Auriol, Emmanuelle; Brilon, Stefanie

Conference Paper

The Good, the Bad, and the Ordinary: Anti-Social Behavior in Profit and Non-Profit Organizations

Proceedings of the German Development Economics Conference, Hannover 2010, No. 40

Provided in cooperation with:
Verein für Socialpolitik

Suggested citation: Auriol, Emmanuelle; Brilon, Stefanie (2010) : The Good, the Bad, and the Ordinary: Anti-Social Behavior in Profit and Non-Profit Organizations, Proceedings of the German Development Economics Conference, Hannover 2010, No. 40, http://hdl.handle.net/10419/40001

The Good, the Bad, and the Ordinary: Anti-Social Behavior in Profit and Nonprofit Organizations *

Emmanuelle Auriol Stefanie Brilon

September 2009

Abstract

Intrinsic motivation of workers may arise from different individual motives. While some workers care about the mission of an organization and derive an intrinsic benefit from advancing this mission ("good" workers), others derive pleasure from some form of destructive or anti-social behavior ("bad" workers). We show that mission-oriented organizations can take advantage of the intrinsic motivation of good workers. Compared to profit-oriented organizations, lower bonus payments and lower monitoring are necessary in order to achieve a high output. However, as soon as there are bad workers, mission-oriented organizations may become more vulnerable to their anti-social behavior than profit-oriented organizations. We analyze the optimal wage contracts and monitoring levels for both types of organization and discuss appropriate measures of ex ante candidate screening to overcome the problems caused by bad workers.

Keywords: motivated agents, non-profit, sabotage, candidate selection.

JEL Classification: D21, D23, L31.

*Emmanuelle Auriol: Toulouse School of Economics, eauriol@cict.fr. Stefanie Brilon: Max-Planck-Institute for Research on Collective Goods, Bonn, and University of Mannheim, brilon@coll.mpg.de. We are indebted to Guido Friebel, Martin Hellwig, Paul Seabright, and Konrad Stahl for helpful comments and discussions. Furthermore, we received helpful comments from Baptiste Massenot, Philipp Weinschenk and seminar participants at University of Mannheim, the Tinbergen Institute Rotterdam, and the MPI in Bonn, as well as conference participants at the ASSET Conference in Padua, the CSAE Conference in Oxford 2008, and the Congrès Annuel de l'AFSE 2009. Research funding by the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

1 Introduction

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.

Most theoretical models on intrinsic motivation suppose that it 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. Intrinsic motivation is hence treated by economists as something generally beneficial to organizations. 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. Take the following example: 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 in a remote location with little control from the outside. Circumstances such as these may also attract workers with quite different intentions (what we will call "bad" workers), as has been 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]

Unfortunately, more or less extreme examples for destructive or anti-social behavior such as this abound: For instance, the Catholic Church is quite obviously an organization that relies on the high intrinsic motivation of its workers, but, as illustrated by the recent scandals of abuse by Catholic priests in the US, has been recurrently targeted by bad workers.[2] Similarly a pyromaniac may best be able to satisfy his urge for fire, while minimizing his risk of being discovered, by working for the firefighters, with the added ad-

[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]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*".

vantage of being perceived as a hero.[3] 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.[4] A pedophile, who derives some intrinsic benefit from working in a job where he is in close contact with children, will target vulnerable children, such as refugees or orphans, simply because they are less likely to expose him. Other examples of anti-social behavior in non-profits are presented in Gibelman and Gelman (2004) who summarize some recent scandals involving US and International non-government organizations (NGOs).[5] Finally, anti-social behaviors are not the monopoly of non-profit organizations. They are also found in for-profits. For instance, a terrorist might want to work in an airport to have a privileged access to planes. Or an industrial 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.

