



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## The spillover effects of monitoring

### Citation for published version:

Belot, M & Schröder, M 2013 'The spillover effects of monitoring: A field experiment' ESE Discussion Papers, no. 238, Edinburgh School of Economics Discussion Paper Series.

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Publisher's PDF, also known as Version of record

### Publisher Rights Statement:

© Belot, M., & Schröder, M. (2013). The Spillover Effects of Monitoring: A Field Experiment. (ESE Discussion Papers; No. 238). Edinburgh School of Economics Discussion Paper Series.

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.





Edinburgh School of Economics  
**Discussion Paper Series**  
Number 238

*The Spillover Effects of Monitoring:  
A Field Experiment*

**Michèle Belot**  
(University of Edinburgh)

**Marina Schröder**  
(University of Magdeburg)

Date  
October 2013

***Published by***

School of Economics  
University of Edinburgh  
30 -31 Buccleuch Place  
Edinburgh EH8 9JT  
+44 (0)131 650 8361  
<http://edin.ac/16ja6A6>



THE UNIVERSITY *of* EDINBURGH

# The Spillover Effects of Monitoring: A Field Experiment\*

Michèle Belot<sup>†</sup> and Marina Schröder<sup>‡</sup>

October 28, 2013

## Abstract

We provide field experimental evidence of the effects of monitoring in a context where productivity is multi-dimensional and only one dimension is monitored and incentivised. We hire students to do a job for us. The job consists of identifying euro coins. We study the effects of monitoring and penalising mistakes on work quality, and evaluate spillovers on non-incentivised dimensions of productivity (punctuality and theft). We find that monitoring improves work quality only if incentives are large, but reduces punctuality substantially irrespectively of the size of incentives. Monitoring does not affect theft, with ten per cent of participants stealing overall. Our setting also allows us to disentangle between possible theoretical mechanisms driving the adverse effects of monitoring. Our findings are supportive of a reciprocity mechanism, whereby workers retaliate for being distrusted.

*Keywords:* counterproductive behaviour, monitoring, experiment

*JEL:* C93, J24, J30, M42, M52

---

\*The authors thank the Institute for Fraud Prevention for financial support.

<sup>†</sup>University of Edinburgh, School of Management, michele.belot@ed.ac.uk, 30 Buccleuch Place, Edinburgh, EH8 9JT, UK

<sup>‡</sup>University of Magdeburg, Faculty of Economics and Management, marina.schroeder@ovgu.de, Postbox 4120, 39106 Magdeburg, Germany

# 1 Introduction

Experts estimate that globally occupational fraud causes annual losses of more than \$3.5 trillion (Association of Certified Fraud Examiners, 2012). The question is what an organisation can do to prevent such behaviour. One straightforward instrument to think of is monitoring workers and incentivising conform behaviour. However, monitoring may entail negative crowding out effects (see Frey, 1993 and Falk and Kosfeld, 2006 for reviews of this literature). In a typical work-relation productivity is multi-dimensional and there are multiple ways in which workers can behave counterproductively: From showing up late to do sloppy work, stealing, bullying or sabotaging other people's work, counterproductive behavior has many possible facets. Thus, crowding out effects may spill over to other non-monitored productivity dimensions. One important question is how monitoring affects the monitored and non-monitored productivity dimensions and ultimately whether monitoring is efficient or not.

So far, the experimental evidence on the effect of monitoring relates to situations where productivity is operationalised with a single measure, such as for example the number of units produced or sold, performance at a test or monetary transfers in an experimental game (see for example Nagin et al., 2002; Fisman and Miguel, 2007; Boly, 2011; Dickinson and Villeval, 2008; Gneezy and Rustichini, 2000; Falk and Kosfeld, 2006). We contribute to this research in studying an experimental setup with multiple observable dimensions of productivity. Studying monitoring in a setting with multiple productivity dimensions allows us to uncover negative spillover effects of monitoring and to disentangle between different mechanisms driving these crowding out effects.

The field experimental setup we use is related to the euro currency. We recruit students to identifying the provenance of euro coins. Every worker receives four boxes of coins and is asked to identify and return the coins by an appointed date. The task has the advantage of offering a menu of observable forms of counterproductive behaviours that are very common in the workplace, i.e. sloppy work, tardiness and theft. These forms of counterproductive behav-

our vary in their nature and perhaps, importantly, in the non-monetary (or moral) costs associated with them (Robinson and Bennett, 1995).

