

Anti-Social Behavior in Profit and Nonprofit Organizations[☆]

Emmanuelle Auriol^{a,*}, Stefanie Brilon^b

^a*Toulouse School of Economics, France*

^b*Goethe Universität Frankfurt am Main, Germany*

Abstract

Two types of intrinsically motivated workers are considered: “good” workers care about the mission of an organization, whereas “bad” workers derive pleasure from destructive behavior. Compared to the case with only good motivated workers, the mission-oriented sector has to resort to higher monitoring to deter bad workers from entering the sector. In equilibrium, bad workers work in the for-profit sector where they behave like “normal” people, while good workers self select into the mission-oriented sector. In the profit-oriented sector, both monitoring and bonus payments for good behavior increase to control the damage caused by bad workers. The results of the paper are illustrated by examples from child care and the threat posed by pedophiles in this sector.

Keywords: intrinsic motivation, incentive contract, non-profit, sabotage, candidate selection.

1. Introduction

Intrinsic motivation is generally treated by economists as something beneficial to organizations. Most theoretical models on the subject suppose that intrinsic

[☆]We are indebted to Roland Bénabou, Maitreesh Ghatak, Guido Friebel, Martin Hellwig, Paul Seabright, and Konrad Stahl for helpful discussions. Furthermore, we thank Christoph Engel, Matthias Lang, Baptiste Massenet, Sarah Sandford and Philipp Weinschenk for their comments, as well as seminar participants at Mannheim, Rotterdam, Bonn, Frankfurt, Oxford, Paris, Hannover and Lugano. Finally we are indebted to two anonymous referees who contributed greatly by their suggestions to improve the paper content and organization. All remaining errors are ours. Emmanuelle Auriol would also like to thank the French Development Agency (AFD) for its financial support. Stefanie Brilon would like to thank the Deutsche Forschungsgemeinschaft, the Max Planck Institute for Research on Collective Goods and the Swiss National Science Foundation (Sinergia grant 130648 and 147668) for their financial support.

*Corresponding author

Email addresses: emmanuelle.auriol@tse-fr.eu (Emmanuelle Auriol), brilon@econ.uni-frankfurt.de (Stefanie Brilon)

motivation arises if workers derive a benefit from doing good - what is often referred to as warm glow utility - or when workers are interested in a certain goal or mission, like for example helping the poor or protecting the environment. An organization that is dedicated to such a mission may find it easier and cheaper to attract workers pursuing similar goals. However, other aspects of a job may also instil intrinsic motivation in certain types of workers. And these other aspects are not necessarily beneficial for the employer. This is illustrated by the United Nations sex-for-food scandal, which was exposed by “Save the Children”, an UK-based nonprofit organization: it showed that in 2006 aid workers were systematically abusing minors in a refugee camp in Liberia, selling food for sex with girls as young as 8.¹ Helping refugees is the kind of mission-oriented work that is likely to attract workers interested in this mission – what we will refer to as good motivated workers. But such a job also involves working with vulnerable children in a remote location with little control from the outside which may also attract workers with quite different intentions – what we will call bad workers. Examples for socially destructive behavior such as this abound, ranging from different forms of sexual misconduct, over terrorism to pyromania and other kinds of abuse.² In this paper, we analyze how different sources of intrinsic motivation of workers may affect labor management and production outcomes both in for-profit and nonprofit organizations.

Psychologists have long recognized and studied anti-social behavior. One strand of the literature, as well as most traditional psychiatry, focuses on so-called internal determinants. Anti-social behaviors, perceived as a pathology, are explained by individual predispositions such as genetics, personality traits, or pathological risk factors rooted in childhood. Another strand of the literature focuses on external determinants. It aims to explain how “ordinary” people can be induced to behave in evil ways by situational variables (see Zimbardo, 2004).³ Our paper is consistent with both views. We assume that the level of negative intrinsic motivation of bad workers is exogenous. That is, anti-social behavior is ultimately driven by internal determinants. However, it depends on the incentives given by an organization whether bad workers will indeed act in an anti-social way or whether they will behave in the organization’s interest. In other words, whether individuals act upon their predisposition for certain behaviors depends on external determinants (i.e., on situational variables). If the punishment they face and/or the reward for good behavior are high enough, most people will not act destructively.

We extend the model by Besley and Ghatak (2005) who consider only good and

¹See the report by Save the Children UK (2006). Similar cases have since been reported from Southern Sudan, Burundi, Ivory Coast, East Timor, Congo, Cambodia, Bosnia and Haiti (see “The U.N. sex-for-food scandal”, Washington Times, Tuesday, May 9, 2006 and the report by Save the Children UK, 2008).

²For a more detailed discussion, see the following section.

³For instance, in a famous experiment on obedience to authority, Milgram (1974) has shown that two thirds of the subjects were willing to inflict lethal electrical shocks upon total strangers.

regular workers, whereas there are three types of workers in our model: good, regular, and bad. Regular workers only care about monetary incentives, good workers care about money and the mission of the organization, and bad workers care about money and whether they can do things they like, but which are harmful to the organization. We further generalize the approach by Besley and Ghatak (2005) by adding monitoring as an additional choice variable of the employer in order to deal with the different incentive issues raised by the presence of different kinds of workers. In contrast to standard incentive theory, where monitoring is treated as a black box (i.e., fixed cost), here the level of monitoring is endogenous and central to the design and management of organizations: while monitoring reinforces the effort incentives of good and regular workers, it makes bad actions or anti-social behavior less attractive as it increases the chances of getting caught and being punished. By disentangling these two roles of supervision, the paper therefore adds a new explanation why, in practice, so many resources are spent on monitoring.

We consider two sectors of the economy, one profit-oriented and one mission-oriented. As in Besley and Ghatak (2005), we assume that in the nonprofit sector, organizations are structured around some mission, for example providing public services, or catering to the needs of disadvantaged groups of society.⁴ These organizations may attract workers who care about this specific mission and derive an intrinsic benefit from their work. Given this setup, we first consider the case with only good and regular workers and find the classic result by Besley and Ghatak (2005) that the mission-oriented sector offers lower wages and makes less use of bonus payments than the profit-oriented sector. We also show that the mission-oriented sector monitors less than the profit-oriented sector, which is not an explicit result in Besley and Ghatak (2005) as there is no monitoring.

We then introduce bad workers who derive utility from behaving in an anti-social way. We take the equilibrium contracts from Besley and Ghatak (2005) as a starting point for our analysis. That is, we use the equilibrium contracts with only good and regular workers, and costly monitoring, as a benchmark against which we compare our results. Since profit-oriented organizations monitor more in equilibrium, they are a priori less vulnerable to anti-social behavior. Bad workers may behave like regular workers in the profit-oriented sector and thus be totally undistinguishable from normal people. By contrast, if bad workers join the mission-oriented sector, then it is only to take advantage of the low level of monitoring and to behave badly. The more organizations in this sector rely on the intrinsic motivation of good workers and the less they make use of monetary

⁴We use the terms mission-oriented and nonprofit organization equivalently since we believe them to be largely congruent in reality. However, there are cases where organizations do not have the legal status of a nonprofit, but still follow a mission. This has recently been highlighted by the literature on corporate social responsibility as discussed, for instance, in Bénabou and Tirole (2010). For a further discussion of mission- vs. profit-oriented organizations, see also Besley and Ghatak (2005).

incentives and control, the more likely they are to become the target of bad workers. This first result helps to explain why, for instance, most scandals of child abuse involve organizations with no or little monitoring in the past, such as some religious institutions or refugee camps (see the discussion in Sections 2 and 5).

We then analyze how contracts in both sectors have to change relative to the situation when there are only regular and good workers if organizations have to deal with bad workers. We consider two possible scenarios: If the damage caused by anti-social behavior is expected to be sufficiently high, then both sectors will want to deter such behavior completely (*full deterrence equilibrium*), whereas a certain amount of destructive behavior may be tolerated in equilibrium if the expected damage is low (*partial deterrence equilibrium*). We show that in an equilibrium with full deterrence, bad workers join the for-profit sector where they are indistinguishable from regular workers, while good workers self-select into the mission-oriented sector. To incentivize the bad workers to behave well, monitoring and bonus payments for good behavior both have to increase in the for-profit sector, combining “the carrot and the stick”. In the nonprofit sector, on the other hand, the focus is on the stick to discourage bad workers from joining and thus prevent the occurrence of bad behavior: monitoring has to go up, while the incentives for good behavior tend to stay the same to avoid losing the benefit from the self-selection of good workers.

Next, we discuss the robustness of our model with respect to variations in the damage caused by bad workers. If the expected damage is relatively low, full deterrence may not be worthwhile. We show that in an equilibrium where there is partial deterrence in the for-profit sector, the mission-oriented sector will generally opt for full deterrence such that all bad workers will again be found in the profit-oriented sector. The only exception occurs when the potential damage by bad workers is sufficiently low and good workers are very motivated such that they do not need any monitoring: In that case the mission-oriented sector may have no interest in introducing costly monitoring just to prevent a relatively low level of sabotage, even though it will thus attract all bad workers. Finally, deterrence, which is costly as it implies higher monitoring, even may become entirely ineffective for workers with very high levels of bad motivation (e.g., extremists, kamikaze). We therefore also discuss ex ante measures of candidate selection, which may help to reduce the occurrence of anti-social behavior by screening out bad workers.

The paper is organized as follows: We first discuss in Section 2 the related literature and some examples that illustrate our model, before we turn to the description of our basic setup with only good and regular workers in Section 3. We then introduce bad workers in Section 4 and show how the optimal contracts have to change. The empirical implications of the theory are discussed in Section 5 in the context of child care services where the threat posed by pedophiles has been widely documented. Section 6 concludes.

2. Examples and Related Literature

There are many examples that illustrate the relevance of anti-social behavior both in mission- and in profit-oriented organizations. Among them, the recent abuse scandals in the Catholic Church stand out both by their shock potential as well as by sheer numbers.

As of April 10, 2012, the U.S. Conference of Catholic Bishops has counted 6,115 clerics “not implausibly” and “credibly” accused of sexually abusing 16,324 minors during the period 1950-2011. The John Jay report (see Terry, 2008) found that “*the problem was indeed widespread and affected more than 95 percent of the dioceses*”. Similar widespread problems of child abuse occurred in Ireland, as documented in the report by the Commission of Inquiry into Child Abuse (see CICA, 2009), and Germany (see Dt. Jugendinstitut, DJI, 2011, on abuse cases in institutions). While the problem is not limited to church organizations it is, however, particularly likely to occur under specific circumstances: A pedophile will preferably target vulnerable children, such as refugees⁵ or orphans, simply because they are less likely to expose him. Cases of physical, sexual or emotional abuse of children in orphanages have been uncovered for instance, at Mount Cashel Orphanage in Canada in the 1980s, or the Haut de la Garenne Children’s home on the channel island of Jersey. Furthermore, foster homes, boarding schools and detention centers seem to be at risk, as documented by the CICA (2009) report and recent cases from Germany and France.⁶

Other examples for anti-social behavior resulting from some form of intrinsic motivation are pyromania, serial killing or sadism. Stambaugh and Styron (2003) show that pyromaniacs may best be able to satisfy their urge for fire by working for the firefighters and provide evidence, mostly from the United States, that shows how serious the problem is.⁷ Along similar lines, the so-called “angels of death” are psychopaths who manage to work in health care facilities, preferably with vulnerable patients (elderly, children, etc), to feed their need to kill.⁸ Finally a sadist might try to work in prisons or detention centers, that are protected by national security secrecy or by their geographical remoteness, to feed his/her need to humiliate and harm others.⁹

Gibelman and Gelman (2004) list further evidence of destructive behavior in mission-oriented organizations which include cases of questionable fund raising practices, mismanagement, embezzlement, theft, money laundering, “personal

⁵See Save the Children UK (2006) and Save the Children UK (2008).

⁶See for instance www.lefigaro.fr/actualite-france/2011/09/15/01016-20110915ARTFIG00583-nouvelle-plainte-contre-le-pere-d-accueil-de-laetitia.php.

⁷Similar cases have been documented elsewhere, see for example www.lexpress.fr/actualite/societe/pompier-pyromane-2-ans-de-prison_459032.html, or www.swiss-firefighters.ch/News-file-article-sid-3427.html.

⁸See http://www.crimelibrary.com/notorious_murders/angels/index.html.

⁹As examples, see the Stanford experiment on prison (see www.prisonexp.org/) and the Abu Ghraib torture scandal (see for instance www.time.com/time/magazine/article/0,9171,1025139,00.html).

lifestyle enhancement” and kickbacks, corruption, as well as sexual misconduct. Note, however, that anti-social behavior is not the monopoly of non-profit organizations, but is also found in for-profits. For instance, a terrorist might want to work in an airport to have a privileged access to planes. Or a spy would be interested in jobs in firms where he is likely to get access to sensitive information, while his risk of being discovered is low.

By considering such destructive behavior and introducing bad workers, we contribute to the literature on intrinsic motivation and its effects on agents’ behavior which has received increasing attention in recent years, as documented, for example, by the papers by Kreps (1997), Bénabou and Tirole (2003), Frey (1997), Murdock (2002) and Akerlof and Kranton (2005).¹⁰

Furthermore, our model is linked to the growing strand of literature on public service motivation¹¹ and its implications for hiring and remuneration schemes, as for example Francois (2000), Francois (2003), Prendergast (2007) and Delfgaauw and Dur (2008). As in this literature, our workers show some form of intrinsic motivation when working in a certain sector or for a particular mission.¹² For instance, Prendergast (2007) shows that intrinsically motivated agents in the public sector should be biased either against or in favor of their clients, depending on circumstances.¹³ While the focus of this paper is quite different from ours, Prendergast (2007) also finds that sometimes the wrong people will be drawn to a certain job.

Our model is closely related to the paper by Besley and Ghatak (2005) who show that matching the mission preferences of principals and agents can enhance organizational efficiency and reduces the need for high-powered incentives.¹⁴ There are hence many sectors where wages are not paid conditional on performance, as for instance the civil service sector or many nonprofit organizations.¹⁵ Depending on circumstances, other factors may also play a role: Nonprofits sometimes are even legally forbidden to pay incentive wages; see, for instance, the discussion in Glaeser (2002). Or, as for example in the judicial sector, there are institutional reasons for low-powered incentives: by minimizing

¹⁰Recently, destructive behavior has also become the subject of experimental economics. See, for instance, Abbink and Herrmann (2009) and Abbink and Sadrieh (2009).

¹¹See Dixit (2002) for a review on incentives in the public sector.

¹²Note, however, that from a technical point of view some of these models are quite different from ours. In Francois (2000), for instance, all workers care for overall output and have no particular preference for the public sector. Differences between the two sectors only come into play through differences in property rights.

¹³That this may indeed be the case has been shown by Heckman, Smith, and Taber (1996) in an empirical study on training programs. Bureaucrats tended to select applicants with lower expected earnings into a training program, even though this negatively affected their own payoff.

¹⁴Agents may also care about other aspects of their work environment. This has, for instance, been analyzed by Kosfeld and Siemens (2011) who show that workers may self-select across firms according to their preferences regarding team work.

¹⁵See also Borzaga and Tortia (2006), Ballou and Weisbrod (2003) and Serra et al. (2011) for empirical studies on the incentives in for-profit and different forms of nonprofit organizations.

economic incentives, the quality and independence of judgement increases (Posner, 1993). Finally, performance may just be too difficult or too expensive to assess. This is the case of development aid, where the costs of monitoring in the field are often prohibitively high. This lack of monitoring may lead to shirking and absenteeism as has been analyzed for example by Chaudhury et al. (2006) and Banerjee and Duflo (2006). However workers may not only just work less. They may also behave in a way that damages the organization for which they work or which is outright criminal. To prevent such destructive behavior, nonprofits therefore may want to engage in a more sophisticated selection process of candidates. The difficulties of such a process have, for instance, been discussed in Goldman (1982) and Greenberg and Haley (1986) for the selection of judges. We will come back to this problem and to the above mentioned examples in Section 5 of our paper.

3. Basic Setup

The model is based on Besley and Ghatak (2005). There are two sectors $i = F, N$, where F indicates for-profit or profit-oriented and N indicates nonprofit or mission-oriented organizations. Furthermore, there are three types of agents $j = g, r, b$, where g stands for good, r for regular and b for bad workers, with shares $x_r + x_g + x_b = 1$ in the population. The last category is new compared to Besley and Ghatak (2005) where there are only good and regular workers. The type of an individual worker is not directly observable, but the distribution of types is common knowledge.

Assumption 1. $x_b < x_g < x_r$.

It is natural to assume that the share of bad workers is smaller than the share of good workers, which is itself smaller than the share of regular workers (i.e., anti-social personalities are fairly rare). Beside its intuitive appeal, this assumption plays a technical role in Section 4.3, when we consider the possibility of bad actions occurring in equilibrium.¹⁶

Each agent produces a basic output q and, depending on his effort $e \in [0, 1]$, an additional output Δq with probability e . His effort cost is $c(e) = ae^2/2$ where a is a constant. In order to induce agent j to work harder, the principal in sector i can offer him a contract consisting of a basic wage w_{ij} plus a bonus payment $t_{ij} \geq 0$ if a high output is observed. However, the principal only observes the agent's output with probability m_i , where m_i is the monitoring level in sector i . This feature is new compared to Besley and Ghatak (2005).¹⁷ Introducing costly

¹⁶Assumption 1 is a sufficient condition to get an interior solution in bonus optimization in Case III in Proposition C.3 in the online appendix. It plays a role only in this proof.

¹⁷In Besley and Ghatak (2005) monitoring is implicitly assumed to be costless (i.e., the principals observe output at 0 cost).

monitoring is a realistic extension of their model once we introduce bad workers in the next section. Organizations need monitoring as another instrument to differentiate between good and bad motivated types.

The cost of monitoring is $M(m_i)$, with $M' > 0$ and $M'' \geq 0$. We assume that $m_i \in \{0, [\underline{m}, 1]\}$, i.e., the principal can choose not to monitor or else he has to choose at least a minimum level of monitoring $\underline{m} > 0$. The idea is that there is some fixed cost to monitoring. For example, the principal may have to hire at least one employee for the task. As will become clear later on, in most cases the principal will want to set the monitoring level positive but as low as possible, which here is normalized at the minimum monitoring level \underline{m} . This result is similar to Becker (1968).¹⁸

We assume that there is a limited liability constraint such that the agent has to receive at least a monetary payoff of 0. Furthermore, the agent's outside utility is assumed to be $\bar{u} \geq 0$.¹⁹

Regular agents react only to extrinsic incentives (i.e., they work for money). Good agents derive an intrinsic benefit $\theta_g > 0$ from working successfully in the nonprofit sector N . In sector F , neither type of agent r or g derives a positive intrinsic benefit. Finally bad workers derive an intrinsic benefit $\theta_b > 0$ from pursuing a destructive action, which may cause a damage to any organization F or N employing them.

As a benchmark case, we first concentrate on the case with only good and regular workers, modeled in Besley and Ghatak (2005). The utility of agent $j = r, g$ in sector $i = F, N$ can be described by

$$u_{ij} = w_{ij} + (m_i t_{ij} + \theta_{ij}) e_{ij} - a e_{ij}^2 / 2, \quad (1)$$

from which follows immediately that the agent will choose his optimal effort level as $e_{ij} = \min \left\{ (m_i t_{ij} + \theta_{ij}) / a, 1 \right\}$. To rule out corner solutions in the following we assume that a is sufficiently large:

Assumption 2. $a > \Delta q + \max\{\theta_g, \theta_b\}$.

Under Assumption 2, we get an interior solution in effort in all cases. Substituting $e_{ij} = (m_i t_{ij} + \theta_{ij}) / a < 1$ in the maximization problem of the principal yields:

$$\max_{w_{ij}, t_{ij}, m_i} \pi_{ij} = q + (\Delta q - m_i t_{ij}) \frac{m_i t_{ij} + \theta_{ij}}{a} - w_{ij} - M(m_i), \quad (2)$$

¹⁸Note that a positive monitoring level also emerges when firms face limits on the penalties they can impose on workers for shirking. See the discussion in Dickens, Katz, Lang, and Summers (1989).

¹⁹Fixing the minimum payoff in the limited liability constraint to 0 is done without loss of generality as long as the agent's reservation utility $\bar{u} \geq 0$ varies.

subject to a limited liability and a participation constraint:

$$(LL) \quad w_{ij} \geq 0, \quad (3)$$

$$(PC) \quad u_{ij} = (m_i t_{ij} + \theta_{ij})^2 / (2a) + w_{ij} \geq \bar{u}. \quad (4)$$

To make sure that inducing effort has some value to the principal in the absence of intrinsic motivation, we make the following assumption:

Assumption 3. $M(\underline{m}) < \min \left\{ \frac{1}{4a} \Delta q^2, q \right\}$

The first part of this assumption ensures that the cost of monitoring is not too high compared to the benefit, i.e., $\frac{1}{4a} \Delta q^2 > M(\underline{m})$, while the second part allows us to concentrate on outcomes with non-negative payoffs for the principal, i.e., $q > M(\underline{m})$.

Let us define \underline{v}_{ij} as the reservation payoff level such that the participation constraint of agent j becomes binding and \tilde{v}_{ij} as the level where the agent's limited liability constraint ceases to be binding. Furthermore, let \bar{v}_{ij} be defined as the level of reservation payoff of agent j such that principal i makes zero profit. Under Assumption 3, the three threshold levels of outside utility \underline{v}_{ij} , \tilde{v}_{ij} and \bar{v}_{ij} , are such that $0 < \underline{v}_{ij} \leq \tilde{v}_{ij} \leq \bar{v}_{ij}$.²⁰

The following proposition characterizes the optimal contract in the absence of bad types.

Proposition 1. : *Suppose Assumptions 2 and 3 hold and that $x_b = 0$. An optimal contract $(m_i^*, t_{ij}^*, w_{ij}^*)$ between a principal in sector $i = F, N$ and an agent of type $j = r, g$ given a reservation payoff $\bar{u} \in [0, \bar{v}_{ij}]$ exists and has the following features:*

(a) *The optimal wage is*

$$w_{ij}^* = \max \left\{ 0, \bar{u}_j - \frac{1}{2a} (\Delta q + \theta_{ij})^2 \right\}$$

(b) *The monitoring level is set at the minimum level whenever extrinsic incentives are necessary, i.e., $m_i^* = \underline{m}$ when $t_{ij} > 0$, and is zero otherwise.*

(c) *The optimal bonus payment is*

$$t_{ij}^* = \begin{cases} \max \{ 0, (\Delta q - \theta_{ij}) / (2\underline{m}) \} & \text{if } \bar{u}_j \in [0, \underline{v}_{ij}] \\ (\sqrt{2a\bar{u}_j} - \theta_{ij}) / \underline{m} & \text{if } \bar{u}_j \in (\underline{v}_{ij}, \tilde{v}_{ij}) \\ \Delta q / \underline{m} & \text{if } \bar{u}_j \in [\tilde{v}_{ij}, \bar{v}_{ij}] \end{cases} .$$

²⁰For a precise definition and derivation of these thresholds see part A.1 of the online appendix.