We therefore face a situation where there are different sources of intrinsic motivation which may affect the production outcome both in the mission- and the profit-oriented sector. To capture this problem, we assume that there are three types of workers, who care for different things: 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 or society. We then consider two sectors of the economy, one profit-oriented and one mission-oriented. As in Besley and Ghatak (2005), on which our model is based, 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.[6] These organizations may attract

[3]Stambaugh and Styron (2003) give an overview over the problem of arson among firefighters and provide evidence, mostly from the United States, showing that the problem is very serious. Similar cases have been documented elsewhere, see for example `http://www.lexpress.fr/actualite/societe/pompier-pyromane-2-ans-de-prison_459032.html`, or `http://www.swiss-firefighters.ch/News-file-article-sid-3427.html`.

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

[5]"Bad" actions by NGO employees mentioned in the paper by Gibelman and Gelman (2004) include questionable fund raising practices, mismanagement, embezzlement, theft, money laundering, "personal lifestyle enhancement" and kickbacks, corruption, as well as sexual misconduct.

[6]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 discussion on corporate social responsibility as discussed, for instance,

workers who care about this specific mission and derive an intrinsic benefit from their work. They can hence offer lower extrinsic incentives and still attract motivated workers. We further generalize the approach by Besley and Ghatak (2005) by introducing "bad" workers and 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.

Given this setup, we first consider the case with only good and regular workers and find the classic result that the mission-oriented sector offers lower wages and makes less use of monitoring than the profit-oriented sector. We then introduce bad workers who derive utility from behaving in an anti-social way. It turns out that profit-oriented organizations are a priori less vulnerable to such behavior. Bad workers may behave like regular workers in the profit-oriented sector and thus be totally undistinguishable from "normal" people. By contrast, if they join the mission-oriented sector, then only in order to follow their destructive instincts. 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. However, deterrence is costly as it implies higher monitoring, and it even may become entirely ineffective for workers with very high levels of bad motivation. We therefore also consider ex ante measures of candidate selection, which may help to reduce the occurrence of anti-social behavior.

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).[7] Our paper is consistent with both views. While it takes the level of bad motivation as exogenous, it depends on the incentives

in Bénabou and Tirole (2009). For a further discussion of mission- vs. profit-oriented organizations, see also Besley and Ghatak (2005).

[7]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 to total strangers.

given by an organization whether bad workers will indeed act in an anti-social way or whether they will behave just like regular workers.

By 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.[8] Furthermore, or model is linked to the growing strand of literature on public service motivation[9] and its implications for hiring and remuneration schemes, as for example Francois (2000), Francois (2002), Prendergast (2007) and, in particular, Delfgaauw and Dur (2008). As in our paper, workers in this literature show some form of intrinsic motivation when working in a certain sector or for a particular mission.[10] For instance, Prendergast (2007) shows that intrinsically motivated agents in the public sector should be biased either against of in favor of their clients, depending on circumstances.[11] As in our paper, he finds that sometimes the wrong people will be drawn to a certain job even though otherwise the paper's focus is quite different.

Furthermore, 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. There are hence many sectors where wages are not paid conditional on performance, as for instance the civil service sector or many nonprofit organizations.[12] Nonprofits sometimes are even legally forbidden to pay incentive wages; see, for instance, the discussion in Glaeser (2002). Depending on the sector, this may have institutional reasons, as for example in the judicial sector, where economic incentives are minimized in order to guarantee high quality independent judgement (Posner, 1993). In other cases, especially in the case of development aid, performance may just be too difficult to assess due to high costs of monitoring in the field. This may lead to shirking and absenteeism as has been analyzed for example by Chaudhury,

[8] See, for example, Bénabou and Tirole (2003), Frey (1997), Murdock (2002) and Akerlof and Kranton (2005). The effects of employees' intrinsic motivation on firm performance are discussed by Kreps (1997).

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

[10] 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.

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

[12] See also Borzaga and Tortia (2006) and Ballou and Weisbrod (2003) for empirical studies on the incentives in for-profit and different forms of nonprofit organizations.

Hammer, Kremer, Muralidharan, and Rogers (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 behaviors, 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.

The following section outlines the basic model with only good and regular workers. We then introduce bad workers in Section 3 and show how the optimal contracts have to change. Section 4 discusses the ex ante selection of job candidates, and Section 5 concludes.

