The Insider Threat: An Economic Analysis of Monitoring Instruments*

Emmanuelle Auriol
Toulouse School of Economics

Stefanie Brilon
Johannes-Gutenberg Universität Mainz

March 2023

Preliminary Draft - Do not quote or circulate

Abstract

We consider employees' incentives for working vs. engaging in criminal or destructive behavior that damages their employer or others. To prevent damages, organizations may want to deter such behavior entirely or at least mitigate it by using a mix of effort monitoring, monetary incentives, policing and ex ante screening of employees. In this setup, we derive optimal contracts and discuss what determines the optimal choice of monitoring instruments.

Keywords: monitoring, screening, employee theft, dishonesty, crowding out, intrinsic motivation, sabotage,

JEL Classification: D23, J33, L31, M54

*Emmanuelle Auriol: Toulouse School of Economics, emmanuelle.auriol@tse-fr.eu; Stefanie Brilon: Johannes Gutenberg-Universität Mainz, sbrilon@uni-mainz.de.

1 Introduction

In the economic literature, monitoring is usually considered as an alternative to financial incentives to increase employees' effort provision. It is rarely explicitly modelled since most monitoring is imperfect and firms may therefore prefer to rather pay a wage premium to incite effort, as suggested by the efficiency wage theory in the tradition of Shapiro and Stiglitz (1984). Nevertheless, firms invest substantial resources in monitoring and controlling workers.[1]

We argue that a considerable parts of these resources go towards what we call policing, i.e., preventing theft, fraud and other kinds of damaging and anti-social behavior.

To make this point, we consider different monitoring instruments and discuss how these instruments are combined under circumstances that differ by (i) the extent of damage done to the firm, (ii) the payoff for deviant or criminal behavior, (iii) the punishment or fines that can be imposed on perpetrators, and (iv) the share of people who are expected to engage in such behaviors.

Our paper thus is linked to the literature on rational cheating, in particular to the field experiment by Nagin et al. (2002), who consider the example of workers in a call center. Like Nagin et al. (2002), we explicitly model workers' benefits from damaging behavior, but we allow for a differentiation in technologies, where monitoring is aimed at "standard" effort provision, whereas policing aims to detect misbehavior.

This employee misbehavior may be anything from theft and reporting fraud to harassment of any form, sabotage or even terrorism. All of these behaviors lead to some form of benefit for the perpetrator, either monetary or non-monetary, can be punished and cause a damage for the firm, other employees

[1] Schmitz (2005) presents an efficiency wage model, where employers may prefer to invest in costly monitoring rather than leaving a higher rent to the employee, and discusses under which circumstances privacy laws that prevent close monitoring may actually be wellfare improving.

and even society at large.[2]

Indeed, employee misbehavior leads to substantial costs. To give a few examples: CNBC reported in 2017, that workplace theft cost US businesses $ 50 billion per year.[3] In their 2022 report, the Association of Certified Fraud Examiners estimates that fraud costs organisations 5% of their revenue, with theft being the most common form of fraud, and financial statement fraud the least common but most costly.[4] Smaller organisations may be at particular risk to suffer losses from this kind of behavior, as reported by the US insurance company Hiscox, who in their 2016 embezzlement study report a majority of cases for firms with fewer than 100 employees, with median losses close to $300,000.[5] This finding is also confirmed by the AFCE, which reports that organizations with the fewest employees report the highest losses (ACFE, 2022).

Another costly form of employee misbehavior is sexual harassment. In its 2019 report for the Australian Human Rights Commission, the consultancy Deloitte estimates that "in 2018, workplace sexual harassment cost $2.6 billion in lost productivity and $0.9 billion in other financial costs. Each case of harassment represents around 4 working days of lost output. Employers bore 70% of the financial costs, government 23% and individuals 7%. Lost well-being for victims was an additional $250m, or nearly $5000 per victim on average."[6] Statistics published by the U.S. Equal Employment Opportunity Commission also suggest that sexual harassment charges may be quite costly

[2]We do not aim to model corporate crime that tends to benefit the organization and damage societies at large. As Van Erp (2018) puts it: "($\cdots$ although corporate crimes are ultimately committed by individual members of an organization, they have more structural roots, as the enabling and justifying organizational context in which they take place plays a defining role" Van Erp (2018), p.1.

[3]See []https://www.cnbc.com/2017/09/12/workplace-crime-costs-us-businesses-50-billion-a-year.html .

[4]See the ACFE Report to the Nations (2022) , downloaded from https://www.acfe.com/report-to-the-nations/2022/ .

[5]See Hiscox (2016), downloaded from https://www.hiscox.com/newsroom/research-and-insights.

[6]See Deloitte (2019), downloaded from https://www2.deloitte.com/au/en/pages/economics/articles/economic-costs-sexual-harassment-workplace.html.

for organisations.[7] Furthermore, Bouzzine and Lueg (2022) show in an event study that in the wake of the MeToo movement, misconduct even by a single individual has had an impact on stock returns of affiliated organizations.

The topic of employee misconduct has received little attention in the theory literature in economics or business studies, with a few notable exceptions. For instance, Greve et al. (2010) provide a very interesting discussion of what encompasses misconduct and how this may be linked to economic theories. Shadnam and Lawrence (2011) consider misconduct through the lens of institutional theory. The extent of employee misconduct may also be affected by factors exogenous to the firm. Krammer et al. (2022) discuss the impact of income inequality in a country on the severity of crimes experienced by businesses.