We compare three treatments with different degrees of monitoring and incentives. A first treatment (*no monitoring*) entails no monitoring at all, a second treatment (*lax monitoring & weak incentives*) introduces monitoring with weak incentives and the third treatment (*strict monitoring & strong incentives*) introduces monitoring with strong incentives. Comparing results across treatments, we are able to differentiate between different mechanisms driving the effects of monitoring and incentives on work behaviour. Specifically we can disentangle between a disciplining effect (Becker, 1968; Grasmik and Bursik, 1990), reciprocal reactions (Frey, 1993; Dufwenberg and Kirchsteiger, 2004) and information effects (Bénabou and Tirole, 2003; Sliwka, 2007).

We find a positive disciplining effect only when incentives are strong. Weak incentives lead to no improvement in work quality at all, while strong incentives reduce the number of mistakes by 40%. However, we also find evidence for negative spillover effects, which appear as soon as monitoring is introduced. Specifically, we find that tardiness increases substantially: the fraction of participants who show up late increases by 35% as soon as monitoring is implemented and the magnitude of the increase is similar in both incentive treatments. Theft, on the other hand, remains constant across treatments: On average, 10% of the participants steal coins. Our results are most supportive of an interpretation related to negative reciprocity, whereby workers wish to punish the principal (for distrusting them) and do so in the least costly manner for themselves (both in monetary and non-monetary terms).

Overall, our experimental results suggest that monitoring and incentives can only be efficient if the incentives are strong. Whether or not monitoring with strong incentives is efficient depends on the ratio of the gains in the monitored productivity dimension to the losses in other non-monitored productivity dimensions.

The rest of the paper is structured as follows. We present the experimental design in Section 2 and derive predictions in Section 3. We discuss the results

in Section 4 and conclude in Section 5.

## 2 Experimental design

The job consisted of identifying the value and country of origin of euro coins that were collected in various countries in the euro zone.<sup>1</sup> Participants had a day to complete the task and were supposed to return the materials by a specific deadline. The job has several methodological advantages. It is a "realistic" job, i.e. it is a job that could realistically be advertised by an economics department. The job has multiple dimensions of productivity that arise naturally: Participants could do a poor job, be late in completing the job or steal some of the coins. Still, it is straightforward for us to design a monitoring scheme targeting only one of these dimensions. Also, participants who failed to comply in either of these three dimensions can be categorised as behaving counterproductively, since we made sure that it is possible for participants to do a perfect job.

### 2.1 Procedure

Each participant received a set of 4 boxes of euro coins collected in 4 different countries of the euro zone. The lid of each box indicated the country the coins were collected in. Within one set, the composition of boxes, with respect to the value and the number of coins varied. Across sets, however, the composition of boxes was similar. Each participant received a total of 780 coins with a value of €114.70.

We recruited student workers via a notice posted at various points on campus. Students contacted us by e-mail and were invited to collect the materials (each of them came separately). At collection, each participant received standardised verbal instructions on how to do the job. All participants received

---

<sup>1</sup>There are currently 17 countries (out of 28 members of the European Union) and three European microstates (Vatican, San Marino and Monaco) that use the euro as their currency. There are 8 euro coin denominations, ranging from one cent to two euro. The coins first came into use in 2002. They have a common reverse, but each country in the eurozone has its own design on the obverse, which means that each coin has a variety of different designs in circulation at once. For a detailed description of this task see Belot and Schröder (2013).

a catalogue illustrating the popular euro coins and four identification tables.<sup>2</sup> Using the catalogue, participants were told to identify each coin by indicating the value and the country it was printed in on the identification table.

Participants were told to identify all coins in each box and were asked to work on one box at a time and to put all coins back into the box once identified. They were told to use a separate identification table for each box and to indicate the country of collection (as indicated in the lid of the box) on the identification table. In the monitoring treatments, we informed participants about the number of boxes that would be checked, the tolerated number of mistakes and the penalty they would incur if the number of mistakes exceeded the tolerated number. Participants were informed about the payment they would receive upon returning the materials and the remaining payment they would receive if they met the quality requirement. If participants had no further questions, we asked them to indicate the exact time at which they would return the coins the next day.<sup>3</sup> Participants were informed that the process of returning coins and collecting payment would only take one minute.

All participants were allowed to take the material with them to work from home. When a participant returned the coins, we noted the exact time the material was returned. We also asked the participants for an estimate of the time they had worked on the task, for their field of study and we recorded the gender.