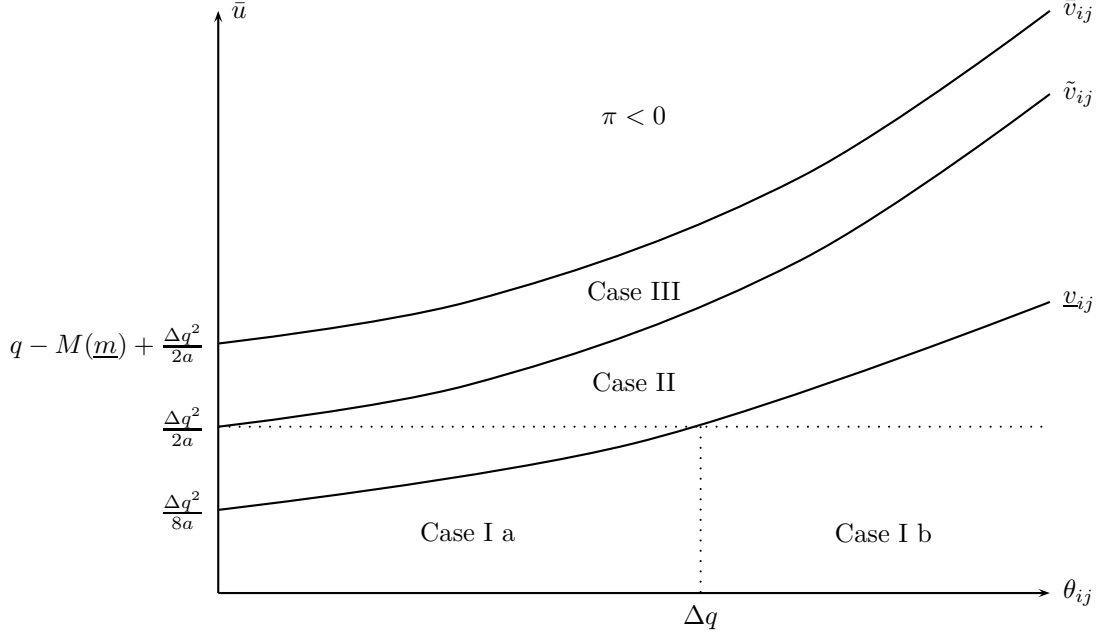


Figure 1: Optimal contract depending on outside utility \bar{u}_j and intrinsic motivation θ_{ij} .

Proof. The proof is in Part A.1 of the online appendix.

We can thus discern three cases:

- *Case I:* For low values of the reservation utility ($\bar{u} \in [0, \underline{v}_{ij}]$), the optimal contract is described by $w_{ij}^* = 0$, $t_{ij}^* = \max\{0, (\Delta q - \theta_{ij})/(2\underline{m})\}$, and $m_i^* = \underline{m}$ if $t_{ij}^* > 0$ and $m_i^* = 0$ otherwise. That is, workers are paid the minimum wage, monitoring is at its minimum level and the expected bonus is relatively low. The bonus goes down as intrinsic motivation increases and may eventually be zero for very high levels of intrinsic motivation, in which case monitoring is no longer needed (Case Ib).
- *Case II:* For intermediary values of the reservation utility ($\bar{u} \in (\underline{v}_{ij}, \tilde{v}_{ij})$), the optimal contract is described by $w_{ij}^* = 0$, $t_{ij}^* = (\sqrt{2a\bar{u}_j} - \theta_{ij})/\underline{m}$, and $m_{ij}^* = \underline{m}$. That is, while the base wage is still at its minimum level, the expected bonus is now higher than in Case I.
- *Case III:* For a high level of reservation utility ($\bar{u} \in [\tilde{v}_{ij}, \bar{v}_{ij}]$), the optimal contract is described by $w_{ij}^* = \bar{u} - (\Delta q + \theta_{ij})^2/(2a)$, $m_i^* = \underline{m}$ and $t_{ij}^* = \Delta q/\underline{m}$, i.e., the base wage is relatively high in order to satisfy the worker's participation constraint.

Figure 1 gives an overview of the results. The first two cases (i.e., the cases below the dotted line in Figure 1) are the cases described in Besley and Ghatak

(2005). The reason why the third case is not relevant in Besley and Ghatak (2005) is that they do not have a basic payoff q which accrues to the principal as base production. As a consequence, whenever the incentive scheme is not profitable because the agent's outside option is too high, then no contract can be made. Here, by contrast, the principal can fulfill the agent's participation condition even for higher outside options (i.e., $\bar{u} > \Delta q^2/2a$, that is the area above the horizontal dotted line in Figure 1) because the resulting costs are still covered by the basic production payoff q . In Sections A.2 and A.3 of the online appendix, we discuss in more detail how Proposition 1 translates into an optimal contract in sectors F and N respectively.

Which case is relevant for the principal in sector F depends only on the agent's outside option \bar{u} (vertical axis). The principal in the profit-oriented sector cannot rely on worker's intrinsic motivation (i.e., $\theta_{Fj} = 0$) and hence always has to provide sufficient extrinsic incentives. In particular, he always has to invest in monitoring: $m_F^* = \underline{m}$.

Which case is relevant for the principal in sector N depends both on the agent's outside option \bar{u} and on his level of intrinsic motivation θ_{Nj} (vertical and horizontal axis). By exploiting the intrinsic motivation of good workers, the principal in N can save on wage and monitoring costs relative to sector F by offering lower incentives and making less use of monitoring. Indeed, for any given level of reservation utility we have: $w_N^* + m_N^* t_N^* \leq w_F^* + m_F^* t_F^*$. As a consequence, the utility of a regular worker in sector N is always smaller than the utility level he can reach under the contract proposed in sector F . Regular workers will hence choose to work in sector F and good workers (i.e., with positive intrinsic motivation $\theta_{Ng} > 0$) will prefer to work in sector N . Moreover, in contrast to sector F , principals in sector N may not need to monitor their workers at all: If $\theta_g > \Delta q$, workers are motivated enough to provide effort even if there is no extrinsic incentive and no monitoring (Case I b).

Since, for any positive level of intrinsic motivation, the mission-oriented sector can save on wage costs, it gives sector N a competitive edge compared to sector F . Comparing π_N with π_F , it is hence straightforward to show that $\pi_N > \pi_F$ if $\theta_g > 0$.²¹

4. Enter the Bad

So far, we have considered the case where intrinsic motivation is necessarily good for the firm. However, this may not always be true. Workers may pursue

²¹See Sections A.2 and A.3 in the online appendix for the exact expressions for π_F and π_N . A nonprofit does not make any profits by definition. So while we sometimes refer to π_N and π_F as profit, it rather measures the relation between personnel costs and production. If the nonprofit has to spend less on its workers, this eases its budget constraint and makes more funds available for other things. This becomes particularly relevant if we take into account that many nonprofits are financed by donations and may have to run their operations on a rather tight budget.

their own private benefit to the detriment of the organization they work for. We model this by allowing workers to choose a destructive effort $d \in [0, 1]$, possibly in addition to the normal effort e . There are some workers who get a private benefit $\theta_b > 0$ from choosing such a destructive effort, and by doing so they may cause a damage $D \geq 0$ to the organization employing them.

In contrast to good workers, bad workers have no preference for one or the other sector. They like to do things that are anti-social independently of the organization they work for. Consider the following utility function for bad workers in sector $i = F, N$:

$$u_{ib} = w_i + m_i t_i e + (\theta_b - K m_i) d - a(e + d)^2/2, \quad (5)$$

where K is an exogenous punishment that can be imposed on a worker if a destructive effort is observed. The idea behind this is that a destructive effort corresponds not just to shirking but is an outright act of sabotage which can be treated as a criminal offense and hence can be punished by a fine or a prison term.

The worker can choose to do some constructive effort (i.e., do his job) and simultaneously some destructive effort by doing what he/she likes and is bad for the organization.²² However, because of the linear structure of the payoffs and the quadratic structure of the costs, at the optimum the worker chooses one of the two options, i.e., he either decides to satisfy his destructive impulse and get intrinsic satisfaction from doing so ($d \geq 0$), or he behaves like a regular worker, chooses $e \geq 0$ and aims at getting monetary rewards. Bad types therefore prefer to exert a constructive effort rather than to follow their destructive impulse if and only if²³

$$m_i t_i \geq \theta_b - m_i K. \quad (6)$$

To avoid introducing additional notation, we assume that the monitoring technology is the same for production and sabotage control. This captures the fact that there are increasing returns to scope in monitoring positive as well as negative behavior in the workplace. Indeed the equilibrium without sabotage is either zero monitoring or the minimum level \underline{m} . This can be interpreted as a fixed cost type of monitoring technology. Once the fixed costs have been paid the organization might choose to increase its controls to specifically fight sabotage. However we could also consider two different monitoring functions/technologies

²²In theory, nothing prevents a pyromaniac to be a brave fire fighter or a pedophile to be a good teacher. We are grateful to Roland Bénabou for pointing out this fact.

²³By contrast if the cost function was $a(e^2 + d^2)/2$, bad workers would simultaneously choose $e_{ib}^* = \frac{m_i t_i}{a}$ and $d_{ib}^* = \frac{\theta_b - K m_i}{a}$. The gist of our results still applies for this cost function, as is shown in part A.4 of the online appendix, but their exposition is more complex as two types of effort need to be considered for bad workers. Here $\partial u_{ib}/\partial e = m_i t_i - a(e + d)$ and $\partial u_{ib}/\partial d = -m_i K + \theta_b - a(e + d)$ cannot both be equal to 0 so that $e_{ib}^* = 0$ and $d_{ib}^* = \frac{\theta_b - K m_i}{a}$ if $\theta_b - m_i K > m_i t$ and $d_{ib}^* = 0$ and $e_{ib}^* = \frac{m_i t_i}{a}$ otherwise.

depending on the behavior in response to oversight. Our results are robust to the introduction of such separate monitoring functions.²⁴

Under Assumption 2 we obtain an interior solution in the effort choice of the agent. Substituting the worker's optimal effort choice, his expected utility is

$$u_{ib} = w_i + \begin{cases} (m_i t_i)^2 / (2a) & \text{if } \theta_b \leq m_i(t_i + K) \\ (\theta_b - m_i K)^2 / (2a) & \text{otherwise.} \end{cases}$$

In the following, we analyze how a bad worker's choice between sector N and F is determined and how the contracts in both sectors have to be adapted to the presence of bad workers. Throughout the paper we assume that if indifferent, bad workers choose sector F rather than sector N . We first analyze the behavior of bad workers given the benchmark contracts described in Proposition 1. Which sector will bad workers choose and how will they behave? When are the benchmark contracts enough to automatically deter anti-social behavior?

4.1. Automatic Deterrence of Bad Workers

We first compare a bad worker's payoff from choosing a constructive effort e or destructive effort d in both sectors given the contracts derived in Section 3. This comparison shows that for a given reservation utility \bar{u} the incentives for choosing a strictly positive constructive effort are always higher in F than in N , i.e., $u_{Fb}(e) > u_{Nb}(e)$. At the same time, the monitoring level in N is always smaller than or equal to that in sector F , thus making it less likely to get caught with destructive behavior in the nonprofit sector and therefore $u_{Nb}(d) \geq u_{Fb}(d)$. From this follows that under the contracts of Proposition 1 bad workers only join N to do harm. One could think that this result holds because the contracts of Proposition 1 have been derived while ignoring the presence of bad workers. However it goes deeper than that. This result holds true under any contract where the mission-oriented sector exploits the intrinsic motivation of good workers. Indeed if it aims to sort out good from regular workers, the mission-oriented sector has to offer lower monetary incentives than the profit oriented sector so that intrinsically motivated workers of good type join, while regular workers choose either not to work or to work in the profit-oriented sector. Since the bad workers have exactly the same preferences as the regular workers regarding the provision of constructive effort (i.e., they work for money), and since by convention they prefer to work in F when their expected utility is the same in F and N , we deduce that they will never join N to do good. Bad workers only join N to follow their destructive impulse, while minimizing the risk of being detected and punished. The next proposition collects this result.

²⁴For instance the fact that organizations must increase monitoring to fight against sabotage holds with a separate technology. Similarly the fact that one organization must increase monitoring more than the other, depending on the parameters of the model, also holds with a separate monitoring function for sabotage.

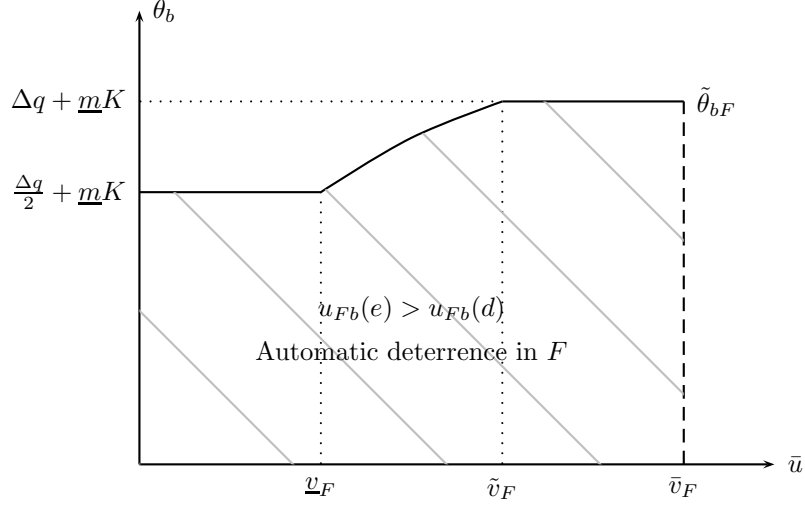


Figure 2: Automatic deterrence in sector F , given outside utility \bar{u} and level of negative motivation θ_b .

Proposition 2. : *Bad workers join the mission-oriented sector only to provide a destructive effort $d > 0$.*

Next, let us look in more detail at what happens in each sector. It is clear from (6) that for low levels of negative motivation θ_b , bad workers are better off if they choose a constructive effort e rather than a destructive effort d . As shown in Figure 2, in sector F , such automatic deterrence of destructive actions by bad workers, i.e., deterrence without any change of the contract as derived in Proposition 1, takes place if θ_b is smaller than

$$\tilde{\theta}_{bF} = \begin{cases} \frac{\Delta q}{2} + \underline{m}K & \text{if } \bar{u} \in [0, \underline{v}_F] \\ \sqrt{2a\bar{u}} + \underline{m}K & \text{if } \bar{u} \in [\underline{v}_F, \tilde{v}_F] \\ \Delta q + \underline{m}K & \text{if } \bar{u} \in [\tilde{v}_F, \bar{v}_F] \end{cases}, \quad (7)$$

where $\underline{v}_F, \tilde{v}_F, \bar{v}_F$ are the same threshold levels of outside utility as in Proposition 1.²⁵

Let us now turn to the nonprofit sector N . By virtue of Proposition 2, bad workers will be discouraged from joining this sector if their utility from sabotage in N is lower than their utility from good behavior in F or their outside utility, i.e., as long as $u_{Nb}(d) \leq u_{Fb}(e)$ or if $u_{Nb}(d) \leq \bar{u}$. Automatic deterrence, i.e.,

²⁵Recall, that the threshold levels are defined as follows: \underline{v}_F is the reservation payoff level such that the participation constraint of a regular worker becomes binding and \tilde{v}_F as the level where a regular agent's limited liability constraint ceases to be binding. Furthermore, \bar{v}_F is defined as the level of reservation payoff of a regular agent such that the principal in sector F makes zero profit. See also Section A.2 of the online appendix.

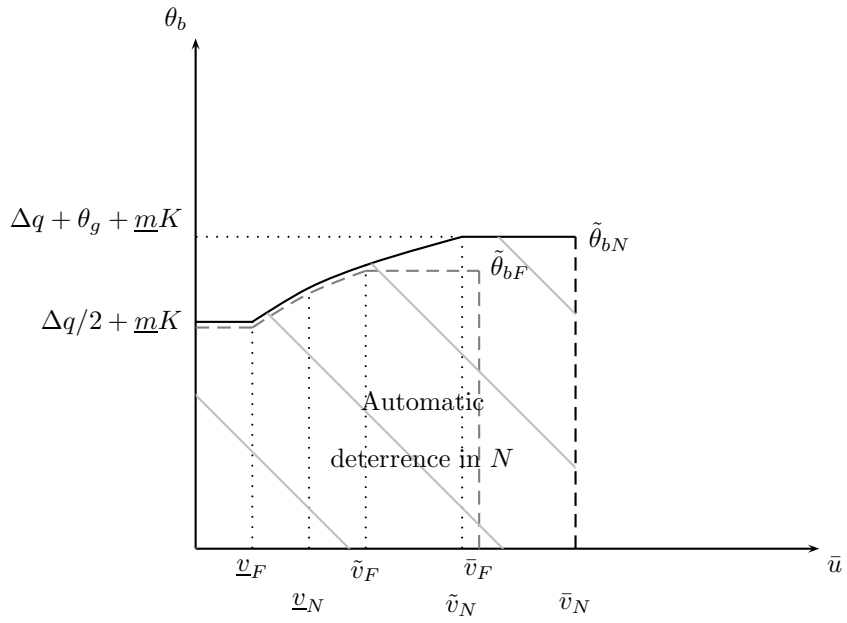
detering bad workers from entering N without any change in the contract (m_N^*, t_N^*, w_N^*) , therefore can be achieved for all θ_b smaller than

$$\tilde{\theta}_{bN} \equiv \begin{cases} \sqrt{2a(w_F - w_N^*) + (m_F t_F)^2} + m_N^* K & \text{if } \bar{u} \leq \bar{v}_F \\ \sqrt{2a(\bar{u} - w_N^*)} + m_N^* K & \text{if } \bar{u} > \bar{v}_F \end{cases}. \quad (8)$$

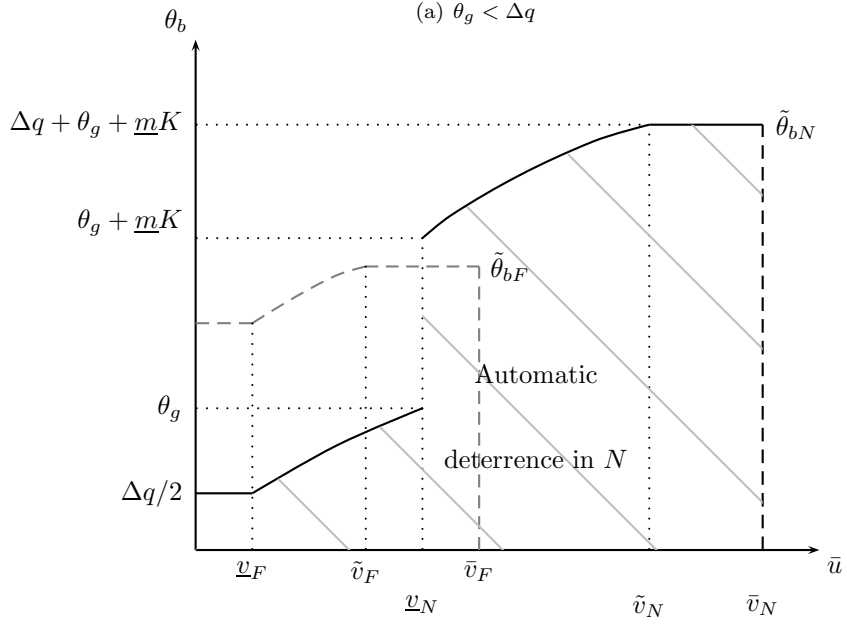
In order to determine the exact level of $\tilde{\theta}_{bN}$, we then have to insert the benchmark contracts in N and F into equation (8). For the sake of brevity, we will skip this exercise here. The interested reader may however find more details in Section A.5 of the online appendix. The results are also shown in Figure 3 which depicts the level of automatic deterrence in the nonprofit sector, θ_{bN} , as a black curve. For (\bar{u}, θ_b) -combinations below this curve, bad workers prefer to work either in sector F or enjoy their outside utility \bar{u} . Furthermore, the level of automatic deterrence in sector F , $\tilde{\theta}_{bF}$, is also featured in Figure 3 and is depicted as a dashed gray line.²⁶ This allows us to see immediately that, depending on the exact values of θ_g , θ_b and \bar{u} , sector N is either better or worse protected from destructive behavior than sector F :

- For $\bar{u} > \bar{v}_F$, i.e., for very high levels of reservation utility, F can no longer offer contracts that would satisfy the worker's participation constraint and at the same time yield a positive payoff to the firm. Therefore, nonprofit organizations are the only possible employer for agents with such a high reservation utility. But even working in N is relatively unattractive due to rather low basic wages. Bad workers will therefore prefer to enjoy their outside utility \bar{u} and only the most motivated will find it worthwhile to work at all. As a result, the level of deterrence in sector N for $\bar{u} > \bar{v}_F$ is rather high, as can be seen both from Figures 3(a) and 3(b).
- A more relevant scenario is one where $\bar{u} \leq \bar{v}_F$, i.e., the outside utility of the agents is such that both types of organizations may attract workers. Let us first consider what happens if $\theta_g < \Delta q$ as shown in Figure 3(a). For such low levels of intrinsic motivation of good workers, the level of automatic deterrence is the same in sector N and F because the monitoring level is the same in both sectors. Only for $\tilde{v}_F < \bar{u} \leq \bar{v}_F$, automatic deterrence is slightly higher in N since the basic wage in N is lower than in F and hence makes working in N less attractive.
- The most interesting case arises for low levels of reservation utility $\bar{u} < \underline{v}_N$ and high intrinsic motivation of good workers $\theta_g \geq \Delta q$ as shown in Figure 3(b). In that case, the nonprofit organization relies entirely on the intrinsic motivation of good workers and hence provides no extrinsic incentives, i.e., $m_N^* = 0$ (Case Ib). It then becomes particularly attractive for bad types.

²⁶Note that the threshold levels of outside utility \underline{v}_N , \tilde{v}_N , and \bar{v}_N shown in Figure 3 are the same as in Proposition 1 and are calculated for good workers. See also Section A.3 of the online appendix.



(a) $\theta_g < \Delta q$



(b) $\theta_g > \Delta q$

Figure 3: Automatic deterrence in N . For combinations of outside utility \bar{u} and negative motivation θ_b in the shaded area, bad workers are automatically deterred from bad actions in sector N .

Therefore, all bad workers with $\theta_b > \Delta q/2$ and $\bar{u} < \underline{v}_N$ will opt for sector N and provide a destructive effort. Bad workers with a lower θ_b will choose sector F and behave like regular workers.²⁷

The analysis in this section provides us with several insights: First, we have seen that bad workers only join sector N in order to behave in a destructive way, whereas they may behave like regular workers in sector F . And second, we have seen that while the low basic wages in N may act as a deterrent for high levels of reservation utility, the nonprofit sector becomes very vulnerable to anti-social behavior if it relies heavily on the intrinsic motivation of its workers and hence does not monitor enough. Finally, whenever $\theta_b \leq \tilde{\theta}_{bF}$ and $\theta_g < \Delta q$ the benchmark contracts derived in Proposition 1 are still optimal: Bad workers join the for profit sector where the extrinsic incentives are strong enough to make them behave like regular workers such that they go undetected. This equilibrium result might help to explain why perfectly integrated and normal looking people might, when their environment and incentives change, start behaving in evil ways (see Zimbardo, 2004).²⁸

In the following, we analyze how the contracts in both sectors have to change when the contracts of Proposition 1 do not lead to automatic deterrence.

