DISCUSSION PAPER SERIES

No. 9009

ANTI-SOCIAL BEHAVIOR IN PROFIT AND NONPROFIT ORGANIZATIONS

Emmanuelle Auriol and Stefanie Brilon

DEVELOPMENT ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP9009.asp

ISSN 0265-8003

ANTI-SOCIAL BEHAVIOR IN PROFIT AND NONPROFIT ORGANIZATIONS

Emmanuelle Auriol, Toulouse School of Economics and CEPR
Stefanie Brilon, Université de Lausanne

Discussion Paper No. 9009

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 9009

ABSTRACT

Anti-Social Behavior in Profit and Nonprofit Organizations*

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. While mission-oriented organizations take advantage of the intrinsic motivation of good workers, they are more vulnerable than profit-oriented organizations to anti-social behavior: bad workers only join them to behave badly. To prevent this, monitoring has to go up in the mission-oriented sector, while the incentives for good behavior stay the same. In the profit-oriented sector, by contrast, both monitoring and bonus payments for good behavior increase to control the damage caused by bad workers. As a result, in equilibrium bad workers are generally working in the for-profit sector where they behave like "normal" people, while good workers self select into the mission-oriented sector.

JEL Classification: D21, D23 and L31
Keywords: candidate selection, motivated agents, non-profit and sabotage

Emmanuelle Auriol
Toulouse School of Economics
21 Allées de Brienne
F31000 Toulouse
FRANCE

Email: eauriol@cict.fr

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=127618

Stefanie Brilon
Université de Lausanne

Email: stefanie.brilon@unil.ch

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=175640

*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 Massenot, Philipp Weinschenk for their comments, as well as seminar participants at Mannheim, Rotterdam, Bonn, Frankfurt, Oxford, Paris, Hannover and Lugano. Research funding by the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

Submitted 1 June 2012

1 Introduction

Intrinsic motivation is generally treated by economists as something beneficial to organizations. Most theoretical models on the subject suppose that intrinsic 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", a 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 mis-conduct, 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 the 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

[1]See 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).

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

[3]For 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.

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, just like regular workers. 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. They will blend in the population of "normal" people and will be indistinguishable from them.

We extend the model by Besley and Ghatak (2005) who consider only good and regular workers, whereas there are three types of workers in our model: good, regular, and bad workers. 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 then 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 bonuses than the profit-oriented sector.

We then introduce bad workers who derive utility from behaving in an anti-social way. 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: 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. The paper hence illuminates the crucial role monitoring plays in the fight against anti-social behaviors. Monitoring not only helps to keep track

[4] We 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).

of workers' good performance but it also helps to contain their destructive actions. In practice, monitoring is therefore central to organizations' design and management to achieve these two purposes, which also explains why so many resources are spent on it. One contribution of the paper is to disentangle the role of monitoring in providing monetary incentives, as postulated by standard incentive theory, where it is generally treated as a black box (e.g., fixed cost), and its role in deterring sabotage and anti-social behaviors.

Since they monitor more, profit-oriented organizations 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 incentives and control, the more likely they are to become the target of bad workers.

We then analyze how contracts have to change in both sectors in order to deter bad workers from their destructive behavior. We show that in an equilibrium with full deterrence, bad workers will work in the for-profit sector where they are indistinguishable from regular workers, while good workers self-select in the mission-oriented sector. To achieve this outcome, monitoring and bonus payments for good behavior in this sector will tend to increase, combining "the carrot and the stick". In the mission-oriented sector, on the other hand, the focus will be more on the stick, which implies that while monitoring has to go up here as well, the incentives for good behavior tend to stay the same.

Furthermore, we discuss the robustness of our model with respect to variations, first, in the damage caused by bad workers and, second, in the level of their motivation. If the potential damage caused by bad workers is sufficiently low, organizations may actually be willing to accept some sabotage in equilibrium and opt for only partial deterrence. 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 be found in the profit-oriented sector. The only exception occurs when 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.

Deterrence is costly as it implies higher monitoring, and it 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 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. Section 5 contains some robustness checks and discusses the ex ante selection of job candidates. 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. The John Jay report (see Terry, 2008) indicated that some 11,000 allegations of sexual abuse of children had been made against 4,392 priests in the USA. This number constituted approximately 4% of the 110,000 priests who had served during the 52-year period covered by the study (1950-2002). The report 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, 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 paedophile will preferably target vulnerable children, such as refugees[5] or orphans,[6] simply because they are less likely to expose him.

Other examples for anti-social behavior resulting from some form of intrinsic motivation are pyromania or sadism. Stambaugh and Styron (2003) show that pyromaniacs may best be able to satisfy their urge for fire by working

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

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

for the firefighters and provide evidence, mostly from the United States, that shows how serious the problem is.[7] Similarly, a sadist might try to work in prisons or detention centers, preferably protected by national security secrecy or by their geographical remoteness, to feed his need to humiliate and harm others.[8]

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 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 a lot of 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).[9]

Furthermore, our model is linked to the growing strand of literature on public service motivation[10] 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.[11] 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.[12] While the focus

[7]Similar 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.

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

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

[10]See Dixit (2002) for a review on incentives in the public sector.

[11]Note, 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.

[12]That this may indeed be the case has been shown by Heckman, Smith, and Taber

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.[13] There are hence many sectors where wages are not paid conditional on performance, as for instance the civil service sector or many nonprofit organizations.[14] 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 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 5.2 of our paper.

3 Basic Setup

The basic 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

(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.

[13]Agents 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.

[14]See 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.

bad workers, with shares $x_g + x_r + x_b = 1$ in the population. While the distribution of types is common knowledge, the type of an individual worker is not directly observable.

As a benchmark case, we first concentrate on good and regular workers only. In contrast to regular agents, good agents derive an intrinsic benefit $\theta_g > 0$ from working in the nonprofit sector N. In sector F, neither type of agent r or g derives a positive intrinsic benefit.

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. 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).

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 $\bar{u} \geq 0$. Given these constraints, the principals in both sectors try to maximize their profits over w_{ij}, t_{ij} and m_i as follows:

$$\pi_{ij} = q + (\Delta q - m_i t_{ij}) e_{ij} - w_{ij} - M(m_i) \ , \tag{1}$$

subject to the following constraints

$$(LL) \qquad w_{ij} \geq 0, \tag{2}$$

$$(PC) \qquad u_{ij} = w_{ij} + (m_i t_{ij} + \theta_{ij}) e_{ij} - a e_{ij}^2/2 \geq \bar{u} \ , \tag{3}$$

$$(IC) \qquad e_{ij} = \arg \max_{e \in [0,1]} \left\{ w_{ij} + (m_i t_{ij} + \theta_{ij}) e_{ij} - a e_{ij}^2/2 \right\} . \tag{4}$$

It follows immediately from the incentive constraint (4) 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 we assume that a is sufficiently large :

Assumption 1 $a > \Delta q + \theta_g$.

Under Assumption 1, we get an interior solution such that $e_{ij} = (m_i t_{ij} + \theta_{ij})/a < 1$. We can hence rewrite the maximization problem as

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

subject to

$$\begin{array}{ll} (LL) & w_{ij} \geq 0 \ , \\ (PC) & u_{ij} = (m_i t_{ij} + \theta_{ij})^2/(2a) + w_{ij} \geq \bar{u} \ . \end{array}$$

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

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

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 for $\bar{u} \geq \underline{v}_{ij}$ 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. That is,[15]

$$\underline{v}_{ij} \equiv \frac{1}{2a}\Big(\max\{0, (\Delta q - \theta_{ij})/2\} + \theta_{ij}\Big)^2 \tag{5}$$

$$\tilde{v}_{ij} \equiv \frac{1}{2a}(\Delta q + \theta_{ij})^2 \tag{6}$$

$$\bar{v}_{ij} \equiv \frac{1}{2a}(\Delta q + \theta_{ij})^2 + q - M(\underline{m}) \ . \tag{7}$$

It is straightforward to check that under Assumption 2: $\underline{v}_{ij} \leq \tilde{v}_{ij} \leq \bar{v}_{ij}$.

Then the following proposition characterizes the optimal contract:

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

[15] For more details on this, see the proof of Proposition 1 in Appendix A.

(a) *The optimal wage is*

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

(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*

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

All proofs can be found in Appendix A.

We can thus discern three cases:

- *Case I:* The limited liability constraint is binding, but not the participation constraint of the agent. This holds for low values of the reservation utility: $\bar{u} \in [0, \underline{v}_{ij}]$. The optimal contract in this case 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, in which case monitoring is no longer needed (Case Ib).

- *Case II:* Both the limited liability and the participation constraint are binding. This holds for intermediary values of the reservation utility: $\bar{u} \in (\underline{v}_{ij}, \tilde{v}_{ij})$. The optimal contract in this case 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:* The participation constraint is binding, but not the limited liability constraint. This corresponds to a case where $\bar{u}$ is relatively high, i.e., for $\bar{u} \in [\tilde{v}_{ij}, \bar{v}_{ij}]$. The optimal contract in this case 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.

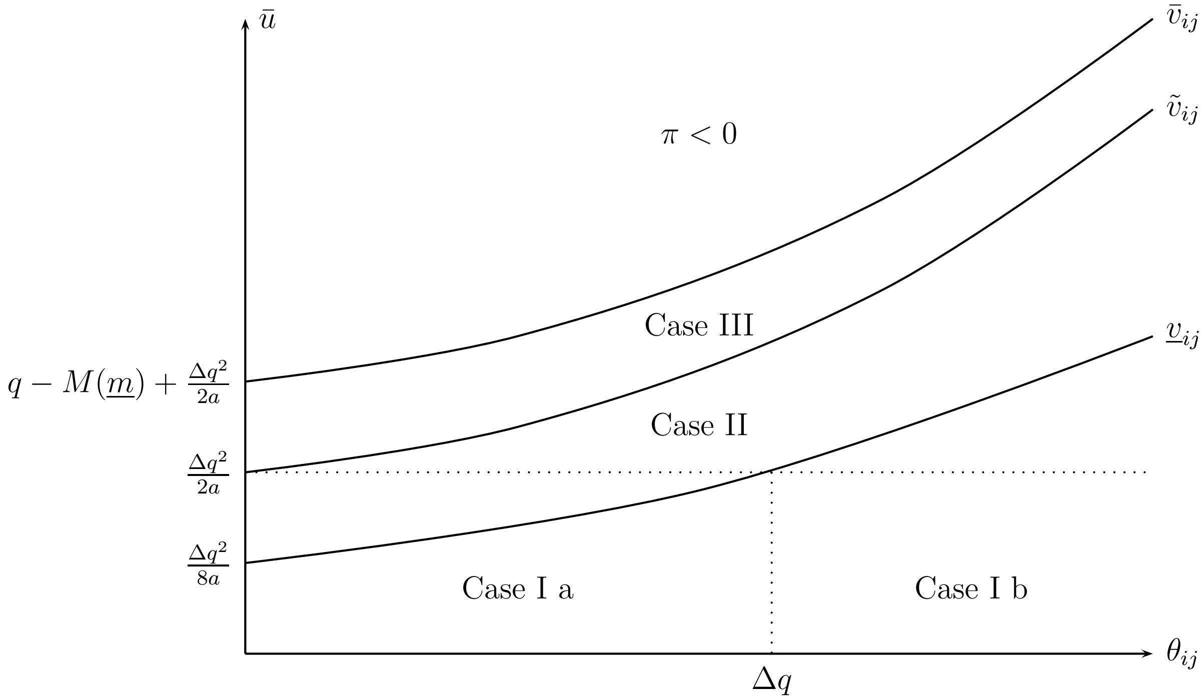


Figure 1: Optimal contract depending on $\bar{u}_j$ and θ_{ij}.