In our paper, we adapt a standard principal agent model to include different forms of employee ethics at work. We consider different types of monitoring with the purpose of measuring workers' performance and preventing misconduct or at least reducing its scope. To do so, we propose a simple model, where agents can engage in productive or destructive behavior. As in standard multitasking models, if one task is monitored more closely and hence likely to be rewarded more, then this may lead to a reallocation of efforts.[8]

Furthermore, we allow for crowding out effects of monitoring. That is, we allow for negative effects of monitoring, when control is perceived as distrust and may thus lead to crowding out of motivation, similar as in von Siemens (2013), Frey and Jegen (2001), Puranam and Vanneste (2009), and Schweitzer et al. (2016). Dickinson and Villeval (2008) have shown experimentally that although monitoring increases agents' effort, it also leads to crowding out, in particular if there is a personal connection between the principal and the agent. Falk and Kosfeld (2006) show that when principals restrict an agent's choice set, the 'controlled' agent responds by being less generous towards the

[7] See https://www.eeoc.gov/statistics/charges-alleging-sex-based-harassment-charges-filed-eeoc-fy-2010-fy-2020 .

[8] Belot and Schröder (2016) test this in a field experiment, and show that monitoring in one dimension indeed may lead to lower performance in other dimensions.

principal.

We discuss our findings in the light of technical innovations that change how workers are monitored, in particular since the Covid pandemic has made remote work more common.[9] As reported by the New York Times in August 2022, 8 out of the 10 biggest employers in the US track productivity metrics of their employees[10], which again highlights the importance of monitoring, but also raises new questions about the effect of control on motivation.

2 Model

2.1 Basic Setup

We consider a principal-agent model, where an organization (the principal) needs to incentivize a worker (the agent) to provide a productive effort, which is not observable. To simplify our analysis of the moral-hazard problem, we assume that agents have an outside utility of zero.[11]

Agents can choose their level of effort for two activities: They can provide a productive work effort $e \in [0, 1]$ that increases the overall output of the organization by Δq, or they can choose a destructive effort $d \in [0, 1]$ that yields benefit θ to the worker, but damage D to the firm. Efforts are not observable to the organization, which leads to a moral-hazard problem. As it is standard in multi-tasking models (see, e.g., Holmström and Milgrom, 1991; Hellmann and Thiele, 2011), we assume that the cost of effort is a function (increasing and convex) of the sum of both efforts: $\psi(e + d)$ with $\psi' > 0,\ \psi'' > 0$.

[9] See, for instance, Nyberg et al. (2021) for an overview of challenges raised by remote work. West (2021) provides a short overview over existing tools and technologies for the surveillance of remote workers.

[10] See []https://www.nytimes.com/interactive/2022/08/14/business/worker-productivity-tracking.html.

[11] We also ignore the possibility of positive intrinsic motivation. Readers who are interested in how optimal contracts change if we let the reservation utility of agents vary and/or we allow for positive intrinsic motivation will find more information on both aspects in Auriol and Brilon (2014).

If the agent engages in destructive behavior, she gets some private benefit θ from this behavior with probability d, but risks receiving a punishment which provides her with disutility K. The private benefit may either be some intrinsic bonus from misbehaving or it may be a monetary benefit for instance from a criminal activity such as stealing. We assume that the punishment K is exogenous. Since the destructive behavior is an imoral, possibly criminal, act, K can be thought of as the punishment administered through the legal system and therefore cannot be determined by the organization.

We allow for two monitoring technologies, one aimed at controlling effort provision, and the other aimed at detecting misbehavior: Positive effort is observed with probability m which corresponds to the level of monitoring in the organization. Destructive efforts are observed with probability p which describes the level of "policing" in the organization. The two forms of monitoring come at a cost described by the function $C(m, p)$, which is increasing and convex for both monitoring technologies: $C_k(m, p) > 0,\ C_{kk}(m, p) \geq 0$ where the index indicates the partial derivatives with respect to $k = m, p$.

To sum up, all agents produce some level of output q. If the agent provides effort e, she produces additional output Δq with probability e. The organization offers a base wage w to the worker and a bonus t in case she is successful at producing Δq, which is observed by the organization with probability m. Destructive behavior, which yields a payoff θ for the agent and a damage D for the organization with probability d, is also monitored by the organization and detected with probability p, which leads to a punishment K for the worker.

A worker's utility then is described by the following function:

$$u = w + mte + (\theta - pK)d - \psi(e + d) \ . \tag{1}$$

The profit of the organization given a worker's effort choice (e, d) is:

$$\pi = q + (\Delta q - mt)e - w - Dd - C(m, p) \ . \tag{2}$$

The principal has to consider the following limited liability *(LL)*, participation *(PC)* and incentive *(IC)* constraints when designing an optimal contract to maximize (2):

$$\begin{aligned} (LL) \quad & w \geq 0, \\ (PC) \quad & u \geq 0 \ . \\ (IC_e) \quad & e = \arg\max_{e \in [0,1]} \{u\} \\ (IC_d) \quad & d = \arg\max_{d \in [0,1]} \{u\} \end{aligned}$$

Lemma 1 *For a reservation utility of zero, the optimal basic wage is zero, i.e.,* $w^* = 0$.

Since the basic wage has no incentive or deterrence effect, it merely impacts the participation constraint of the agent. For an outside utility of zero, an agent is willing to accept the contract as long as the expected bonus is greater or equal to zero, thus making it optimal from the point of view of the firm to offer no fixed wage component.

Since this Lemma also holds for the maximization problems in the following sections, we set $w = 0$ to save on notation.

2.2 Benchmark contracts without bad workers

Let us start by analysing our model if agents do not get any benefit from destructive behavior, i.e., for $\theta = 0$.[12] Since in this case, there is no benefit from misbehaving, the negative effort is $d = 0$.

Then the utility function of the worker is given by

$$u = mte - \psi(e) \ . \tag{3}$$

[12]This follows the analysis in Auriol and Brilon (2014).

