes'	Shirked
Clicked 'Yes'[b]	-0.03*	0.04*		
Shirked[c]	-0.13*	-0.02*	0.12*	
Time Spent[d]	0.11*	0.04*	-0.12*	-0.74*
Correct[e]	0.05*	-0.06*	-0.002	-0.23*

[a] ⋆ 1% significance.
[b] The answer clicked is 'yes'.
[c] The item is solved within 5 seconds.
[d] Total amount of time (sec.) spent to click.
[e] The code is correctly solved.
[f] The certainty equivalent rescaled, divided by -1,000.
[g] The difficulty of each item is the sum of all the numbers in the 15 digits.

Table 23: Subjects' Performance Across Tasks.

	Task 0[†]	Task 1	Task 2	Task 3
Task level				
Number of items reviewed	19.21	29.00	31.25	34.78
	(0.47)	(0.66)	(0.82)	(0.81)
Number of items correctly reviewed	16.98	24.16	24.33	24.84
	(0.43)	(0.46)	(0.46)	(0.47)
Number of items reviewed in less than 5s	3.13	11.13	15.15	20.79
	(0.39)	(0.76)	(1.05)	(1.18)
Payment received (€)	.	2.42	2.81	3.29
	(.)	(0.05)	(0.08)	(0.08)
Observations	263	263	263	263
Item level				
Time spent on an item (secs)	10.91	7.33	6.16	4.99
	(0.11)	(0.06)	(0.06)	(0.06)
Item correctly reviewed	0.88	0.83	0.78	0.71
	(0.00)	(0.00)	(0.00)	(0.00)
Item reviewed in less than 5s	0.16	0.38	0.48	0.60
	(0.01)	(0.01)	(0.01)	(0.01)
Observations	5053	7627	8220	9147

Standard errors are reported in parentheses.
[†] Task 0 is the practice session. Participants get paid 5 euros as a fixed rate.

Table 24: Task Level, Fixed Effect.

	(1) % Correct[b]	(2) # Clicks[c]	(3) # Correct[d]	(4) # Shirk[e]	(5) Gains	(6) d Shirk[f]
Monitoring Rate = 0.6	0.088	-6.608	-1.660	-14.482*	-1.125*	-4.948
	(0.06)	(4.25)	(2.14)	(6.92)	(0.45)	(3.49)
Monitoring Rate = 1	0.213***	-12.233**	0.481	-27.112***	-1.673***	-12.714***
	(0.06)	(4.25)	(2.14)	(6.92)	(0.45)	(3.49)
RA2xT2	-0.029	3.446	1.278	9.559	0.730	2.168
	(0.07)	(4.69)	(2.37)	(7.63)	(0.49)	(3.85)
RA2xT3	-0.127	7.212	-1.298	19.483*	0.898	8.509*
	(0.07)	(4.69)	(2.37)	(7.63)	(0.49)	(3.85)
Constant	0.755***	34.779***	24.844***	12.886***	3.285***	9.935***
	(0.01)	(0.48)	(0.24)	(0.78)	(0.05)	(0.39)
r2_w	0.166	0.128	0.010	0.140	0.229	0.148
r2_b	0.018	0.039	0.021	0.051	0.034	0.037
r2_o	0.065	0.023	0.002	0.045	0.070	0.051
N	789	789	789	789	789	789

[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 5 seconds: the number of shirking codes.
[f] The difference between the number of clicks and the number of correct answers: shirking proportion.
[g] Standard errors are clustered at the individual level.

Table 25: Item Level, Fixed Effect.

	(1) Shirk[c]	(2) Time[d]	(3) Correct[e]
Monitoring Rate = 0.6	-0.315***	1.663***	0.083*
	(0.03)	(0.43)	(0.04)
Monitoring Rate = 1	-0.523***	3.440***	0.230***
	(0.03)	(0.44)	(0.04)
RA2xT2	0.189***	-0.614	-0.020
	(0.03)	(0.48)	(0.04)
RA2xT3	0.321***	-1.436**	-0.136**
	(0.03)	(0.48)	(0.04)
difficulty	0.001***	0.031***	-0.002***
	(0.00)	(0.00)	(0.00)
Constant	0.313***	8.239***	0.778***
	(0.01)	(0.09)	(0.01)
r2_w	0.087	0.041	0.017
r2_b	0.001	0.000	0.020
r2_o	0.055	0.027	0.016
N	24994	24994	24994

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

Table 26: Gender Test: Task Level.