Figure 1 gives an overview of the result. 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 7.2 and 7.3 of the Appendix, we discuss in more detail how Proposition 1 translates into an optimal contract in the sector F and N respectively.

Which case is relevant for the principal in sector F depends 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: $m_F^* = \underline{m}$. The utility of a worker in sector F in Case I is $u_F = \Delta q^2/(8a)$. In Cases II and III it is equal to $\bar{u}$.[16] The principal's

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

payoff is

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

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_{ij} (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" motivated workers (i.e., with $\theta_{Ng} > 0$) 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). The utility of a motivated agent in Cases II and III corresponds to his reservation utility $\bar{u}$, whereas in Case I he gets $u_{Ng} = \frac{1}{2a}(\Delta q + \theta_g)^2/4$ if $\theta_g \leq \Delta q$ (Case Ia) and $u_{Ng} = \theta_g^2$ if $\theta_g > \Delta q$ (Case Ib).[17] The principal's profit in sector N is

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

Comparing π_N with π_F in each case, it is straightforward to see that $\pi_N > \pi_F$ if $\theta_g > 0$. In contrast to the profit-oriented sector, the mission-oriented sector N can save on wage costs by exploiting the intrinsic motivation of good workers who self-select themselves into the nonprofit sector. It gives sector N a competitive edge compared to sector F.[18]

[17]This is higher or equal to what he would get in sector F. "Good" agents with low reservation utility hence prefer the contract proposed in N to the contract offered in sector F. Low reservation utility typically corresponds to junior workers with no or little experience and thus relatively low outside opportunity. We thus expect young idealistic people to join the nonprofit sector. Empirically they should be over-represented compared to other workers.

[18]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.

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 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 θ_b from choosing such a negative effort, and by doing so they may cause a damage D to the organization they are working for. In order to get interesting results we assume that the damage D is sufficiently large so that the organization wants to fight it in equilibrium. In Section 5.1 we discuss the robustness of our results to the case where the damage is sufficiently low so that the organization might tolerate sabotage in equilibrium.

In contrast to good workers, bad workers have no intrinsic preference for one or the other sector. Consider the following utility function for bad workers in sector $i = F, N$:

$$u_{ib} = w_i + m_i t_i e + (\theta_b - K m_i) d - a(e + d)^2/2 \,, \tag{8}$$

where K is an exogenous punishment that can be imposed on a worker if a negative effort is observed. The idea behind this is that a negative 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.[19] However, as this is beyond the influence of the firm, we treat the punishment as exogenous.

The worker can choose to do some positive effort (i.e., do his/her job) and simultaneously to spend some negative effort by doing what he/she likes and is destructive for the organization.[20] However, because of the linear structure of the payoffs and the quadratic structure of the costs, at the optimum the worker chooses either one of the two options, i.e., he either decides to satisfy

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.

[19]The examples we have in mind are a paedophile working in a teaching institution, a refugee camp, or an orphanage; a terrorist working in an airport, a spy working in an intelligence service; a pyromaniac working in a fire fighter crew; a sadistic working in a detention center.

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

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.[21] Bad types therefore prefer to exert a positive effort rather than to follow their destructive impulse if and only if

$$m_i t_i \geq \theta_b - m_i K \ . \tag{9}$$

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 depending on the behavior in response to oversight. Our results are robust to the introduction of such separate monitoring functions.[22]

In order to get an interior solution, we have to amend Assumption 1 as follows:

Assumption 3 $a > \Delta q + \max\{\theta_g, \theta_b\}$.

Given this assumption, we can calculate the worker's optimal effort choice and his expected utility, which is

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

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

[21]Optimizing (8) with respect to e and d yields the first order conditions: $\frac{\partial u_{ib}}{\partial e} = m_i t_i - a(e+d)$ and $\frac{\partial u_{ib}}{\partial d} = -m_i K + \theta_b - a(e+d)$. It is generally impossible that the two conditions are simultaneously equal to 0. The problem is concave which implies that $e^*_{ib} = 0$ when $d^*_{ib} > 0$ (i.e., when $\theta_b - m_i K > m_i t$) or, symmetrically, $d^*_{ib} = 0$ when $e^*_{ib} > 0$ (i.e., when $\theta_b - m_i K \leq m_i t$).

[22]For 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.

analyze the behavior of bad workers given the optimal contracts derived in Section 3. Which sector will bad workers choose and how will they behave? When are the benchmark contracts described in Proposition 1 enough to "automatically" deter anti-social behavior?

4.1 Automatic Deterrence of Bad Workers

Since all bad workers face the same tradeoffs, sector $i = N, F$ attracts either all or none of the bad workers depending on the utility that they can achieve from the different options available.[23] We first compare a bad worker's payoff from choosing effort e or d in both sectors given the optimal contracts derived in Section 3. This comparison shows that for a given reservation utility $\bar{u}$ the incentives for choosing a positive effort e 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 or equal than that in sector F, thus making it less likely to get caught with bad actions in the nonprofit sector and therefore $u_{Nb}(d) \geq u_{Fb}(d)$. From this follows that under the optimal contracts of Proposition 1 bad workers only join N to do harm. More generally this result holds true under any contracts 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 intrinsic motivated workers of good type join and provide a positive effort, while regular workers choose either not to work or to work in the profit-oriented sector. Since the bad workers have exactly the same preference as the regular workers regarding the provision of a positive effort (i.e., they work for money), and since by assumption 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. The next lemma collects this result.

Lemma 1 *: Bad workers join the mission-oriented sector only to provide a negative effort d.*

This result is especially true under the optimal contracts of Proposition 1: Bad workers will only join N to follow their destructive impulse, while minimizing the risk of being detected and punished.

Next, let us look in more detail at what happens in each sector. It is clear from (9) that for low levels of negative motivation θ_b, bad workers are better off if they choose a positive rather than a destructive effort. In sector F, such

[23] By assumption, they all choose F if indifferent between the two sectors.

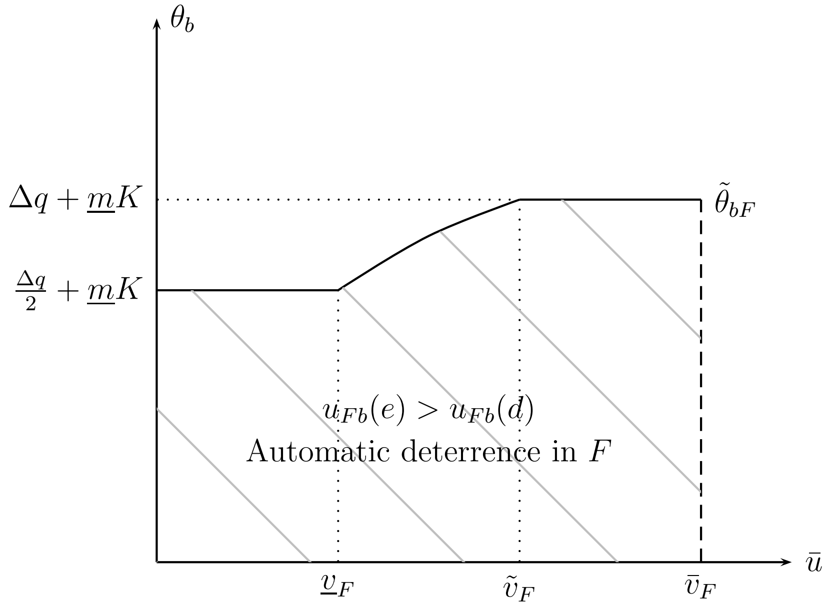


Figure 2: Automatic deterrence in sector F.

"automatic" deterrence of bad workers, i.e., deterrence without any change of the optimal contract as derived in Proposition 1, takes place if θ_b is smaller than

$$\tilde{\theta}_{bF} = \begin{cases} \Delta q/2 + \underline{m}K & \text{if } \bar{u} \in [0, \underline{v}_F] \\ \sqrt{2a\bar{u}} + \underline{m}K & \text{if } \bar{u} \in [\underline{v}_F, \tilde{v}_F] \\ \Delta q + \underline{m}K & \text{if } \bar{u} \in [\tilde{v}_F, \bar{v}_F] \end{cases}, \tag{10}$$

where $\underline{v}_F, \tilde{v}_F, \bar{v}_F$ are defined in (5), (6) and (7) respectively.[24] If θ_b is smaller than the values above, then a bad worker's payoff from choosing a normal effort e is anyway higher than his payoff from choosing a destructive effort d in F. As shown in Figure 2, bad workers with $\theta_b < \tilde{\theta}_{bF}$ are therefore automatically deterred from anti-social behavior.

Let us now turn to the nonprofit sector N. By virtue of Lemma 1 bad workers will be discouraged from joining this sector as long as $u_{Nb}(d) \leq u_{Fb}(e)$ or if $u_{Nb}(d) \leq \bar{u}$. Automatic deterrence, i.e., deterrence of bad workers without any change in the optimal contract (m_N^*, t_N^*, w_N^*), therefore can be achieved for $w_N^* + \frac{1}{2a}(\theta_b - m_N^* K)^2 \leq \max\{w_F + \frac{1}{2a}(m_F t_F)^2, \bar{u}\}$, i.e., for all θ_b smaller than

$$\tilde{\theta}_{bN} \equiv \begin{cases} \sqrt{2a(w_F - w_N^*) + (m_F t_F)^2} + m_N^* K & \text{if } \bar{u} \leq \bar{v}_F \\ \sqrt{2a(\bar{u} - w_N^*)} + m_N^* K & \text{if } \bar{u} > \bar{v}_F \end{cases}. \tag{11}$$

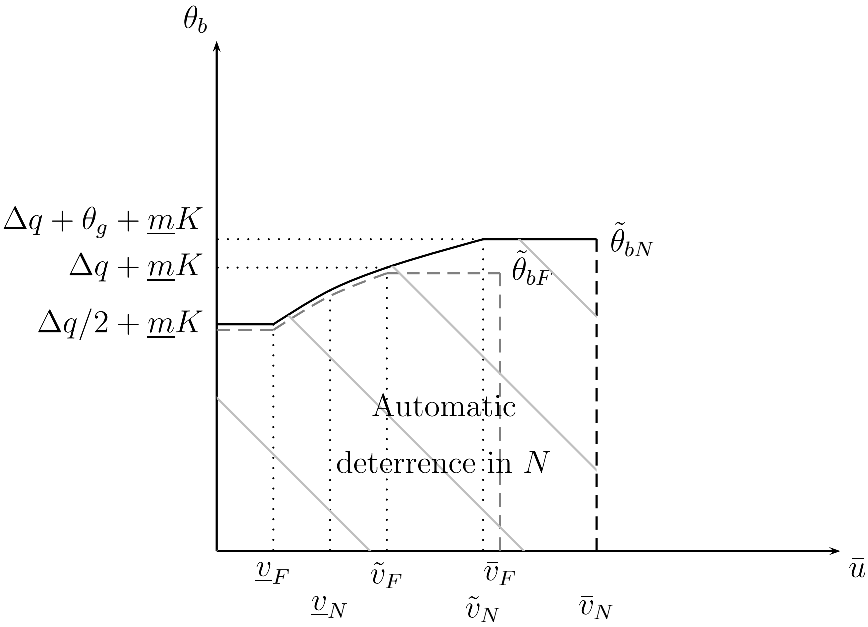
[24]See Corollary 4 in Section 7.2 of the appendix for the details of the contract and the exact value of the thresholds.