Optimizing this utility function with respect to effort e, the worker chooses an optimal work effort $\epsilon(mt)$ solution to the following necessary and sufficient first order condition:[13] $\psi'(e) = mt$. We deduce that

$$\epsilon(mt) = \psi'^{-1}(mt). \tag{4}$$

The principal maximizes the following objective function when hiring an individual worker:

$$\pi = q + (\Delta q - mt)\epsilon(mt) - C(m, p) \ , \tag{5}$$

where $q \geq 0$ is the basic output produced by the worker, $\Delta q \geq 0$ the extra output that depends on the worker's effort, $\epsilon(mt)$ is the optimal effort choice of the worker and $C(m, p) \geq 0$ is the per capita cost of monitoring and policing. We show the following result.

Proposition 1 *Let $\epsilon(x)$ be defined as in (4). When there is no threat of destructive behavior, both policing and monitoring are set at the minimum level and the optimal wage contract is: $\theta = 0 \ \Rightarrow \ (p^*, m^*) = (0, \underline{m})$ and t^* such that*

$$\Delta q = \underline{m}t + \frac{\epsilon(\underline{m}t)}{\epsilon'(\underline{m}t)} \tag{6}$$

Proof. See Appendix A.1 ■

To save on costs, the principal prefers to provide incentives through the bonus rather than through monitoring so that $m^* = \underline{m}$. The optimal bonus t^* solution to (6) is such that the marginal benefit of stimulating effort through t from the organization's point of view is equal to the marginal cost of providing the incentive. Moreover since workers do not have any benefit from misbehaving, the negative effort is $d = 0$, and policing is not needed, i.e., the optimal level of policing is zero: $p^* = 0$.

[13] Since ψ is convex, the utility function is concave in effort e and the FOC is sufficient.

2.2.1 Quadratic Example

Consider the effort function $\psi(e+d) = (e+d)^2 a/2$. To rule out corner solutions in the following, we assume that a is sufficiently large:

Assumption 1 $a > \Delta q$.

The optimal effort of a motivated individual then is[14]

$$e^* = \Delta q / a \,. \tag{7}$$

Plugging this into the principal's maximization problem, the resulting optimal contract is $w^* = 0, \quad t^* = \Delta q/(2\underline{m}), \quad m^* = \underline{m}, \, p^* = 0$.

To make sure that this contract is profitable, we have to assume that monitoring costs are not too high compared to the payoff from producing:

Assumption 2 $C(\underline{m}) < min\{q, \Delta q^2/(4a)\}$.

2.3 Bad Workers and Automatic Deterrence

In the following, we allow for an individual benefit $\theta > 0$ from destructive actions, which can be interpreted as negative intrinsic motivation or payoff from criminal behavior such as theft. We assume that there is a share β of bad workers in the population with this characteristic. The rest of the population, i.e., share $(1-\beta)$, has no such intrinsic benefit from behaving in a destructive way.

Given our assumptions on the effort cost function, a worker of type θ will maximize her utility by choosing effort vector (e, d) as a solution to:

$$\max_{(e,d)} u = mte + (\theta - pK)d - \psi(e+d) \,. \tag{8}$$

[14] Assumption 1 implies that we obtain an interior solution for the optimal effort level.

The worker chooses an optimal work effort e^o such that the following first order condition holds:

$$mt - \psi'(e+d) \leq 0 \ . \tag{9}$$

The optimal level of destructive effort d^o is chosen such that

$$\theta - pK - \psi'(e+d) \leq 0 \ . \tag{10}$$

Unless $mt = \theta - pK$ it is impossible that (9) and (10) bind simultaneously. At the optimum the worker chooses either $e^o > 0$ or $d^o > 0$

A worker will choose to provide effort e rather than effort d if his net benefit from doing so is higher, i.e., if $u(e^o, d = 0) \geq u(e = 0, d^o)$. If we assume the effort cost function $\psi(e + d)$ to be symmetric in e and d then workers will choose a normal work effort rather than a destructive effort if

$$mt \geq \theta - pK \ . \tag{11}$$

If this deterrence constraint holds, then there is automatic deterrence of bad behavior.

Consider the benchmark contract derived in the previous section, i.e., the optimal contract without bad workers. It has no policing and will lead to automatic deterrence if

$$\theta \leq \tilde{\theta} := \underline{mt}^* \ . \tag{12}$$

For higher levels of θ, either the level of monitoring, policing or the bonus payment for high effort will have to increase to deter bad behavior. We analyze this problem in the next section.

3 Ex Post Monitoring

3.1 Full Deterrence

If $\theta > \tilde{\theta}$, bad behavior is not automatically deterred. Instead, the principal has to offer a contract that implements the deterrence constraint derived

in (11) in order to make bad behavior unattractive. Hence, the principal's maximization problem can be described as follows:

$$\begin{aligned} \max_{w,m,t,p} \quad & \pi = q + (\Delta q - mt)e - C(m,p) \\ s.t. \quad (PC) \quad & u \geq 0 \ , \\ (IC) \quad & e = \arg\max_{e \in [0,1]} \{u\} \\ (DET) \quad & mt \geq \theta - pK \ , \end{aligned}$$

The following proposition summarizes the optimal contract with full deterrence.

Proposition 2 ***(Carrot and stick)*** *Let $\epsilon(x)$ be defined as in (4) and let $\tilde{\theta}$ be defined in (12). For $\theta > \tilde{\theta}$, the optimal contract with full deterrence is given by $m^{det} = m^* = \underline{m}$ and t^{det} and p^{det} such that*

$$\Delta q = \underline{m}t + \frac{\epsilon(\underline{m}t)}{\epsilon'(\underline{m}t)} - \frac{\partial C}{\partial p}\frac{1}{K}\frac{1}{\epsilon'(\underline{m}t)} \tag{13}$$

$$p^{det} = \frac{\theta - \underline{m}t^{det}}{K} \ . \tag{14}$$

Proof. See Appendix A.2 ■

By comparing the optimality condition without bad workers as given in (6) and with full deterrence of bad workers as given in (13), we can see that the latter has an additional negative term on the right hand side which decreases in the marginal cost of policing. That is, the resulting bonus level t for which the condition is binding is decreasing in p. Otherwise, the same reasoning as before applies, namely that the bonus level is chosen such that the marginal benefit from effort is equal to the marginal cost. However, the benefit now includes the damage avoided through the deterrence of bad workers, so if we rearrange the terms of Equation (13) accordingly, we get

$$\Delta q + \frac{\partial C}{\partial p}\frac{1}{K}\frac{1}{\epsilon'(\underline{m}t)} = \underline{m}t + \frac{\epsilon(\underline{m}t)}{\epsilon'(\underline{m}t)} \ .$$

It is easy to see, that benefits on the left hand side now are higher than in (6) and thus justify a higher bonus level than without bad workers.