We checked all returned materials with respect to coin composition and mistakes in the identification task. Whenever we observed deviations in the composition of coins, we replaced coins by identical coins or coins with similar collector's value before handing the materials to the next participant. In the monitoring treatments, participants were informed whether they had met the work quality requirement and could collect the remaining amount of money.

---

<sup>2</sup>The catalogue and the tables did not include special coins and coins from the microstates Monaco, San Marino and the Vatican. If any box contained such a coin, participants were told to identify the coin as "not in the catalogue."

<sup>3</sup>We gave participants enough time to check their calendar for the best suitable time in the time horizon between 3:30 p.m. and 6:00 p.m. Once a participant had decided on the exact return time, we wrote the time on a sheet of paper handed out to the participant.

## 2.2 Treatments

Table 1 summarises the three treatments of the experiment. In the *no monitoring* treatment, there was no monitoring at all. When returning the work material, participants in the *no monitoring* treatment immediately received the full payment of €20 in cash. In both monitoring treatments, participants knew that 1 out of the 4 boxes would be checked after returning the coins. In the *monitoring & weak incentives* treatment participants were allowed to make 10 mistakes. If we found more than 10 mistakes in the box randomly chosen for checking, the participant would only receive €19 instead of €20. In the *monitoring & strong incentives* treatment, the tolerated number of mistakes was only 2. If we found more than 2 mistakes in the checked box, the participants' payment was only €5 instead of €20. Participants receive the sure part of the payment first and could collect the remaining part later (usually a day later) if they met the work quality requirements.

**Table 1** Treatments of the experiment

	no. of boxes checked	tolerated no. of mistakes	penalty
no monitoring	0	-	-
monitoring & weak incentives	1	10	€1
monitoring & strong incentives	1	2	€15

## 2.3 Sample

Overall, 91 students participated in this study, 30 in the *no monitoring* and *monitoring & weak incentives* treatments and 31 in the *monitoring & strong incentives* treatment. We recruited participants via a notice posted at various points on campus. The notice informed students that we needed support for a research project in economics, that all students could participate, that the task would last for 2 to 3 hours and could be fulfilled from home, and that the average payment was €20. Interested students were asked to contact the research team via mail. Those students who had not participated in any previous related studies, received a response mail shortly explaining the task. Further, we suggested two collection dates with the corresponding return dates and asked

students to choose one date and to indicate at what time they would collect the working material.<sup>4</sup>

### 3 Predictions

We now derive predictions of how monitoring and incentives may affect behaviour in this setting. Importantly, our goal is to be able to disentangle between the different mechanisms that have been proposed in the literature.

We assume that workers are risk neutral and that the correct identification of coins and punctuality require the provision of costly effort. In the absence of monitoring, standard economic theory predicts that workers provide zero effort (both in the identification task and in punctuality) and steal all the coins. However, it is often observed that workers provide positive effort even when they have no monetary incentive to do so. We will focus on two possible explanations on why workers provide effort that have been proposed in the literature. The first explanation has to do with other-regarding preferences. An inequity averse or altruistic worker, for example, may care both about her own and about the principal's payoff (Fehr and Schmidt, 1999; Andreoni and Miller, 2002). To increase the principal's payoff, an other-regarding worker may provide effort in the identification task, show up on time and return all coins.

The second explanation for why workers provide effort even when this is not monetarily beneficial has to do with social norms. Doing a proper job, showing up on time and returning the coins are associated with social norms that are costly to violate. It is interesting to point out that these costs of violating norms may differ across productivity dimensions. We conjecture that showing up late is not as costly as theft, which is in fact illegal (Robinson and Bennett, 1995).

Now, let us turn to the possible effects of introducing monitoring. The standard effect predicted by principal agent models is the *disciplining effect*. Following the prediction of standard economic theory, monitoring and incentives effectively increase the marginal benefit of providing effort in the monitored

---

<sup>4</sup>Collection was always in the morning between 10:00 a.m. and 12:30 p.m. and return was the next day between 3:30 p.m. and 6:00 p.m.

productivity dimension. Therefore, work quality should increase in the level of monitoring and incentives (Becker, 1968).