In order to determine the exact level of $\tilde{\theta}_{bN}$, we then have to insert the optimal contracts in N and F into equation (11). For the sake of shortness, we will skip this exercise here. The interested reader may however find more details in Section 7.4 of the appendix. The results are also shown in Figure 3 which depicts the level of automatic deterrence in the nonprofit sector, $\tilde{\theta}_{bN}$, as a black curve. For $(\bar{u}, \theta_b)$-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. 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:

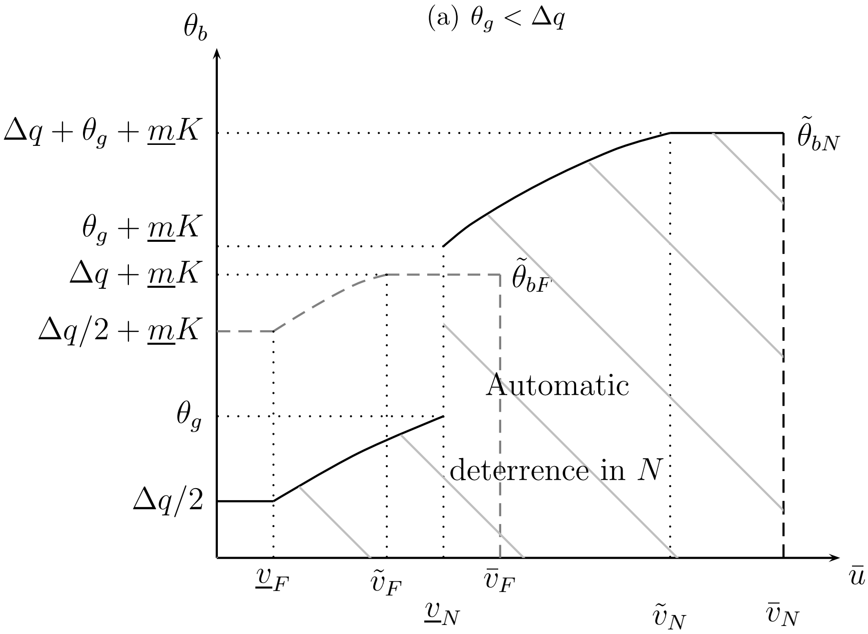
- For $\bar{u} > \bar{v}_F$, i.e., for very high levels of reservation utility, F can no longer offer contracts that would satisfy the worker's participation constraint and at the same time yield a positive payoff to the firm. Therefore, nonprofit organizations are the only possible employer for agents 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 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 $\bar{u}$ and high intrinsic motivation of good workers ($\theta_g \geq \Delta q$) as shown in Figure 3(b). In that case, the nonprofit firm relies entirely on the intrinsic motivation of good workers and hence provides no extrinsic incentives, i.e., $m_N^* = 0$ (Case Ib). The nonprofit firm then becomes particularly attractive for bad types. They can get $u_{Nb}(d) = \theta_b^2/(2a)$ from choosing a negative effort in sector N, whereas they would get utility $u_{Fb}(e) = \Delta q^2/(8a)$ from choosing a positive effort in sector F. Therefore, all bad workers with $\theta_b > \Delta q/2$ will opt for sector N and



(a) $\theta_g < \Delta q$



(b) $\theta_g > \Delta q$

Figure 3: Automatic deterrence in N. For $(\bar{u}, \theta_b)$-combinations in the shaded area, bad workers are automatically deterred from bad actions in sector N.

provide a destructive effort. Bad workers with a lower θ_b will choose sector F and behave like regular workers.

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 $\theta_b \leq \tilde{\theta}_{bF}$ the optimal contracts derived in Proposition 1 are still optimal as bad types of workers join exclusively the for profit sector where they are deterred from behaving badly. 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).

Sustaining the constructive equilibrium depends on how bad the destructive impulses are and on the optimality of the carrot and stick offered to contain them. In the following, we analyze how the optimal contracts in both sectors have to change in order to account for the presence of bad motivated workers if the contracts of Proposition 1 do not lead to automatic deterrence.

4.2 Full Deterrence of Bad Workers: large D

As we have seen in the previous section, it is not necessary to adjust the optimal contracts described in Proposition 1 as long as the intrinsic motivation of bad workers θ_b is sufficiently low. However for larger values of θ_b the optimal contracts need to be adjusted. In this section we focus on cases where the damage D is sufficiently large so that both organizations want to fully deter bad action. By virtue of Lemma 1, it is never attractive for bad workers to choose a positive 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 in the for-profit sector and provide a positive effort e. The next proposition collects this result.

Proposition 2 *If bad workers choose to work rather than to enjoy their reservation utility then in the equilibrium with full deterrence they are all in sector F and behave like regular workers.*

We now turn to the optimal reaction of the for-profit sector to the presence of bad workers.

4.2.1 Deterring Bad Actions from Bad Workers in the Profit-Oriented Sector

In this section we consider what happens if θ_b is higher than the automatic deterrence threshold $\tilde{\theta}_{bF}$ as defined in (10) and how the optimal contracts described in Proposition 1 then should be adjusted. By virtue of Proposition 2, if the principal wants to deter bad workers from being destructive, he has to make sure that $u_{Fb}(d) \leq u_{Fb}(e)$. Hence, the principal's maximization problem in sector F becomes[25]

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

subject to

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

where the last constraint is new. This deterrence constraint ensures that bad workers prefer to make a positive rather than a destructive effort. For $\theta_b > \tilde{\theta}_{bF}$ defined in (10) the deterrence constraint becomes binding and we can rewrite the principal's maximization problem as

$$\max_{w_F,m_F} \pi_F = q + (\Delta q - \theta_b + m_F K)(\theta_b - m_F K)\frac{1}{a} - w_F - M(m_F) ,$$

subject to

$$(LL) \quad w_F \geq 0 , \tag{12}$$

$$(PC) \quad (\theta_b - m_F K)^2/(2a) + w_F \geq \bar{u}_j . \tag{13}$$

As before, the solution of this maximization problem gives rise to three different cases, depending on the reservation utility of the workers. We define $\underline{v}_F^{det}$ as the outside utility for which the modified participation constraint as given in (13) becomes binding. Furthermore, let us define $\tilde{v}_F^{det}$ as the level of outside utility at which the limited liability constraint ceases to be binding and $\bar{v}_F^{det}$ as the highest level of outside utility at which the for-profit firm still

[25]The agent's incentive constraint is already taken into account here.

makes a nonnegative profit. The value of these thresholds are formally given in Appendix 7.5.

The optimal contract with full deterrence in sector F then is described by the following proposition:

Proposition 3 *: 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 $m_F^{det} = \min\{\max\{\underline{m}, \widetilde{m}_F^{det}\}, 1\}$, where $\widetilde{m}_F^{det}$ is such that the following conditions hold:*

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

The proof of Proposition 3 is in Section 7.5 in Appendix A. The appendix also shows that, although we still may get three cases, depending on the outside utility of the agents, the borders between these three cases have shifted relative to those in Proposition 1. In particular, $\underline{v}_F^{det} > \underline{v}_F$, and $\tilde{v}_F^{det} > \tilde{v}_F$, but $\bar{v}_F^{det} < \bar{v}_F$. Also, depending on the exact parameter values, profits can become negative in all three subcases, i.e. $\bar{v}_F^{det} < \tilde{v}_F^{det}$ or $\bar{v}_F^{det} < \underline{v}_F^{det}$ is possible.

Comparing the optimal contract in Proposition 3 with the benchmark contract of Proposition 1 yields the following result.[26]

Corollary 1 *Let $\theta_b \in (\tilde{\theta}_{bF}, \bar{\theta}_{bF})$. In order to fully deter sabotage by bad workers in F monitoring increases, $m_F^{det} \geq m_F^*$, and incentives become steeper, $m_F^{det}t_F^{det} \geq m_F^* t_F^*$.*

Figure 4 shows for a given outside utility how the monitoring level of Proposition 3 and profits develop as θ_b increases. It is drawn for the monitoring function $M(m) = m^2/2$.[27] The graph also illustrates that profits may become zero or negative even for low levels of outside utility[28] if the motivation of bad workers θ_b is sufficiently high.

[26]For more details, see Section 7.5 of the Appendix.

[27]The graph is based on the following parameter values: $q = 2, \Delta q = 2, a = 10, \underline{m} = 0.1, K = 3$, $\bar{u} = 0$. Computations for this example are in Appendix B.

[28]The graph is drawn for $\bar{u} = 0$, i.e. we are in Case I.

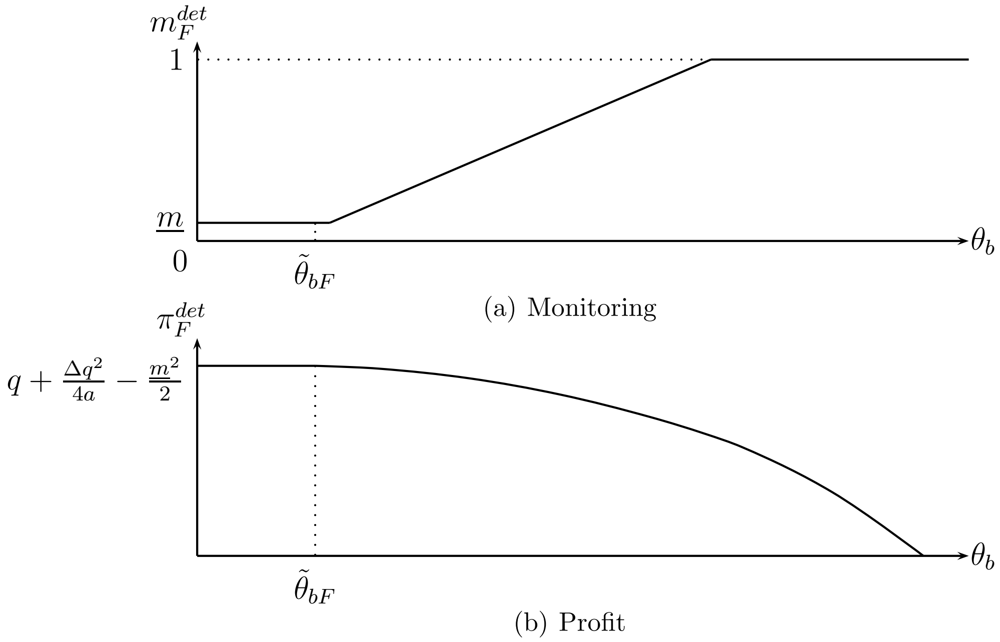


Figure 4: Monitoring level and profit in F for $\bar{u} = 0$ and $M(m) = m^2/2$.