A contract with full deterrence of bad behavior thus implies both an optimal bonus and the optimal policing level above the benchmark contract without bad behavior. In other words, it combines carrot and stick to keep workers in line. By contrast, monitoring of work effort continues to be as low as possible.

We would therefore expect firms to spend the least amount possible on monitoring of work effort, a result that is in accordance with the literature going back to Becker (1968). Both the optimal bonus payment t^{det} and the level of policing decreases in K and increases in θ. If the exogenous fine K is higher and hence a bigger deterrent, then even a lower level of policing leads to complete deterrence. That is, for given θ, we would expect smaller incentives for good behavior and less policing on offenses that carry large fines or even imprisonment. At the same time we would expect higher incentives for good behavior and more policing if offenses lead to little exogenous punishment.

Lemma 2 *If one can distinguish monitoring technologies by their intention – inciting work effort vs. policing – then only the latter increases above its minimal level if the payoff for bad behavior is high.*

That is, even though $C(m, p)$ is increasing and convex in both its arguments, only policing is used as a tool against bad behavior. If we see high monitoring costs in firms, these are likely to actually be policing costs.

How do our results change if there are spill-overs between monitoring technologies? For instance, electronic surveillance to reduce fraud may at the same time make it easier to monitor work effort. Or, the other way round, implementing performance measures may make it easier to prevent fraud.

In our model, these spill-overs could be introduced through the monitoring cost function. If the cross-derivative of the cost function $C''_{mp} < 0$, then this implies a lower cost of policing the higher the monitoring level and vice versa.

Lemma 3 *If there are spill-overs between monitoring technologies in the sense that increasing m lowers the marginal cost of p and vice versa, then the optimal contract with full deterrence is characterized by minimal monitoring $\underline{m}$, but more policing and a lower bonus payment compared to the case without cost spillovers.*

To see this, consider again the proof of Proposition 2. There, we have shown that the optimal monitoring level with full deterrence corresponds to the minimal level of monitoring $\underline{m}$, no matter the precise form of the monitoring cost function. However, if there are spill-overs, the cost of policing for a given level of monitoring is smaller than without spill-overs, leading to an increase in the optimal level of policing. At the same time, if monitoring stays the same and policing increases, then the deterrence constraint is still binding even with lower bonus payments for good behavior, i.e., t^{det} decreases compared to the case without spill-overs. In the next subsection, we provide an example to illustrate this point.

3.1.1 The Quadratic Example

Let us again consider our results so far for the specific example from Section 2.2.1. Given the effort cost function $\psi(e+d) = (e+d)^2 a/2$, the optimal effort levels are $e^o = mt/a$ and $d^o = (\theta - pK)/a$. Hence, the benchmark contract $m^* = \underline{m}, t^* = \Delta q/(2\underline{m}), p^* = 0$ is enough to deter bad behavior as long as the payoff from such behavior is $\theta < \tilde{\theta} = \Delta q/2$. For higher θ, the optimal contract with full deterrence is given by $m^{det} = \underline{m}$, and

$$\begin{aligned} t^{det} &= \frac{\Delta q}{2\underline{m}} + \frac{a}{2\underline{m}K}\frac{\partial C}{\partial p}\,, \\ p^{det} \quad \text{s.t.} \quad & \frac{a}{2K}\frac{\partial C}{\partial p} + pK = \theta - \frac{\Delta q}{2}\,. \end{aligned}$$

That is, both the bonus payment t and the policing level p are higher than in the benchmark contract, whereas effort monitoring stays the same.

To make our example even more specific, let us assume the following monitoring cost function:

$$C(m,p) = \frac{1}{2}(m^2 + p^2 - 2\gamma mp) \ , \tag{15}$$

where the parameter γ measures the spillovers between m and p. The higher γ, the lower is the cost of p for a given level of m and vice versa. With this cost function, the optimal contract with full deterrence is characterized by

$$\begin{aligned} m^{det} &= \underline{m} \ , \\ p^{det} &= \Big(\theta - \frac{\Delta q}{2}\Big)\frac{2K}{a+2K^2} + \frac{a\gamma\underline{m}}{a+2K^2} \ , \\ t^{det} &= \frac{\Delta q}{2\underline{m}} + \frac{a}{2\underline{m}K}(p^{det} - \gamma\underline{m}) \\ &= \frac{\Delta q}{2\underline{m}} + \frac{a}{a+2K^2}\Big(\frac{2\theta - \Delta q}{2\underline{m}} - \gamma K\Big) \end{aligned}$$

.

That is, the higher the cost spill-over γ, the higher is the optimal policing level, since policing becomes relatively cheaper for a given level of standard monitoring m. At the same time, the optimal bonus needed for deterrence is decreasing in the level of policing, and thus becomes lower if there are cost spillovers.[15]

3.2 Partial Deterrence

If the damage D is relatively low or if there are very few workers who are inclined to misbehave, then some destructive behavior may be tolerable in equilibrium. Let β be the share of "bad" workers, i.e., workers with a benefit from destructive effort $\theta > \tilde{\theta}$ as defined in (12). The payoff of the principal then corresponds to

$$\pi = q + (1-\beta)(\Delta q - mt)e - \beta Dd - C(m,p) \ .$$

[15]Note that we need to make a number of assumptions on the parameters to get an interior solution for the level of policing p. In particular, a and K need to be sufficiently large, and $0 < \gamma < 1/\underline{m}$.