4.2. Full Deterrence of Bad Workers: large D

In this section, we focus on cases where the damage D is sufficiently large so that both organizations want to fully deter bad behavior. By virtue of Proposition 2, it is never attractive for bad workers to choose an effort e in sector N since this sector offers lower monetary incentives to exploit the intrinsic motivation of good workers. This implies that in an equilibrium with full deterrence of destructive actions, all the bad workers are either in the for-profit sector and provide a constructive effort e or they do not work.

Let us first consider what happens in sector F . In the presence of bad workers, the principal's maximization problem in sector F becomes²⁹

$$\max_{w_F, t_F, m_F} \pi_F = q + (\Delta q - m_F t_F) \frac{m_F t_F}{a} - w_F - M(m_F), \quad (9)$$

²⁷Note that for $\theta_g \geq \Delta q$ and $\bar{u} > \underline{v}_N$, the nonprofit sector is again better protected against bad workers than the for profit sector (cf. Figure 3(b)): For $\bar{u} > \underline{v}_N$, the nonprofit sector has to implement a minimum level of monitoring even without bad workers, in order to fulfill the participation constraint of good workers. Combined with the lower basic wage in sector N , this makes working there less attractive for bad workers.

²⁸Note that for $\theta_g > \Delta q$, the benchmark contracts are still optimal if $\theta_b \leq \min\{\tilde{\theta}_{bF}, \tilde{\theta}_{bN}\}$. Bad workers then work in sector F and behave like regular workers.

²⁹The agent's incentive constraint is already taken into account here.

subject to

$$\begin{aligned} (LL) \quad & w_F \geq 0, \\ (PC) \quad & (m_F t_F)^2 / (2a) + w_F \geq \bar{u}, \\ (DET) \quad & m_F t_F \geq \theta_b - m_F K, \end{aligned}$$

where the last constraint is new. This deterrence constraint ensures that bad workers prefer to make a constructive rather than a destructive effort, i.e., it implies that $u_{Fb}(d) \leq u_{Fb}(e)$. Taking into account the new constraint (*DET*) in the optimization problem (9) is challenging. Indeed, by introducing the bad workers, we have to move from an optimization problem with two parameters of static comparative, \underline{u} and θ_g , to three parameters, \underline{u} , θ_g , and θ_b . Depending on the value of the parameters, different constraints bind and different subcases have to be considered. For the sake of simplicity, we first derive the optimal contracts in the case where the outside option of the workers is zero. We next discuss the robustness of the results in the general case with positive outside option.

4.2.1. A Simple Illustrative Case: No Reservation Utility

When $\bar{u} = 0$, the participation constraint (*PC*) always holds and the limited liability constraint (*LL*) binds, so that $w_F = w_N = 0$. This only leaves the deterrence constraint (*DET*) to be considered.

For $\theta_b < \tilde{\theta}_{bF}$, there is no need to change the benchmark contracts in sector F .³⁰ By contrast, if $\theta_b > \tilde{\theta}_{bF}$, the (*DET*) constraint binds so that $m_F t_F = \theta_b - m_F K$ or equivalently $t_F = \theta_b / m_F - K$. Substituting this value in (9), the optimal monitoring level in sector F then solves:

$$\max_{m_F} \pi_F = q + [\Delta q - (\theta_b - m_F K)] \frac{\theta_b - m_F K}{a} - M(m_F) \quad (10)$$

Since $M(m)$ is convex the objective function (10) is strictly concave. The first order condition is sufficient and the optimal monitoring level, denoted \tilde{m}_F , solves:

$$2\tilde{m}_F K + M'(\tilde{m}_F) a / K = 2\theta_b - \Delta q. \quad (11)$$

Depending on the functional form of $M(m)$ and the value of the different parameters, the optimal monitoring level calculated above may be smaller than the minimal monitoring level \underline{m} or larger than 1. Since both of these cases are excluded by assumption, the optimal monitoring level is $m_F^{det} = \min\{\max\{\underline{m}, \tilde{m}_F\}, 1\}$. It is easy to check that, since \tilde{m}_F increases with θ_b , m_F^{det} increases with θ_b . Figure 4 illustrates this result. The figure shows for an outside utility $\bar{u} = 0$ how the monitoring level and expected bonus develop as θ_b increases.³¹

³⁰For $\bar{u} = 0$, the benchmark contracts in F are given by: $w_F^* = 0$, $m_F^* = \underline{m}$, $t_F^* = \Delta q / (2\underline{m})$. There is automatic deterrence up to a level of negative motivation $\tilde{\theta}_{bF} = \Delta q / 2 + \underline{m}K$.

³¹Figure 4 is drawn for the monitoring function $M(m) = m^2 / 2$, which implies that $\tilde{m}_F = (2\theta_b - \Delta q) / (2K + a/K)$. Furthermore, it is based on $q = 2$, $\Delta q = 1.4$, $a = 15$, $\underline{m} = 0.1$, and $K = 3$.

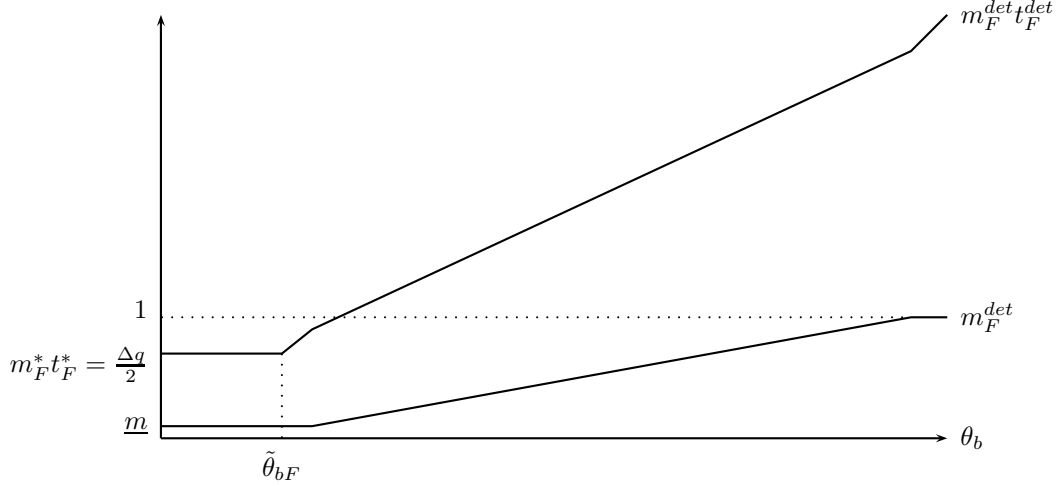


Figure 4: Monitoring and expected bonus level for outside utility $\bar{u} = 0$ and monitoring cost function $M(m) = m^2/2$, depending on the level of negative intrinsic motivation θ_b .

To sum up, when $\bar{u} = 0$, the optimal contract with full deterrence of destructive behavior in F is

$$\{w_F^{det}, m_F^{det}, t_F^{det}\} \equiv \begin{cases} 0, \underline{m}, \Delta q/(2\underline{m}) & \text{if } \theta_b \leq \tilde{\theta}_{bF} \\ 0, \min\{\max\{\underline{m}, \tilde{m}_F\}, 1\}, \theta_b/m_F^{det} - K & \text{if } \theta_b > \tilde{\theta}_{bF} \end{cases} \quad (12)$$

To fully deter bad workers from bad actions, the principal in sector F can use two tools. He can increase rewards for good behavior and/or punishment for bad behavior. By combining stick and carrot he optimizes his profit. One can check that \tilde{m}_F in (11) is strictly lower than \underline{m} when $\theta_b = \tilde{\theta}_{bF}$.³² For θ_b close to the threshold $\tilde{\theta}_{bF}$, the principal keeps monitoring at its minimum level \underline{m} , and just increases the rewards for good behavior so that $t_F^{det} > t^*$. In other words, for moderate levels of θ_b the carrot is sufficient (see Figure 4).

However, such a scheme will not entice workers with a high negative motivation to behave well. To deter such extreme types from anti-social behavior, it is not enough to make a positive effort more attractive, but also the expected punishment for bad behavior has to increase. The principal therefore has to raise the monitoring level beyond its minimum level to \tilde{m}_F^{det} . Equation (11) then implies that $m_F^{det} t_F^{det} = \theta_b - K m_F^{det} = \Delta q/2 + \frac{M'(\tilde{m}_F)a}{2K} > \Delta q/2 = m_F^* t_F^*$.

We deduce that in the for-profit sector the expected bonus payment for good effort increases relative to the benchmark case without bad workers: if $\theta_b > \tilde{\theta}_{bF}$ then $m_F^{det} t_F^{det} > m_F^* t_F^*$. This result is illustrated in Figure 4. As a consequence,

³²For instance with $M(m) = m^2/2$, if $\tilde{\theta}_{bF} < \theta_b \leq \tilde{\theta}_{bF} + a\underline{m}/(2K)$ then $\tilde{m}_F = (2\theta_b - \Delta q)/(2K + a/K) \leq \underline{m}$.

besides deterring bad workers from bad actions, this contract will also induce regular workers to choose a higher effort level, in some cases above the first best level (i.e., for θ_b very large $m_F^{det} t_F^{det} > \Delta q$).

We next turn to the non-profit sector. The wage scheme defined in Proposition 1 does not need to be adapted as long as $\theta_b \leq \tilde{\theta}_{bN}$.³³ Now if $\theta_b > \tilde{\theta}_{bN}$, two cases hold:

- *Case Ia:* For low levels of positive intrinsic motivation (i.e., $\theta_g < \Delta q$), the monitoring in N in the benchmark case is the same as in F and hence $\tilde{\theta}_{bN} = \tilde{\theta}_{bF}$. If θ_b exceeds this threshold level, F implements the optimal contract $\{m_F^{det}, t_F^{det}, w_F^{det}\}$ described in (12) to deter bad workers from sabotage. Given this contract, bad workers will prefer to switch to N and misbehave there if $u_{bN}(d) > u_{bF}^{det}(e)$. To prevent this, it is necessary and sufficient that N raises its monitoring level to the one used in F : $m_N^{det} = m_F^{det}$.
- *Case Ib:* For high levels of positive intrinsic motivation (i.e., $\theta_g \geq \Delta q$), N does not monitor at all in the benchmark case and as a result the level of automatic deterrence is lower than in sector F , i.e., $\tilde{\theta}_{bN} < \tilde{\theta}_{bF}$. If the level of negative motivation is somewhere between these two values,³⁴ then $u_{bN}(d) > u_{bF}(e) > u_{bF}(d)$. That is, given the benchmark contracts, all bad guys will be in sector N , making a destructive effort since there is no monitoring in N . However, N can fully deter bad workers from joining by raising m_N from 0 to the level of monitoring in F , i.e., $m_N^{det} = \underline{m} = m_F^*$.³⁵ If $\theta_b > \tilde{\theta}_{bF}$, then the same results as before apply. That is, F has to choose the optimal contract $\{m_F^{det}, t_F^{det}, w_F^{det}\}$ described in (12) to deter bad actions by bad workers, whereas N optimally deters bad workers from joining by imitating F 's choice of monitoring level: $m_N^{det} = m_F^{det}$.

We conclude that if the level of motivation of bad workers is higher than $\tilde{\theta}_{bN}$, then in an equilibrium with full deterrence in *both* sectors, $m_N^{det} = m_F^{det}$. With bad workers, the mission-oriented sector thus loses much of its cost advantage compared to the for-profit sector. The loss is particularly high when good workers are very motivated, i.e., if $\theta_g \geq \Delta q$: in this case, the presence of bad workers means that N has to go from no monitoring at all to whatever monitoring there is in the for-profit sector. By raising the level of monitoring, destructive behavior in N becomes sufficiently unattractive and bad workers prefer to behave like regular workers in sector F .

How do these changes in incentive schemes affect good workers? For high enough intrinsic motivation of good workers ($\theta_g > m_F^{det} t_F^{det} - m_N^* t_N^*$), N has no need

³³With $\bar{u} = 0$ the benchmark contract in sector N is: $w_N^* = 0$, $m_N^* = \underline{m}$, $t_N^* = (\Delta q - \theta_g)/(2\underline{m})$ if $\theta_g < \Delta q$ (i.e., *Case Ia*) and $w_N^* = m_N^* = t_N^* = 0$ if $\theta_g \geq \Delta q$ (i.e., *Case Ib*).

³⁴That is, if $\tilde{\theta}_{bN} = \Delta q/2 < \theta_b < \tilde{\theta}_{bF} = \Delta q/2 + \underline{m}K$.

³⁵Note that F has no need to change its benchmark contracts since bad workers have no incentive to behave badly in F anyway.

to increase the incentives for good behavior. As m_N goes up, t_N can decrease (Case Ia) or stay at $t_N = 0$ (Case Ib) such that the overall incentives for good workers stay the same. Only if θ_g is small, good workers may be tempted to switch to sector F since this sector now offers higher rewards for good behavior. In that case, N needs also to increase the bonus payment to retain good workers, but less so than F . Even with full deterrence of bad workers, the wage bill in sector N therefore is still lower than in sector F .

4.2.2. Equilibrium Results in the General Case: $\bar{u} \geq 0$

We contrast the results obtained in the particular case where $\bar{u} = 0$ with the results obtained in the general case where $\bar{u} \geq 0$. The details of the computation of the optimal full deterrence contracts can be found in Section B of the online appendix.

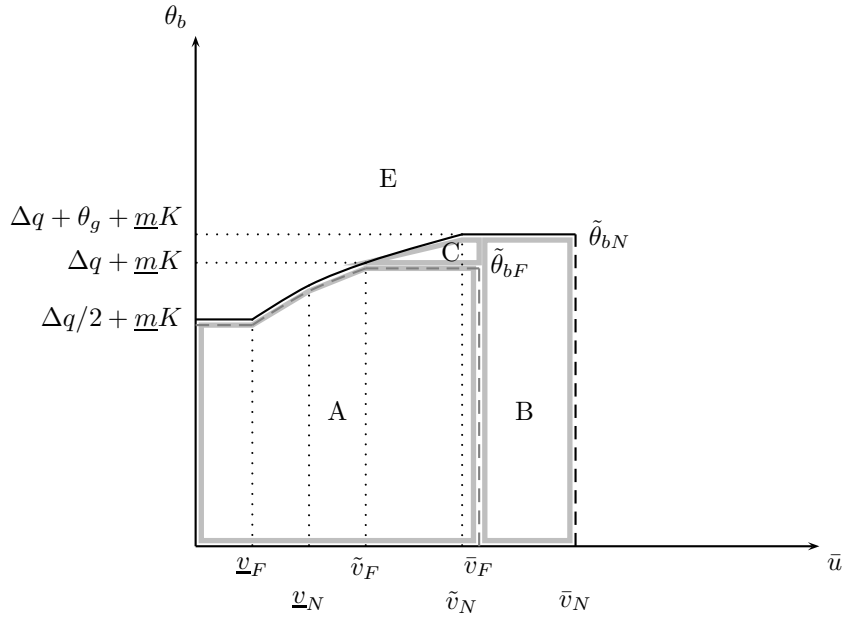
For high values of negative intrinsic motivation θ_b , the deterrence constraint binds. The principal in F needs to raise the expected bonus to make a constructive effort more attractive and/or the monitoring level to make a destructive effort less attractive. For parameter values such that the (LL) and (PC) constraints are binding simultaneously (i.e., kif Case II holds), both in the benchmark scenario and in the scenario with full deterrence, the optimal monitoring level is strictly higher than \underline{m} , whereas the expected bonus is $m_F^{det} t_F^{det} = m_F^* t_F^*$. In all other cases, the principal raises the bonus payment first (the “carrot”), before he starts to raise the monitoring level (the “stick”) so that $m_F^{det} t_F^{det} > m_F^* t_F^*$. The exact details of the resulting optimal contracts are derived in Proposition B.1 in the online appendix.

Sector N can achieve full deterrence by raising the monitoring level just high enough that another option - either moving to F or enjoying their outside utility - becomes more attractive for bad workers. For low levels of outside utility ($\bar{u} < \bar{v}_F$), this condition translates into choosing the same monitoring level as in sector F , whereas for high levels of outside utility ($\bar{u} \geq \bar{v}_F$) the monitoring level in N has to increase, but slightly less so than in F . This is due to the fact that for high reservation utility, the basic wage in N is smaller than in F , making work in N less attractive anyway (see Section B.3 of the online appendix).

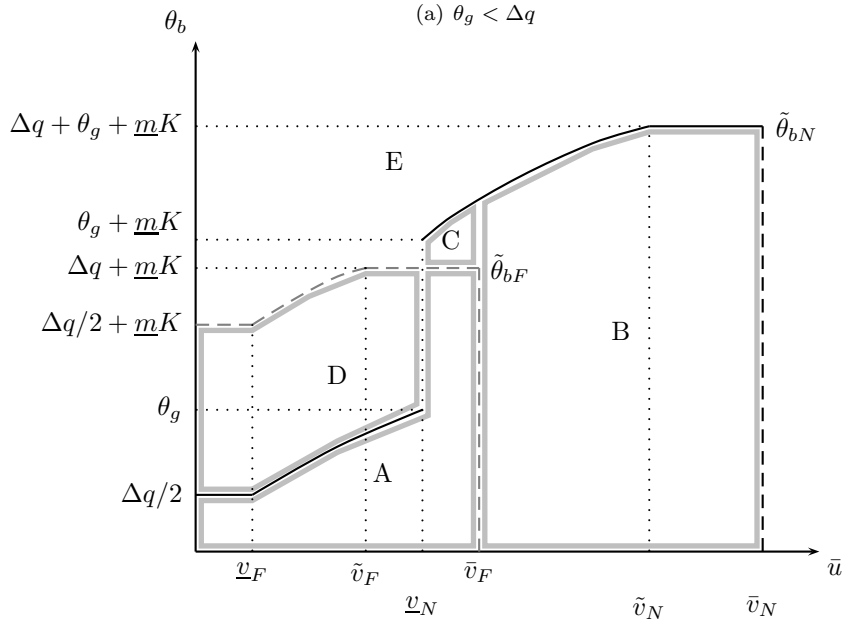
The next proposition collects these results.

Proposition 3. *Assume $\bar{u} \geq \underline{v}_F$. In an equilibrium with full deterrence of destructive behavior:*

- *The bad workers are all in sector F and behave like regular workers.*
- *In F , monitoring increases, $m_F^{det} \geq m_F^*$, and incentives become steeper, $m_F^{det} t_F^{det} \geq m_F^* t_F^*$.*
- *In N , monitoring increases, $m_N^{det} \geq m_N^*$, while monetary incentives stay the same, $m_N^{det} t_N^{det} = m_N^* t_N^*$.*



(a) $\theta_g < \Delta q$



(b) $\theta_g > \Delta q$

Figure 5: Equilibria with full deterrence in both sectors, depending on outside utility \bar{u} and the level of negative intrinsic motivation θ_b .

Sector F optimally combines the carrot and the stick to induce bad workers to choose a constructive effort and behave like regular workers, while sector N relies on the stick to keep them away. This last result is obtained under the assumption that \bar{u} is not too small (i.e., $\bar{u} \geq \underline{v}_F$).

By contrast, when \bar{u} is smaller than \underline{v}_F , for instance when $\bar{u} = 0$, the individual rationality constraint does not bind and the limited liability constraint binds, which implies that the workers in F earn strictly positive rents. Sector N has to make sure that good workers still get a higher payoff from working in N rather than in F . If good workers have a weak intrinsic motivation (i.e., if $\theta_g < m_F^{det} t_F^{det} - m_N^* t_N^*$) and a low reservation utility (i.e., if $\bar{u} < \underline{v}_F$), they will prefer to switch sectors, unless N also raises its bonus level. Yet, as before, N does not have to go all the way in imitating F since it still can count on the intrinsic motivation of good workers. In other words, monetary incentives in N have to be raised such that $m_F^{det} t_F^{det} > m_N^{det} t_N^{det} > m_N^* t_N^*$ if and only if $\theta_b > \tilde{\theta}_{bN}$, $\bar{u} < \underline{v}_F$ and $\theta_g < m_F^{det} t_F^{det} - m_N^* t_N^*$. In all other cases there is no need to adjust monetary incentives: $m_N^{det} t_N^{det} = m_N^* t_N^*$.

Figures 5(a) and 5(b) illustrate the equilibrium outcomes depending on \bar{u} and θ_b . Overall, we distinguish five different cases characterized as follows:

- Area A:** With the benchmark contracts $u_{Fb}(e) > u_{Fb}(d)$ and $u_{Fb}(e) > u_{Nb}(d)$. Therefore, bad workers are in F and behave like regular workers.
- Area B:** For workers with such a high reservation utility, F is not active and only N is a possible employer. However, with the benchmark contracts, $u_{Nb}(d) < \bar{u}$. Joining N is not attractive and bad workers prefer to enjoy their outside utility.
- Area C:** With the benchmark contracts $u_{Fb}(d) > u_{Fb}(e)$ and $u_{Fb}(e) > u_{Nb}(d)$. There is automatic deterrence in sector N , but not in sector F . In equilibrium, F opts for the full deterrence contracts described in Proposition B.1 in the online appendix, whereas N can keep its benchmark contract. All bad workers are in sector F where they choose e .
- Area D:** With the benchmark contracts $u_{Fb}(e) > u_{Fb}(d)$ and $u_{Nb}(d) > u_{Fb}(e)$. There is automatic deterrence in F , but N has to introduce minimal monitoring \underline{m} . Bad workers are in F where they choose e .
- Area E:** With the benchmark contracts $u_{Fb}(d) > u_{Fb}(e)$ and $u_{Nb}(d) > u_{Fb}(e)$. To achieve full deterrence, contracts in both sectors have to be adapted as described in Propositions B.1 and B.2 in the online appendix. Bad workers choose either (F, e) or \bar{u} .³⁶

³⁶The latter holds if only N is a possible employer.

4.3. Partial Deterrence: low D

So far we have considered that the damage D is so large that full deterrence of destructive behavior is the best option. However, under certain circumstances, some destructive behavior may induce only limited damage. In this section, we therefore consider the possibility that destructive behavior occurs in equilibrium.

Let $\beta_i \in [0, 1)$ be the share of bad workers in sector $i = N, F$. Recall that $\theta_{Ng} = \theta_g$ and $\theta_{Fg} = 0$. Taking into account the agent's optimal effort choice, the principal's maximization problem in sector $i = N, F$ corresponds to

$$\max_{w_i, t_i, m_i} \pi_i = (1 - \beta_i)(\Delta q - m_i t_i) \frac{m_i t_i + \theta_{ig}}{a} - \beta_i D \frac{\theta_b - m_i K}{a} + q - w_i - M(m_i) \quad (13)$$

subject to the worker's limited liability and participation constraint as stated in (3) and (4). There is no need to adapt the benchmark contracts of Proposition 1 when $\theta_b \leq \tilde{\theta}_{bi}$. We focus on the case where $\theta_b > \tilde{\theta}_{bi}$. Optimizing (13), using the first order condition on t_i , yields for m_i :³⁷

$$M'(\tilde{m}_i^{part}) = \frac{\beta_i D}{a} K. \quad (14)$$

Since \tilde{m}_i^{part} is not necessarily comprised between \underline{m} and 1, if it is strictly positive, the optimal monitoring level under partial deterrence is

$$m_i^{part} = \min\{\max\{\underline{m}, \tilde{m}_i^{part}\}, 1\}.$$

The level of monitoring, which is independent of θ_b , increases with the damage done by bad workers, D , and their share in the population of sector i , β_i . Moreover, when the reservation utility is not too high (i.e., so that the limited liability constraint binds as in Cases *I* and *II* in Proposition C.3 in the online appendix), the optimal bonus solves $m_i^{part} t_i^{part} = m_i^* t_i^*$. The optimal contract with partial deterrence has thus two main features: The organization tries to limit damages and to discourage destructive behavior through higher monitoring, $m_i^{part} \geq m_i^*$, while keeping the same expected payoff from constructive effort as in the benchmark contract, $m_i^{part} t_i^{part} = m_i^* t_i^*$.