2 Basic Setup

There are two sectors $i = F, N$, where F stands for for-profit or profit-oriented and N for nonprofit or mission-oriented. Furthermore, there are three types of agents $j = g, r, b$, where g stands for good, r for regular and b for bad workers, with shares $x_g + x_r + x_b = 1$ in the population. 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, an additional output Δq with probability e. His effort cost is $c(e) = a \cdot e^2/2$. 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 $M(m_i)$. 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 $\underline{m} > 0$. As will become clear later on, in most cases the principal will want to set the monitoring level as low as possible. This result is similar to Becker (1968). For the sake of clarity we therefore introduce a minimum monitoring level $\underline{m}$. 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.

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

$$\begin{aligned}
&(LL) \qquad w_{ij} \geq \underline{w} \ , &(2)\\
&(PC) \qquad u_{ij} = w_{ij} + (m_i t_{ij} + \theta_{ij}) e_{ij} - a e_{ij}^2/2 \geq \bar{u}_j \ , &(3)\\
&(IC) \qquad e_{ij} = \arg\max_{e \in [0,1]} \left\{ w_{ij} + (m_i t_{ij} + \theta_{ij}) e_{ij} - a e_{ij}^2/2 \right\} . &(4)
\end{aligned}$$

It follows immediately from the incentive constraint (4) that the agent will choose his optimal effort level as $e_{ij} = (m_i t_{ij} + \theta_{ij})/a$. We assume that a is sufficiently large to make sure that we get an interior solution $e_{ij} \leq 1$:

Assumption 1 $a \geq \Delta q + \max\{\theta_g, \theta_b\}$.

Under Assumption 1, we can 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

$$\begin{aligned}
&(LL) \qquad w_{ij} \geq \underline{w} \ ,\\
&(PC) \qquad u_{ij} = (m_i t_{ij} + \theta_{ij})^2/(2a) + w_{ij} \geq \bar{u}_j \ .
\end{aligned}$$

We aim to study cases where in the absence of intrinsic motivation, inducing effort has some value to the principal. This requires that the cost of monitoring is not too high compared to the benefit: $\frac{1}{4a}\Delta q^2 \geq M(\underline{m})$. Moreover, we concentrate on outcomes with non-negative payoffs for the principal which is assured by $q - \underline{w} - M(\underline{m}) > 0$. The following assumption assures that the solutions derived in the paper are hence optimal:

Assumption 2 $\Delta q^2 \geq 4aM(\underline{m})$ *and* $q > \underline{w} + M(\underline{m})$.

Let us define $\underline{v}_{ij}$ as the reservation payoff level such that for $\bar{u}_j \geq \underline{v}_{ij}$ the participation constraint of agent j becomes binding and $\tilde{v}_{ij}$ as the level of outside utility 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.[13] That is,

$$\begin{aligned} \underline{v}_{ij} &\equiv \frac{1}{2a}\Big(max\{0,(\Delta q-\theta_{ij})/2\}+\theta_{ij}\Big)^2+\underline{w}\ , \\ \tilde{v}_{ij} &\equiv \frac{1}{2a}(\Delta q+\theta_{ij})^2+\underline{w}\ , \\ \bar{v}_{ij} &\equiv \frac{1}{2a}(\Delta q+\theta_{ij})^2+q-M(\underline{m})\ . \end{aligned}$$

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 $\bar{u}_j \in [0, \bar{v}_{ij}]$ exists and has the following features:*

(a) *The optimal fixed wage is*

$$w_{ij}^* = \begin{cases} \underline{w} & \text{if } \bar{u}_j \in [0, \tilde{v}_{ij}] \\ \bar{u}_j - \frac{1}{2a}(\Delta q+\theta_{ij})^2 & \text{if } \bar{u}_j \in [\tilde{v}_{ij}, \bar{v}_{ij}] \end{cases} ,$$

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

(c) *The optimal bonus payment is*

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

All proofs can be found in the Appendix.