Workers will choose their optimal level of effort such that the marginal benefit of effort corresponds to the marginal effort cost. That is, the share $1-\beta$ of workers who provide normal work effort will choose their optimal effort level e^o such that $mt = \psi'$, whereas the share of workers who provide a destructive effort will choose their optimal effort level d^o such that $\theta - pK = \psi'$.

Under these conditions, the following proposition describes the optimal contract with partial deterrence:

Proposition 3 ***(Stick without carrot)*** *The optimal contract with partial deterrence of employee misbehavior is given by* $m^{part} = \underline{m} = m^*$, $t^{part} = t^*$ *and* p^{part} *such that*

$$-\beta D \frac{\partial d^o}{\partial p} = \frac{\partial C}{\partial p}$$

Proof. See Appendix A.3 ■

Under partial deterrence, the overall expected payoff for good behavior is thus as in the benchmark model without bad workers, i.e., $m^{part}t^{part} = m^*t^*$. But in contrast to this benchmark, there is now a positive level of policing to at least partially deter destructive behavior. This leads to lower expected damage, as the optimal destructive effort d^o is decreasing in p, but not zero damage.

3.2.1 Example

Recall that the optimal effort levels are $e^o = mt/a$ and $d^o = (\theta - pK)/a$, given the effort cost function $\psi(e+d) = (e+d)^2 a/2$. The resulting optimal contract with partial deterrence therefore is $m^{part} = \underline{m} = m^*$, $t^{part} = t^* = \Delta q/(2\underline{m})$ and p^{part} such that

$$\frac{\beta DK}{a} = \frac{\partial C}{\partial p}$$

If we furthermore assume the monitoring cost function to be as in (15), then the optimal level of policing with partial deterrence corresponds to $p^{part} = (\beta DK)/a + \gamma \underline{m}$.

4 Crowding Out of Effort

Previous research (see, among others, e.g., Falk and Kosfeld, 2006; Dickinson and Villeval, 2008) has shown that, to the extent that control is perceived as a sign of lack of trust, it can lower workers' motivation to provide effort. In our model, this effect is particularly relevant for the policing technology which by definition is implemented to prevent destructive behavior. Especially since these technologies (keystroke control, camera, voice recording, etc) often infringe on privacy. Workers may react to more policing by lowering their effort which we capture by adapting the utility function of workers. The new utility function is

$$u^{crowd} = w + mte + (\theta - pK)d - \psi(e + d) - f(p)e \; . \tag{16}$$

The function $f(p)$ captures the worker's disutility from policing. We assume that $f' > 0$. If $f(p) = 0$, then there is no crowding out of effort through control.

As a result of this modification, the optimal effort level e^o of the worker is such that

$$mt - f(p) = \frac{\partial \psi}{\partial e} \; . \tag{17}$$

That is, effort will be lower if p is higher, since the benefit from providing effort is lower. This also has an impact on the deterrence constraint which becomes

$$mt - f(p) \geq \theta - pK \; . \tag{18}$$

If this constraint holds, then the worker prefers to provide work effort e rather than a destructive effort d. This implies that if crowding out of effort is a real concern, then policing affects the deterrence constraint by making bad behavior less attractive, but also by decreasing the benefit of good behavior. For any positive level of policing, this constraint is more likely to hold the higher the punishment K, the lower the benefit from destructive behavior θ and the smaller the crowding out effect $f(p)$.

4.1 Full Deterrence and Crowding Out

So how does crowding out affect our results on optimal contracts? In the case of full deterrence, the principal's maximization problem can be stated as

$$\begin{aligned} \max_{w,m,t,p} \quad & \pi = q + (\Delta q - mt)e - C(m,p) \\ s.t. \quad (PC) \quad & u \geq 0 \ , \\ (IC) \quad & e = \arg\max_{e \in [0,1]} \{u^{crowd}\} \\ (DET) \quad & mt - f(p) \geq \theta - pK \ , \end{aligned}$$

where the incentive constraint is now referring to the modified utility function with crowding out as described in (16). Furthermore, the deterrence constraint has changed as explained above.

Taking into account Lemma 1 and all constraints, the resulting Lagrangian is

$$\max_{m,t,p} \quad \mathcal{L} = q + (\Delta q - mt)e^o - C(m,p) - w + \lambda(mt - f(p) - \theta + pK) \ ,$$

and we get these Kuhn Tucker conditions:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial t} &= (\Delta q - mt)\frac{\partial e^o}{\partial t} - me^o + \lambda m \leq 0 \\ \frac{\partial \mathcal{L}}{\partial m} &= (\Delta q - mt)\frac{\partial e^o}{\partial m} - te^o - \frac{\partial C}{\partial m} + \lambda t \leq 0 \\ \frac{\partial \mathcal{L}}{\partial p} &= (\Delta q - mt)\frac{\partial e^o}{\partial p} - \frac{\partial C}{\partial p} + \lambda\Big(K - \frac{df(p)}{dp}\Big) \leq 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= mt - f(p) - \theta + pK \geq 0 \\ \lambda &\geq 0 \\ & \lambda\Big(mt - f(p) - \theta + pK\Big) = 0 \ . \end{aligned}$$

(to be completed)

4.2 Partial Deterrence and Crowding Out

4.3 Discussion

Organizations may have an incentive to either camouflage their policing efforts for instance as technical or control requirements that are imposed by regulatory authorities. Or they may at least want to consider carefully how they communicate about any kind of control with employees.

That the framing of control efforts may have an effect on the motivation and work effort of employees has for instance been considered in an experiment by Christ et al. (2012) who show that positively framed formal controls may prevent the deterioration of trust.