In a context with multiple productivity dimensions, monitoring one productivity dimension may additionally lead to a shift of effort away from non-monitored dimensions and towards the monitored dimension (Holmstrom and Milgrom, 1991; Baker, 1992). This *distortion effect* only applies if efforts associated with different productivity dimensions are substitutes. In the context of our job, it is plausible that doing the task well and showing up on time are substitutes. Doing the identification task well presumably requires more time and may therefore make it harder for workers to complete the task on time. On the other hand, sloppy work and theft are unlikely to be substitutes. So the first prediction we propose is the following:

*Prediction 1 (disciplining effect) - Monitoring and incentives will increase effort in the monitored dimension and have negative spillover effects on dimensions for which efforts are substitutes.*

That is, we would expect monitoring to reduce sloppy work, possibly increase tardiness and to have no effect on theft.

Monitoring and incentives may also have a crowding-out effect on intrinsic motivation. One mechanism driving this crowding-out is through *reciprocity*. For a given level of effort, monitoring and incentives effectively reduce the expected payoff of a worker. Indeed, in the absence of monitoring, workers get the full payment for sure, while in the presence of monitoring, they only get it with a probability (which is lower or equal to 1). Compared to the *no monitoring* treatment, workers in the two monitoring treatments also infer additional costs because they have to collect the sure payoff and, if applicable, the remaining payoff at different dates. Workers may want to reduce the principal's payoff in order to reciprocate this reduction in their own expected payoff (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004). If reciprocity is due to a reduction in expected payoff, we expect the negative effect of monitoring and incentives to be increasing in the level of monitoring and incentives.

Monitoring and incentives (independent of the level) may also be perceived

as a signal of distrust and workers may reciprocate this distrust by caring less about the utility of the principal (Frey, 1993). If reciprocity is due to distrust, we expect the negative effect of monitoring to be independent of the level of monitoring and incentives.

In a multi-dimensional context, we expect workers to pick the least costly manner to retaliate. There are three ways in which workers can retaliate: (1) they can put less effort, (2) they can steal coins, (3) they can be late in returning the work material.<sup>5</sup> If they put less effort in the monitored task, they also get paid less in expectation. If they retaliate through other dimensions, they incur no monetary costs (in the case of theft they even incur a monetary gain), but may incur costs associated to breaking social norms. We conjecture that the costs of theft are higher than the costs of tardiness (due to a strong social and legal norm). Thus, the prediction we associate with reciprocity is the following:

*Prediction 2 (reciprocity) - Monitoring will not affect or may improve productivity in the monitored dimension (work quality), but will increase tardiness. If workers reciprocate a reduction in monetary payoff, we expect the negative effect of monitoring to increase in the level of monitoring, i.e. to be highest in the monitoring & strong incentives treatment. If workers reciprocate being distrusted, we expect the negative effect of monitoring to be independent of the level of monitoring and incentives.*

Another mechanism that has been proposed to explain crowding out effects is through *information*. There are two theories of what information monitoring could reveal. The first by Bénabou and Tirole (2003), who argue that monitoring could serve as a signal of task difficulty. Workers who are monitored infer that the task is difficult and as a consequence put less effort into it (Bénabou and Tirole, 2003). In a multi-dimensional effect, this would generate positive spillovers on the dimensions that are close substitutes. We therefore predict the following:

*Prediction 3 (information about task difficulty) - If monitoring provides in-*

---

<sup>5</sup>All of the experiments were conducted by the authors of this paper. Thus, tardiness is associated with costs to the authority deciding on monitoring mechanisms and can be considered as a method of reciprocating.

*formation about the task difficulty to workers, monitoring should decrease work quality. Substitution effects may lead to a decrease in tardiness, while theft is not affected by monitoring.*

Sliwka (2007) proposes that monitoring could reveal information about the peers' behaviour. In his model, monitoring signals the principal's expectations of the worker's propensity to behave counterproductively. Thus, monitoring work quality could signal that the principal expects a large fraction of workers to work sloppily. Workers who aim at behaving conform to their peers respond to this signal and choose to behave counterproductively as well, that is, they will also work sloppily. But how does this affect counterproductive behaviour in the other dimensions? The signal is obviously directly relevant for the dimension that is monitored, but the question is whether monitoring could also give a signal to workers about their peers' propensity to show up late or steal. This will very much depend on whether there is a positive correlation in behaviour across the different dimensions. If people believe that showing up late and stealing money are correlated with doing a bad job at the identification task, then it could be that monitoring one dimension (work quality) provides information about the behaviour of the peers along the other dimensions as well.