	(1) # Shirk[b]	(2) # Clicks[c]	(3) # Correct[d]	(4) % Correct[e]	(5) Gains	(6) d Shirk[f]	(7) # Shirk[b]	(8) # Clicks[c]	(9) # Correct[d]	(10) # Correct[e]	(11) Gains	(12) d Shirk[f]
Monitoring Rate = 0.6	-9.309***	-5.252***	-0.619	0.093***	-0.630***	-4.633***	-9.309***	-5.252***	-0.619	0.093***	-0.630***	-4.633***
	(1.49)	(0.92)	(0.47)	(0.01)	(0.10)	(0.75)	(1.49)	(0.92)	(0.47)	(0.01)	(0.10)	(0.75)
Monitoring Rate = 1	-14.532***	-7.986***	-0.504	0.144***	-1.036***	-7.482***	-14.532***	-7.986***	-0.504	0.144***	-1.036***	-7.482***
	(1.49)	(0.92)	(0.47)	(0.01)	(0.10)	(0.75)	(1.49)	(0.92)	(0.47)	(0.01)	(0.10)	(0.75)
FemxT2	7.172***	3.663**	0.215	-0.063**	0.337*	3.448**	7.172**	3.663**	0.215	-0.063**	0.337*	3.448**
	(2.17)	(1.34)	(0.68)	(0.02)	(0.14)	(1.09)	(2.17)	(1.34)	(0.68)	(0.02)	(0.14)	(1.09)
FemxT3	10.299***	4.679***	-0.375	-0.094***	0.353*	5.055***	10.299***	4.679***	-0.375	-0.094***	0.353*	5.055***
	(2.17)	(1.34)	(0.68)	(0.02)	(0.14)	(1.09)	(2.17)	(1.34)	(0.68)	(0.02)	(0.14)	(1.09)
Subject is a woman	-11.947***	-9.063***	-3.900***	0.070***	-0.785***	-5.163***						
	(1.75)	(1.49)	(0.90)	(0.02)	(0.14)	(0.96)						
Age	-0.045	0.096	0.054	-0.001	0.008	0.042						
	(0.12)	(0.12)	(0.08)	(0.00)	(0.01)	(0.07)						
Highest diploma	0.421	1.097	1.001	0.009	0.142	0.096						
	(1.23)	(1.28)	(0.81)	(0.01)	(0.11)	(0.73)						
Constant	19.004***	35.242***	24.007***	0.723***	3.261***	11.235***	12.886***	34.779***	24.844***	0.755***	3.285***	9.935***
	(2.98)	(3.04)	(1.91)	(0.03)	(0.26)	(1.74)	(0.77)	(0.47)	(0.24)	(0.01)	(0.05)	(0.39)
r2_w	0.168	0.147	0.010	0.196	0.235	0.175	0.168	0.147	0.010	0.196	0.235	0.175
r2_b	0.093	0.090	0.092	0.008	0.098	0.039	0.092	0.084	0.083	0.006	0.087	0.038
r2_o	0.133	0.106	0.077	0.094	0.153	0.104	0.020	0.004	0.004	0.054	0.050	0.031
chi2	131.466	115.304	31.472	129.437	188.632	121.315						
N	789	789	789	789	789	789	789	789	789	789	789	789

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

Table 27: Gender Test: Item Level.

	(1) Shirking[c]	(2) Time[d]	(3) Correct[e]	(4) Shirking	(5) Time	(6) Correct
Monitoring Rate = 0.6	-0.216***	1.611***	0.094***	-0.216***	1.607***	0.094***
	(0.01)	(0.09)	(0.01)	(0.01)	(0.09)	(0.01)
Monitoring Rate = 1	-0.330***	2.748***	0.152***	-0.329***	2.743***	0.150***
	(0.01)	(0.10)	(0.01)	(0.01)	(0.10)	(0.01)
FemxT2	0.163***	-1.170***	-0.069***	0.164***	-1.173***	-0.069***
	(0.01)	(0.15)	(0.01)	(0.01)	(0.14)	(0.01)
FemxT3	0.216***	-1.368***	-0.097***	0.215***	-1.369***	-0.095***
	(0.01)	(0.15)	(0.01)	(0.01)	(0.15)	(0.01)
difficulty	0.001***	0.031***	-0.002***	0.001***	0.031***	-0.002***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Subject is a woman	-0.253***	3.037***	0.078***			
	(0.02)	(0.36)	(0.01)			
Age	-0.002	0.022	-0.001			
	(0.00)	(0.03)	(0.00)			
Highest diploma	0.010	-0.560	0.009			
	(0.02)	(0.36)	(0.01)			
Constant	0.406***	8.007***	0.766***	0.311***	8.253***	0.779***
	(0.05)	(0.83)	(0.03)	(0.01)	(0.09)	(0.01)
r2_w	0.101	0.045	0.019	0.101	0.045	0.019
r2_b	0.145	0.102	0.069	0.045	0.067	0.000
r2_o	0.111	0.071	0.021	0.030	0.010	0.014
chi2	2767.108	1172.361	478.684			
Observations	24994	24994	24994	24994	24994	24994

[a] * 10%, ** 5%, and *** 1% significance.

[b] Standard errors are clustered at the individual level.

[c] The item is solved within 5 seconds: being shirked.

[d] Time spent on the item (seconds).

[e] The answer is correct (=1).