5 Ex Ante Screening

If the damage from destructive behavior is particularly high, or if agents are difficult to monitor once they have been hired, organizations may opt for ex ante screening of agents, in order to avoid hiring agents that are particularly likely to engage in destructive behavior.

Examples for ex ante screening can be found in jobs ranging from airport baggage handling to intelligence agencies and varies in intensity and scope. Sometimes, ex ante screening is often furthermore combined with recurring background checks and ex post monitoring of employees once they have been hired.

To incorporate ex ante screening in our model, we introduce a screening technology, which detects "bad" workers with probability s at a cost $S(s)$.

(Following sections: revision in process)

5.1 Perfect Ex Ante Screening

Let us suppose for the moment that the screening technology works perfectly. That is, if job candidates are screened before hiring, bad types are detected with certainty, that is $s = 1$. Ex post policing of bad workers therefore becomes superfluous ($p = 0$).

With screening, the optimization problem of the principal then corresponds to

$$\begin{aligned} \max_{w,m,t} \quad & \pi = q + (\Delta q - mt)e - w - C(m) - S(1)\frac{1}{1-\beta} \\ s.t. \quad (LL) \quad & w \geq 0, \\ (PC) \quad & u \geq \underline{u}\ , \\ (IC) \quad & e = \arg\max_{e \in [0,1]} \{u\}\ , \end{aligned}$$

where $S(1)$ is the cost of perfect screening and β is the share of bad workers in the population. If a job candidate is found out to be a bad type, the organisation has to go through a new round of screening candidates, such that screening costs increase in the share of bad workers.

As can be seen from the above optimization problem, screening only enters as a fixed cost, meaning that there is no reason to change the incentives after hiring a worker compared to the bench mark case. That is, the optimal contract with screening for $\underline{u} = 0$ still corresponds to the optimal contract derived in Section 2.2. For the example discussed previously, the optimal contract corresponds to $w^* = 0, \quad t^* = \Delta q/(2\underline{m}), \quad m^* = \underline{m}$ and $p^* = 0$.

Both with full deterrence and perfect screening, there will be no sabotage in equilibrium. However, both measures have different effects. While screening has a high upfront cost, the incentives of regular workers are not distorted. Monitoring and bonus payments are at the same level as in the benchmark case. Full deterrence, on the other hand, may have lower fixed costs, but distorts the incentives of regular workers as well, since overall incentives for positive effort go up.

If screening costs are sufficiently low and the share of bad guys is low, then screening becomes more attractive. Also, screening may yield better results if the motivation of bad guys, θ, is very high, because then it is costly or even impossible to deter them through monetary incentives after hiring.

5.2 Imperfect Screening

In general, most screening technologies are not perfect such that some bad workers will not be detected in the process. Let us assume that the screening technology comes at a cost $S(s) = \sigma 0.5 s^2$ and detects bad workers with probability $s \in [0, 1]$. In order to fill one open position, the firm may therefore have to go through several rounds of screening. When a firm screens one candidate, with probability $(1-s)$, they learn nothing and hire the guy, who is bad with probability β. With probability s, they find out that the candidate is good (happens with probility $(1-\beta)$) and hire him. With probability $s\beta$, the candidate is found out to be bad, in which case the position cannot be filled and the screening process moves on to the next candidate.

The profit generated by a regular worker is

$$\pi_r = q - w + (\Delta q - mt)e - C(m, p) \; . \tag{19}$$

The profit from hiring a bad worker is

$$\pi_b = q - w - Dd - C(m, p) \; . \tag{20}$$

The optimization problem of the principal then corresponds to

$$\begin{aligned} \max_{w,m,t,s} E\pi \;\; &= \;\; \frac{1}{1-s\beta}\Big((1-\beta)[q - w + (\Delta q - mt)e - C(m)] \\ &+ \;\; (1-s)\beta[q - w - Dd - C(m)] - S(s)\Big) \end{aligned} \tag{21}$$

subject to the following constraints

$$\begin{aligned}
(LL) \quad & w \geq 0, \\
(PC) \quad & u = w + mte - \psi(e+d) \geq \bar{u} \ , \\
(IC_r) \quad & e = \arg\max_{e \in [0,1]} \{u\} \\
(IC_b) \quad & d = \arg\max_{d \in [0,1]} \{u\} \ .
\end{aligned}$$

Solving this for s, we get that s has to be such that

$$\beta(1-\beta)\Delta\pi \quad = \quad \beta S(s) + (1 - s\beta)S'(s) \ ,$$

where $\Delta\pi = (\pi_r - \pi_b)$.

Since we assumed a specific function for $S(s)$, we can rewrite the optimality condition as

$$0.5\beta\sigma s^2 - \sigma s + \beta(1-\beta)\Delta\pi = 0 \ ,$$

such that, when solving for s, we get

$$s = \frac{1}{\sigma\beta}\Big[\sigma \pm \sqrt{\sigma^2 - 2\sigma\beta^2(1-\beta)\Delta\pi}\Big] \ , \tag{22}$$

where $\Delta\pi = \pi_r - \pi_b = (\Delta q - mt)\frac{mt}{a} + D\frac{\theta_b - mK}{a}$. To get a solution for (22), the term under the square root has to be positive. This is the case if the following assumption holds:

Assumption 3 $\sigma > 2\beta^2(1-\beta)\Big((\Delta q - mt)\frac{mt}{a} + D\frac{\theta_b - mK}{a}\Big)$.