*Prediction 4 (information about peers) - If monitoring provides information about peers, monitoring should decrease work quality but may also increase tardiness and theft.*

It is worth pointing out that it is the multi-dimensional aspect of productivity, combined with the asymmetric monitoring, that allows us to differentiate between these mechanisms. In a standard set up with only one dimension of productivity, we would only observe the net effect of monitoring (disciplining minus crowding out effects) and we would not be able to say whether the crowding effects are driven by negative reciprocity or by information.

## 4 Results

### 4.1 Summary statistics

Table 2 shows summary statistics for the behaviours of interest across the three treatments. First, the quality of work is on average better in the *monitoring & strong incentives* treatment than in the *no monitoring* and *monitoring & weak incentives* treatments. In fact, quality in the *no monitoring* and the *monitoring & weak incentives* treatments is very similar. In these two treatments, workers make 10 mistakes on average, while they make on average 7 mistakes in the *monitoring & strong incentives* treatment. The proportion of workers making more than 10 mistakes is largest in the monitoring treatment with *weak incentives* (30%), followed by the *no monitoring* treatment (23%) and is lowest in the *monitoring & strong incentives* treatment (16%). These results indicate that monitoring only improves work quality when the incentives are high enough.

Second, tardiness varies substantially across treatments. The average delay is larger and the average advance smaller in the two monitoring treatments compared to the no monitoring treatment. The average delay is highest in the *monitoring & strong incentives* treatment.

Finally, 10% of the participants (9 people out of 91) steal money. Overall, it seems that theft is motivated by the collectors' value of coins, rather than the nominal value of circulating coins. Subjects especially steal coins that are only rarely found in Germany, such as coins from the Vatican, Slovenia, or Slovakia. These are coins that have a higher collectors' value than their actual nominal value. For example, in three cases a 50 cent coin from the Vatican is stolen. On the German ebay platform this coin is sold for €3 (plus shipping). In two cases subjects replaced coins with a higher collectors' value by other coins with the same nominal value. We categorise these acts as theft as they did not inform us that they replaced the coins. In addition to the two coins that were replaced, 12 coins were stolen, resulting in a nominal loss of €1.53. We observe no variation in the prevalence of theft across treatments.

We do not observe a significant correlation between the different forms of

counterproductive behaviour. Comparing individuals who steal to those who do not steal, we do not detect any significant difference in tardiness (U-test,  $p > 0.10$ , two-tailed) or the number of mistakes (U-test,  $p > 0.10$ , two-tailed). Further, we find a positive but insignificant correlation between the number of mistakes and the delay in minutes (Spearman Correlation,  $\rho = 0.144$ ,  $p = 0.17$ , two-tailed).

**Table 2 Summary of the results**

	no monitoring	monitoring & weak incentives	monitoring & strong incentives
<b>Work quality</b>			
avg. no. of mistakes	10.23 (16.23)	9.97 (13.45)	6.90 (10.93)
no. of subjects with 0-2 mistakes	37%	40%	35%
no. of subjects with 3 mistakes or more	63%	60%	65%
no. of subjects with 10 mistakes or more	23%	30%	16%
<b>Tardiness</b>			
avg. delay in minutes	0.77 (6.29)	4.63 (15.48)	9.84 (38.93)
avg. advance in minutes	152.60 (584.90)	7.50 (17.04)	26.29 (130.31)
<b>Theft</b>			
no. of subjects who stole coins	3	3	3
<b>Work time</b>			
avg. work time in minutes	111.83 (42.58)	112.50 (45.04)	124.45 (47.69)

Mean values with standard deviation in parenthesis.

## 4.2 Regression analysis

We now turn to a regression analysis of the number of mistakes and tardiness (we do not analyse theft since there is no variation across treatments), which allows us to control for some observable characteristics of the workers. Starting with work quality, Col. (1) shows the results of a Poisson regression.<sup>6</sup> We find that there are 40% less mistakes under the *monitoring & strong incentives* treatment than under *no monitoring*. On the other hand, we observe no significant differences between *monitoring & weak incentives* and *no monitoring*.