All bad workers face the same tradeoffs. In equilibrium either all or none of them are in sector $i = N, F$ such that $\beta_i \in \{0, \bar{\beta}_i\}$, where $\bar{\beta}_F \equiv x_b/(x_b + x_r)$ and $\bar{\beta}_N \equiv x_b/(x_b + x_g)$. Since by Assumption 1, there are more regular workers than good ones, $\bar{\beta}_F = \frac{x_b}{x_b + x_r} < \bar{\beta}_N = \frac{x_b}{x_b + x_g}$ such that $\tilde{m}_F^{part} \leq \tilde{m}_N^{part}$ in (14). Bad workers will rather work in F where their probability of detection is lower than in N . We deduce that, as long as F is active, sector N never implements partial deterrence. The only case where the principal in N might opt for partial

³⁷To see this point let $\bar{u} = 0$ so that the (PC) constraint never binds while the (LL) constraint binds. Setting $w_i = 0$ and optimizing (13) with respect to t_i and m_i yields: $m_i^{part} t_i^{part} = \max\{0, (\Delta q - \theta_{ig})/2\}$ and $m_i^{part} = \min\{\max\{\underline{m}, \tilde{m}_i^{part}\}, 1\}$, where \tilde{m}_i^{part} solves (14). For the general case see the proof of Proposition C.3 in the online appendix.

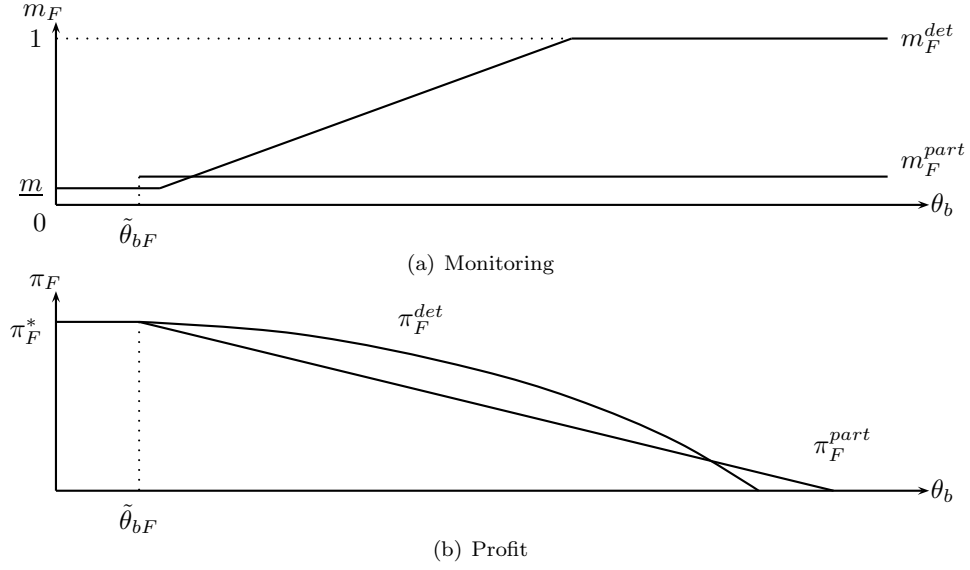


Figure 6: Comparison of monitoring levels and profit in F under full and partial deterrence, depending on the level of negative intrinsic motivation θ_b . The graphs are drawn for $M(m) = m^2/2$ and $\underline{u} = 0$. \underline{m} denotes the minimum monitoring level, and $\tilde{\theta}_{bF}$ the level of negative motivation above which bad workers are no longer automatically deterred from bad behavior under the benchmark contract in F .

deterrence is when the reservation utility of regular workers is so high that only sector N is a possible employer. In all other cases, N implements either full deterrence (i.e., by choosing the same/lower monitoring level as in sector F depending on the value of \bar{u}), or no deterrence at all. To see this point consider the case where $\theta_g \geq \Delta q$ and $\bar{u} \in [0, \underline{v}_N]$ so that the benchmark contract in N is $w_N^* = m_N^* = t_N^* = 0$. In presence of bad workers this contract yields the profit $\Pi_N^* = q + (1 - \bar{\beta}_N)\Delta q\theta_g/a - \bar{\beta}_N D\theta_b/a$, while the full deterrence profit is $\Pi_N^{det} = q + \Delta q\theta_g/a - M(m_N^{det})$. The former is higher than the latter if $[aM(m_N^{det})(1 + x_g/x_b) - D\theta_b]/\Delta q \leq \theta_g$. Then, N sticks to its benchmark contracts even if it attracts all bad workers. The next proposition collects this result.³⁸

Proposition 4. : *As long as both sectors are active, full deterrence in N is optimal unless $\bar{u} \in [0, \underline{v}_N]$ and $[aM(m_N^{det})(1 + x_g/x_b) - D\theta_b]/\Delta q > \theta_g \geq \Delta q$, in which case sector N sticks to the benchmark contract $w_N^* = m_N^* = t_N^* = 0$.*

Except when the intrinsic motivation of good workers is very high and the expected damage caused by bad workers is sufficiently low, sector N adapts its monitoring to deter bad workers from joining and the for-profit sector is again

³⁸The full description of the equilibrium with partial deterrence is derived in Section C.2 of the online appendix.

confronted with bad workers. Whether the principal in sector F then prefers full deterrence of destructive behavior or whether he opts for partial deterrence depends on his respective expected profit in the two cases. Under the former regime, his expected profit is

$$\pi_F^{det} = q + (\Delta q - m_F^{det} t_F^{det}) \frac{m_F^{det} t_F^{det}}{a} - w_F^{det} - M(m_F^{det}),$$

whereas in the latter case his profit becomes

$$\begin{aligned} \pi_F^{part} = & (1 - \bar{\beta}_F)(\Delta q - m_F^{part} t_F^{part}) \frac{m_F^{part} t_F^{part}}{a} - \bar{\beta}_F D \frac{\theta_b - m_F^{part} K}{a} \\ & + q - w_F^{part} - M(m_F^{part}). \end{aligned}$$

As can be seen easily from the second function, the expected profit with partial deterrence is strictly decreasing in the share of bad workers in sector F , $\bar{\beta}_F$ and in the damage these workers may cause D . If, for instance, the number of regular workers in the population x_r is very high, this implies that the relative share of bad workers in sector F , $\bar{\beta}_F$, is low and full deterrence hence is less attractive. Furthermore, the monitoring technology also plays a role. If the marginal cost of an increased level of monitoring is high, then full deterrence may be too costly.

These considerations are illustrated in Figure 6.³⁹ The graph shows that for a given level of $\bar{\beta}_F$ and D , full deterrence is optimal as long as the intrinsic motivation of bad workers, θ_b , is low. However, as θ_b increases, full deterrence becomes too costly and the principal hence prefers partial deterrence. Note, however, that this conclusion depends on the exact level of $\bar{\beta}_F$ and D . For extreme cases such as very low levels of $\bar{\beta}_F$ and/or D , i.e., when m_F^{part} is close to \underline{m} , the profit under partial deterrence would be higher than the profit under full deterrence over the entire relevant range of θ_b values, whereas the opposite is true if $\bar{\beta}_F$ and/or D are so high that m_F^{part} is close to 1.

5. Discussion and Implications in Child Care

The analysis in Section 4 has important implications for labor management, especially in nonprofit organizations. The presence of ill motivated people always spells trouble for these organizations since such workers never join nonprofits to do good. It is not easy to find empirical evidence to illustrate this result in economic studies as they focus on standard moral hazard issues (e.g., free-riding, shirking at work), but do not generally consider negative intrinsic motivation. By contrast, anti-social behaviors have been the focus of researchers in other fields of social science and in medicine. In particular, there is a vast literature

³⁹The graphs are drawn for $\underline{u} = 0$, $q = 2$, $\Delta q = 1.4$, $a = 15$, $\underline{m} = 0.1$, $K = 3$, $x_b = 0.2$, $x_r = 0.5$, $D = 3$ and $M(m) = m^2/2$.

in modern psychiatry, psychology and sociology on the issue of pedophilia. We thus rely on these studies to illustrate some of our results in the context of child care services.

The example of child care is relevant as it is an area where both for-profit and nonprofit organizations coexist to offer services (e.g., in daycare, education, sports training, music, hobbies, transport, vacation). It is also an area that tends to instill high intrinsic motivation in people, generally good, but sometimes also bad, as in the case of pedophiles.

Pedophilia is defined by health professionals as a mental disorder, which has yet no cure.⁴⁰ Recently, researchers began reporting a series of findings linking pedophilia with brain structure and function, suggesting that genetic factors play a role in the development of this paraphilia (Gaffney et al. 1984; Schiffer et al. 2007; Schiltz et al. 2007; Walter et al. 2007; Cantor et al. 2008; Schiffer et al. 2008). The typical offender is male, begins molesting by age 15, engages in a variety of deviant behaviors, and molests an average of 117 youngsters (Abel, 1985). However the most serious and chronic offenders often show signs of antisocial behavior as early as the preschool years (American Psychiatric Association, 1994).⁴¹ These findings are consistent with the assumption made in the paper that negative intrinsic motivation is exogenous.

Contrary to common belief, pedophiles generally know their victims. Acquaintance perpetrators are hence the most common abusers, constituting approximately 70-90% of all reported perpetrators (Finkelhor, 1994; Whealin and Barnett, 2007). For instance in US schools “nearly 9.6% of students are targets of educator sexual misconduct”. Teachers, coaches, substitute teachers, bus drivers and teacher’s aides are in decreasing order totaling 69% of the offenders (Shakeshaft, 2004, p.24-25).⁴² Consistently with our theory, studies show that

⁴⁰The International Classification of Diseases (ICD) defines pedophilia as a “disorder of adult personality and behavior” in which there is a sexual preference for children of prepubertal or early pubertal age. Similarly the Diagnostic and Statistical Manual of Mental Disorders (DSM), defines it as a paraphilia in which a person has intense and recurrent sexual urges towards and fantasies about prepubescent children and on which feelings they have either acted or which cause distress or interpersonal difficulty.

⁴¹A Mayo Clinic report states that approximately 95% of incidents of sexual abuse of children age 12 and younger are committed by offenders who meet the diagnostic criteria for pedophilia. While pedophilic child molesters make up “only” 65% of child molestation offenders, they commit ten times more sexual acts against children than non-pedophilic child molesters explaining the discrepancy between these two statistics (Hall and Hall, 2007).

⁴²According to the FBI, child molestation is with rape one of the most underreported crimes. Although it is difficult to measure with accuracy because of the massive underreporting, all existing studies show that child sexual abuse is quite common in advanced economies. Studies in the United States hence show that 17% to 31% of females and 7% to 16% of males experienced unwanted sexual contact before the age of 18 years (Hall and Hall, 2007; Finkelhor, 1994). Similarly, two Canadian studies, which randomly sampled 750 women and 750 men between the ages of 18 and 27 years, found that 32% of the women and 15.6% of the men had experienced “unwanted sexual contact” before the age of 17 (Hall and Hall, 2007). Studies in the UK estimate prevalence at about 5% to 8% for boys and 12% to 18% for girls (Baker and Duncan, 1985; Radford, Corral, Bradley, Fisher, Bassett, Howat, and Collishaw,

pedophiles tend to choose occupations where they have the opportunity to interact with children in an unsupervised way. They often have good interpersonal skills and can easily gain the trust of children and their parents. If they do not work directly with children, they volunteer their services to sport teams, Scout troops, or religious or civic organizations that serve youth.⁴³ They sometimes also become foster parents to have an easy access to vulnerable children (Hall and Hall, 2007).

The analysis in Section 4 shows that the mission-oriented sector is the target of bad people whenever it relies heavily on the intrinsic motivation of its workers and does not monitor enough. We thus predict an over-representation of pedophiles in organizations that deal with youth and exert a lax control over their members. This helps to explain the magnitude of sexual abuse scandals that hit the Catholic Church, particularly in the US and in Ireland. According to the last statistic of the Bishop Accountability.org web site, 6,115 priests have been exposed in the US by their archdioceses as pedophile, representing 5.6% of the 109,694 priests in ministry during the period 1950-2002. This number appears to be a lower bound. Indeed in US dioceses compelled to release internal abuse files to law enforcement authorities or the public (i.e., in Providence, Philadelphia, Manchester, Covington) the percentage of accused priests is close to 10 percent.⁴⁴ Although it is difficult to establish the exact percentage of child abusers in the general population, it is much lower than 10 or even 6 percent, as pedophilia is a rare disorder. For instance in their 1993 nationwide survey on sexual practices for the US, Janus and Janus (1993) found that 2 percent of sampled males reported having had sex with a child. In another study, 1 percent of males, who were anonymously surveyed, reported having sexually assaulted a child themselves since they became an adult (Bagley, Wood, and Young, 1994). Finally, Seto (2008) suggests that perhaps 3% of men are sexually attracted to prepubescent children.

According to the theory above, the over-representation of child molesters in the catholic clergy is the result of a policy of *laissez-faire* by the church's dioceses and parishes. Indeed, data in the 2011 John Jay report indicate that, when allegations of abuse were made, the majority of diocesan leaders took actions to help "rehabilitate" the abusive priests rather than to focus on the victims (Terry et al., 2011). Most other cases involving the Catholic church were handled in a similar way. In the scandals involving the Catholic Brothers in Ireland, abusers had hence nothing to fear because everything was covered up and there was no punishment to be expected, just possibly a transfer to a different school (CICA,

2011).

⁴³This point is illustrated by the case of the late BBC presenter Jimmy Savile, who publicly was seen as a modern days saint for his efforts to help children and youth through his TV shows and extensive fund raising, while at the same time he apparently used his position to molest and abuse minors and young women. See the so-called Yewtree Report by Gray and Watt (2013).

⁴⁴See also the Sipe Report at http://www.richardsipe.com/reports/sipe_report_X.htm entitled "Public disclosure cause no. 141-198356-03 in the district court of Tarrant county, Texas"

2009). This explains that in addition to the criminal prosecutions of the abusers there have been many civil lawsuits against the Catholic hierarchy who failed to report sex abuse allegations to the legal authorities and misled victims by deliberately relocating the accused priests.⁴⁵ In the United States alone the various Roman Catholic Dioceses have paid out approximately \$1 billion to settle hundreds of such lawsuits since the early 1990s.

An interesting question is how malevolently motivated individuals, such as pedophiles, could have been tolerated by mission-oriented organizations along the equilibrium path.⁴⁶ According to the theory this can only happen when the cost - to the organization - of such actions is relatively low.⁴⁷ The implication then is that these cases of sexual abuse of children were not costly to church-based organizations in the not so distant past because so much secrecy persisted, and victims were too ashamed to speak out.⁴⁸ Without public reckoning, the organizations could simply carry on, while exposing the pedophile would have had a dramatic impact on the number of churchgoers and on the Church's funding. For the organizations, the costs of the scandal were higher than the costs in terms of adverse effects on human lives by tolerating acts of pedophilia. This is illustrated by the CICA report on Ireland: "Cases of sexual abuse were managed with a view to minimising the risk of public disclosure and consequent damage to the institution and the Congregation. (...) The desire to protect the reputation of the Congregation and institution was paramount. Congregations asserted that knowledge of sexual abuse was not available in society at the time and that it was seen as a moral failing on the part of the Brother or priest. This assertion, however, ignores the fact that sexual abuse of children was a criminal offence", (CICA, 2009, pp. 21-22). What is puzzling though, is that mission-motivated individuals, who themselves were not abusive, allowed this to continue and covered things up. One explanation for this may be loyalty within organizations as described in the CICA report: "Congregational loyalty enjoyed priority over other considerations including safety and protection of children," (CICA, 2009, p. 22). Another motive may be a prevailing norm that wrongdoers deserve a second chance to redeem themselves.⁴⁹ This position is summarized by

⁴⁵Charges have also been brought against the government. In January 2014, the European Court of Human Rights ruled that the Irish State had failed to put in place "any mechanism of effective State control against the risks of such abuse occurring"(cf. Grand chamber judgement O'Keefe vs. Ireland, January 28, 2014.

⁴⁶We are grateful to an anonymous referee for suggesting this discussion.

⁴⁷Only if the damage and the relative share of bad workers are very small, it is cheaper for N to accept the presence of bad workers rather than to deter them.

⁴⁸According to the CICA report on Ireland (CICA, 2009, p. 14): "[Contemporary] complaints were ignored, witnesses were punished, or pressure was brought to bear on the child and family to deny the complaint and/or to remain silent. Witnesses reported that their sense of shame, the power of the abuser, the culture of secrecy and isolation and the fear of physical punishment inhibited them in disclosing abuse."

⁴⁹In a report on the incidents in the Archdiocese of Chicago, the Chicago Tribune summarizes: "The clergymen (...) were shielded by Roman Catholic Church officials who thought the men could be cured with counseling, by bishops blinded by a belief in second chances and forgiveness, and by leaders concerned a public scandal would damage the church's image."

Salesian Cardinal Oscar Rodriguez of Honduras: “For me it would be a tragedy to reduce the role of a pastor to that of a cop. I’d be prepared to go to jail rather than harm one of my priests.”

In sharp contrast with previous studies, our results show that nonprofit organizations cannot save on monitoring costs if they are willing to fulfill their missions. Monitoring not only helps to keep track of workers’ good performances, but it is also the main tool to control sabotage and anti-social behaviors. In the theory above, we did not discuss the exact nature of the monitoring technology. However, in practice it can take different forms from standard monitoring by supervisors at the work site to camera and audio recording (e.g., in police cars). The problem is that in some cases (i.e., in time of crisis, in remote locations, when job requirements imply working outside the organization) it is very difficult to monitor workers. To prevent destructive behavior, organizations might choose to promote peer monitoring.⁵⁰ Empirical studies such as Knez and Simester (2001) and Hamilton, Nickerson, and Owan (2003) have found that team incentives and mutual monitoring may indeed have positive effects on workers’ effort. Sometimes information campaigns may be enough to raise the awareness of potential misbehavior and thus alert coworkers and employers to keep their eyes open. This approach has for instance been proposed by the South Carolina Forestry Commission to address fire fighter arson which is not an uncommon problem (see Stambaugh and Styron (2003)). In other cases, peer monitoring can be induced through simple institutional features, such as letting employees work pairwise, as it is common for police officers, hiring couples, or providing joint housing for aid workers.⁵¹ While this may give rise to collusion among evil-doers as in the United Nation “sex for food” scandals, such a scheme is likely to work reasonably well if there are enough good motivated workers who care about the mission of the organization they work for. In this case promoting heterogeneity in the workforce composition is a good way to avoid collusion of the bad.

Finally, while our model suggests that, in general, organizations should increase their monitoring efforts to deal with the problem of negative intrinsic motivation, this increase of ex post monitoring may be very costly, and, as illustrated in Figure 6, for high levels of negative motivation it becomes even entirely ineffective. Firms therefore may want to invest in ex ante measures of screening to reduce the probability of hiring a bad worker in the first place. A better candidate selection process thus can serve as a (partial) substitute for worker

(Chicago Tribune, January 21, 2014).

⁵⁰There are relatively few theoretical papers on peer monitoring, exceptions being Barron and Gjerde (1997) and Kandel and Lazear (1992), who both analyze the interaction between peer pressure and the provision of incentives in teams.

⁵¹The French service for teaching abroad prefers to hire couples, not only for monitoring reasons, but also because they have been found to withstand stress caused by a new environment better. “German Doctors”, a German NGO provides joint housing to doctors doing short term volunteer work.

monitoring.⁵² The higher the expected damage of hiring a bad worker, the more an organization or firm will be inclined to invest in applicant screening. This is observed in sectors where candidates, once hired, are difficult to fire, as for example civil servants⁵³ or where the stakes are high as for instance in intelligence services.⁵⁴ In other cases, establishing a clearer profile of bad workers may help. This has, for instance, been done in the US to prevent fire fighter arson.⁵⁵ In the case of child care, legal requirements may also play an important role in order to help employers screen out bad workers.⁵⁶

6. Conclusion

The existence of destructive workers who derive satisfaction from actions that are detrimental to their employer or others has an impact on the optimal monitoring and wage contracts offered in organizations. In particular, we discussed how this affects nonprofit organizations that rely on the intrinsic motivation of their workers. Without bad workers, the mission-oriented sector N can save on wage and monitoring costs compared to the profit-oriented sector F . If the intrinsic motivation of good workers is high enough, it may even forego bonus payments and monitoring altogether. However, the lack of monitoring and extrinsic incentives makes N particularly vulnerable to destructive behavior by bad workers. Indeed, when bad workers join the nonprofit sector it is only to follow their destructive instincts and not because they want to provide a constructive effort.

Unless the damage and the relative share of bad workers are very small, in which case it is cheaper for N to accept their presence, the equilibrium solutions both under full and partial deterrence show that N will always want to change its contracts to deter bad workers from joining. To do so, sector N needs to increase its monitoring level, which erodes its cost advantage compared to the for-profit sector. To what extent this is the case also depends on the level of motivation

⁵²See Huang (2007) and Huang and Cappelli (2006) for a discussion on the possible tradeoff between worker monitoring and ex ante applicant screening.

⁵³Goldman (1982) and Greenberg and Haley (1986) discuss this issue for the case of judges in the United States.

⁵⁴The selection process in these cases can be quite lengthy and generally involves all kinds of tests and background checks. See for instance the CIA web site at www.cia.gov/careers/faq/index.html#a3, or the FBI one www.fbijobs.gov/61.asp#3

⁵⁵Studies by the South Carolina Forestry Commission and the FBI (see Stambaugh and Styron (2003) for a summary of both studies) have found that arsonists are typically white males between 17 and 26 years of age, with a difficult family background, lacking social and interpersonal skills, often of average intelligence but with poor academic performance. Also, arson seems to be more likely with volunteer fire fighters than with professionals ones.

⁵⁶In Germany, for instance, employers can ask applicants to a job that may bring them in contact with children or youths for an “extended police clearance certificate” (“erweitertes Führungszeugnis”), documenting any offence including cases of molestation, child pornography, exhibitionism etc.

of good workers: For high enough motivation of good workers, the mission-oriented sector can achieve full deterrence by choosing the same monitoring level as in sector F , but otherwise keeping extrinsic incentives at the same level as before. That is, to the same extent that the monitoring level increases, the bonus payment decreases such that the overall effort incentives are still at their optimal level. The mission-oriented sector therefore still may enjoy a certain cost advantage, since it is cheaper to get already motivated workers to provide effort.

Given this optimal reaction of sector N , sector F will be confronted with bad workers. In order to reduce their negative impact, the profit-oriented sector has to increase its bonuses and its monitoring levels. We showed that to achieve full deterrence of destructive actions by bad workers, F may even have to increase effort incentives beyond the first best level. Overall, the optimal incentive policies of sectors N and F imply that, in equilibrium, all bad workers are generally in F where they behave like regular workers. It is important to note, however, that to achieve this equilibrium outcome it is not enough to reward good behavior better, but both kinds of organization have to invest in monitoring to deter bad behavior.