Since $s \in [0,1]$, we can neglect s^+, such that the optimal s corresponds to

$$s = \frac{1}{\sigma\beta}\Big[\sigma - \sqrt{\sigma^2 - 2\sigma\beta^2(1-\beta)\Delta\pi}\Big] \ , \tag{23}$$

How does the optimal screening level change with the share of bad guys? To check this, let us assume for the moment that $\sigma = 1$ to save on notation. The optimal screening level then corresponds to

$$s = \frac{1}{\beta}\Big[1 - \sqrt{1 - 2\beta^2(1-\beta)\Delta\pi}\Big] \ , \tag{24}$$

Taking the derivative with respect to β, we get

$$\frac{\partial s}{\partial \beta} = \frac{1}{\beta^2}\Big(- \frac{\beta}{2\sqrt{\cdot}}(-2\Delta\pi)(2\beta - 3\beta^2) - (1 - \sqrt{\cdot})\Big) \ ,$$

where $\sqrt{\cdot} = \sqrt{\sigma^2 - 2\sigma\beta^2(1-\beta)\Delta\pi}$. That is,

$$\frac{\partial s}{\partial \beta} = \frac{1}{\beta^2}\Big(\underbrace{\frac{\beta\Delta\pi}{\sqrt{\cdot}}}_{>0} \underbrace{(2\beta - 3\beta^2)}_{>0 \ if \beta<2/3} - \underbrace{(1 - \sqrt{\cdot})}_{>0}\Big) \ .$$

Note that the last term has to be positive because the optimal s as given in (24) corresponds to $(1 - \sqrt{\cdot})/\beta > 0$ by definition, hence $(1 - \sqrt{\cdot}) > 0$.

Therefore, $\partial s/\partial\beta < 0$ if $\beta \geq 2/3$. For $\beta < 2/3$, we get $\partial s/\partial\beta > 0$ if

$$\Delta\pi(2\beta^2 - 3\beta^3) > (1 - \sqrt{\cdot})\sqrt{\cdot} \ .$$

6 Conclusion

Discuss:

- Role of peer monitoring
- Distinction between damage to the firm and social damage (example: corruption) and implications for our results

References

ACFE. Report to the nations: 2022 global study on occupational fraud and abuse. Technical report, Association of Certified Fraud Examiners, 2022.

Emmanuelle Auriol and Stefanie Brilon. Anti-social behavior in profit and nonprofit organizations. *Journal of Public Economics*, 117:149–161, 2014.

Gary Becker. Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2):169–217, 1968.

Michèle Belot and Marina Schröder. The spillover effects of monitoring: A field experiment. *Management Science*, 62(1):37–45, 2016.

Yassin Denis Bouzzine and Rainer Lueg. The reputation costs of executive misconduct accusations: Evidence from the# metoo movement. *Scandinavian Journal of Management*, 38(1):101196, 2022.

Margaret H Christ, Karen L Sedatole, and Kristy L Towry. Sticks and carrots: The effect of contract frame on effort in incomplete contracts. *The Accounting Review*, 87(6):1913–1938, 2012.

Deloitte. The economic costs of sexual harassment in the workplace. Technical report, Deloitte Australia, 2019.

David Dickinson and Marie-Claire Villeval. Does monitoring decrease work effort?: The complementarity between agency and crowding-out theories. *Games and Economic behavior*, 63(1):56–76, 2008.

Armin Falk and Michael Kosfeld. The hidden costs of control. *American Economic Review*, 96(5):1611–1630, December 2006.

Bruno Frey and Reto Jegen. Motivation crowding theory. *Journal of Economic Surveys*, 15(5):589–611, 2001.

Henrich R Greve, Donald Palmer, and Jo-Ellen Pozner. Organizations gone wild: The causes, processes, and consequences of organizational misconduct. *Academy of Management annals*, 4(1):53–107, 2010.

Thomas Hellmann and Veikko Thiele. Incentives and innovation: A multitasking approach. *American Economic Journal: Microeconomics*, 3(1): 78–128, 2011.

Hiscox. The 2016 hiscox embezzlement study: A report on white collar crime in america. Technical report, Hiscox, 520 Madison Avenue, New York, NY 10022, www.hiscoxbroker.com, 2016.

Bengt Holmström and Paul Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org.*, 7: 24, 1991.

Sorin MS Krammer, Addisu A Lashitew, Jonathan P Doh, and Hari Bapuji. Income inequality, social cohesion, and crime against businesses: Evidence from a global sample of firms. *Journal of International Business Studies*, pages 1–16, 2022.

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

Anthony J Nyberg, Jason D Shaw, and Jing Zhu. The people still make the (remote work-) place: lessons from a pandemic. *Journal of Management*, 47(8):1967–1976, 2021.

Phanish Puranam and Bart S. Vanneste. Trust and governance: Untangling a tangled web. *Academy of Management Review*, 34(1):11–31, 2009.

Patrick W Schmitz. Workplace surveillance, privacy protection, and efficiency wages. *Labour Economics*, 12(6):727–738, 2005.

Maurice E Schweitzer, Teck-Hua Ho, and Xing Zhang. How monitoring influences trust: A tale of two faces. *Management Science*, 64(1):253–270, 2016.

Masoud Shadnam and Thomas B. Lawrence. Understanding widespread misconduct in organizations: An institutional theory of moral collapse. *Business Ethics Quarterly*, 21(3):379–407, 2011.

Carl Shapiro and Joseph E Stiglitz. Equilibrium unemployment as a worker discipline device. *The American economic review*, 74(3):433–444, 1984.

Judith Van Erp. The organization of corporate crime: Introduction to special issue of administrative sciences, 2018.

Ferdinand A. von Siemens. Intention-based reciprocity and the hidden costs of control. *Journal of Economic Behavior & Organization*, 92:55 – 65, 2013. ISSN 0167-2681.

Darrell M West. How employers use technology to surveil employees. Technical report, 2021.