We can thus discern three cases:

- *Case I:* The limited liability constraint is binding, but not the participation constraint of the agent. This corresponds to a case where $\underline{w}$ is relatively high compared to $\bar{u}_j$. In other words, this holds for low values of the reservation utility: $\bar{u}_j \in [0, \underline{v}_{ij}]$. The optimal contract in this case is described by $w_{ij}^* = \underline{w}$, $t_{ij}^* = \max\{0,(\Delta q-\theta_{ij})/(2\underline{m})\}$, and $m_i^* = \underline{m}$ if $t_{ij}^* > 0$ and zero otherwise.

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

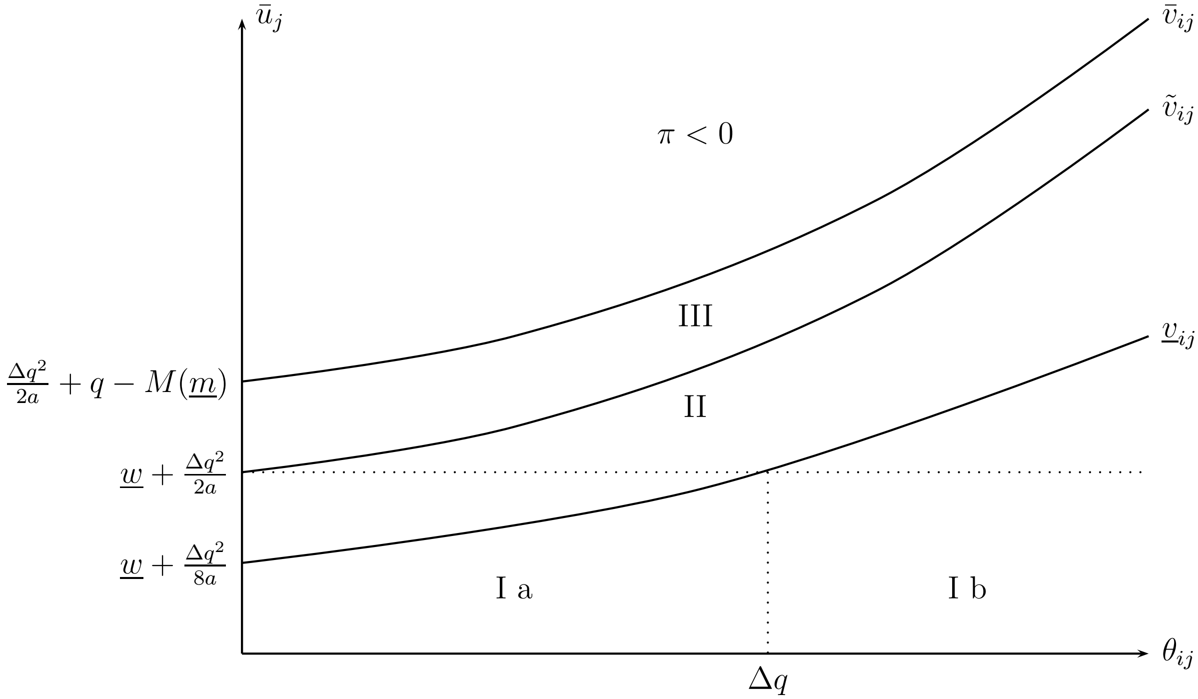


Figure 1: Optimal contract depending on $\bar{u}$ and θ_{ij}.

- *Case II:* Both the limited liability and the participation constraint are binding . This holds for intermediary values of the reservation utility: $\bar{u}_j \in [\underline{v}_{ij}, \tilde{v}_{ij}]$. The optimal contract in this case is described by $w^*_{ij} = \underline{w}$, $m^*_i t^*_{ij} = \sqrt{2a(\bar{u}_j - \underline{w})} - \theta_{ij}$, and $m^*_{ij} = \underline{m}$ if $m^*_i t^*_{ij} > 0$ and 0 otherwise.
- *Case III:* The participation constraint is binding, but not the limited liability constraint. This corresponds to a case where $\bar{u}_j$ is relatively high, i.e., for $\bar{u}_j \in [\tilde{v}_{ij}, \bar{v}_{ij}]$. The optimal contract in this case is described by $w^*_{ij} = \bar{u}_j - (\Delta q + \theta_{ij})^2/(2a)$, $m^*_i = \underline{m}$ and $t^*_{ij} = \Delta q/\underline{m}$.