A question that remains open is why some organizations in the mission oriented sector are not implementing the optimal contract in the first place and, even worse, hide crimes from the general public, thereby protecting the bad agents. Although answering this question is beyond the scope of the paper, we want to discuss an hypothesis that we wish to investigate in future work. In order to focus on the incentive problems raised by the presence of bad workers, we have not taken into account other differences between profit- and mission-oriented organizations. Yet it may be worthwhile to take a look at those differences, in particular the way organizations are financed: While profit-oriented organizations usually have to survive on the proceeds from their sales, which provides them with strong incentives to please their customers, many mission-oriented organizations are run as non-governmental organizations or associations that essentially depend on donations. For them, the scandal caused by bad workers may hence also have considerable negative consequences for their funding, thus making deterrence of bad workers all the more important. However, it also provides strong incentives to hide bad actions by their workers, hence implicitly encouraging them to continue.

Another aspect that needs to be discussed is the effect of control on the intrinsic motivation of good workers. There is a recent literature on the crowding out of intrinsic motivation by extrinsic incentives or control.⁵⁷ Taking into account such effects would mean that the more the mission-oriented sector N increases monitoring in order to prevent damage from bad workers, the lower would be the intrinsic motivation of good workers. N would therefore also have to increase his monetary incentives in order to induce good workers to work hard enough,

⁵⁷See Seabright (2009), Frey and Jegen (2001), Frey and Oberholzer-Gee (1997).

thus losing its cost advantage. Eventually, good intrinsic motivation would disappear all together and organizations in sector N would operate under the same conditions as firms in the profit-oriented sector F and also offer the same contracts.

However, it is unclear to what extent such crowding out of intrinsic motivation actually exists in the context considered here. Motivation crowding out seems to be affected by other factors than the level of monitoring, such as framing and general treatment by the employer (Nagin et al., 2002). As Akerlof and Kranton (2008) underline, “What matters is not more or less monitoring per se, but how employees think of themselves in relation to the firm” (Akerlof and Kranton (2008), p. 212). If it is made clear that monitoring is increased in order to reduce fraud and anti-social behavior, the motivation of good workers should not be too much affected.

ABBINK, K., AND B. HERRMANN (2009): “Pointless Vendettas,” Discussion paper, CBESS Discussion Paper 09-10, University of East Anglia, Centre for Behavioural and Experimental Social Science.

ABBINK, K., AND A. SADRIEH (2009): “The pleasure of being nasty,” *Economics Letters*, 105, 306–308.

ABEL, G. (1985): *The Evaluation of Child Molesters: Final Report to the Center on Antisocial and Violent Behavior*. National Institute of Mental Health, Rockville, MD.

AKERLOF, G., AND R. KRANTON (2005): “Identity and the Economics of Organizations,” *Journal of Economic Perspectives*, 19(1), 9–32.

——— (2008): “Identity, Supervision, and Work Groups,” *American Economic Review*, 98(2), 212–217.

AMERICAN PSYCHIATRIC ASSOCIATION (1994): *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association, Washington, DC, 4th edn.

BAGLEY, C., M. WOOD, AND L. YOUNG (1994): “Victim to abuser: Mental health and behavioral sequels of child sexual abuse in a community survey of young adult males,” *Child Abuse & Neglect*, 18(8), 683–697.

BAKER, A., AND S. DUNCAN (1985): “Child sexual abuse: a study of prevalence in Great Britain,” *Child Abuse & Neglect*, 9(4), 457–67.

BALLOU, J., AND B. WEISBROD (2003): “Managerial rewards and the behavior of for-profit, governmental, and nonprofit organizations: evidence from the hospital industry,” *Journal of Public Economics*, 87, 1895–1920.

BANERJEE, A., AND E. DUFLO (2006): “Addressing Absence,” *Journal of Economic Perspectives*, 20(1), 117–132.

- BARRON, J., AND K. GJERDE (1997): "Peer Pressure in an Agency Relationship," *Journal of Labor Economics*, 15(2), 234–254.
- BECKER, G. (1968): "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76(2), 169–217.
- BÉNABOU, R., AND J. TIROLE (2003): "Intrinsic and Extrinsic Motivation," *Review of Economic Studies*, 70, 489–520.
- (2010): "Individual and Corporate Social Responsibility," *Economica*, 77(305), 1–19.
- BESLEY, T., AND M. GHATAK (2005): "Competition and Incentives with Motivated Agents," *American Economic Review*, 95(3), 616–636.
- BORZAGA, C., AND E. TORTIA (2006): "Worker Motivations, Job Satisfaction, and Loyalty in Public and Nonprofit Social Services," *Nonprofit and Voluntary Sector Quarterly*, 35(2), 225–248.
- CANTOR, J. M., N. KABANI, B. K. CHRISTENSEN, R. B. ZIPURSKY, H. E. BARBAREE, R. DICKEY, P. E. KLASSEN, D. J. MIKULIS, M. E. KUBAN, AND T. BLAK (2008): "Cerebral white matter deficiencies in pedophilic men," *Journal of Psychiatric Research*, 42(3), 167–183.
- CHAUDHURY, N., J. HAMMER, M. KREMER, K. MURALIDHARAN, AND F. H. ROGERS (2006): "Missing in Action: Teacher and Health Worker Absence in Developing Countries," *Journal of Economic Perspectives*, 20(1), 91–116.
- CICA (2009): "Report of the Commission to Inquire into Child Abuse - Executive Summary," Discussion paper, Commission to Inquire into Child Abuse.
- DELFGAAUW, J., AND R. DUR (2008): "Incentives and Workers' Motivation in the Public Sector," *Economic Journal*, 118, 171–191.
- DICKENS, W., L. KATZ, K. LANG, AND L. SUMMERS (1989): "Employee Crime and the Monitoring Puzzle," *Journal of Labor Economics*, 7(3), 331–347.
- DIXIT, A. (2002): "Incentives and Organizations in the Public Sector," *Journal of Human Resources*, 37(4), 696–727.
- FINKELHOR, D. (1994): "Current information on the scope and nature of child sexual abuse," *The Future of Children*, 4(2), 31–53.
- FRANCOIS, P. (2000): "'Public service motivation' as an argument for government provision," *Journal of Public Economics*, 78, 275–299.
- (2003): "Not-for-profit Provision of Public Services," *Economic Journal*, 113(486), C53–C61.
- FREY, B. (1997): *Not Just for the Money: an economic theory of personal motivation*. Edward Elgar Publishing.

- FREY, B., AND R. JEGEN (2001): "Motivation Crowding Theory," *Journal of Economic Surveys*, 15(5), 589–611.
- FREY, B., AND F. OBERHOLZER-GEE (1997): "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out," *American Economic Review*, 87(4), 746–755.
- GAFFNEY, G., S. LURIE, AND F. BERLIN (1984): "Is there familial transmission of pedophilia?," *Journal of Nervous and Mental Disease*, 172(9), 546–8.
- GIBELMAN, M., AND S. GELMAN (2004): "A Loss of Credibility: Patterns of Wrongdoing Among Nongovernmental Organisations," *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, 15(4), 355–381.
- GLAESER, E. (2002): "The Governance of Not-for-Profit Firms," Harvard Institute of Economic Research Discussion Paper 1954.
- GOLDMAN, S. (1982): "Judicial Selection and the Qualities That Make a "Good" Judge," *Annals of the American Academy of Political and Social Science*, 462, 112–124.
- GRAY, D., AND P. WATT (2013): "Giving Victims a Voice," Discussion paper, MPS and NSPPC.
- GREENBERG, P., AND J. HALEY (1986): "The Role of the Compensation Structure in Enhancing Judicial Quality," *Journal of Legal Studies*, 15(2), 417–426.
- HALL, R. C., AND R. C. HALL (2007): "A Profile of Pedophilia: Definition, Characteristics of Offenders, Recidivism, Treatment Outcomes, and Forensic Issues," *Mayo Clin Proceedings (MAYO Foundation for medical education and research)*, 82(4), 457–471.
- HAMILTON, B., J. NICKERSON, AND H. OWAN (2003): "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation," *Journal of Political Economy*, 111(3), 465–497.
- HECKMAN, J., J. SMITH, AND C. TABER (1996): "What Do Bureaucrats Do? The Effects of Performance Standards and Bureaucratic Preferences on Acceptance into the JTPA Program," NBER Working Paper 5535.
- HELMING, E., H. KINDLER, A. LANGMEYER, M. MAYER, C. ENTLEITNER, P. MOSSER, AND M. WOLFF (2011): "Sexuelle Gewalt gegen Mädchen und Jungen in Institutionen - Rohdatenbericht," Discussion paper, Deutsches Jugendinstitut, München, Germany.
- HUANG, F. (2007): "To Trust or to Monitor: A Dynamic Analysis," SMU Economics & Statistics Working Paper No. 11-2007.
- HUANG, F., AND P. CAPPELLI (2006): "Employee Screening: Theory and Evidence," NBER Working Paper 12071.

- JANUS, S. S., AND C. L. JANUS (1993): *The Janus Report on Sexual Behavior*. Wiley, Oxford, England.
- KANDEL, E., AND E. LAZEAR (1992): “Peer Pressure and Partnerships,” *Journal of Political Economy*, 100(4), 801–817.
- KNEZ, M., AND D. SIMESTER (2001): “Firm Wide Incentives and Mutual Monitoring at Continental Airlines,” *Journal of Labor Economics*, 19(4), 743–772.
- KOSFELD, M., AND F. V. SIEMENS (2011): “Competition, cooperation, and corporate culture,” *RAND Journal of Economics*, 42(1), 23–43.
- KREPS, D. (1997): “Intrinsic Motivation and Extrinsic Incentives,” *American Economic Review*, 87(2), 359–364.
- MILGRAM, S. (1974): *Obedience to Authority*. Harper & Row, New York.
- MURDOCK, K. (2002): “Intrinsic Motivation and Optimal Incentive Contracts,” *RAND Journal of Economics*, 33(4), 650–671.
- 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.
- POSNER, R. (1993): “What Do Judges and Justices Maximize? (The Same Thing Everybody Else Does),” *Supreme Court Economic Review*, 3, 1–41.
- PRENDERGAST, C. (2007): “The Motivation and Bias of Bureaucrats,” *American Economic Review*, 97(1), 180–196.
- RADFORD, L., S. CORRAL, C. BRADLEY, H. FISHER, C. BASSETT, N. HOWAT, AND S. COLLISHAW (2011): “Child abuse and neglect in the UK today,” Discussion paper, NSPCC.
- SAVE THE CHILDREN UK (2006): “From Camp to Community: Liberia Study of Exploitation of Children,” Discussion paper, Save the Children UK.
- (2008): “No One to Turn To: The under-reporting of child sexual exploitation and abuse by aid workers and peacekeepers,” Discussion paper, Save the Children UK.
- SCHIFFER, B., T. PAUL, E. GIZEWSKIB, M. FORSTING, N. LEYGRAF, M. SCHEDLOWSKI, AND T. H. KRUGER (2008): “Functional brain correlates of heterosexual paedophilia,” *NeuroImage*, 41(1), 80–91.
- SCHIFFER, B., T. PESCHEL, T. PAUL, E. GIZEWSKI, M. FORSTING, N. LEYGRAF, M. SCHEDLOWSKI, AND T. H. C. KRUEGER (2007): “Structural brain abnormalities in the frontostriatal system and cerebellum in pedophilia,” *Journal of Psychiatric Research*, 41, 753–762.

- SCHILTZ, K., J. WITZEL, G. NORTHOFF, K. ZIERHUT, U. GUBKA, H. FELLMAN, J. KAUFMANN, C. TEMPELMANN, C. WIEBKING, AND B. BOGERTS (2007): "Brain pathology in pedophilic offenders: Evidence of volume reduction in the right amygdala and related diencephalic structures," *Archives of General Psychiatry*, 64(6), 737–746.
- SEABRIGHT, P. (2009): "Continuous Preferences and Discontinuous Choices: How Altruists Respond to Incentives," *The Berkeley Electronic Journal of Theoretical Economics*, 9(1), Article 14.
- SERRA, D., P. SERNEELS, AND A. BARR (2011): "Intrinsic motivations and the non-profit health sector: Evidence from Ethiopia," *Personality and Individual Differences*, 51, 309–314.
- SETO, M. (2008): *Understanding pedophilia and sexual offending against children: Theory, assessment, and intervention*. American Psychological Association, Washington, DC.
- SHAKESHAFT, C. (2004): "Educator Sexual Misconduct: A Synthesis of the Literature," Discussion paper, U.S. Department of Education.
- STAMBAUGH, H., AND H. STYRON (2003): "Special Report: Firefighter Arson," Discussion Paper USFA-TR-141, U.S. Fire Administration.
- TERRY, K. (2008): "Stained Glass: The Nature and Scope of Child Sexual Abuse in the Catholic Church," *Criminal Justice and Behavior*, 35(5), 549–569.
- TERRY, K. J., M. L. SMITH, K. SCHUTH, J. R. KELLY, B. VOLLMAN, AND C. MASSEY (2011): "The Causes and Context of Sexual Abuse of Minors by Catholic Priests in the United States, 1950-2010," Discussion paper, A Report Presented to the United States Conference of Catholic Bishops by the John Jay College Research Team, ISBN 978-1-60137-201-7.
- WALTER, M., J. WITZEL, C. WIEBKING, U. GUBKA, M. ROTTE, K. SCHILTZ, F. BERMPOHL, C. TEMPELMANN, B. BOGERTS, H. J. HEINZE, AND G. NORTHOFF (2007): "Pedophilia is Linked to Reduced Activation in Hypothalamus and Lateral Prefrontal Cortex During Visual Erotic Stimulation," *Biological Psychiatry*, 62(6), 698–701.
- WHEALIN, J., AND E. BARNETT (2007): "Child Sexual Abuse," Discussion paper, National Center for Post Traumatic Stress Disorder, US Department of Veterans Affairs.
- ZIMBARDO, P. (2004): "A Situationist Perspective on the Psychology of Evil: Understanding How Good People Are Transformed into Perpetrators," in *The social psychology of good and evil: Understanding our capacity of kindness and cruelty*, ed. by A. Miller. Guilford, New York.

Anti-Social Behavior in Profit and Nonprofit Organizations: Online Appendix

Emmanuelle Auriol Stefanie Brilon

May 6, 2014

A Proofs for Basic Setup

A.1 Proof of Proposition 1

After inserting the incentive constraint, the maximization problem stated in (2) to (4) can be rewritten as the following Lagrangian:

$$\begin{aligned} \max_{w_{ij}, m_i, t_{ij}, \lambda_{LL}, \lambda_{PC}} \quad & L = q + (\Delta q - m_i t_{ij})(m_i t_{ij} + \theta_{ij}) \frac{1}{a} - w_{ij} - M(m_i) \\ & + \lambda_{LL} w_{ij} + \lambda_{PC} (w_{ij} + (m_i t_{ij} + \theta_{ij})^2 / (2a) - \bar{u}_j) , \end{aligned}$$

where λ_{LL} and λ_{PC} are the respective Lagrange multipliers of the limited liability and the participation constraint. and the corresponding first-order conditions are

$$\frac{\partial L}{\partial w_{ij}} = -1 + \lambda_{LL} + \lambda_{PC} \leq 0 , \quad (\text{A.1})$$

$$\frac{\partial L}{\partial t_{ij}} = \frac{m_i}{a} [\Delta q - 2m_i t_{ij} - \theta_{ij} + \lambda_{PC} (m_i t_{ij} + \theta_{ij})] \leq 0 , \quad (\text{A.2})$$

$$\frac{\partial L}{\partial m_i} = \frac{t_{ij}}{a} [\Delta q - 2m_i t_{ij} - \theta_{ij} + \lambda_{PC} (m_i t_{ij} + \theta_{ij})] - M'(m_i) \leq 0 \quad (\text{A.3})$$

$$\frac{\partial L}{\partial \lambda_{LL}} = w_{ij} \geq 0 , \quad (\text{A.4})$$

$$\frac{\partial L}{\partial \lambda_{PC}} = w_{ij} + (m_i t_{ij} + \theta_{ij})^2 / (2a) - \bar{u}_j \geq 0 , \quad (\text{A.5})$$

$$0 = \lambda_{LL} w_{ij} , \quad (\text{A.6})$$

$$0 = \lambda_{PC} (w_{ij} + (m_i t_{ij} + \theta_{ij})^2 / (2a) - \bar{u}_j) , \quad (\text{A.7})$$

From (A.1) follows immediately that at least one of the two constraints has to be binding, i.e., it is not possible that $\lambda_{LL} = \lambda_{PC} = 0$. Indeed, if both $\lambda_{LL} = \lambda_{PC} = 0$, (A.1) implies that the profit of the principal could be increased by reducing w_{ij} to its minimum level $w = 0$, a contradiction with $\lambda_{LL} = 0$.

Furthermore, if (A.2) is binding, then (A.3) cannot be, unless $m_i = t_{ij} = 0$. The first-order condition with respect to m is always smaller or equal to zero, (i.e., $\frac{\partial L}{\partial m_i} \leq 0$) so that the principal wants to set m as low as possible. We deduce that $m_i^* = \underline{m}$ if extrinsic incentives for effort are needed and $m_i^* = 0$ if no such incentives are needed.

We then get three cases:¹

Case I: (LL) binding, (PC) not binding: If the (LL) constraint is binding then $\lambda_{LL} > 0$ and $w_{ij} = 0$. If the (PC) is not binding then $\lambda_{PC} = 0$. By Assumption 3, namely that $\Delta q^2 \geq 4aM(\underline{m})$, the principal always wants to induce some effort from the worker. Extrinsic incentives are necessary only if θ_{ij} is small. To be more specific, from (A.2) it follows that $m_i t_{ij} = \max\{0, (\Delta q - \theta_{ij})/2\}$ is optimal.

The principal's payoff then is

$$\pi_{ij}^I = q + \begin{cases} \frac{1}{a}\Delta q\theta_{ij} & \text{if } \Delta q < \theta_{ij} \\ \frac{1}{a}\left(\frac{\Delta q + \theta_{ij}}{2}\right)^2 - M(\underline{m}) & \text{if } \Delta q \geq \theta_{ij} \end{cases},$$

and the agent's payoff is

$$u_{ij} = \frac{1}{2a} \begin{cases} \theta_{ij}^2 & \text{if } \Delta q < \theta_{ij} \\ (\Delta q + \theta_{ij})^2/4 & \text{if } \Delta q \geq \theta_{ij} \end{cases}.$$

In the limit, if the agent's reservation utility is equal to this payoff, his reservation utility becomes binding. This is true if $\bar{u}_j = \underline{v}_j$ where

$$\underline{v}_j \equiv \frac{1}{2a} \left(\max\{0, (\Delta q - \theta_{ij})/2\} + \theta_{ij} \right)^2.$$

This means that Case I is only relevant when the agent's reservation utility is $\bar{u}_j \in [0, \underline{v}_j]$.

Case II: (LL) binding, (PC) binding: If the (LL) constraint is binding ($\lambda_{LL} > 0$), then $w_{ij} = 0$. If the (PC) is also binding ($\lambda_{PC} > 0$), then from (A.5) follows that $m_i t_{ij} = \sqrt{2a\bar{u}_j} - \theta_{ij}$ is optimal. For this to be a solution, it

¹For clarity, we use case indices I to III in the following.

is necessary that $m_i t_{ij} \geq 0$ which is equivalent to $\bar{u}_j \geq \theta_{ij}^2 / (2a)$. The agent's payoff is by construction $u_{ij} = \bar{u}_j$. The principal's payoff is

$$\pi_{ij}^{II} = q + \frac{1}{a} \left(\Delta q + \theta_{ij} - \sqrt{2a\bar{u}_j} \right) \sqrt{2a\bar{u}_j} - M(\underline{m}) .$$

It is easy to check that $\pi_{ij}^I = \pi_{ij}^{II}$ if $\bar{u}_j = \underline{v}_{ij}$.

Case III: (LL) not binding, (PC) binding: If the (LL) constraint is not binding ($\lambda_{LL} = 0$), then $w_{ij} > 0$. This implies in (A.1) an interior solution so that $\lambda_{PC} = 1$. We deduce that if $m_i = \underline{m} > 0$, by (A.2), we get $m_i t_{ij} = \Delta q$. Plugging that into the participation constraint which is binding we get $w_{ij} = \bar{u}_j - (\Delta q + \theta_{ij})^2 / (2a)$.

Note that for this it has to hold that $\bar{u}_j - (\Delta q + \theta_{ij})^2 / (2a) > 0$. That is, Case III is only relevant for agents with a reservation utility above

$$\tilde{v}_{ij} \equiv \frac{1}{2a} (\Delta q + \theta_{ij})^2 .$$

The principal's payoff then is

$$\pi_{ij}^{III} = q - \left[\bar{u}_j - \frac{1}{2a} (\Delta q + \theta_{ij})^2 \right] - M(\underline{m}) ,$$

which, under the assumption that $\Delta q^2 \geq 4aM(\underline{m})$, is higher than the profit achieved without monitoring (i.e., without extrinsic incentives $\pi_{ij} = q - [\bar{u}_j - \frac{1}{2a} (\theta_{ij})^2]$). The agent's payoff is by construction $u_{ij} = \bar{u}_j$.

The principal's payoff from Case III becomes negative if the agent's outside utility exceeds

$$\bar{v}_{ij} \equiv \frac{1}{2a} (\Delta q + \theta_{ij})^2 + q - M(\underline{m}) .$$

Finally comparing π_{ij}^{II} with π_{ij}^{III} it is easy to check that $\pi_{ij}^{II} = \pi_{ij}^{III}$ iff $\bar{u}_j = \tilde{v}_{ij}$. The principal prefers Case III over Case II whenever the agent's outside utility exceeds \tilde{v}_{ij} .

That is, Case III is relevant when the agent's reservation utility is $\bar{u}_j \in [\tilde{v}_{ij}, \bar{v}_{ij}]$, Case II is relevant when the agent's reservation utility is $\bar{u}_j \in [\underline{v}_{ij}, \tilde{v}_{ij}]$, and Case I is relevant when the agent's reservation utility is $\bar{u}_j \in [0, \underline{v}_{ij}]$.

To finish the proof, we have to make sure that the principal's payoff from each scenario is positive. For this, $q - M(\underline{m}) > 0$ is a sufficient assumption. It also ensures that $\underline{v}_{ij} \leq \tilde{v}_{ij} \leq \bar{v}_{ij}$. QED

A.2 Implications of Proposition 1 for the For-Profit Sector

Since $\theta_{Fj} = 0$ we can deduce the following corollary from Proposition 1 for the for-profit sector:

Corollary A.1 : *Depending on the size of the agent's reservation utility, the optimal contract in F takes the following form:*

- **Case I**: For $\bar{u} \in [0, \underline{v}_F]$, $w_F^* = 0$, $m_F^* = \underline{m}$, $t_F^* = \Delta q / (2\underline{m})$;
- **Case II**: For $\bar{u} \in (\underline{v}_F, \tilde{v}_F)$, $w_F^* = 0$, $m_F^* = \underline{m}$, $t_F^* = \sqrt{2a\bar{u}} / \underline{m}$;
- **Case III**: For $\bar{u} \in [\tilde{v}_F, \bar{v}_F]$, $w_F^* = \bar{u} - \Delta q^2 / (2a)$, $m_F^* = \underline{m}$, $t_F^* = \Delta q / \underline{m}$,

where $\underline{v}_F = \Delta q^2 / (8a)$, $\tilde{v}_F = \Delta q^2 / (2a)$, and $\bar{v}_F = \Delta q^2 / (2a) + q - M(\underline{m})$.