We find that the time difference (i.e. difference between the actual and the appointed return time) is significantly larger in both monitoring treatments compared to the treatment without monitoring (Col. (2)). One question here is

<sup>6</sup>The distribution of the number of mistakes is not normal. There is a substantial fraction of zeros and small positive values. In those cases, count data models are more appropriate. This is why we use a Poisson regression.

whether the difference is driven by a distortion effect, i.e. workers show up late because they put more effort into the identification task. We asked participants how much time they spent on the task and the average reported working time was 112 minutes for the *no monitoring* treatment, 113 minutes for the *monitoring & weak incentives* treatment and 124 minutes for the *monitoring & strong incentives*, with none of these differences being statistically significant (U-test,  $p > 0.10$ , two-tailed). We find that differences between treatments with respect to punctuality exist even when controlling for the total number of mistakes and the reported work time (Col. (3)). There is some evidence that part of the delay in the *monitoring & strong incentives* treatment could be due to extra care in the task (the difference in delay falls from 143 minutes to 132 minutes, which corresponds exactly to the additional amount of time spent on the task.) But neither the total number of mistakes, nor the reported working time appear to be correlated with the delay at all.

Col. (4-7) look at the probability of completing the task early or late. We only find significant differences in the probability of being late. Participants are 35% and 36% more likely to be late under *monitoring & weak incentives* and *monitoring & strong incentives*, respectively. The effects of monitoring remains identical if we control for the total number of mistakes and the reported work time (Col. 5 and 7), which shows that there is no relationship between effort in the identification task and tardiness. Thus, our results are most supportive of a reciprocity interpretation. Workers perceive monitoring as unkind and retaliate by putting less effort in the dimension that is the least costly for themselves (both in monetary and non-monetary terms). Since we find that the extent of retaliation is independent of the strength of the incentives, it seems that workers respond more to the mere presence of a monitoring technology than to the loss in expected monetary terms. It does not matter how strong the incentives are, the workers just dislike being distrusted.

One important question is whether it pays to monitor workers. Clearly, this is not the case when we compare *monitoring & weak incentives* to *no monitoring*. There are no significant differences in the quality of work and tardiness

increases with monitoring. Introducing monitoring with strong incentives, on the other hand, improves work quality, but also increases tardiness. In that case, depending on the opportunity cost of time, it could be that monitoring pays off.

**Table 3 Regression analysis**

	Number of mistakes (Poisson)	Time difference (OLS)	Early (Probit)	Late (Probit)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
monitoring & weak	.003 (.082)	145.754 (65.137)***	145.154 (64.896)**	-.137 (.119)	-.136 (.119)	.348 (.132)**	.356 (.132)**
monitoring & strong	-.407 (.089)***	143.140 (64.791)***	131.754 (65.157)**	-.105 (.120)	.066 (.106)	.363 (.129)***	.389 (.131)**
female	-.298 (.074)***	37.005 (54.098)	36.556 (53.897)	.070 (.105)	.066 (.106)	.000 (.103)	.015 (.104)
total mistakes	-	-	1.341 (1.997)	-	-.001 (.004)	-	.005 (.004)
reported work time	-	-	.772 (.602)	-	.000 (.001)	-	.000 (.001)
constant	2.435 (.062)***	-185.343 (55.355)***	-266.523 (83.951)***	-	-	-	-
(Pseudo) R <sup>2</sup>	.027	.082	.047	.014	.016	.081	.098

\*\*significance at  $p < 0.05$ , \*\*\*significance at  $p < 0.001$

Marginal effects are reported for Probit estimates in Col. (4)-(7)

## 5 Discussion and conclusions

This paper provides field evidence on the effect of monitoring in a context where productivity is multi-dimensional and only one of the dimensions (work quality) is monitored. We find that introducing monitoring with weak incentives is inefficient. There is no significant improvement in work quality and tardiness increases significantly. Monitoring with strong incentives is more effective. The number of mistakes falls substantially, but at the same time the adverse effects on the other dimensions are as large as in the monitoring treatment with weak incentives.

Overall, these results are in line with a model of reciprocal behaviour. Workers choose to punish the principal for monitoring (and therefore distrusting) them, but choose to do this through dimensions that have low costs for them.

Theft is presumably much more costly to the workers (in moral terms) than tardiness and putting less effort in the monitored task involves direct costs. Tardiness on the other hand does not involve high moral costs and has no financial consequences for the workers.

Based on these results, we conclude that introducing a monitoring technology only pays off if (1) the incentives associated with passing the checks are high and (2) the dimensions that cannot be monitored either entail high moral costs or the relative gains in productivity in the monitored dimension more than compensate for the losses in other dimensions.