Proposition 3 implies that, in order 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. For moderate levels of θ_b, the principal will keep monitoring at its minimum level $\underline{m}$ (see illustration Figure 4) and just increase the rewards for good behavior t_F^{det} (i.e., the carrot is sufficient). However, such a scheme will not entice workers with a very 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 $\widetilde{m}_F^{det}$. In all cases, i.e., no matter what the outside utility or motivation level is, as long as $\theta_b > \tilde{\theta}_{bF}$, the expected bonus payment for good effort increases relative to the benchmark case without bad workers. That is, $m_F^{det} t_F^{det} \geq m_F^* t_F^*$. As a consequence, besides deterring bad workers from bad actions, this contract will also induce regular workers to choose a higher effort level.

4.2.2 Deterring Bad Workers from Joining the Mission-Oriented Sector

For low levels of θ_b, the non-profit sector is protected from destructive behavior by the higher effort incentives offered in the profit-oriented sector or by a relatively high outside utility of workers. The principal in the mission-oriented sector does not need to adapt his optimal wage policies as defined in Proposition 1 as long as $\theta_b \leq \tilde{\theta}_{bN}$ where $\tilde{\theta}_{bN}$ is defined in (11) given a contract (w_F, m_F, t_F) in sector F. If the level of motivation of bad workers is higher than $\tilde{\theta}_{bN}$, then the principal in sector N will have to increase his monitoring level to deter bad workers from joining and choosing a destructive effort. At the same time, we have to make sure that the good workers will still want to work in N. The following proposition holds:

Proposition 4 *: 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 $u^{max} \equiv \max\{u_F(e), u_F(d), \bar{u}\}$.

(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.

That is, 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 $\bar{u} < \tilde{v}_F$ this condition translates to choosing the same monitoring level as in sector F, whereas for $\bar{u} \geq \tilde{v}_F$ the monitoring level in N has to increase but slightly less so than in F. This is due to the fact that the basic wage in N is smaller than in F, making work in N less attractive anyway.

On the other hand, N has to make sure that good workers still get a higher payoff from working in N rather than in F. For $\bar{u} \geq \underline{v}_F$ this condition is fulfilled even with N's benchmark contracts since $u_{Ng}(e) \geq u_F(e) = \bar{u}$. However, for $\bar{u} < \underline{v}_F$, F achieves full deterrence by raising t_F and keeping m_F at its minimum level. In this case if good workers have a weak intrinsic motivation (i.e., such that $\theta_g < m_F^{det} t_F^{det} - m_N^* t_N^*$), they may 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. The next Corollary collects this result.

Corollary 2 *Let $\theta_b \in (\tilde{\theta}_{bN}, \bar{\theta}_{bN})$. In order to fully deter bad workers to join N, monitoring increases,i.e., $m_N^{det} \geq m_N^*$. Monetary incentives increase such that $m_F^{det} t_F^{det} > m_N^{det} t_N^{det} > m_N^* t_N^*$ if and only if $\bar{u} < \underline{v}_F$ and $\theta_g < m_F^{det} t_F^{det} - m_N^* t_N^*$. They remain unchanged otherwise,i.e., $m_N^{det} t_N^{det} = m_N^* t_N^*$.*

With bad workers, the mission-oriented sector hence looses much of its wage cost advantage compared to the for-profit sector. The loss is particularly high when $\theta_g > \Delta q$: in this case, the presence of bad workers means that firms have to go from no monitoring at all to whatever monitoring there is in the for-profit sector. That is, 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. However, the optimal basic wage stays the same as before, and overall incentives will remain lower than in sector F. Even with full deterrence of bad workers, the profit in sector N therefore may still be higher than in sector F.

The various equilibrium outcomes depending on $\bar{u}$ and θ_b are illustrated in Figures 8(a) and 8(b) in Section 7.7 of the appendix. To illustrate our main results in the (θ_g, θ_b)-space we also derive a simple example where the workers' outside utility is $\bar{u} = 0$. We refer the interested reader to this section.

5 Discussion and Robustness

5.1 Partial Deterrence: low D

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

The principal in sector $i = N, F$ may accept the possibility that destructive behavior occurs. Let β_i be the share of bad workers in sector $i = N, F$ in this case. Since all bad workers face the same tradeoffs, sector i attracts either all or none of the bad workers depending on the utility that they can achieve from the different options available.[29] That is, either all or none of the bad workers will be 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 $\bar{\beta}_N \equiv x_b/(x_b+x_g)$. In order to get an interior solution, we assume that the share of bad workers is relatively limited compared to good and regular workers:

Assumption 4 $\quad x_b < x_g < x_r.$

Assumption 4 implies that $\bar{\beta}_F = \frac{x_b}{x_b+x_r} < \bar{\beta}_N = \frac{x_b}{x_b+x_g} < 0.5$.

Recall that $\theta_{Ng} = \theta_g$ and $\theta_{Fg} = 0$. Taking into account the agent's optimal effort choice, the principal's maximization problem then corresponds to

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

subject to the worker's limited liability and participation constraint as stated in (2) and (3). Section 7.8 of the appendix shows that the optimal contracts with positive monitoring then take the following form:

Lemma 2 *: For $\theta_b > \tilde{\theta}_{bi}$, 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*

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

[29] By assumption, they all choose F if indifferent between the two sectors.

where

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

Note that, depending on the exact parameter values, $\bar{v}_i^{part} \leq \tilde{v}_i^{part}$ is possible such that the third case in the above basically disappears. For more details on this special case see Section 7.8 of the appendix.

Since there are more regular than good workers, i.e., $x_r > x_g$ and hence $\bar{\beta}_F < \bar{\beta}_N$, Lemma 2 implies that $\tilde{m}_F^{part} \leq \tilde{m}_N^{part}$. Everything else being equal, bad workers will rather choose to work in F where their probability of detection will generally be lower than in N. The only two exceptions are, first, when sector N chooses to stick to its zero monitoring policy (i.e., when θ_g is large and D is small), in which case all the bad workers will be in N, and, second, when the reservation utility of regular workers is so high that only sector N is a possible employer. The next proposition summarizes these results. The details can be found in Section 7.9 in the appendix.

Proposition 5 *: 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.,* $m_N^* = t_N^* = 0$ *and* $w_N = w_N^*$*. 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* $\tilde{\theta}_{bN} < \theta_b \leq [2a\bar{u} - \Delta q^2 + (m_F K)^2]/(2m_F K)$*.*

Otherwise, the optimal contract in F is given by Lemma 2.

(b) *If* $\theta_g \geq [aM(m_F^{part})(1 + \frac{x_g}{x_b}) - D\theta_b]/\Delta q$*, and* $\bar{u} \in [0, \bar{v}_F^{part}]$*, the optimal contract in F is given by Lemma 2. 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 Lemma 2.*

In case (a) it is optimal for sector N to stick to its first best contract and not to monitor, even if in most cases this will imply that all bad workers will be in sector N, doing d.[30] When the damage D and the proportion of bad workers are small enough, introducing monitoring is indeed more costly than tolerating the rare and relatively harmless actions of bad workers.

In case (b), all bad workers are in sector F, doing d, as long as $\bar{u} \leq \bar{v}_F^{part}$. Sector N achieves full deterrence of bad workers by raising its monitoring level just high enough to make working in N unattractive. When $\bar{u} \in [0, \tilde{v}_F^{part}]$ then N and F offer the same fixed wage $\underline{w}$ so that N has to set its monitoring level to $m_N = m_F^{part}$ to repel bad types, while for $\bar{u} \in [\tilde{v}_F^{part}, \bar{v}_F^{part}]$ N can even achieve full deterrence of bad workers with a slightly lower monitoring level than F since the lower basic wage makes N already less attractive.

By virtue of Proposition 5 and Lemma 2, in sector F monitoring increases relative to its optimal level without bad workers, $\underline{m}$, if the share of bad workers and the damage they cause are high enough (i.e., if $\tilde{m}_F^{part} > \underline{m}$ which is equivalent to $M'^{-1}\big(\beta_F DK/a\big) > \underline{m}$). A higher monitoring level means that the incentive to provide negative effort goes down, which limits the damage caused by bad workers under partial deterrence. The incentives for positive effort, on the other hand, stay the same for reservation utilities that are not too high: $t_F^{part} m_F^{part} = t_F^* m_F^*$ for $\bar{u} \leq \tilde{v}_F^{part}$. Therefore the bonus payment t_F^{part} can be lower than without bad workers. We deduce that if the share of bad workers, respectively the damage they cause, are sufficiently low (i.e., if $\tilde{m}_F^{part} \leq \underline{m}$) and the workers' reservation utility is not too high then monitoring and bonus payments in F stay exactly the same as in the benchmark case. By ignoring the bad workers, the firm makes lower profits due to the damage they cause, but it would be more costly to raise the monitoring level to fight them. Only for very high levels of outside utility the overall incentives for good behavior go up: $t_F^{part} m_F^{part} \geq t_F^* m_F^* = \Delta q$ for $\bar{u} > \tilde{v}_F^{part}$. This result is counter-intuitive because in incentive theory it is usually not optimal to provide a bonus bigger than Δq to the workers: each time they are successful the principal looses money as she gives them in bonus more than what they actually produce. However when the binding constraint is the participation constraint of the workers, the principal prefers to meet this constraint by increasing the bonus rather than the fixed wage because the later benefits also the bad workers, while the former benefits only the workers who provide a positive effort. As a result, the expected bonus is

[30]The exception being the following: Depending on the exact parameter values, it is possible that the basic wage w_F in F is high enough to attract all the bad workers even if there is no monitoring in sector N. This possibility arises only if the conditions described under (a) and (ii) in Proposition 5 hold.

higher than Δq, the optimal expected bonus in the absence of bad workers.

Finally, in case (c), when F is no longer an option because $\bar{u} > \bar{v}_F^{part}$ all bad workers will be either in sector N and choose d or will enjoy their reservation utility. If their intrinsic motivation is large enough so that they choose to join N, the optimal contracts for N are then derived in Lemma 2. For more details see Section 7.9 in the appendix.

So when is partial deterrence better than full deterrence? Considering Propositions 4 and 5, and focusing on an equilibrium where both sectors are active for the sake of realism, we find:

Corollary 3 *: As long as both sectors are active, full deterrence in N is always optimal unless $[aM(m_N^{det})(1 + x_g/x_b) - D\theta_b]/\Delta q > \theta_g > \Delta q$ and $\bar{u} \in [0, \underline{v}_N]$.*

When both sectors are active our findings can be summarized as followed: Both under full and partial deterrence it is optimal for N to change its contracts to deter bad workers from joining. Only if the damage and the relative share of bad workers are very small (Case (a) in Proposition 5), it is cheaper for N to accept the presence of bad workers rather than to deter them. Symmetrically sector F is always confronted with bad workers in equilibrium unless all the conditions of Proposition 5(a) hold.