As a consequence, the utility of a worker, no matter whether good or regular, in sector F in Case I is $u_F = \Delta q^2 / (8a)$. In Cases II and III it is equal to \bar{u} .²

The principal's payoff is

$$\pi_F = q - M(\underline{m}) + \begin{cases} \Delta q^2 / (4a) & \text{in Case I} \\ (\Delta q - \sqrt{2a\bar{u}})\sqrt{2a\bar{u}} / a & \text{in Case II} \\ \Delta q^2 / (2a) - \bar{u} & \text{in Case III} \end{cases} .$$

A.3 Implications of Proposition 1 for the Non-Profit Sector

Let us define $\theta_{Ng} := \theta_g$. For $\theta_g \geq 0$ we get the following corollary from Proposition 1 for the non-profit sector:

Corollary A.2 : *Depending on the size of the agent's reservation utility, the optimal contract in sector N is characterized as follows:*

- **Case I**: If $\bar{u} \in [0, \underline{v}_N]$, we get two subcases:
 - (a) If $\theta_g < \Delta q$, then $w_N^* = 0$, $m_N^* = \underline{m}$, and $t_N^* = (\Delta q - \theta_g) / (2\underline{m})$.
 - (b) If $\theta_g \geq \Delta q$, then $w_N^* = 0$, $m_N^* = 0$, and $t_N^* = 0$.

²By Assumption 3, $\tilde{v}_F = \Delta q^2 / (2a) < \bar{v}_F = \Delta q^2 / (2a) + q - M(\underline{m})$.

- **Case II:** If $\bar{u} \in (\underline{v}_N, \tilde{v}_N)$, then $w_N^* = 0$, $t_N^* = \frac{1}{\underline{m}}(\sqrt{2a\bar{u}} - \theta_g)$ and $m_N^* = \underline{m}$.
- **Case III:** If $\bar{u} \in [\tilde{v}_N, \bar{v}_N]$, then $w_N^* = \bar{u} - (\Delta q + \theta_g)^2/(2a)$, $m_N^* = \underline{m}$, and $t_N^* = \Delta q/\underline{m}$.

Furthermore note that $\underline{v}_N = (\max\{0, (\Delta q - \theta_g)/2\} + \theta_g)^2/(2a)$, $\tilde{v}_N = (\Delta q + \theta_g)^2/(2a)$, and $\bar{v}_N = (\Delta q + \theta_g)^2 + q - M(\underline{m})/(2a)$.

The utility of a motivated agent in Cases II and III corresponds to his reservation utility \bar{u} , whereas in Case I he gets

$$u_{Ng} = \frac{1}{2a} \begin{cases} \theta_g^2 & \text{if } \Delta q < \theta_g \\ (\Delta q + \theta_g)^2/4 & \text{if } \Delta q \geq \theta_g \end{cases},$$

which is higher or equal to what he would get in sector F . Regular agents, on the other hand, do not derive any intrinsic satisfaction from working in the mission-oriented sector, but only care about monetary incentives. Since $w_N^* + m_N^*t_N^* \leq w_F^* + m_F^*t_F^*$ for any level of \bar{u} , the utility of a regular worker in N is always lower than in F .

The principal's profit in sector N hence is

$$\pi_N = q - M(\underline{m}) + \frac{1}{a} \begin{cases} (\Delta q + \theta_g)^2/4 & \text{if } \theta_g \leq \Delta q \\ \Delta q\theta_g + aM(\underline{m}) & \text{if } \theta_g > \Delta q \end{cases} \quad \text{in Case I} \\ \begin{cases} (\Delta q + \theta_g - \sqrt{2a\bar{u}})\sqrt{2a\bar{u}} & \text{in Case II} \\ (\Delta q + \theta_g)^2/2 - a\bar{u} & \text{in Case III} \end{cases}$$

A.4 Robustness of Results to a Different Effort Cost Function of Bad Workers

In our paper, we made the assumption that the effort costs of bad workers take the form $c(e, d) = a(e+d)^2/2$. As a result, bad workers choose to provide either a positive effort e or a destructive effort d .

Alternatively, we could also consider a cost function of the form $c(e, d) = a(e^2 + d^2)/2$ such that a bad worker with a level of negative motivation θ_b would choose his optimal effort as $e^* = m_i t_i/a$ and $d^* = (\theta_b - m_i K)/a$, i.e., he might provide both a constructive and a destructive effort at the same time.

The gist of our results still applies for this cost function. To keep the comparison short, let us look at the case where $\bar{u} = 0$. Without bad workers,

the optimal contracts are as described in Case I in the two previous sections. Given these contracts, and given the optimal effort choice of bad workers, their expected utility in sector $i = F, N$ hence corresponds to

$$\begin{aligned} u_{Fb} &= \frac{\Delta q^2}{8a} + \frac{(\theta_b - \underline{m}K)^2}{2a} \\ u_{Nb} &= \begin{cases} \frac{(\Delta q - \theta_g)^2}{8a} + \frac{(\theta_b - \underline{m}K)^2}{2a} & \text{if } \theta_g < \Delta q \\ \theta_b^2 / (2a) & \text{if } \theta_g \geq \Delta q \end{cases} . \end{aligned}$$

From this we can deduct the level of automatic deterrence in each sector (i.e., without any change in contract): In Sector F , $d^* = 0$ if $\theta_b \leq \underline{m}K$. If $\theta_g < \Delta q$, there are no bad workers in N since their payoff from good behavior is higher in F and their payoff from bad behavior is the same. If $\theta_g \geq \Delta q$, then bad workers with $\theta_b \leq \Delta q^2 / (8\underline{m}K) + \underline{m}K/2$ prefer sector F , whereas if they have a higher negative motivation, they prefer sector N where they provide *only* a destructive effort.

Full Deterrence:

To achieve full deterrence of bad behavior, the principal in the for-profit sector F maximizes the following objective function:³

$$\max_{w_F, t_F, m_F} \pi_F = q + (\Delta q - m_F t_F) \frac{m_F t_F}{a} - w_F - M(m_F) ,$$

subject to the limited liability and participation constraints as given in (3) and (4), plus the deterrence constraint

$$(DET) \quad \theta_b \leq m_F K .$$

From this we can see immediately, that in contrast to the model outlined in the paper, we now can use only the monitoring level to prevent destructive behavior. The bonus payment t_F has no impact on deterrence. Hence $m_F^{det} = \theta_b / K$ and $t_F^{det} = m_F^* t_F^* / m_F^{det}$. That is, while monitoring goes up, the overall incentives for good behavior can stay the same.

In the nonprofit sector N , the principal has to set m_N^{det} such that $u_{Nb}^{det} \leq u_{Fb}^{det}$. Like in the for-profit sector, the extrinsic incentives for good behavior can stay the same, i.e. $t_N^{det} = m_N^* t_N^* / m_N^{det}$. Since the expected bonus for good behavior in N is already lower than in F , full deterrence in N can be achieved with a slightly lower monitoring level. More specifically, for $\bar{u} = 0$ and $\theta_g < \Delta q$,

³The objective function is the same as in the main part of the paper, only the deterrence constraint (DET) is different. In the paper, it is $m_F t_F \geq \theta_b - m_F K$.

$u_{Nb}^{det} \leq u_{Nb}^{det}$ if $m_N^{det} = m_F^{det} - \sqrt{(\theta_b - m_F^{det}K)^2 + \theta_g(2\Delta q - \theta_g)/4}/K$. For $\bar{u} = 0$ and $\theta_g \geq \Delta q$, $u_{Nb}^{det} \leq u_{Nb}^{det}$ if $m_N^{det} = m_F^{det} - \sqrt{(\theta_b - m_F^{det}K)^2 + 4\Delta q}/K$.

Partial Deterrence:

The objective function of the principal under partial deterrence stays very similar to the one in the paper. With the new cost function, the principal in the for-profit sector now maximizes⁴

$$\max_{w_F, t_F, m_F} \pi_F = q + (\Delta q - m_F t_F) \frac{m_F t_F}{a} - w_F - M(m_F) - \beta_F D \frac{\theta_b - m_F K}{a},$$

and the principal in sector N maximizes⁵

$$\begin{aligned} \max_{w_N, t_N, m_N} \pi_N = & q + (\Delta q - m_N t_N) \frac{1}{a} (m_N t_N + (1 - \beta_N) \theta_g) - w_N \\ & - M(m_N) - \beta_N D \frac{1}{a} (\theta_b - m_N K). \end{aligned}$$

Since bad workers react like regular workers to the bonus payment, there is no distinction between the two in sector F concerning the provision of positive effort, whereas in sector N , bad workers provide a lower positive effort than good workers due to the lower extrinsic incentives. The level of negative effort of bad workers in both sectors depends only on the prevailing monitoring level, as before.

The maximization problem in F is essentially the same as before, in particular the optimal monitoring level with partial deterrence m_F^{part} is such that $M'(m_F^{part}) = \beta_F DK/a$.⁶ The same is true for sector N . Also, N can even achieve full deterrence of bad behavior by setting a monitoring level slightly lower than in F . To do so, N has to choose m_N such that $u_{Nb}(m_N^{part}) \leq u_{Fb}(m_F^{part})$. For $\bar{u} = 0$ this corresponds to a monitoring level

$$m_N^{part} = \frac{\theta_b}{K} - \frac{1}{K} \begin{cases} \sqrt{(\theta_b - m_F^{part}K)^2 + \theta_g(2\Delta q - \theta_g)/4} & \text{if } \theta_g < \Delta q \\ \sqrt{(\theta_b - m_F^{part}K)^2 + 4\Delta q^2} & \text{if } \theta_g \geq \Delta q \end{cases}.$$

Finally, the result that destructive behavior may be tolerated along the equilibrium path still holds. To see this, consider the case where $\bar{u} = 0$ and $\theta_g > \Delta q$. If the nonprofit sector does not change its benchmark contracts

⁴For comparison: in the paper, the profit function in F is $\pi_F = q + (1 - \beta_F)(\Delta q - m_F t_F) \frac{m_F t_F}{a} - w_F - M(m_F) - \beta_F D \frac{\theta_b - m_F K}{a}$.

⁵For comparison: in the paper, the profit function in N is $\pi_N = q + (1 - \beta_N)(\Delta q - m_N t_N) \frac{m_N t_N + \theta_g}{a} - w_N - M(m_N) - \beta_N D \frac{\theta_b - m_N K}{a}$.

⁶See proof of Proposition C.3 in Part C of the Appendix for details.

($w_N^* = m_N^* = t_N^* = 0$), i.e., if the existence of bad workers is just ignored, then the profit is

$$\pi_N^* = q + \Delta q \frac{1}{a} (1 - \beta_N) \theta_g - \beta_N D \frac{1}{a} \theta_b .$$

If N changes its contracts such that there is full deterrence ($w_N = 0, m_N = m_N^{part}, t_N = 0$), then its profit is given by

$$\pi_N^{part} = q + \Delta q \frac{1}{a} \theta_g - M(m_N^{part}) .$$

The former option is more attractive if the cost of additional monitoring is higher than the damage done by bad workers, i.e., if

$$M(m_N^{part}) > \frac{\beta_N}{a} (\Delta q \theta_g + D \theta_b) ,$$

or if $(aM(m_N^{part})/\beta_N - D\theta_b)/\Delta q > \theta_g > \Delta q$. That is, it may be cheaper for the nonprofit sector to accept occasional destructive behavior if the expected damage is sufficiently low.

A.5 Calculating Automatic Deterrence in N

In order to calculate the level of automatic deterrence in N , $\tilde{\theta}_{bN}$, we have to insert the relevant contracts both in sector N and F into (8). This is equivalent to comparing the utility of a bad worker from effort e in F with his utility from effort d in N .

Let us first consider the case where $\theta_g \geq \Delta q$. Depending on the level of reservation utility of the agents, Figure A.2 indicates which of the cases derived in Corollaries A.1 and A.2 is relevant in each sector and summarizes the resulting utility levels $u_{Nb}(d)$ and $u_{Fb}(e)$ that can be achieved by bad workers. We then have to compare each possible combination of utility levels in order to determine the relevant level of automatic deterrence. For instance, Case Ib in sector N overlaps with Cases I, II and III in sector F . If we insert the relevant values for m_N, t_N, w_N as well as m_F, t_F, w_F into (8), we find that $\tilde{\theta}_{bN} = \Delta q/2$ if $\bar{u} < \underline{v}_F$ and $\tilde{\theta}_{bN} = \sqrt{2a\bar{u}}$ if $\underline{v}_F < \bar{u} < \underline{v}_N$.

Similar comparisons have to be made for the remainder of cases, as well as for a setting where $\theta_g < \Delta q$, which is illustrated in Figure A.1.

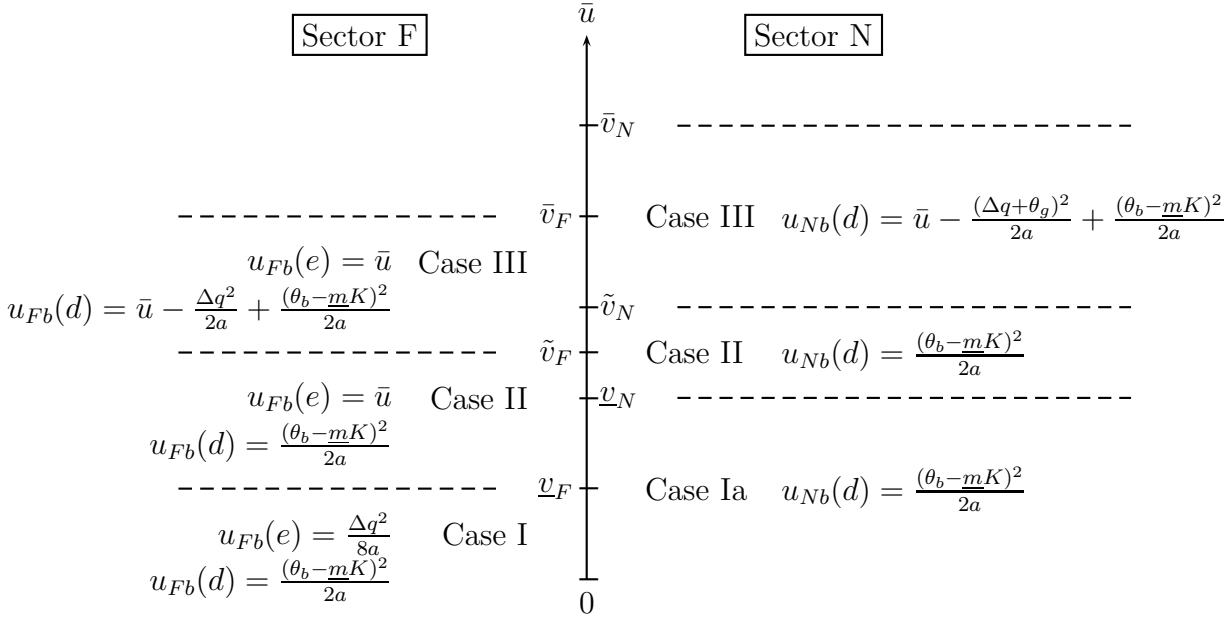


Figure A.1: Bad workers' utility from constructive and destructive effort in F and destructive effort in N for $\theta_g < \Delta q$.

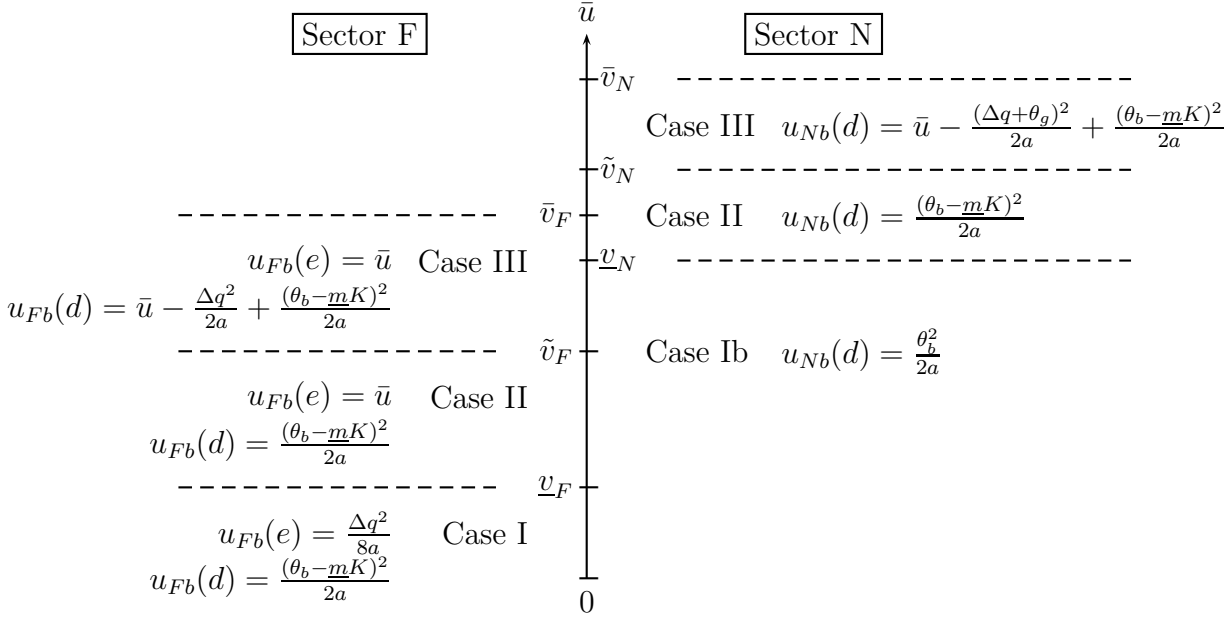


Figure A.2: Bad workers' utility from constructive and destructive effort in F and destructive effort in N for $\theta_g \geq \Delta q$.

B Full Deterrence Contracts

B.1 Full Deterrence in F

Proposition B.1 : For a given $\theta_b > \tilde{\theta}_{bF}$ and $\bar{u} \in [0, \bar{v}_F^{det}]$, the optimal contract with full deterrence $(m_F^{det}, t_F^{det}, w_F^{det})$ in sector F has the following features:

- (a) The optimal fixed wage is $w_F^{det} = \max\{0, \bar{u} - \frac{1}{2a}(\theta_b - m_F^{det}K)^2\}$,
- (b) The optimal bonus payment is $t_F^{det} = \theta_b/m_F^{det} - K$.
- (c) The optimal monitoring level is $m_F^{det} = \min\{\max\{\underline{m}, \tilde{m}_F^{det}\}, 1\}$, where \tilde{m}_F^{det} is such that the following conditions hold:

$$\begin{aligned} 2\tilde{m}_F^{det}K + M'(\tilde{m}_F^{det})a/K &= 2\theta_b - \Delta q & \text{if } \bar{u} \in [0, \underline{v}_F^{det}], \\ \tilde{m}_F^{det} &= \frac{1}{K}(\theta_b - \sqrt{2a\bar{u}}) & \text{if } \bar{u} \in (\underline{v}_F^{det}, \tilde{v}_F^{det}), \\ \tilde{m}_F^{det}K + M'(\tilde{m}_F^{det})a/K &= \theta_b - \Delta q & \text{if } \bar{u} \in [\tilde{v}_F^{det}, \bar{v}_F^{det}]. \end{aligned}$$

Proof. To solve the principal's maximization problem with full deterrence of bad workers we can formulate the following Lagrangian:

$$\begin{aligned} \max_{w_F, m_F, \lambda_{LL}, \lambda_{PC}} \quad & L(w_F, m_F, \lambda_{LL}, \lambda_{PC}) \\ &= q + (\Delta q - \theta_b + m_F K) \cdot \frac{\theta_b - m_F K}{a} - w_F - M(m_F) \\ &+ \lambda_{LL} w_F + \lambda_{PC} \left(w_F + \frac{(\theta_b - m_F K)^2}{2a} - \bar{u}_j \right), \end{aligned}$$

The corresponding first-order conditions are

$$\frac{\partial L}{\partial w_F} = -1 + \lambda_{LL} + \lambda_{PC} = 0, \quad (\text{B.8})$$

$$\frac{\partial L}{\partial m_F} = \frac{K}{a}[2\theta_b - 2m_F K - \Delta q] - M'(m_F) - \lambda_{PC} \frac{K}{a}(\theta_b - m_F K) \quad (\text{B.9})$$

Furthermore it has to hold that

$$0 = \lambda_{LL} w_F \quad (\text{B.10})$$

$$0 = \lambda_{PC} (w_F + (\theta_b - m_F K)^2 / (2a) - \bar{u}_j). \quad (\text{B.11})$$

As before, we get three cases:⁷

Case I: (LL) binding, (PC) not binding: From condition (B.10) and $\lambda_{LL} > 0$ follows that the optimal basic wage in Case I $w_F^I = 0$. From condition (B.9) and $\lambda_{PC} = 0$ it follows that the optimal monitoring level \tilde{m}_F^I has to be such that

$$2\theta_b - \Delta q = \frac{a}{K}M'(\tilde{m}_F^I) + 2\tilde{m}_F^I K .$$

Case II: (LL) and (PC) binding: If both conditions are binding, then $\lambda_{LL} > 0$ and $\lambda_{PC} > 0$. Again, by condition (B.10) we therefore have that the optimal wage in Case II $w_F^{II} = 0$. Furthermore, condition (B.11) is fulfilled iff

$$\tilde{m}_F^{II} = \frac{\theta_b - \sqrt{2a\bar{u}}}{K} .$$

Case III: (LL) not binding, (PC) binding: Since $\lambda_{LL} = 0$, by (B.8) we get $\lambda_{PC} = 1$. Inserting this in (B.9), we get that the monitoring level in Case III \tilde{m}_F^{III} has to be such that the following holds:

$$\theta_b - \Delta q = \frac{a}{K}M'(\tilde{m}_F^{III}) + \tilde{m}_F^{III} K .$$

Furthermore, since $\lambda_{PC} = 1$, the corresponding optimal basic wage is $w_F^{III} = \bar{u} - (\theta_b - \tilde{m}_F^{III}K)^2/(2a)$.

Note that depending on the functional form of $M(m)$, the optimal monitoring levels calculated above may not be contained in $[\underline{m}, 1]$. Since this is excluded by assumption, the optimal monitoring level is $m_F^{det} = \min\{\max\{\underline{m}, \tilde{m}_F\}, 1\}$, where \tilde{m}_F is such that the following conditions are fulfilled:

$$\begin{aligned} 2\tilde{m}_F K + M'(\tilde{m}_F)a/K &= 2\theta_b - \Delta q && \text{in Case I ,} \\ \tilde{m}_F &= (\theta_b - \sqrt{2a\bar{u}})/K && \text{in Case II ,} \\ \tilde{m}_F K + M'(\tilde{m}_F)a/K &= \theta_b - \Delta q && \text{in Case III .} \end{aligned}$$

The appropriate monitoring level has to be plugged into every expression containing m . Therefore the basic wage in Case III has to be rewritten as $w_F^{III} = \bar{u} - (\theta_b - m_F^{detIII}K)^2/(2a)$, where $m_F^{detIII} = \min\{\max\{\underline{m}, \tilde{m}_F^{III}\}, 1\}$.