These findings relate more broadly to the literature on adverse effects of incentives (see Gneezy et al., 2011 for a recent review) and the adverse effects of control (Falk and Kosfeld, 2006) and monitoring (Frey, 1993). In line with this literature, we find that weak monitoring and weak incentives are less effective than no monitoring and no incentives.

*Acknowledgments:* We thank the team of the chair in e-Business at the University of Magdeburg, in particular Claudia Gorylla, for helping us in the task of checking identified coins and recruiting participants. We also benefited from valuable comments from participants at the European Workshop on Experimental and Behavioral Economics in Frankfurt 2013, the Royal Economic Society 2013 Conference, the 2013 Florence Workshop on Behavioural and Experimental Economics, and the Dijon Workshop on Understanding Employees' Dishonesty Behaviours in the Workplace. We further thank Uri Gneezy, Bernd Irlenbusch and Karim Sadrieh for valuable suggestions. The support and hospitality of the Rady School of Management at the UC San Diego are acknowledged, where the second author was a visiting scholar while part of this research was conducted.

## References

- [1] Association of Certified Fraud Examiners, 2012. 2012 Report to the Nations on Occupational Fraud and Abuse. Available at [http://www.acfe.com/uploadedFiles/ACFE\\_Website/Content/rtn/2012-report-to-nations.pdf](http://www.acfe.com/uploadedFiles/ACFE_Website/Content/rtn/2012-report-to-nations.pdf), last access 07.05.2013.
- [2] Andreoni, J., J. Miller. Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2) 737–753.
- [3] Baker, G. P. 1992. Incentive Contracts and Performance Measurement. *Journal of Political Economy* 100(3) 598–614.
- [4] Becker, G. S. 1968. Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76(2) 169–217.
- [5] Belot, M., M. Schröder. 2013. Sloppy Work, Lies and Theft: A Novel Experimental Design to Study Counterproductive Behaviour. *Journal of Economic Behavior and Organization*, Available online.
- [6] Bénabou, R., J. Tirole. 2003. Intrinsic and Extrinsic Motivation. *Review of Economic Studies* 70 489–520.
- [7] Boly, A. 2011. On the Incentive Effects of Monitoring: Evidence from the Lab and the Field. *Experimental Economics* 14(2) 241–253.
- [8] Dickinson, D., M.-C. Villeval. 2008. Does Monitoring Decrease Work Effort? The Complementary Between Agency and Crowding-Out Theories. *Games and Economic Behavior* 63(1) 56–76.
- [9] Dufwenberg, M., G. Kirchsteiger. 2004. A Theory of Sequential Reciprocity. *Games and Economic Behavior* 47 268–98.
- [10] Falk, A., M. Kosfeld. 2006. The Hidden Costs of Control. *American Economic Review* 96(5) 1611–1630.

- [11] Fehr, E., K. M. Schmidt. 1999. A Theory of Fairness, Competition and Cooperation. *Quarterly Journal of Economics* 114 817-868.
- [12] Fisman, R., E. Miguel. 2007. Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets. *Journal of Political Economy* 115(6) 1020–1048.
- [13] Frey, B. S. 1993. Does Monitoring Increase Work Effort? The Rivalry with Trust and Loyalty. *Economic Inquiry* 31(4) 663–670.
- [14] Gneezy, U., S. Meier and P. Rey-Biel. 2011. When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives* 25(4) 191-210.
- [15] Gneezy, U., A. Rustichini. 2000. A Fine is a Price. *Journal of Legal Studies* 29(1) 1-18.
- [16] Grasmick, H. G., R. J. Bursik, JR. 1990. Conscience, Significant Others, and Rational Choice: Extending the Deterrence Model. *Law and Society Review* 24(3) 837–861.
- [17] Holmstrom, B., P. Milgrom. 1991. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics and Organization* 7 24-52.
- [18] Nagin, D. S., J. B. Rebitzer, S. Sanders, L. J. Taylor. 2002. Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment. *American Economic Review* 92(2) 850-873.
- [19] Rabin, M. 1993. Incorporating Fairness into Game Theory and Economics. *American Economic Review* 83(5) 1281–302.
- [20] Robinson, S. L., R. J. Bennett. 1995. A Typology of Deviant Workplace Behaviors: A Multidimensional Scaling Study. *The Academy of Management Journal* 38(2) 555–572.

- [21] Sliwka, D. 2007. Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes. *American Economic Review* 97(3) 999–1012.