Which case is relevant for the principal depends on the agent's outside option $\bar{u}_j$ and his level of intrinsic motivation θ_{ij}. Figure 1 gives an overview.

The first two cases are the cases described in Besley and Ghatak (2005). The reason why the third case is not relevant in Besley and Ghatak's (2005) model is that they do not have a basic payoff q which accrues to the principal even if the agent makes no special effort. 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., $\bar{u}_j > \underline{w} + \Delta q^2/2a$, 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 the following, we will discuss in more detail how Proposition 1 translates into an optimal contract in each of the two sectors N and F.

2.1 For-Profit Sector

Let us first consider the implications of Proposition 1 for the profit-oriented sector. 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. As a corollary from Proposition 1 we then get the following:

Corollary 1 *: Depending on the size of the agent's reservation utility, we can discern the following cases:*

- ***Case I:*** *For $\bar{u} \in [0, \underline{v}_F]$, the optimal contract in F is given by*

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

- ***Case II:*** *For $\bar{u} \in [\underline{v}_F, \tilde{v}_F]$, the optimal contract in F is given by*

$$w_F^* = \underline{w} \ , \ m_F^* = \underline{m} \ , \ t_F^* = \sqrt{2a(\bar{u} - \underline{w})}/\underline{m} \ ;$$

- ***Case III:*** *For $\bar{u} \in [\tilde{v}_F, \bar{v}_F]$, the optimal contract in F is given by*

$$w_F^* = \bar{u} - \Delta q^2/(2a) \ , \ m_F^* = \underline{m} \ , \ t_F^* = \Delta q/\underline{m} \ ;$$

where

$$\begin{aligned}
\underline{v}_F &= \Delta q^2/(8a) + \underline{w} \ , \\
\tilde{v}_F &= \Delta q^2/(2a) + \underline{w} \ , \\
\bar{v}_F &= \Delta q^2/(2a) + q - M(\underline{m}) \ .
\end{aligned}$$

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

The principal's payoff is

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

2.2 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. By offering a lower basic wage and/or lower effort incentives, N can still attract good workers (i.e., with $\theta_{Ng} \geq 0$) whereas regular workers will prefer their outside option or work in F.

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 2 *: Depending on the size of the agent's outside option, the optimal contract between a "good" agent and the principal in sector N is characterized as follows:*

- ***Case I:*** *For $\bar{u} \in [0, \underline{v}_N]$, we get two subcases:*

  (a) *If θ_g is low, i.e., if $\theta_g < \Delta q$ (case Ia in Fig. 1), then the optimal contract is given by*

$$w_N^* = \underline{w} \ , \ m_N^* = \underline{m} \ , \ t_N^* = (\Delta q - \theta_g)/(2\underline{m})$$

  (b) *If θ_g is high, i.e., if $\theta_g > \Delta q$ (case Ib in Fig. 1), then*

$$w_N^* = \underline{w} \ , \ m_N^* = 0 \ , \ t_N^* = 0$$

- ***Case II:*** *If $\bar{u} \in [\underline{v}_N, \tilde{v}_N]$, then the optimal contract is given by*

$$w_N^* = \underline{w} \ , \ t_N^* = \tfrac{1}{\underline{m}}(\sqrt{2a(\bar{u} - \underline{w})} - \theta_g),$$
$$m_N^* = \underline{m} \ , \ \textit{if } t_N^* > 0 \textit{ and } m_N^* = 0 \textit{ otherwise} .$$