We now can determine the transfer payment that rewards positive effort: Since the deterrence constraint $m_F t_F = \theta_b - m_F K$ is binding, the optimal

⁷For the sake of shortness, the index ‘‘det’’ is omitted unless needed for clarity. Instead, a case index is added.

transfer level t_F^{det} is always calculated as $t_F^{det} = \theta_b/m_F^{det} - K$, where again the appropriate value of m_F^{det} has to be plugged in.

To determine when each of the Cases I to III is relevant, we have to calculate the critical values \underline{v}_F^{det} , \tilde{v}_F^{det} , and \bar{v}_F^{det} of the agent's outside utility delimiting the above three cases.

Case I holds until the (PC) becomes binding. That is, it holds if the level of outside utility is $\bar{u} \in [0, \underline{v}_F^{det}]$, where \underline{v}_F^{det} is defined as

$$\underline{v}_F^{det} \equiv (\theta_b - m_F^{detI} K)^2 / (2a) ,$$

and $m_F^{detI} = \min\{\max\{\underline{m}, \tilde{m}_F^I\}, 1\}$ as defined above. Recall that $\underline{v}_F = (m_F^* t_F^*)^2 / (2a)$ and that $(\theta_b - m_F^{detI} K) > m_F^* t_F^*$ since $\theta_b > \tilde{\theta}_{bF}$. Therefore $\underline{v}_F^{det} > \underline{v}_F$.

Next, let us consider \tilde{v}_F^{det} , which defines the border between Cases II and III. Case III is only relevant if $\bar{u}_j - (\theta_b - m_F^{III} K)^2 / (2a) > 0$, where $m_F^{III} = \min\{\max\{\underline{m}, m_F^{III}\}, 1\}$. That is, Case III is only relevant for agents with a reservation utility above

$$\tilde{v}_F^{det} \equiv \frac{1}{2a} (\theta_b - m_F^{III} K)^2 .$$

Note that the limited liability constraint is trivially fulfilled if $\bar{u} > \tilde{v}_F^{det}$. Again, since we consider only cases where $\theta_b > \tilde{\theta}_{bF}$ and hence $(\theta_b - m_F^{III} K) > m_F^* t_F^*$, we get that $\tilde{v}_F^{det} > \tilde{v}_F$.

Finally, \bar{v}_F^{det} is defined as the outside utility of the agent for which the principal's profit becomes zero. This may happen in either of the three cases, depending on the relative size of θ_b . In Case III, $\pi_F^{III} = 0$ if:

$$\begin{aligned} \bar{v}_F^{det} &\equiv q + \frac{1}{2a} (\theta_b - m_F^{III} K)^2 - M(m_F^{III}) \\ &+ (\Delta q - \theta_b + m_F^{III} K)(\theta_b - m_F^{III} K) \frac{1}{a} , \end{aligned}$$

where $m_F^{III} = \min\{\max\{\underline{m}, \tilde{m}_F^{III}\}, 1\}$ and \tilde{m}_F^{III} is such that $\theta_b - \Delta q = a/KM'(m_F) + m_F K$ as derived above.

In Cases I and II, profits become zero or negative if

$$q + (\Delta q - \theta_b + m_F^{det} K)(\theta_b - m_F^{det} K) / a - M(m_F^{det}) \geq 0 . \quad (\text{B.12})$$

If this last condition holds, then also $\bar{v}_F^{det} < \tilde{v}_F^{det}$ such that Case III is no longer relevant. Instead, profits become zero for any outside utility below \tilde{v}_F^{det} if θ_b is sufficiently high to fulfill condition (B.12).

The derivative of \bar{v}_F^{det} with respect to θ_b is smaller than zero if $(\Delta q - \theta_b + m_F^{III}K)M''(m_F^{III}) \leq aM'(m_F^{III})$. Since $M'(m) > 0$ and $M''(m) > 0$ and since we consider only cases where $\theta_b > \tilde{\theta}_{bF}$ it holds that $(\theta_b - m_F^{III}K) > m_F^*t_F^* = \Delta q$. Hence the expression in brackets is negative and the above inequality is fulfilled. We therefore know that \bar{v}_F^{det} is decreasing in θ_b .

How high is \bar{v}_F^{det} relative to \bar{v}_F ? Recall that $\bar{v}_F = q + \frac{1}{2a}\Delta q^2 - M(\underline{m})$. Hence $\bar{v}_F^{det} < \bar{v}_F$ if

$$\begin{aligned} \frac{1}{2a}[(\theta_b - m_F^{III}K)^2 - \Delta q^2] - M(m_F^{III}) + M(\underline{m}) \\ + (\Delta q - \theta_b + m_F^{III}K)(\theta_b - m_F^{III}K)\frac{1}{a} < 0. \end{aligned}$$

This is equivalent to:

$$-(\theta_b - m_F^{III}K - \Delta q)^2 - 2a(M(m_F^{III}) - M(\underline{m})) < 0.$$

Since $m_F^{III} \geq \underline{m}$ and $M(\cdot)$ is an increasing function of m , the left-hand side of this inequality is negative, and we hence have shown that $\bar{v}_F^{det} < \bar{v}_F$. QED

B.2 Proof of Proposition 3: Comparison with Benchmark Contracts

How does the monitoring level and the expected bonus under full deterrence change compared to the benchmark contracts?

Lemma B.1 *As θ_b increases, the principal first keeps monitoring at its minimal level and rather increases the bonus payment, before he also starts to monitor more. The only exception is Case II in Proposition B.1, where $m_F^{det} > \underline{m}$ for any $\theta_b > \tilde{\theta}_{bF}$.*

Proof. Consider the value of θ_b for which $\tilde{m}_F^{det} > \underline{m}$ in each case of Proposition B.1, i.e., $\tilde{m}_F^{det} > \underline{m}$ if

$$\theta_b > \begin{cases} \Delta q/2 + \frac{\underline{m}K + M'(\tilde{m}_F^I)a/K}{\sqrt{2a\bar{u}}} & \text{in Case I ,} \\ \frac{\underline{m}K + \sqrt{2a\bar{u}}}{\Delta q + \frac{\underline{m}K + M'(\tilde{m}_F^{III})a/K}{\sqrt{2a\bar{u}}}} & \text{in Case II ,} \\ \Delta q + \frac{\underline{m}K + M'(\tilde{m}_F^{III})a/K}{\sqrt{2a\bar{u}}} & \text{in Case III .} \end{cases}$$

Comparing this to the definition of $\tilde{\theta}_{bF}$ in each case as given in (7), we can easily see that the critical level of θ_b for which F starts to increase his monitoring level beyond \underline{m} is above $\tilde{\theta}_{bF}$, except in Case II, where F starts to increase monitoring right away as soon as $\theta_b > \tilde{\theta}_{bF}$. QED.

Lemma B.2 For $\theta_b > \tilde{\theta}_{bF}$ and any $\bar{u} \geq 0$, the expected bonus under full deterrence is $m_F^{det} t_F^{det} \geq m_F^* t_F^*$.

Proof Suppose $m_F^{det} = m_F^* = \underline{m}$. Then from (7) and Proposition B.1 follows that $t_F^{det} > t_F^*$ and therefore, $m_F^{det} t_F^{det} > m_F^* t_F^*$.

Suppose $m_F^{det} > \underline{m}$. If we solve the conditions for \tilde{m}_F^{det} given in Proposition B.1 for $\tilde{m}_F^{det} K$ and plug them into the deterrence constraint, we get

$$m_F^{det} t_F^{det} = \theta_b - m_F^{det} K = \begin{cases} \Delta q/2 + M'(m_F^{det})a/K & \text{in Case I ,} \\ \sqrt{2a\bar{u}} & \text{in Case II ,} \\ \Delta q + M'(m_F^{det})a/K & \text{in Case III .} \end{cases} \quad (\text{B.13})$$

Recall that

$$m_F^* t_F^* = \begin{cases} \Delta q/2 & \text{in Case I ,} \\ \sqrt{2a\bar{u}} & \text{in Case II ,} \\ \Delta q & \text{in Case III .} \end{cases} \quad (\text{B.14})$$

Comparing (B.13) and (B.14), we can see right away that $m_F^{det} t_F^{det} > m_F^* t_F^*$ in Cases I and III since $M'(m_F^{det})a/K > 0$. If Case II applies both with full deterrence and in the benchmark scenario, then $m_F^{det} t_F^{det} = m_F^* t_F^*$. Hence, overall $m_F^{det} t_F^{det} \geq m_F^* t_F^*$. QED

An illustrative example for these results can be found in Section D of the Appendix.

B.3 Full Deterrence in N

Proposition B.2 : If $\theta_b > \tilde{\theta}_{bN}$ then, for a given contract (m_F, t_F, w_F) in sector F and a reservation payoff $\bar{u} \in [0, \bar{v}_N^{det}]$, the principal in sector N can achieve full deterrence of bad workers by offering a contract $(m_N^{det}, t_N^{det}, w_N^{det})$ with the following features:

- (a) The fixed wage is $w_N^{det} = w_N^*$ with w_N^* as defined in Proposition 1.
- (b) The monitoring level is $m_N^{det} = \min\{\max\{\underline{m}, \tilde{m}_N^{det}\}, 1\}$, with

$$\tilde{m}_N^{det} = \left(\theta_b - \sqrt{2a(u^{max} - w_N^*)} \right) / K ,$$

where $u^{max} := \max\{u_F(e), u_F(d), \bar{u}\}$.

(c) The bonus payment is

$$t_N^{det} = \begin{cases} t_F - \theta_g/m_F, & \text{if } \bar{u} < \underline{v}_F \text{ and } m_F t_F > m_N^* t_N^* + \theta_g \\ m_N^* t_N^*/m_N^{det} & \text{otherwise} \end{cases},$$

with $m_N^* t_N^*$ as defined in Proposition 1.

Proof. To achieve full deterrence, N has to make sure that $u_{Nb}(d) \leq u^{max} := \max\{u_F(e), u_F(d), \bar{u}\}$, that is, $w_N^{det} + (\theta_b - m_N^{det} K)^2/(2a) \leq u^{max}$. For any given outside utility \bar{u} , the basic wage in N is smaller or equal than the basic wage in F . This acts as a deterring element to bad, but not to good workers. Taking this into account, there is no reason to change w_N compared to the benchmark contracts, i.e., $w_N^{det} = w_N^*$. The above condition therefore can be rewritten as $m_N^{det} \geq (\theta_b - \sqrt{2a(u^{max} - w_N^*)})/K$, which gives us a general formula for the optimal level of monitoring. More precisely, the optimal monitoring level will be given by

$$m_N^{det} = \begin{cases} m_F & \text{if } \bar{u} < \tilde{v}_F \\ \theta_b/K - \sqrt{2a(w_F - w_N^*) + (\theta_b - m_F K)^2}/K & \text{if } \tilde{v}_F \leq \bar{u} < \bar{v}_F \\ \theta_b/K - \sqrt{2a(\bar{u} - w_N^*)}/K & \text{if } \bar{u} \geq \bar{v}_F \end{cases}$$

That is, N will set the same monitoring level as in F when the basic wage is the same in both sectors or a just slightly lower level when the basic wage in F is higher than in N . When F is no longer active, N has to set monitoring such that bad workers prefer not to work.

Since F may increase incentives for good behavior above their optimal level as part of its deterrence strategy, we have to check whether good workers in that case will still prefer to work in N . For $\bar{u} \geq \underline{v}_F$, it is easy to show that $u_{Ng}(e) \geq u_F(e) = \bar{u}$ and hence $m_N^{det} t_N^{det} = m_N^* t_N^*$ is a sufficient incentive. That is, as m_N^* increases, t_N^* goes down such that the overall incentives for good workers stay the same.

However, for $\bar{u} < \underline{v}_F$, $u_F(e)$ may be greater than in the benchmark case and hence $m_N^{det} t_N^{det} = m_F^{det} t_F^{det} - \theta_g$ has to hold. Since in this range of values $m_N^{det} = m_F^{det}$, the optimal bonus in N is $t_N^{det} = t_F^{det} - \theta_g/m_F^{det}$. Yet, this case is only relevant if θ_g is small enough, i.e., if $\theta_g < m_F^{det} t_F^{det} - m_N^* t_N^*$. If the level of intrinsic motivation is higher, then good workers will prefer N even with the benchmark contracts.

B.4 Equilibrium with Full Deterrence

To determine the equilibrium outcome, we have to check where the bad guys are, given the benchmark contracts. We do this by comparing the utility

they can achieve from choosing (F, e) , (F, d) , or (N, d) .⁸

For a low level of positive intrinsic motivation ($\theta_g < \Delta q$), we can distinguish the following cases, depending on the level of outside utility:

(i) $\bar{u} < \tilde{v}_F$: For low levels of negative intrinsic motivation, there is automatic deterrence in both sectors. However, for $\theta_b > \tilde{\theta}_{bN} = \tilde{\theta}_{bF}$, $u_{Nb}(d) = u_{Fb}(d) > u_{Fb}(e) > \bar{u}$. With the benchmark contracts, bad workers are thus indifferent between N and F , but will behave badly in any case. To deter such behavior, F will have to adopt the contracts described in Proposition B.1. Given F 's choice of contract, N can deter bad workers by raising m_N as described in Proposition B.2. In equilibrium, all bad workers are in F and behave like regular types.

(ii) $\tilde{v}_F < \bar{u} < \bar{v}_F$: In this range of values, the level of automatic deterrence in N is higher than in F . For $\theta_b \leq \tilde{\theta}_{bF}$, there is automatic deterrence in both sectors, meaning that bad workers will prefer F , but will be indistinguishable from regular workers. For $\tilde{\theta}_{bF} < \theta_b \leq \tilde{\theta}_{bN}$, their utility from choosing (F, d) will be higher than from (F, e) or (N, d) under the benchmark contracts, meaning that F will have to switch to the full deterrence contract outlined in Proposition B.1, whereas N only has to adjust its contracts if $\theta_b > \tilde{\theta}_{bN}$.

(iii) $\bar{v}_F < \bar{u}$: For $\theta_b \leq \tilde{\theta}_{bN}$, $u_{Nb}(d) \leq \bar{u}$ holds and bad workers hence prefer to enjoy their outside utility rather than work in N . If $\theta_b > \tilde{\theta}_{bN}$, then N will have to raise m_N as described in Proposition B.2 to achieve the same effect.

For a high level of positive intrinsic motivation ($\theta_g \geq \Delta q$), we can distinguish the following cases:⁹

(i) $\bar{u} < \underline{v}_N$: For very low levels of negative intrinsic motivation ($\theta_b \leq \tilde{\theta}_{bN}$), there is automatic deterrence in both sector N and F . However, as θ_b goes up, the complete lack of monitoring in sector N will make working in N the most attractive option of bad workers. To get rid of the bad guys, it is sufficient for N to introduce monitoring and set it to the same level as in sector F . This latter sector can count on automatic deterrence of bad workers up to $\theta_b \leq \tilde{\theta}_{bF}$. If the negative motivation of bad workers is higher than that, F can achieve full deterrence by introducing the contracts described in Proposition B.1. In an equilibrium with full deterrence in both sectors, bad workers will always work in F and behave like regular workers.

(ii) $\underline{v}_N < \bar{u} < \bar{v}_F$: Due to the lower basic wage, automatic deterrence is higher in N than in F . As a consequence, for $\theta_b \leq \tilde{\theta}_{bF}$ bad workers choose (F, e) whereas for $\tilde{\theta}_{bF} < \theta_b \leq \tilde{\theta}_{bN}$, they prefer (F, d) . To achieve full deterrence, F will have to introduce the contract described in Proposition B.1 for any $\theta_b > \tilde{\theta}_{bF}$. N can keep its benchmark contract as long as $\theta_b \leq \tilde{\theta}_{bN}$, but will

⁸See Figures A.1 and A.2. By virtue of Proposition 2 these are the only relevant options.

⁹See Figure A.2 for an overview of the utility of bad workers in both sectors.

have to raise m_N for higher levels of θ_b such that $u_{bN}(d) \leq \bar{u}$, i.e., by setting $m_N = m_N^{det}$ as defined in Proposition B.2. In equilibrium, all the bad guys will behave regularly in sector F .

(iii) $\bar{v}_F < \bar{u}$: For a very high reservation utility, sector N is the only possible employment option for bad workers. If $\theta_b < \hat{\theta}_{bN}$ bad workers are automatically deterred, otherwise N can achieve full deterrence by setting m_N such that $u_{bN}(d) \leq \bar{u}$, i.e., by setting $m_N = m_N^{det}$ as defined in Proposition B.2. As a result, bad workers prefer not to work at all in equilibrium.

C Partial Deterrence Contracts

C.1 Optimal Contracts with Partial Deterrence

Proposition C.3 : For $\theta_b > \hat{\theta}_{bi}$, the optimal contract with partial deterrence and strictly positive monitoring in sector $i = N, F$ given a reservation payoff $\bar{u} \in [0, \bar{v}_i^{part}]$ has the following features:

- (a) The optimal fixed wage is $w_i^{part} = \max \left\{ 0, \bar{u} - \frac{1}{2a} \left(\frac{1-\beta_i}{1-2\beta_i} (\Delta q + \theta_{ig}) \right)^2 \right\}$,
- (b) The monitoring level is $m_i^{part} = \min \{ \max \{ \underline{m}, \tilde{m}_i^{part} \}, 1 \}$, where \tilde{m}_i^{part} is such that $M'(\tilde{m}_i^{part}) = \beta_i DK/a$.
- (c) The optimal bonus payment is

$$t_i^{part} = \begin{cases} (\Delta q - \theta_{ig}) / (2m_i^{part}) & \text{if } \bar{u} \in [0, \underline{v}_i^{part}] \\ (\sqrt{2a\bar{u}} - \theta_{ig}) / m_i^{part} & \text{if } \bar{u} \in (\underline{v}_i^{part}, \tilde{v}_i^{part}) \\ (\Delta q - \beta_i(\Delta q - \theta_{ig})) / ((1 - 2\beta_i)m_i^{part}) & \text{if } \bar{u} \in [\tilde{v}_i^{part}, \bar{v}_i^{part}] \end{cases} .$$

where

$$\begin{aligned} \underline{v}_i^{part} &= \underline{v}_i^* = (m_i^{part} t_i^{part} + \theta_{ig})^2 / (2a) \\ \tilde{v}_i^{part} &= \frac{1}{2a} \left(\frac{1 - \beta_i}{1 - 2\beta_i} (\Delta q + \theta_{ig}) \right)^2 \\ \bar{v}_i^{part} &\text{ s.t. } \Pi_i^{part} = 0. \end{aligned}$$

Proof. To solve the principal's maximization problem in sector $i = \{N, F\}$ with only partial deterrence, we can formulate the following Lagrangian:

$$\begin{aligned} \max_{w_i, m_i, t_i, \lambda_{LL}, \lambda_{PC}} \quad & L(w_i, m_i, t_i, \lambda_{LL}, \lambda_{PC}) = q - w_i - M(m_i) \\ & + (1 - \beta_i)(\Delta q - m_i t_i) \frac{(m_i t_i + \theta_{ig})}{a} - \beta_i D(\theta_b - m_i K) \frac{1}{a} \\ & + \lambda_{LL} w_i + \lambda_{PC} (w_i + (m_i t_i + \theta_{ig})^2 / (2a) - \bar{u}_j) , \end{aligned}$$

and the corresponding first-order conditions are

$$\frac{\partial L}{\partial w_i} = -1 + \lambda_{LL} + \lambda_{PC} = 0, \quad (\text{C.15})$$

$$\frac{\partial L}{\partial t_i} = \frac{m_i}{a} [(1 - \beta_i)(\Delta q - 2m_i t_i - \theta_{ig}) + \lambda_{PC}(m_i t_i + \theta_{ig})] = 0, \quad (\text{C.16})$$

$$\begin{aligned} \frac{\partial L}{\partial m_i} &= \frac{t_i}{a} [(1 - \beta_i)(\Delta q - 2m_i t_i - \theta_{ig}) + \lambda_{PC}(m_i t_i + \theta_{ig})] \\ &\quad - M'(m_i) + \frac{\beta_i DK}{a} = 0. \end{aligned} \quad (\text{C.17})$$

Furthermore, the following has to be true:

$$0 = \lambda_{LL} w_i, \quad (\text{C.18})$$

$$0 = \lambda_{PC} (w_i + (m_i t_i + \theta_{ig})^2 / (2a) - \bar{u}_j). \quad (\text{C.19})$$

Equation (C.16), i.e., the FOC with respect to t_i , is fulfilled if the expression in square brackets is equal to zero. This implies that (C.17), the FOC with respect to m_i , simplifies to $-M'(m_i) + \beta_i DK/a = 0$, and hence the optimal level of monitoring without full deterrence of bad workers is such that $M'(m_i^{part}) = \beta_i DK/a$. However, we have to make sure that $m_i^{part} \in \{0, [\underline{m}, 1]\}$, i.e., that we get an interior solution. This problem is the same for all possible parameter values discussed below and we therefore can write immediately that the optimal monitoring level with partial deterrence is

$$\begin{aligned} m_i^{part} &= \min\{\max\{\underline{m}, \tilde{m}_i^{part}\}, 1\}, \\ \text{where } \tilde{m}_i^{part} &\text{ s.t. } M'(\tilde{m}_i^{part}) = \beta_i DK/a. \end{aligned} \quad (\text{C.20})$$

As before, we get three possible cases:

Case I: (LL) binding, but not (PC): In this case, $\lambda_{LL} > 0$, and therefore, by (C.18), $w_i^{part} = 0$. Furthermore, $\lambda_{PC} = 0$, such that condition (C.16), i.e., the FOC with respect to t_i is fulfilled if $m_i^{part} t_i^{part} = (\Delta q - \theta_{ig})/2 = m_i^* t_i^*$. The optimal contract in Case I therefore is given by $w_i^{part} = 0$, m_i^{part} as defined in (C.20), and $t_i^{part} = m_i^* t_i^* / m_i^{part} = (\Delta q - \theta_{ig}) / (2m_i^{part})$. The resulting profit in Case I is:

$$\pi_i^{part} = (1 - \beta_i) \frac{(\Delta q + \theta_{ig})^2}{4a} - \frac{\beta_i D}{a} (\theta_b - m_i^{part} K) + q - M(m_i^{part}).$$

Case I is valid as long as the participation constraint is not binding, which is true for

$$\bar{u} < w_i^{part} + (m_i^* t_i^* + \theta_{ig})^2 / (2a) = 0 + (\Delta q + \theta_{ig})^2 / (2a) = \underline{v}_i^{part} = \underline{v}_i^*.$$

The boundary for Case I is thus the same as in the benchmark case without bad workers.

Case II: (LL) and (PC) binding: Since $\lambda_{LL} > 0$, by (C.18), $w_i^{part} = 0$. Furthermore, since $\lambda_{PC} > 0$, from (C.19) follows that $m_i t_i = \sqrt{2a\bar{u}} - \theta_{ig}$ which is exactly the same condition as in Case II without bad workers. Hence $m_i^{part} t_i^{part} = m_i^* t_i^*$ also in Case II.