Whether the principal in sector F then prefers full deterrence 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

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

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, β_F, in the damage these workers may cause D and in their intrinsic motivation θ_b. This means that the larger the share of bad workers in sector F and the higher the expected damage, the more likely it is that $\pi_F^{det} > \pi_F^{part}$, i.e., that the principal in sector F will prefer to fully deter bad workers. If, for

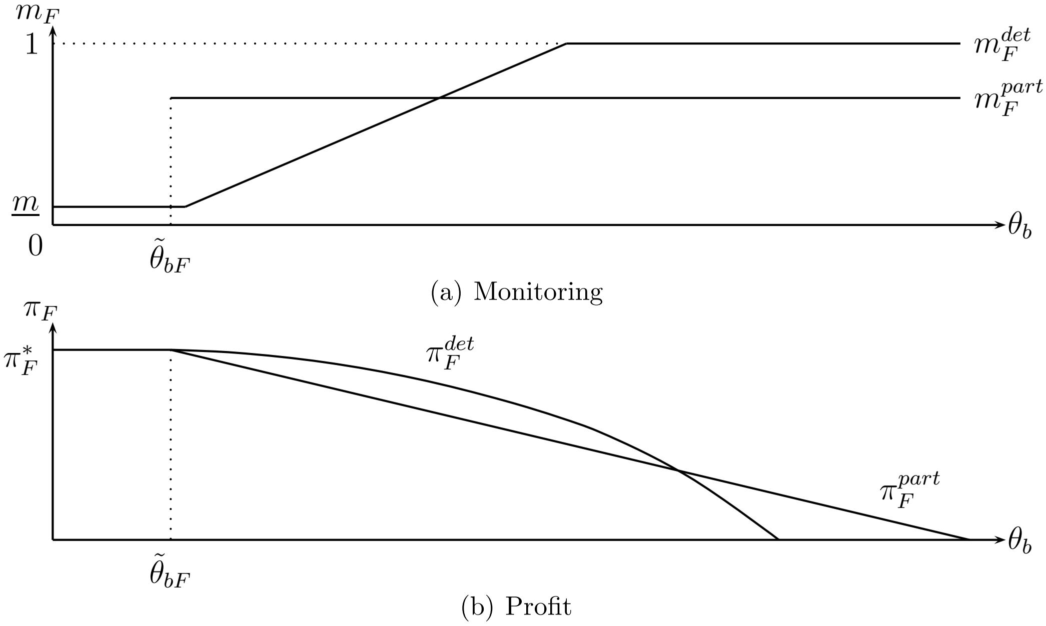


Figure 5: Comparison of monitoring levels and profit in F under full and partial deterrence

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, β_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 5.[31] The graph shows that for a given level of β_F and D, partial deterrence is less costly as long as the intrinsic motivation of bad workers, θ_b, is low, but full deterrence becomes more attractive as θ_b rises. For extreme cases such as very low levels of β_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 β_F and/or D are high such that m_F^{part} is close to 1.

5.2 Ex Ante Control: very large θ_b

Section 4 has shown that depending on the level of negative motivation of bad workers, θ_b, firms may be able to deal with the problem by adapting

[31]The graphs are drawn for $q = 2, \Delta q = 2, a = 10, \underline{m} = 0.1, K = 3, x_b = 0.2, x_r = 0.5, D = 8$ and $M(m) = m^2/2$.

their incentive schemes and in particular their monitoring levels. However, this increase of ex post monitoring may be very costly, and for high levels of negative motivation it becomes even entirely ineffective. Firms therefore may want to invest in ex ante measures to reduce the probability of hiring a bad worker in the first place.

Some form of applicant screening, which may serve to filter out more trustworthy or motivated workers, is quite common in most firms. The higher the expected damage of hiring a bad worker, the more an organization or firm will be inclined to invest in a more sophisticated selection process of applicants. This is observed in sectors where candidates, once hired, are difficult to fire, as for example civil servants[32] or where the stakes are high as for instance in intelligence services. The selection process in these cases can be quite lengthy and generally involves all kinds of tests and background checks. For instance, the CIA states on its web site:[33] "Depending on an applicant's specific circumstances, the [application] process may take as little as two months or more than a year. [. . .] Applicants must undergo a thorough background investigation examining their life history, character, trustworthiness, reliability and soundness of judgment [. . .], [their] freedom from conflicting allegiances, potential to be coerced and willingness and ability to abide by regulations governing the use, handling and the protection of sensitive information. The Agency uses the polygraph to check the veracity of this information. The hiring process also entails a thorough medical examination of one's mental and physical fitness to perform essential job functions." The FBI states that "The clearance process can take anywhere from several months to a year or more",[34] and lists as part of the background check "a polygraph examination; a test for illegal drugs; credit and records checks; and extensive interviews with former and current colleagues, neighbors, friends, professors, etc.".

Similarly, many nonprofit organizations require a lot of previous experience and conduct extensive interviews before hiring someone, especially in cases where monitoring in the field is difficult (e.g., Médecins sans Frontières).

A better candidate selection process can thus serve as a (partial) substitute for worker monitoring.[35] However, checking each applicant thoroughly is costly, and therefore has to be seen in relation to the potential damage of hiring a bad worker.

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

[33] See www.cia.gov/careers/faq/index.html#a3

[34] See www.fbijobs.gov/61.asp#3

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

In this context, legal requirements may play an important role in order to help employers screen out bad workers. In Germany, for instance, employers can ask applicants for a police clearance certificate ("Führungszeugnis"), which, however, only documents offenses that are punishable beyond a certain degree of penalty in order to give offenders a second chance. Unfortunately, until recently, many potentially relevant cases of molestation, child pornography, exhibitionism etc. did thus not appear in the records. This came under discussion with the occurrence of several cases of child molestation where the employer was unaware of his employee's history, although the employee had been convicted for similar behavior before. To prevent cases like this in the future, the government introduced an "extended police clearance certificate" ("erweitertes Führungszeugnis"), which can be requested for anyone seeking employment in a job that may bring him or her in contact with children or youths.[36]

In other cases, establishing a clearer profile of bad workers may help. This has, for example, been done in the US to prevent fire fighter arson. Studies by the South Carolina Forestry Commission and the FBI[37] 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 who, in the U.S. as well as in many European countries make up for only 25% of all fire fighters. The South Carolina Forestry Commission hence has designed an "Arson Screening and Prediction System" which is supposed to help field level administrators to evaluate candidates. It attributes a numeric score to the answers to a questionnaire covering areas such as the candidate's family background, his social skills, capacity for self control, intelligence, self-esteem and academic performance, stress and attitudes towards the fire service.

Yet another measure to prevent destructive behavior may be to promote peer monitoring, which is especially attractive if ex ante candidate screening is less than perfect and monitoring of workers is difficult. There 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. However, 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.

[36]See press release of the German Ministry of Justice from 14 May 2009, www.bmj.bund.de/enid/Nationales_Strafrecht/Erweitertes_Fuehrungszeugnis_1js.html .

[37]See Stambaugh and Styron (2003) for a summary of both studies.

Depending on circumstances, different practical measures may be appropriate in order to introduce some extent of mutual monitoring. In the case of fire fighter arson, for example, promoting peer monitoring consists of awareness programs that are supposed to alert fire departments to the problem and keep their eyes open. 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,[38] or providing joint housing for aid workers.[39] While this may give rise to collusion among evil-doers, 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.

6 Conclusion

The existence of "destructive" workers who derive satisfaction from actions that are detrimental to their employer or others affects 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 positive 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 of good workers: For high enough motivation

[38]There is anecdotal evidence that, for example, the French service for teaching abroad prefers to hire couples, not only for monitoring reasons, but mainly because they have been found to withstand stress caused by a new environment better.

[39]This is for example the approach of Ärzte für die Dritte Welt (Doctors for Developing Countries), a German NGO that runs several permanent projects in Africa, Asia and Central America with the help of doctors doing short term volunteer work. Again, this rather has practical reasons and is not necessarily intended as a measure to promote peer monitoring, but still it may act in such a way.

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 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.

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 business, many mission-oriented organizations are run as non-government 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. It also provides strong incentives to hide bad actions by their workers, hence implicitly encouraging them to continue. For instance in the case of the abuse scandals involving the Catholic Brothers in Ireland, abusers had 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, 2009).[40]

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.[41] Taking into account such effects would mean that the more the mission-oriented sector N

[40]Many other cases involving the Catholic church were handled in a similar way, possibly not only to avoid a scandal but also because of a prevailing norm that wrongdoers deserve a second chance to redeem themselves.

[41]See Seabright (2009), Frey and Jegen (2001), Frey and Oberholzer-Gee (1997).

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 effort incentives t_N in order to induce good workers to work hard enough, 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, Rebitzer, Sanders, and Taylor, 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.

7 Appendix A: Proofs

7.1 Proof of Proposition 1

After inserting the incentive constraint, the maximization problem stated in (1) to (4) can be rewritten as the following Lagrangian:

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

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

$$
\begin{aligned}
\frac{\partial L}{\partial w_{ij}} &= -1 + \lambda_{LL} + \lambda_{PC} \leq 0 \ , && (14)\\
\frac{\partial L}{\partial t_{ij}} &= \frac{m_i}{a}[\Delta q - 2m_i t_{ij} - \theta_{ij} + \lambda_{PC}(m_i t_{ij} + \theta_{ij})] \leq 0 \ , && (15)\\
\frac{\partial L}{\partial m_i} &= \frac{t_{ij}}{a}[\Delta q - 2m_i t_{ij} - \theta_{ij} + \lambda_{PC}(m_i t_{ij} + \theta_{ij})] - M'(m_i) \leq 0 \ , && (16)\\
\frac{\partial L}{\partial \lambda_{LL}} &= w_{ij} \geq 0 \ , && (17)\\
\frac{\partial L}{\partial \lambda_{PC}} &= w_{ij} + (m_i t_{ij} + \theta_{ij})^2/(2a) - \bar{u}_j \geq 0 \ , && (18)\\
0 &= \lambda_{LL} w_{ij} \ , && (19)\\
0 &= \lambda_{PC}(w_{ij} + (m_i t_{ij} + \theta_{ij})^2/(2a) - \bar{u}_j) \ , && (20)
\end{aligned}
$$

From (14) 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 $\lambda_{LL} = \lambda_{PC} = 0$, (14) 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 (15) is binding, then (16) 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:

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 2, 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 (15) 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 $\bar{u}_j = \underline{v}(\theta_{ij})$ where

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

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

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 (18) follows that $m_i t_{ij} = \sqrt{2a\bar{u}_j} - \theta_{ij}$ is optimal. For this to be a solution, it 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

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

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

Case III: (LL) not binding, (PC) binding

If the (LL) constraint is not binding ($\lambda_{LL} = 0$), then $w_{ij} > 0$. This implies in (14) an interior solution so that $\lambda_{PC} = 1$. We deduce that if $m_i = \underline{m} > 0$, by (15), 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

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

The principal's payoff then is

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

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

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

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

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

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 $\underline{v}(\theta_{ij}) \leq \tilde{v}(\theta_{ij}) \leq \bar{v}(\theta_{ij})$. QED

7.2 Implications of Proposition 1 for the For-Profit Sector

The principal in the profit-oriented sector cannot rely on worker's intrinsic motivation (i.e., $\theta_{Fj} = 0$) and hence has to provide sufficient extrinsic incentives. In particular, he always has to invest in monitoring. We can therefore deduce the following corollary from Proposition 1 for the for-profit sector:

Corollary 4 *: 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 $\bar{u}$.[42]

The principal's payoff is

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

7.3 Implications of Proposition 1 for the Non-Profit Sector

In contrast to the profit-oriented sector, the mission-oriented sector N can save on wage costs by exploiting the intrinsic motivation of "good" workers.