- ***Case III:*** *If $\bar{u} \in [\tilde{v}_N, \bar{v}_N]$, then the optimal contract is given by*

$$w_N^* = \bar{u} - (\Delta q + \theta_g)^2/(2a) \ , \ m_N^* = \underline{m} \ , \ t_N^* = \Delta q/\underline{m} \ .$$

where

$$\begin{aligned}
\underline{v}_N &= \frac{1}{2a}\Big(max\{0, (\Delta q - \theta_g)/2\} + \theta_g\Big)^2 + \underline{w} \ , \\
\tilde{v}_N &= \frac{1}{2a}(\Delta q + \theta_g)^2 + \underline{w} \ , \\
\bar{v}_N &= \frac{1}{2a}(\Delta q + \theta_g)^2 + q - M(\underline{m}) \ .
\end{aligned}$$

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

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

which is higher or equal to what he would get in sector F. "Good" agents with low reservation utility hence prefer the contract proposed in Corollary 2 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.

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. It turns out that 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 under the above contract is always smaller than the utility level he can reach under the contract proposed in sector F.

As a result, regular workers will choose to work in sector F and "good" motivated workers will prefer to work in sector N.

The principal's profit in case I hence is

$$\pi_N = q - \underline{w} + \frac{1}{a} \begin{cases} \Delta q \theta_g & \text{if } \Delta q < \theta_g \\ \left(\frac{\Delta q + \theta_g}{2}\right)^2 - M(\underline{m}) & \text{if } \Delta q \geq \theta_g \end{cases} .$$

In case II, it is

$$\pi_N = q + \frac{1}{a}(\Delta q + \theta_g - \sqrt{(2a(\bar{u} - \underline{w})})\sqrt{(2a(\bar{u} - \underline{w})} - \underline{w} - M(\underline{m}) ,$$

and in case III

$$\pi_N = q - \left[\bar{u} - \frac{1}{2a}(\Delta q + \theta_g)^2\right] - M(\underline{m}) .$$

By exploiting the intrinsic motivation of "good" workers, the principal in N can hence save on wage and monitoring costs relative to sector F by offering lower incentives and making less use of monitoring.[14] Indeed, comparing π_N

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

with π_F in each case, it is straightforward to see that $\pi_N > \pi_F$ if $\theta_g > 0$. 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.

3 Bad Motivation

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" $d \in [0, 1]$ rather than the "normal" effort e considered so far. 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 firm they are working for. We denote the probability that a job candidate in sector $i = F, N$ is a "bad" type as β_i.

Consider the following utility function for bad workers in sector $i = F, N$:

$$u_{ib} = w_i + \begin{cases} m_i t_i e - ae^2/2 & \text{if } e \geq 0 \\ (\theta_b - m_i K)d - ad^2/2 & \text{if } d \geq 0 \end{cases},$$

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. However, as this is beyond the influence of the firm, we treat the punishment as exogenous.

Note that the worker chooses either one of the two options, i.e., he either decides to satisfy his destructive impulse and get intrinsic satisfaction from doing so ($d \geq 0$). Or he behaves like a regular worker, chooses $e \geq 0$ and aims at getting monetary rewards.

As can be seen from this utility function, bad guys may be willing to make a "good" effort if given the right incentives.

Taking into account the worker's optimal effort choice, which under Assumption 1 is still lower than 1 (i.e., there is an interior solution), we can rewrite his expected utility as

$$u_{ib} = w_i + \begin{cases} (m_i t_i)^2/(2a) & \text{if } e \geq 0 \\ (\theta_b - m_i K)^2/(2a) & \text{if } d \geq 0 \end{cases}.$$

Bad types therefore prefer to exert a positive effort rather than to follow their destructive impulse and sabotage if

$$m_i t_i \geq \theta_b - m_i K \ . \tag{5}$$

In the following, we first analyze how a bad worker's choice between sector N and F is determined before we look at the implications of this choice for the optimal contracts in each sector.