The same reasoning as outlined above then can be made. The optimal contract in Case II is $w_i^{part} = 0$, m_i^{part} as defined in (C.20) and $t_i^{part} = m_i^* t_i^* / m_i^{part} = (\sqrt{2a\bar{u}} - \theta_{ig}) / m_i^{part}$. The resulting profit in Case II is

$$\pi_i^{part} = (1 - \beta_i)(\Delta q - \sqrt{2a\bar{u}} + \theta_{ig}) \frac{\sqrt{2a\bar{u}}}{a} - \frac{\beta_i D}{a} (\theta_b - m_i^{part} K) + q - M(m_i^{part}).$$

Case III: (PC) binding, but not (LL): Since $\lambda_{LL} = 0$, by (C.18), $w_i^{part} > 0$. Furthermore, $\lambda_{PC} > 0$ and hence $w_i^{part} = \bar{u} - (m_i t_i + \theta_{ig})^2 / (2a)$ must hold.

By condition (C.15), $\lambda_{LL} = 0$ also implies that $\lambda_{PC} = 1$. Condition (C.16), therefore is fulfilled if $m_i t_i = [(1 - \beta_i)\Delta q + \beta_i \theta_{ig}] / (1 - 2\beta_i)$. Note that this last expression is positive if $\beta_i < 0.5$ and smaller or equal zero if $\beta_i \geq 0.5$. The latter would, however, imply a destructive effort incentive for workers which does not make much sense.

From this, we get as optimal contract in Case III $w_i^{part} = \bar{u}[(\Delta q + \theta_{ig})(1 - \beta_i) / (1 - 2\beta_i)]^2$, m_i^{part} as defined in (C.20) and $t_i^{part} = [\Delta q - \beta_i(\Delta q - \theta_{ig})] / [(1 - 2\beta_i)m_i^{part}]$. The profit in Case III hence is given by

$$\pi_i^{part} = q - \bar{u} + \frac{[(1 - \beta_i)(\Delta q + \theta_{ig})]^2}{(1 - 2\beta_i)2a} - \frac{\beta_i D}{a} (\theta_b - m_i^{part} K) - M(m_i^{part}).$$

Furthermore, from the above results we can deduce the frontier between Cases II and III: If we plug w_i^{part} into the limited liability constraint we find that Case III is only valid for an outside utility $\bar{u} > \tilde{v}_i^{part}$ where

$$\tilde{v}_i^{part} \equiv \frac{1}{2a} \left(\frac{1 - \beta_i}{1 - 2\beta_i} (\Delta q + \theta_{ig}) \right)^2.$$

Recall that in the benchmark case without bad workers $\tilde{v}_i = (\Delta q + \theta_{ig})^2 / (2a)$. Comparing these two values, we find that $\tilde{v}_i^{part} > \tilde{v}_i$ if $\beta_i < 0.5$.

Finally, we have to determine \bar{v}_i^{part} , i.e., when profits become negative. In contrast to the benchmark model, this may happen in any of the three subcases, provided that q is low and/ or θ_b is high enough. Let us define

$Z := q - \frac{\beta_i D}{a}(\theta_b - m_i^{part} K) - M(m_i^{part})$. Then, in Case I, profit becomes negative if

$$Z \leq \frac{1 - \beta_i}{a}(\Delta q + \theta_{ig})^2$$

for all $\bar{u} \in [0, \underline{v}_i]$. If Z is larger, profits become negative for

$$\bar{u} \geq \min\{\bar{v}_i^{partII}, \bar{v}_i^{partIII}\},$$

where

$$\bar{v}_i^{partII} \equiv \frac{1}{4a} \left[(\Delta q + \theta_{ig})^2 + \frac{2aZ}{1 - \beta_i} + \sqrt{(\Delta q + \theta_{ig})^4 + 4(\Delta q + \theta_{ig})^2 \frac{aZ}{1 - \beta_i}} \right]$$

is the value of outside utility for which profits in Case II become negative, and

$$\bar{v}_i^{partIII} \equiv Z + \frac{1 - 2\beta_i}{2a} \left(\frac{1 - \beta_i}{1 - 2\beta_i} (\Delta q + \theta_{ig}) \right)^2$$

is the value of outside utility for which profits in Case III become negative. The former is smaller than the latter if

$$Z \leq \frac{\beta_i}{a} \left(\frac{1 - \beta_i}{1 - 2\beta_i} (\Delta q + \theta_{ig}) \right)^2.$$

That is, if the last inequality is fulfilled, i.e., if q is small and/ or θ_b high enough, then Case III basically disappears, since profits in this case will be negative.

C.2 Equilibrium with Partial Deterrence

Proposition C.4 : *In an equilibrium with partial deterrence, the optimal contracts in N and F can be described as follows:*

(a) *If $[aM(m_N^{det})(1 + x_g/x_b) - D\theta_b]/\Delta q > \theta_g > \Delta q$ and $\bar{u} \in [0, \underline{v}_N]$, then N sticks to the benchmark contracts, i.e., $m_N^* = t_N^* = w_N^* = 0$. F also keeps its benchmark contracts if furthermore one of the following conditions holds:*

- (i) $\bar{u} \in [0, \tilde{v}_F^{part}]$, or
- (ii) $\bar{u} \in [\tilde{v}_F^{part}, \underline{v}_N]$ and $\tilde{\theta}_{bF} < \theta_b \leq [2a\bar{u} - \Delta q^2 + (m_F K)^2]/(2m_F K)$.

Otherwise, the optimal contract in F is given by Proposition C.3.

- (b) If either $\theta_g < \Delta q$ or $\theta_g \geq [aM(m_F^{part})(1 + x_g/x_b) - D\theta_b]/\Delta q$, and $\bar{u} \in [0, \bar{v}_F^{part}]$, the optimal contract in F is given by Proposition C.3. N sets its monitoring level $m_N^{part} \geq \underline{m}$ such that $u_{Nb}(d) \leq u_{Fb}(d)$ while keeping the incentives for positive effort as in the benchmark contracts: $m_N^{part} t_N^{part} = m_N^* t_N^*$ and $w_N^{part} = w_N^*$.
- (c) If $\bar{u} \in [\bar{v}_F^{part}, \bar{v}_N^{part}]$, then only N is active and the optimal contract in N is given by Proposition C.3.

Proof. If $\bar{u} \in [\bar{v}_F^{part}, \bar{v}_N]$, then only sector N is active. To at least partially deter bad workers, the principal in sector N hence has to adapt the contracts according to Proposition C.3.

Next, let us consider all cases, where both sectors offer contracts and the optimal monitoring in sector N in the benchmark case is $m_N^* = \underline{m}$, i.e., if $\theta_g < \Delta q$ and $\bar{u} \in [0, \bar{v}_F^{part}]$ or if $\theta_g \geq \Delta q$ and $\bar{u} \in [\underline{v}_N, \bar{v}_F^{part}]$. Given the benchmark contracts and $\theta_b > \tilde{\theta}_{bF}$, all bad workers derive the highest possibility from sabotage in sector F such that $\beta_F = \bar{\beta}_F$. F will therefore have to adapt his contracts according to Proposition C.3. In that case, if $m_F^{part} > \underline{m}$, all bad workers would switch to N such that $\beta_N = \bar{\beta}_N$. Since by Assumption ?? $\bar{\beta}_N > \bar{\beta}_F$, the resulting optimal monitoring level in N according to Proposition C.3 would be equal or higher than in F . However, N can achieve full deterrence of bad workers by adopting a monitoring level $m_N \leq m_F^{part}$ such that $u_N(d) \leq u_F(d)$.

Now we are only left with the case where $\theta_g \geq \Delta q$ and $\bar{u} \in [0, \underline{v}_N]$. The benchmark contract in N implies zero monitoring and hence all bad guys choose (N, d) for $\theta_b > \tilde{\theta}_{bN}$. The principal in N then faces two options:

- (a) Stick with the benchmark contracts, even if he thus attracts all bad guys, which yields profit

$$\Pi_N(w_N = 0, m_N = t_N = 0) = q + (1 - \bar{\beta}_N)\Delta q\theta_g/a - \bar{\beta}_N D\theta_b/a .$$

- (b) Introduce monitoring. If he chooses this option he can even achieve full deterrence by setting $m_N = m_N^{det}$, where $m_N^{det} = m_F^{part}$ for $\bar{u} \in [0, \bar{v}_F^{part}]$, whereas for $\bar{u} \in [\bar{v}_F^{part}, \underline{v}_N]$, m_N^{det} such that $u_{Nb}(d) \leq u_{Fb}(d)$. That is, the profit in N becomes

$$\Pi_N(w_N = 0, m_N = m_N^{det}, t_N = 0) = q + \Delta q\theta_g/a - M(m_N^{det}) .$$

If damage D is low enough, the former option may yield a higher profit. Then, N will prefer to stick to its first best contracts even if it thus attracts all bad

workers. That is, he prefers option (a) if $[aM(m_N^{det})(1 + x_g/x_b) - D\theta_b]/\Delta q > \theta_g \geq \Delta q$. This corresponds to case (a) in Proposition C.4.

Given these considerations for sector N , how should F react? Obviously, if there is no monitoring in N , all bad workers will choose that sector and hence F can keep its benchmark contracts. This is true if $[aM(m_N^{det})(1 + x_g/x_b) - D\theta_b]/\Delta q > \theta_g$ and if either (i) $\bar{u} \in [0, \tilde{v}_F^{part}]$ or (ii) $\bar{u} \in [\tilde{v}_F^{part}, \underline{v}_N]$ and $\theta_{bF} < \theta_b \leq [2a\bar{u} - \Delta q^2 + (m_F K)^2]/(2m_F K)$. If $\bar{u} \in [\tilde{v}_F^{part}, \underline{v}_N]$ and $\theta_b > [2a\bar{u} - \Delta q^2 + (m_F K)^2]/(2m_F K)$ then the basic wage in F is sufficiently high to attract all bad workers even if there is zero monitoring in N . Therefore, F will have to adapt its contracts according to Proposition C.3.

If, on the other hand, option (b) yields a higher profit for N than option (a), i.e., if $[aM(m_N^{det})(1 + x_g/x_b) - D\theta_b]/\Delta q < \theta_g$, then N will want to introduce monitoring. In equilibrium, F will then change its contracts according to Proposition C.3 and N will set the monitoring level $m_N \geq \underline{m}$ such that $u_{Nb}(d) \leq u_{Fb}(d)$. For $\bar{u} \in [0, \tilde{v}_F^{part}]$ this means setting $m_N = m_F^{part}$, for $\bar{u} \in [\tilde{v}_F^{part}, \underline{v}_N]$ m_N can be slightly lower than m_F^{part} due to the higher basic wage in F .

D Example

In the following, let us consider an example where the monitoring function is $M(m) = m^2/2$, which may give the interested reader a better intuition of the results in Section 4. Note that the results of the benchmark case without bad workers do not change substantially and therefore can be directly gathered from Section 3.

Full Deterrence in the For-Profit Sector

As before, we can distinguish three cases:

Case I: For $\theta_b < \tilde{\theta}_{bF}^I := \Delta q/2 + \underline{m}K$, there is automatic deterrence of bad workers. The optimal contract then corresponds to the contract outlined in Corollary A.1, i.e., the basic wage is 0, monitoring is at its minimum level \underline{m} , and the bonus payment is given by $\Delta q/(2\underline{m})$.

If $\theta_b > \tilde{\theta}_{bF}^I$, we have to calculate the optimal level of monitoring according to Proposition B.1 which gives us

$$\tilde{m}_F^{det} = (2\theta_b - \Delta q)K/(a + 2K^2) .$$

However, this term is smaller than \underline{m} if $\theta_b < \tilde{\theta}_b^I + \underline{am}/(2K)$ and larger than 1 if $\theta_b > \Delta q/2 + K + a/(2K)$. Hence, for $\theta_b > \theta_{bF}^I$, the optimal contract with full deterrence is given by

$$w_F^{det} = 0, \quad t_F^{det} = \theta_b/m_F^{det} - K,$$

$$m_F^{det} = \begin{cases} \underline{m} & \text{if } \theta_b \in (\tilde{\theta}_{bF}^I, \tilde{\theta}_{bF}^I + \underline{am}/(2K)) \\ (2\theta_b - \Delta q)K/(a + 2K^2) & \text{if } \theta_b \in [\tilde{\theta}_{bF}^I + \underline{am}/(2K), \\ \Delta q/2 + K + a/(2K)] \\ 1 & \text{if } \theta_b > \Delta q/2 + K + a/(2K) \end{cases}.$$

That is, even when bad workers are not automatically deterred (i.e., when $\theta_b > \tilde{\theta}_{bF}^I$), the monitoring level stays low. It is cheaper to just increase the bonus payment. However, as θ_b increases further, the principal in F also has to increase monitoring, although this comes at a higher cost than just raising the bonus. For very high levels of θ_b , we may get a corner solution such that m hits its maximum level $m = 1$. In that case, the principal's only option to deter bad workers is to pay a higher bonus for good behavior.

Case II: In Case II, bad workers are automatically deterred from bad actions if $\theta_b < \tilde{\theta}_{bF}^{II} := \sqrt{2a\bar{u}} + \underline{m}K$, and the optimal contract remains unchanged compared to the case without bad workers. That is, the basic wage is 0, the monitoring level is \underline{m} , and the bonus payment is given by $\sqrt{2a\bar{u}}/\underline{m}$.

If $\theta_b > \tilde{\theta}_{bF}^{II}$, $\tilde{m}_F^{det} = (\theta_b - \sqrt{2a\bar{u}})/K$, which is smaller than \underline{m} for $\theta_b < \tilde{\theta}_{bF}^{II}$ and larger than 1 if $\theta_b > K + \sqrt{2a\bar{u}}$. Therefore the optimal contract with full deterrence is given by

$$w_F^{det} = 0, \quad t_F^{det} = \theta_b/m_F^{det} - K,$$

$$m_F^{det} = \begin{cases} \underline{m} & \text{if } \theta_b < \tilde{\theta}_{bF}^{II} \\ (\theta_b - \sqrt{2a\bar{u}})/K & \text{if } \theta_b \in [\tilde{\theta}_{bF}^{II}, K + \sqrt{2a\bar{u}}] \\ 1 & \text{if } \theta_b > K + \sqrt{2a\bar{u}} \end{cases}.$$

Case III: Finally, in Case III, bad workers are automatically deterred from bad actions if $\theta_b \leq \tilde{\theta}_{bF}^{III} := \Delta q + \underline{m}K$. That is, the optimal contract is as if there were no bad workers with a basic wage equal to $\bar{u} - \Delta q^2/(2a)$, monitoring at its minimal level \underline{m} , and the bonus payment equal to $\Delta q/\underline{m}$.

If $\theta_b > \tilde{\theta}_{bF}^{III}$, then according to Proposition B.1, monitoring is calculated as

$$\tilde{m}_F^{det} = (\theta_b - \Delta q)K/(a + 2K^2).$$

However, this value is smaller than \underline{m} if $\theta_b < \tilde{\theta}_{bF}^I + \underline{am}/K$ and larger than one if $\theta_b > \Delta q + K + a/K$. Therefore, for $\theta_b > \tilde{\theta}_{bF}^{III}$, the optimal contract

with full deterrence is given by

$$w_F^{det} = \bar{u} - (\theta_b - m_F^{det} K)^2 / (2a), \quad t_F^{det} = \theta_b / m_F^{det} - K,$$

$$m_F^{det} = \begin{cases} \underline{m} & \text{if } \theta_b \in (\tilde{\theta}_{bF}^{III}, \tilde{\theta}_{bF}^I + a\underline{m}/K) \\ (\theta_b - \Delta q)K / (a + 2K^2) & \text{if } \theta_b \in [\tilde{\theta}_{bF}^{II} + a\underline{m}/K, \\ \quad \Delta q + K + a/K] \\ 1 & \text{if } \theta_b > \Delta q + K + a/K \end{cases}.$$

That is, as in Case I, the optimal monitoring level is constrained by the corner solutions \underline{m} and 1. In both of these cases, the principal has to increase the bonus payment for good behavior in order to achieve full deterrence. However, for intermediate values of θ_b it is better to increase the monitoring level rather than the bonus payment, even though monitoring is costly.

Which case is relevant when? Case I is valid if the agents' outside utility $\bar{u} < \underline{v}_F$, where

$$\underline{v}_F = \frac{1}{2a} \cdot \begin{cases} \Delta q^2 / 4 & \text{if } \theta_b \leq \tilde{\theta}_{bF}^I \\ (\theta_b - \underline{m}K)^2 & \text{if } \theta_b \in (\tilde{\theta}_{bF}^I, \tilde{\theta}_{bF}^I + a\underline{m}/(2K)) \\ ((a\theta_b + \Delta qK^2)/(a + 2K^2))^2 & \text{if } \theta_b \in [\tilde{\theta}_{bF}^I + a\underline{m}/(2K), \\ \quad K + a/(2K) + \Delta q/2] \\ (\theta_b - K)^2 & \text{if } \theta_b \geq K + a/(2K) + \Delta q/2 \end{cases}$$

where $\tilde{\theta}_{bF}^{III} := \Delta q/2 + \underline{m}K$.

Case II is valid if the agents' outside utility $\bar{u} \in (\underline{v}_F, \tilde{v}_F)$, where \underline{v}_F is defined above and

$$\tilde{v}_F = \frac{1}{2a} \cdot \begin{cases} \Delta q^2 & \text{if } \theta_b \leq \tilde{\theta}_{bF}^{II} \\ (\theta_b - \underline{m}K)^2 & \text{if } \theta_b \in (\tilde{\theta}_{bF}^{II}, \tilde{\theta}_{bF}^{II} + a\underline{m}/K) \\ (a\theta_b + \Delta qK^2)^2 / (a + K^2)^2 & \text{if } \theta_b \in (\tilde{\theta}_{bF}^{III} + a\underline{m}/K), \\ \quad K + a/K + \Delta q) \\ (\theta_b - K)^2 & \text{if } \theta_b \geq K + a/K + \Delta q \end{cases}$$

where $\tilde{\theta}_{bF}^{III} := \Delta q + \underline{m}K$.

Case III is valid if the agents' outside utility $\bar{u} > \tilde{v}_F$, where \tilde{v}_F is defined above.

Note that, depending on the exact parameter values, the profit of organization F may become zero in any of the three cases.

Figure D.3 illustrates the above considerations and shows, which case is relevant for which combinations of outside utility \bar{u} and negative intrinsic motivation θ_b .¹⁰

¹⁰The graph is based on the following parameter values: $q = 2, \Delta q = 2, a = 10, \underline{m} = 0.1, K = 3$. The monitoring function takes the form $M(m) = m^2/2$.

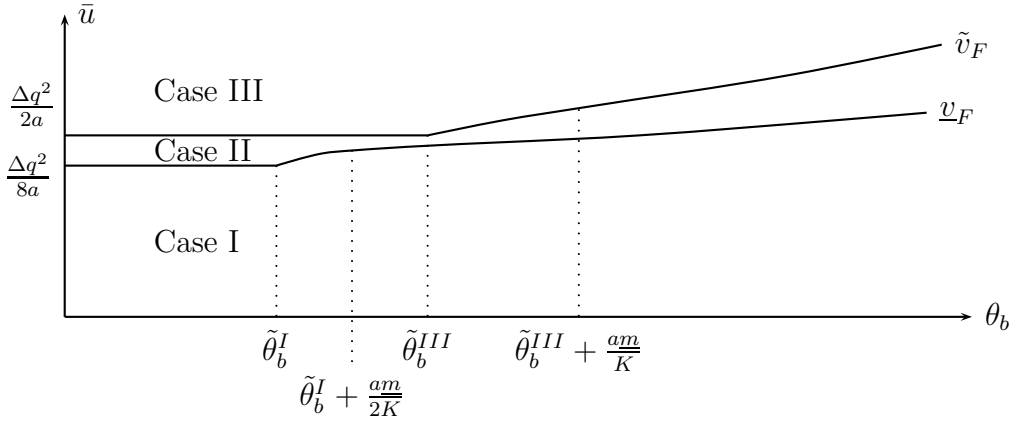


Figure D.3: Relevant cases in F depending on outside utility \bar{u} and negative intrinsic motivation θ_b .

Furthermore, Figure 6 shows for a given outside utility how the monitoring level and profits develop as θ_b increases. Figure 4 shows how the expected bonus develops.

Full Deterrence in the Non-Profit Sector

According to Proposition B.2, N can achieve full deterrence by setting \tilde{m}_N^{det} such that $u_{Nb}(d) \leq \max\{u_{Fb}(e), u_{Fb}(d), \bar{u}\}$.

Depending on which is the relevant comparison, \tilde{m}_N^{det} takes the following form:

- If $\max\{u_{Fb}(e), u_{Fb}(d), \bar{u}\} = u_{Fb}(e)$:

$$\begin{aligned} u_{Nb}(d) &\leq u_{Fb}(e) \\ w_N + (\theta_b - m_N K)^2 / (2a) &\leq w_F + (m_F t_F)^2 / (2a) \\ \Rightarrow \tilde{m}_N^{det} &= (\theta_b - \sqrt{2a(w_F - w_N) + (m_F t_F)^2}) / K \end{aligned}$$

- If $\max\{u_{Fb}(e), u_{Fb}(d), \bar{u}\} = u_{Fb}(d)$:

$$\begin{aligned} u_{Nb}(d) &\leq u_{Fb}(d) \\ w_N + (\theta_b - m_N K)^2 / (2a) &\leq w_F + (\theta_b - m_F K)^2 / (2a) \\ \Rightarrow \tilde{m}_N^{det} &= (\theta_b - \sqrt{2a(w_F - w_N) + (\theta_b - m_F K)^2}) / K \end{aligned}$$

- If $\max\{u_{Fb}(e), u_{Fb}(d), \bar{u}\} = \bar{u}$:

$$\begin{aligned} u_{Nb}(d) &\leq \bar{u} \\ w_N + (\theta_b - m_N K)^2 / (2a) &\leq \bar{u} \\ \Rightarrow \tilde{m}_N^{det} &= (\theta_b - \sqrt{2a(\bar{u} - w_N)}) / K \end{aligned}$$

Suppose that F does not adapt its initial contracts to the presence of bad workers, but sticks to (w_F^*, m_F^*, t_F^*) . Then depending on the level of negative motivation θ_b and outside utility \bar{u} , the following inequalities hold:

- (a) For $\theta_b \leq \tilde{\theta}_{bF}$ and $\bar{u} < \bar{v}_F$: $u_{Fb}(e) \geq \bar{u} \geq u_{Fb}(d)$.
- (b) For $\theta_b > \tilde{\theta}_{bF}$ and $\bar{u} < \bar{v}_F$: $u_{Fb}(d) > u_{Fb}(e) \geq \bar{u}$.
- (c) For $\bar{u} > \bar{v}_F$: only \bar{u} relevant, since F is no longer active.

That is, full deterrence can be achieved if $m_N^{det} = \min\{\max\{\underline{m}, \tilde{m}_N^{det}\}, 1\}$ where \tilde{m}_N^{det} is given by

$$\tilde{m}_N^{det} = \frac{\theta_b}{K} - \frac{1}{K} \cdot \begin{cases} \sqrt{2a(w_F - w_N) + (m_F t_F)^2} & \text{in (a)} \\ \sqrt{2a(w_F - w_N) + (\theta_b - m_F K)^2} & \text{in (b)} \\ \sqrt{2a(\bar{u} - w_N)} & \text{in (c)} \end{cases},$$

where the relevant values for (w_F, m_F, t_F) as well as for w_N , θ_b and \bar{u} have to be plugged in.