Suppose the level of intrinsic motivation of good workers is $\theta_{Ng} \equiv \theta_g$. Then we get the following corollary from Proposition 1 for the non-profit sector:

Corollary 5 *: 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 $t_N^* = 0$.
- ***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

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

[42] By Assumption 2 $\tilde{v}_F = \frac{\Delta q^2}{2a} < \bar{v}_F = \frac{\Delta q^2}{2a} + q - M(\underline{m})$.

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

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

7.4 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 (11). 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 7 indicates which of the cases derived in Corollaries 4 and 5 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 (11), 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 6.

7.5 Proof of Proposition 3: Full Deterrence in F

The solution to the principal's maximization problem with full deterrence of bad workers is similar to the solution in the benchmark model. We can formulate the following Lagrangian:

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

The corresponding first-order conditions are

$$\frac{\partial L}{\partial w_F} = -1 + \lambda_{LL} + \lambda_{PC} = 0 , \qquad (21)$$

$$\frac{\partial L}{\partial m_F} = \frac{K}{a}[2\theta_b - 2m_F K - \Delta q] - M'(m_F) - \lambda_{PC} \frac{K}{a}(\theta_b - m_F K) . \qquad (22)$$

Furthermore it has to hold that

$$0 = \lambda_{LL} w_F \qquad (23)$$

$$0 = \lambda_{PC}(w_F + (\theta_b - m_F K)^2/(2a) - \bar{u}_j) . \qquad (24)$$

As before, we get three cases:[43]

Case I: (LL) binding, (PC) not binding

When (LL) is binding, then $\lambda_{LL} > 0$. From condition (23) therefore follows that the optimal basic wage in Case I $w_F^I = 0$. If the (PC) is not binding, then $\lambda_{PC} = 0$. Hence, from condition (22) 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 (23) we therefore have that the optimal wage in Case II $w_F^{II} = 0$. Furthermore, condition (24) is fulfilled iff

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

[43]For the sake of shortness, the index "det" is omitted unless needed for clarity. Instead, a case index is added.

Case III: (LL) not binding, (PC) binding

Since the limited liability constraint is not binding, $\lambda_{LL} = 0$ and hence by (21) $\lambda_{PC} = 1$. Inserting this in (22), 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 the participation constraint is binding the corresponding optimal basic wage is $w_F^{III} = \bar{u} - (\theta_b - \tilde{m}_F^{III} K)^2/(2a)$.

We thus have derived the optimal fixed wage and monitoring level for all three cases. Note, however, that depending on the functional form of $M(m)$, 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 impossible by assumption, the optimal monitoring level is $m_F^{det} = \min\{\max\{\underline{m}, \widetilde{m}_F\}, 1\}$, where $\widetilde{m}_F$ is such that the following conditions are fulfilled:

$$\begin{aligned} 2\widetilde{m}_F K + M'(\widetilde{m}_F) a/K = 2\theta_b - \Delta q & \quad \text{in Case I} \ , \\ \widetilde{m}_F = \frac{1}{K}(\theta_b - \sqrt{2a\bar{u}}) & \quad \text{in Case II} \ , \\ \widetilde{m}_F K + M'(\widetilde{m}_F) a/K = \theta_b - \Delta q & \quad \text{in Case III} \ . \end{aligned}$$

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 follows:

$$w_F^{III} = \bar{u} - \frac{(\theta_b - m_F^{detIII} K)^2}{2a} \ ,$$

where $m_F^{detIII} = \min\{\max\{\underline{m}, \widetilde{m}_F^{III}\}, 1\}$.

Finally, we still have to 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 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.

The question remains when each of the Cases I to III is relevant. That is, 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

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

For outside values above this one, Case III holds. 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

$$q + (\Delta q - \theta_b + m_F^{det} K)(\theta_b - m_F^{det} K)/a - M(m_F^{det} \geq 0 \ . \tag{25}$$

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 (25).

The derivative of $\bar{v}_F^{det}$ with respect to θ_b is smaller than zero if

$$(\Delta q - \theta_b + m_F^{III} K) M''(m_F^{III}) \ \leq \ a M'(m_F^{III}) \ .$$

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

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

This is equivalent to:

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

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

Note on Corollary 1:

Since we are only looking at cases where $\theta_b \geq \tilde{\theta}_{bF}$, by definition of $\tilde{\theta}_{bF}$ we know that $\theta_b - m_F^* K \geq m_F^* t_F^*$. Furthermore the deterrence constraint has to hold.[44] Hence $m_F^{det} t_F^{det} \geq \theta_b - m_F^* K \geq m_F^* t_F^*$ has to hold.

7.6 Proof of Proposition 4

To achieve full deterrence, N has to make sure that $u_{Nb}(d) \leq u^{max} \equiv \max\{u_F(e), u_F(d), \bar{u}\}$, i.e., $w_N^{det} + (\theta_b - m_N^{det}K)^2/(2a) \leq u^{max}$. 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

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

[44] The deterrence constraint also shows that monitoring is increasing in θ_b.

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.

7.7 Equilibrium with Full Deterrence

Where are the bad guys? To answer this question, we start with the benchmark contracts by comparing workers' utility levels.[45]

Low Level of Positive Intrinsic Motivation

Let us first consider what happens if $\theta_g < \Delta q$. Given the benchmark contracts which do not account for the presence of bad workers, Figure 6 summarizes the utility such a worker can achieve by choosing a destructive effort in either of the two sectors or by behaving like a regular worker in sector F (by virtue of Corollary 1 these are the only relevant options). Depending on the level of reservation utility, we can distinguish the following cases:

$\bar{u} < \tilde{v}_F$: 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 3. Given F's choice of contract, N can deter bad workers

[45] See Figures 6 and 7.

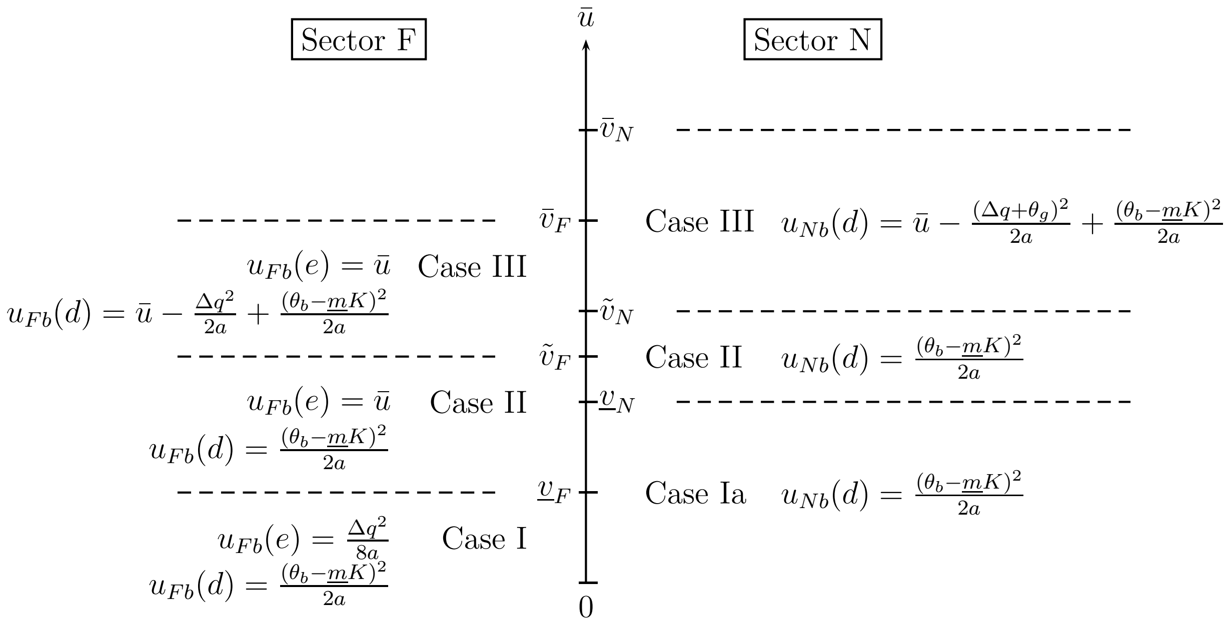


Figure 6: Bad workers' utility from positive and negative effort in F and negative effort in N for $\theta_g < \Delta q$.

by raising m_N as described in Proposition 4. In equilibrium, all bad workers are in F and behave like regular types.

$\tilde{v}_F < \bar{u} < \bar{v}_F$: 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 3, whereas N only has to adjust its contracts if $\theta_b > \tilde{\theta}_{bN}$.

$\bar{v}_F < \bar{u}$: 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 4 to achieve the same effect.

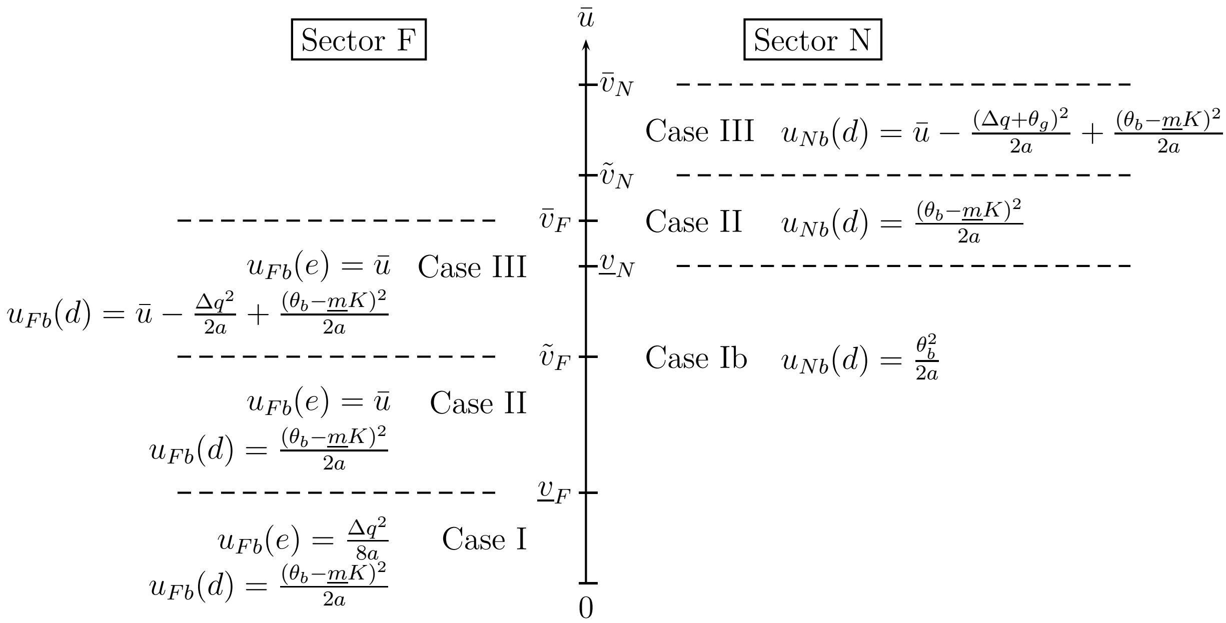


Figure 7: Bad workers' utility from positive and negative effort in F and negative effort in N for $\theta_g \geq \Delta q$.