A Appendix

A.1 Proof of Proposition 1

Taking into account the IC constraint, $e^o = \psi'^{-1}(mt)$, in the principal's optimization problem (see Equation 5), we get the following first order conditions:

$$\frac{\partial \pi}{\partial t} = (\Delta q - mt)\frac{\partial e^o}{\partial t} - me^o \leq 0 \tag{25}$$

$$\frac{\partial \pi}{\partial m} = (\Delta q - mt)\frac{\partial e^o}{\partial m} - te^o - \frac{\partial C}{\partial m} \leq 0 \tag{26}$$

It is easy to check that if $\frac{\partial \pi}{\partial t} = 0$ then $\frac{\partial \pi}{\partial m} = -\frac{\partial C}{\partial m} < 0$. To see this, set (25) to zero, solve for e^o and plug the resulting expression into the second constraint. By rearranging terms, we get

$$\frac{\partial \pi}{\partial m} = (\Delta q - mt)\left[\frac{\partial e^o}{\partial m} - \frac{t}{m}\frac{\partial e^o}{\partial t}\right] - \frac{\partial C}{\partial m} \leq 0 \tag{27}$$

Since $e^o = \psi'^{-1}(mt)$, we can apply the chain rule to get to the partial derivatives. Let $e^o = f(mt)$ describe our function and define $z = mt$, then

$$\frac{\partial e^o}{\partial t} = \frac{\partial f(z)}{\partial z}\frac{\partial z}{\partial t} = f'(z)m$$

$$\frac{\partial e^o}{\partial m} = \frac{\partial f(z)}{\partial z}\frac{\partial z}{\partial m} = f'(z)t \ .$$

Therefore the term in square brackets in (27) is zero. This means that, if $\partial \pi / \partial t = 0$, then $\partial \pi / \partial t = -\partial C / \partial m$ which is automatically smaller than zero. The principal therefore wants to set the monitoring level as low as possible.

With this result in mind, we can plug $\underline{m}$ into (25) to get to the optimal bonus payment.

QED.

A.2 Proof of Proposition 2

Since we assume that the agent's reservation utility is zero, the argument from Lemma 1 still holds, and the optimal basic wage $w^{det} = 0$.

If we plug in the optimal effort e^o, the principal's maximization problem then corresponds to solving the following Lagrangian

$$\max_{m,t,p} \quad \mathcal{L} = q + (\Delta q - mt)e^o - C(m,p) - w + \lambda(mt - \theta + pK) \ ,$$

and we get these Kuhn Tucker conditions:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial t} &= (\Delta q - mt)\frac{\partial e^o}{\partial t} - me^o + \lambda m \leq 0 && (28)\\
\frac{\partial \mathcal{L}}{\partial m} &= (\Delta q - mt)\frac{\partial e^o}{\partial m} - te^o - \frac{\partial C}{\partial m} + \lambda m \leq 0 && (29)\\
\frac{\partial \mathcal{L}}{\partial p} &= -\frac{\partial C}{\partial p} + \lambda K \leq 0 && (30)\\
\frac{\partial \mathcal{L}}{\partial \lambda} &= mt - \theta + pK \geq 0 && (31)\\
\lambda &\geq 0 && (32)\\
& \lambda\Big(mt - \theta + pK\Big) = 0 \ . && (33)
\end{aligned}$$

Equation (28) holds with equality if

$$e^o = (\Delta q - mt)\frac{\partial e^o}{\partial t} + \lambda \ . \tag{34}$$

Plugging this expression into Equation (29), we get

$$\frac{\partial \mathcal{L}}{\partial m} = (\Delta q - mt)\Big[\frac{\partial e^o}{\partial m} - t\frac{\partial e^o}{\partial t}\Big] - \lambda t + \lambda t - \frac{\partial C}{\partial m} \leq 0 \ ,$$

The term in square brackets is the same as in Equation (27) in the proof of Proposition 1 (see A.1). Hence the same argument can be made, and the equation simplifies to

$$\frac{\partial \mathcal{L}}{\partial m} = -\frac{\partial C}{\partial m} \leq 0 \ , \tag{35}$$

which is automatically fulfilled for $C'_m > 0$. From this, we can conclude that m is going to be chosen as low as possible, *independent of the specification of both the effort cost and the monitoring cost functions.* Hence $m^{det} = \underline{m} = m^*$.

Furthermore, we are only interested in cases without automatic deterrence, i.e., cases where $\theta > \tilde{\theta} = \underline{m}t^*$. In these cases, the deterrence constraint is binding and hence $\lambda > 0$. From (30), we get that

$$\lambda = \frac{\partial C}{\partial p}\frac{1}{K} \ .$$

Taking this into account, as well as our result on m, (34) can be rewritten as

$$\Delta q + \frac{\partial C}{\partial p}\frac{1}{K}\frac{1}{\partial e^o/\partial t} = \underline{m}t + \frac{e^o}{\partial e^o/\partial t} \ .$$

The optimal bonus payment t^{det} has to be such that this condition holds.

If we plug our results for t and m into (31), we get that p^{det} has to be such that

$$p^{det} \geq (\theta - \underline{m}t^{det})/K \ .$$

QED.

A.3 Proof of Proposition 3

Taking into account the optimal effort levels e^o and d^o of the agents, the principal maximizes the following profit function:

$$\pi = q + (1-\beta)(\Delta q - mt)e^o - \beta D d^o - C(m,p) \ .$$

The corresponding first order conditions are

$$\frac{\partial \pi}{\partial t} = (1-\beta)(\Delta q - mt)\frac{\partial e^o}{\partial t} \leq 0 \tag{36}$$

$$\frac{\partial \pi}{\partial m} = (1-\beta)(\Delta q - mt)\frac{\partial e^o}{\partial m} - \frac{\partial C}{\partial p} \leq 0 \tag{37}$$

$$\frac{\partial \pi}{\partial p} = -\beta D\frac{\partial d^o}{\partial p} - \frac{\partial C}{\partial p} = 0 \ . \tag{38}$$

By the same argument as in the proof of Proposition 1, one can show that if (36) holds with equality, then (37) is smaller or equal zero and as a result, the optimal monitoring level is as low as possible. Moreover, if (36) holds with equality, it is essentially the same condition as in the case without bad workers and therefore leads to the same optimal bonus level t.

QED.