High Level of Positive Intrinsic Motivation

If the intrinsic motivation of good workers is high, i.e. if $\theta_g \geq \Delta q$, we can distinguish the following cases:[46]

$\bar{u} < \underline{v}_N$: For very low levels of negative intrinsic motivation ($\theta_b \leq \tilde{\theta}_{bN}$), there is automatic deterrence in both sector N and F, and the benchmark contracts are sufficient to contain bad behavior in both sectors. 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, unless the contracts are adapted. 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 3. In an equilibrium with full deterrence in both sectors, bad workers will always work in F and behave like regular workers.

$\underline{v}_N < \bar{u} < \bar{v}_F$: 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

[46]See Figure 7 for an overview of the utility of bad workers in both sectors.

(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 3 for any $\theta_b > \tilde{\theta}_{bF}$. N can keep its benchmark contract as long as $\theta_b \leq \tilde{\theta}_{bN}$, but will 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 4. In equilibrium, all the bad guys will behave regularly in sector F.

$\bar{v}_F < \bar{u}$: For a very high reservation utility, sector N is the only possible employment option for bad workers. If $\theta_b < \tilde{\theta}_{bN}$ 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 4. As a result, bad workers prefer not to work at all in equilibrium.

7.7.1 Equilibrium Results

Figures 8(a) and 8(b) illustrate the various equilibrium outcomes depending on $\bar{u}$ and θ_b. Overall, we can distinguish five different cases which are illustrated in Figure 8 and characterized as follows:

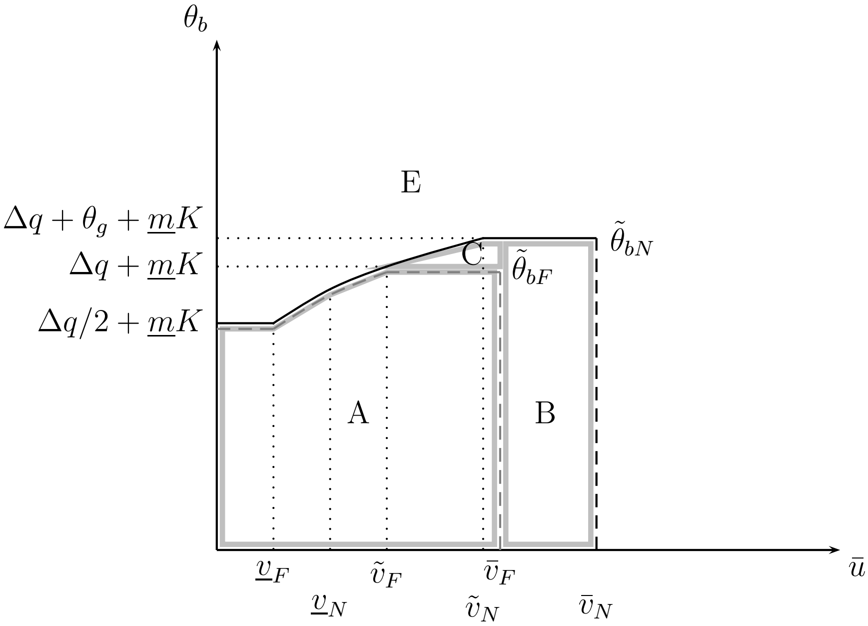
`Area A`: With the benchmark contracts, $u_{Fb}(e) > u_{Fb}(d)$ and $u_{Fb}(e) > u_{Nb}(d)$. Even without a change in contracts, all bad workers are in sector F and behave like regular workers.

`Area B`: With the benchmark contracts, $u_{Nb}(d) < \bar{u}$. Only N is a possible employer for bad workers with such a high reservation utility. However, joining N is not attractive and bad workers will prefer to enjoy their outside utility.

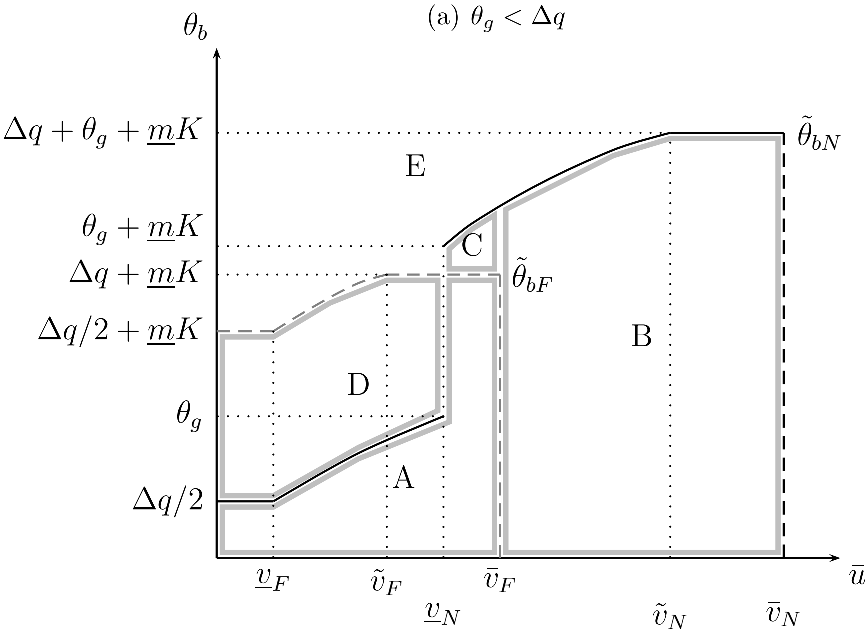
`Area C`: With the benchmark contracts, $u_{Fb}(d) > u_{Fb}(e)$ and $u_{Fb}(e) > u_{Nb}(d)$. In this case, there is automatic deterrence in sector N, but not in sector F. In equilibrium, the optimal contracts in F are given by Proposition 3, whereas N can keep its benchmark contract. All bad workers will be 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 full deterrence in F, but N will have to introduce minimal monitoring $\underline{m}$. Then, all bad workers will be in sector 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



(a) $\theta_g < \Delta q$



(b) $\theta_g > \Delta q$

Figure 8: Equilibria with full deterrence in both sectors.

be adapted. The optimal contracts in this case are given by Propositions 3 and 4. All bad workers choose either (F, e) or $\bar{u}$.[47]

7.7.2 A Simple Illustrative Case: No Reservation Utility

To illustrate our main results in the (θ_g, θ_b)-space we derive a simple example where the workers' outside utility is $\bar{u} = 0$. If $\bar{u} = 0$ and there are only good and regular workers, then the optimal contract in F is given by *Case I* in Section 3:

$$w_F^* = 0,\ m_F^* = \underline{m},\ t_F^* = \Delta q/(2\underline{m})\ ,$$

and the optimal contract in N is given by[48]

$$\begin{array}{lll} \textit{Case Ia} & w_N^* = 0,\ m_N^* = \underline{m},\ t_F^* = (\Delta q - \theta_g)/(2\underline{m}) & \text{if } \theta_g < \Delta q\ , \\ \textit{Case Ib} & w_N^* = m_N^* = t_F^* = 0 & \text{if } \theta_g \geq \Delta q\ . \end{array}$$

Given these contracts, bad workers will compare their respective utility levels from choosing a positive effort e or a negative effort d in either of the two sectors. In sector F, the above contract is sufficient to deter bad workers from bad actions as long as their negative intrinsic motivation θ_b is smaller than $\tilde{\theta}_{bF} = \Delta q/2 + \underline{m}K$, i.e., for $\theta_b \leq \Delta q/2 + \underline{m}K$ it holds that $u_{Fb}(e) \geq u_{Fb}(d)$. In sector N, the above benchmark contract is enough to deter bad workers from joining if

$$\theta_b \leq \tilde{\theta}_{bN} = \begin{cases} \Delta q/2 + \underline{m}K & \text{if } \theta_g < \Delta q \\ \Delta q/2 & \text{if } \theta_g \geq \Delta q \end{cases}\ ,$$

i.e., for $\theta_b \leq \tilde{\theta}_{bN}$ it holds that $u_{Nb}(d) < u_{Fb}(e)$. The area where this *automatic deterrence* result holds is illustrated in Figure 9. Now, what happens if the motivation of bad workers exceeds these thresholds?

Case Ia: $\theta_g < \Delta q$. For $\theta_b > \tilde{\theta}_{bF} = \tilde{\theta}_{bN} = \Delta q/2 + \underline{m}K$, F will have to increase the expected bonus for good behavior and/or raise the monitoring level in order to deter bad behavior. As has been shown in Proposition 3, this can be achieved by setting $w_F = 0$, $t_F = \theta_b/m_F^{det} - K$ and $m_F^{det} = \min\{\max\{\underline{m}, \tilde{m}_F^{det}\}, 1\}$ where $\tilde{m}_F^{det}$ is such that $2\tilde{m}_F^{det}K + M'(\tilde{m}_F^{det})a/K = 2\theta_b - \Delta q$. Given this contract, bad workers will prefer to behave like regular types in F or switch to N and misbehave there. To prevent the latter, it is

[47] The latter holds especially if only N is a possible employer.

[48] For the derivation of these contracts see Corollaries 4 and 5 in Section 7.2 and 7.3 of the appendix.

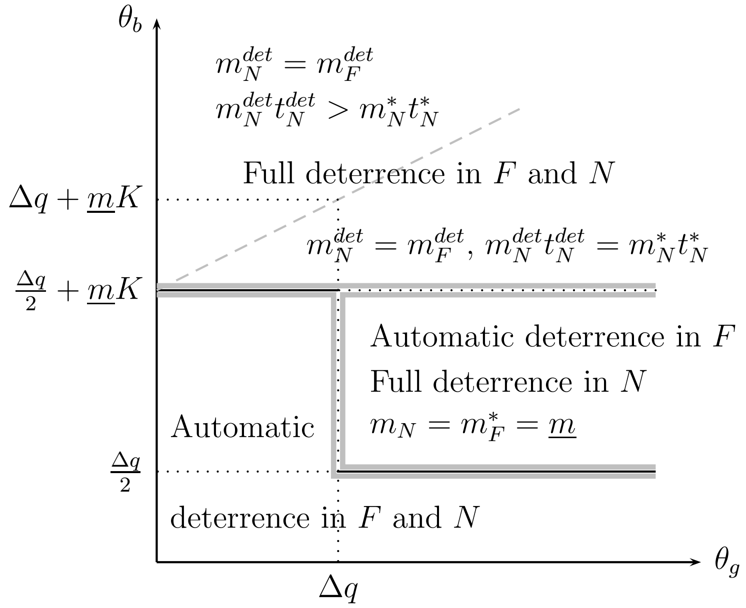


Figure 9: Equilibrium with $\bar{u} = 0$.

sufficient that N raises its monitoring level to the one used in F. As a result, in an equilibrium with full deterrence in *both* sectors, $m_N = m_F = m_F^{det}$.

Case Ib: $\theta_g \geq \Delta q$. If $\tilde{\theta}_{bN} = \Delta q/2 < \theta_b < \tilde{\theta}_{bF} = \Delta q/2 + \underline{m}K$, then $u_N(d) > u_F(e) > u_F(d)$. That is, given the benchmark contracts, all bad guys will be in sector N, making a destructive effort. However, N can achieve full deterrence by introducing monitoring, i.e., by raising m_N from 0 to $\underline{m}$. Note that F has no need to change its benchmark contracts since bad workers have no incentive to behave badly in F anyway. If $\theta_b > \tilde{\theta}_{bF} = \Delta q/2 + \underline{m}K$, 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 above to deter bad workers, whereas N can achieve full deterrence by imitating F's choice of monitoring level.

Turning to monetary incentives in sector N, for high enough intrinsic motivation of good workers ($\theta_g > m_F^{det} t_F^{det} - m_N^* t_N^*$), N has no need 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 relatively small, good workers may be tempted to switch to sector F since this sector 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. These considerations are also illustrated in Figure 9.[49] Above the diagonal line, sector N has to raise the incentives

[49]Figure 9 is drawn for $\theta_b < K + \Delta q/2 + a\underline{m}/(2k)$, which implies that $m_F^{det} = \underline{m}$ and

for good behavior above their benchmark level to keep good workers around; below this line, good workers stay in N even if the incentives stay the same as in the benchmark contracts.

7.8 Proof of Lemma 2

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

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

and the corresponding first-order conditions are

$$\frac{\partial L}{\partial w_i} = -1 + \lambda_{LL} + \lambda_{PC} = 0 , \tag{26}$$

$$\frac{\partial L}{\partial t_i} = \frac{m_i}{a}[(1-\beta_i)(\Delta q - 2m_i t_i - \theta_{ig}) + \lambda_{PC}(m_i t_i + \theta_{ig})] = 0 , \tag{27}$$

$$\begin{aligned}\frac{\partial L}{\partial m_i} = \; & \frac{t_i}{a}[(1-\beta_i)(\Delta q - 2m_i t_i - \theta_{ig}) + \lambda_{PC}(m_i t_i + \theta_{ig})] \\ & -M'(m_i) + \frac{\beta_i DK}{a} = 0 .\end{aligned} \tag{28}$$

Furthermore, the following has to be true:

$$0 = \lambda_{LL} w_i , \tag{29}$$

$$0 = \lambda_{PC}(w_i + (m_i t_i + \theta_{ig})^2/(2a) - \bar{u}_j) . \tag{30}$$

Equation (27), i.e., the FOC with respect to t_i, is fulfilled if the expression in square brackets is equal to zero. This implies that (28), the FOC with respect to m_i, simplifies to

$$-M'(m_i) + \frac{\beta_i DK}{a} = 0 ,$$

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

$t_F^{det} = (\theta_b - K)/\underline{m}$. For higher values of θ_b, $m_F^{det} > \underline{m}$ and the diagonal line shifts upwards.

that $m_i^{part} \in \{0, [\underline{m}, 1]\}$, i.e., that we get an interior solution.[50] 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

$$m_i^{part} = \min\{\max\{\underline{m}, \tilde{m}_i^{part}\}, 1\} \ , \qquad (31)$$
$$\text{where } \tilde{m}_i^{part} s.t. M'(\tilde{m}_i^{part}) = \beta_i DK/a \ .$$

As before, we get three possible cases:

Case I: (LL) binding, but not (PC)

In this case, $\lambda_{LL} > 0$, and therefore, by (29), $w_i^{part} = 0$. Furthermore, $\lambda_{PC} = 0$, such that condition (27), 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 (31), 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

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

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 (29), $w_i^{part} = 0$. Furthermore, since $\lambda_{PC} > 0$, the (PC) becomes binding, such that $(m_i t_i + \theta_{ig})^2/(2a) = \bar{u}$ must hold. Hence it must be true 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 (31) 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

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

[50] Since we are considering cases where automatic deterrence no longer works, $m_i^{part} = 0$ is no longer an option.

Case III: (PC) binding, but not (LL)

Since $\lambda_{LL} = 0$, by (29), $w_i^{part} > 0$. Furthermore, $\lambda_{PC} > 0$, i.e. the (PC) is binding. In this case,

$$w_i^{part} = \bar{u} - (m_i t_i + \theta_{ig})^2/(2a)$$

must hold.

By condition (26), $\lambda_{LL} = 0$ also implies that $\lambda_{PC} = 1$. Inserted in (27), we get the following expression:

$$\frac{m_i}{a}[(1-\beta_i)(\Delta q - 2m_i t_i - \theta_{ig}) + m_i t_i + \theta_{ig}] = 0 \ .$$

This condition is fulfilled if the term in square brackets is equal to zero, which is the case if

$$m_i t_i = \frac{1-\beta_i}{1-2\beta_i}\Delta q + \frac{\beta_i}{1-2\beta_i}\theta_{ig} \ . \qquad (32)$$

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 negative 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 (31) 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

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

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 $\beta_i < 0.5$.

Last but not least, 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 $Z \equiv q - \frac{\beta_i D}{a}(\theta_b - m_i^{part} K) - M(m_i^{part})$. 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

$$\bar{u} \geq \min\{\bar{v}_i^{part^{II}}, \bar{v}_i^{part^{III}}\} \ ,$$

where

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

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

$$\bar{v}_i^{part^{III}} \equiv Z + \frac{1-2\beta_i}{2a}\Big(\frac{1-\beta_i}{1-2\beta_i}(\Delta q + \theta_{ig})\Big)^2$$

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

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

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.

7.9 Proof of Proposition 5

If $\bar{\in}[\bar{v}_F^{part}, \bar{v}_N]$, 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 Lemma 2.

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 $\theta_b > \tilde{\theta}_{bF}$, in theses cases 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 Lemma 2. In that case, however, all bad workers would switch to N such that $\beta_N = \bar{\beta}_N$. Since

$\bar{\beta}_N > \bar{\beta}_F$, the resulting optimal monitoring level in N according to Lemma 2 would be equal or higher than in F. However, N can even achieve full deterrence of bad workers by adapting 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 $\bar{u} \in [0, \tilde{v}_F^{part}]$, whereas for $\bar{u} \in [\tilde{v}_F^{part}, \underline{v}_N]$, 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 5.

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 $\tilde{\theta}_{bF} < \theta_b \leq [2a\bar{u} - \Delta q^2 + (m_F K)^2]/(2m_F K)$. 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 Lemma 2.

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 Lemma 2 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.

8 Appendix B: 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 $\theta_b < \tilde{\theta}^I_{bF} := \Delta q/2 + \underline{m}K$, there is automatic deterrence of bad workers. The optimal contract then corresponds to the contract outlined in Corollary 4, 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 $\theta_b > \tilde{\theta}^I_{bF}$, we have to calculate the optimal level of monitoring according to Proposition 3 which gives us

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

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

$$\begin{aligned} w^{det}_F &= 0 \; , \; t^{det}_F = \theta_b/m^{det}_F - K \; , \\ m^{det}_F &= \begin{cases} \underline{m} & \text{if } \theta_b \in (\tilde{\theta}^I_{bF}, \tilde{\theta}^I_{bF} + a\underline{m}/(2K)) \\ (2\theta_b - \Delta q)K/(a + 2K^2) & \text{if } \theta_b \in [\tilde{\theta}^I_{bF} + a\underline{m}/(2K), \\ & \qquad \Delta q/2 + K + a/(2K)] \\ 1 & \text{if } \theta_b > \Delta q/2 + K + a/(2K) \end{cases} . \end{aligned}$$

That is, even when bad workers are not automatically deterred (i.e., when $\theta_b > \tilde{\theta}^I_{bF}$), 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 $\theta_b < \tilde{\theta}^{II}_{bF} := \sqrt{2a\bar{u}} + \underline{m}K$, 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 $\theta_b < \tilde{theta}_b^{II}$ and larger than 1 if $\theta_b > K + \sqrt{2a\bar{u}}$. Therefore the optimal contract with full deterrence is given by

$$\begin{aligned} w_F^{det} &= 0 \; , \; t_F^{det} = \theta_b/m_F^{det} - K \; , \\ m_F^{det} &= \begin{cases} \underline{m} & \text{if } \theta_b < \tilde{\theta}_{bF}^{II} \\ (\theta_b - \sqrt{2a\bar{u}})/K & \text{if } \theta_b \in [\tilde{\theta}_{bF}^{II}, K + \sqrt{2a\bar{u}}] \\ 1 & \text{if } \theta_b > K + \sqrt{2a\bar{u}} \end{cases} \; . \end{aligned}$$

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 3, 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 $\theta_b < \tilde{\theta}_{bF}^{I} + a\underline{m}/K)$ 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

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

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

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

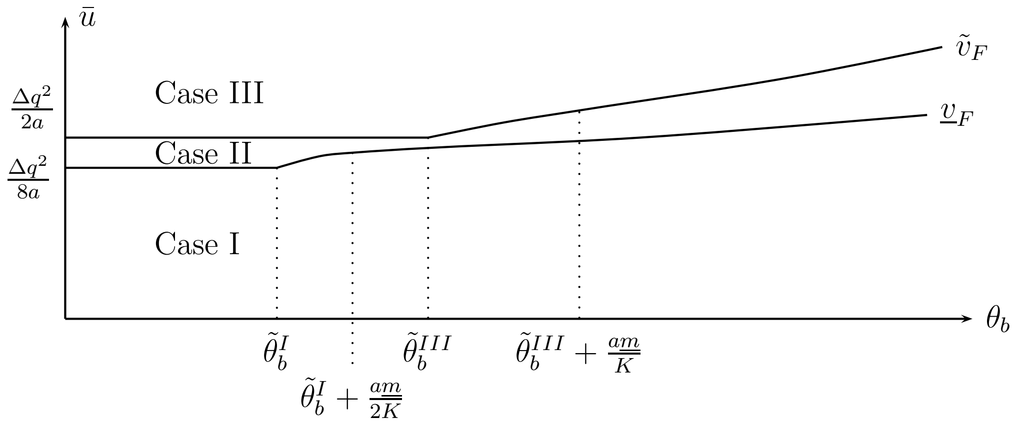


Figure 10: Relevant cases in F depending on $\bar{u}$ and θ_b.

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

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

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 10 illustrates the above considerations and shows, which case is relevant for which combinations of outside utility $\bar{u}$ and negative intrinsic motivation θ_b.[51]

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

Full Deterrence in the Non-Profit Sector

According to Proposition 4, N can achieve full deterrence by setting $\tilde{m}_N^{det}$ such that $u_{Nb}(d) \leq \max\{u_{Fb}(e), u_{Fb}(d), \bar{u}\}$.

[51]The 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 $M(m) = m^2/2$.

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.

References

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.

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.

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.

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.

Dixit, A. (2002): "Incentives and Organizations in the Public Sector," *Journal of Human Resources*, 37(4), 696–727.

DJI (2011): "Sexuelle Gewalt gegen Mädchen und Jungen in Institutionen - Rohdatenbericht," Discussion paper, Deutsches Jugendinstitut.

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.

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.

Greenberg, P., and J. Haley (1986): "The Role of the Compensation Structure in Enhancing Judicial Quality," *Journal of Legal Studies*, 15(2), 417–426.

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.

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.

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.

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.

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.

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.

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.