3.1 Automatic Deterrence of Bad Workers

In this section, we analyze the behavior of bad workers for a given set of contracts, namely the optimal contracts derived in the previous section. This also allows us to determine under what circumstances organizations in sectors N and F have to adapt their incentive schemes to the presence of bad workers and when there is "automatic" deterrence of anti-social behavior, i.e., without any change in the optimal contracts.

Suppose, for the moment, that the contracts in both sectors stay as calculated in Section 2, i.e., that they are not adapted to the presence of "bad" workers. How will bad workers behave under these circumstances? When will they opt for sector N, when for sector F?

To answer these questions, we need to compare a bad worker's payoff from choosing effort e or d in both sectors given the optimal contracts derived in Section 2. This comparison shows that for a given reservation utility $\bar{u}$ the incentive 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:

Corollary 3 *: Under the optimal contracts as proposed in Proposition 1, "bad" workers never join the mission-oriented sector to do good, i.e., in order to provide a positive effort e.*

That is, 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 (5) 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

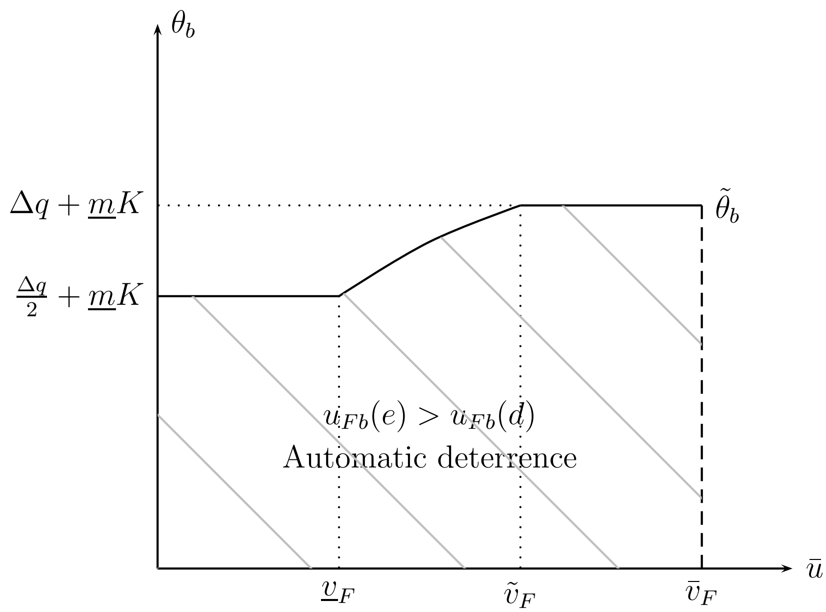


Figure 2: Automatic Deterrence in Sector F.

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

$$\tilde{\theta}_b = \begin{cases} \Delta q/2 + \underline{m}K & \text{if } \bar{u} \in [0, \underline{v}_F] \\ \sqrt{2a(\bar{u} - \underline{w})} + \underline{m}K & \text{if } \bar{u} \in [\underline{v}_F, \tilde{v}_F] \\ \Delta q + \underline{m}K & \text{if } \bar{u} \in [\tilde{v}_F, \bar{v}_F] \end{cases} , \tag{6}$$

where $\underline{v}_F, \tilde{v}_F, \bar{v}_F$ are defined as in Corollary 1. If the above holds true, 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 $\theta_b < \tilde{\theta}_b$ are therefore "automatically" deterred from anti-social behavior.

Next, let us consider what happens in the nonprofit sector N. Bad workers will be discouraged from joining this sector as long as $u_{Nb}(d) \leq u_{Fb}(e)$.[15] 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

$$w_N^* + \frac{1}{2a}(\theta_b - m_N^* K)^2 \leq w_F + \frac{1}{2a}(m_F t_F)^2 , \tag{7}$$

i.e., for all θ_b smaller than

$$\tilde{\tilde{\theta}}_b \equiv \sqrt{2a(w_F - w_N^*) + (m_F t_F)^2} + m_N^* K . \tag{8}$$

[15] That this is the relevant comparison follows from Corollary 3.

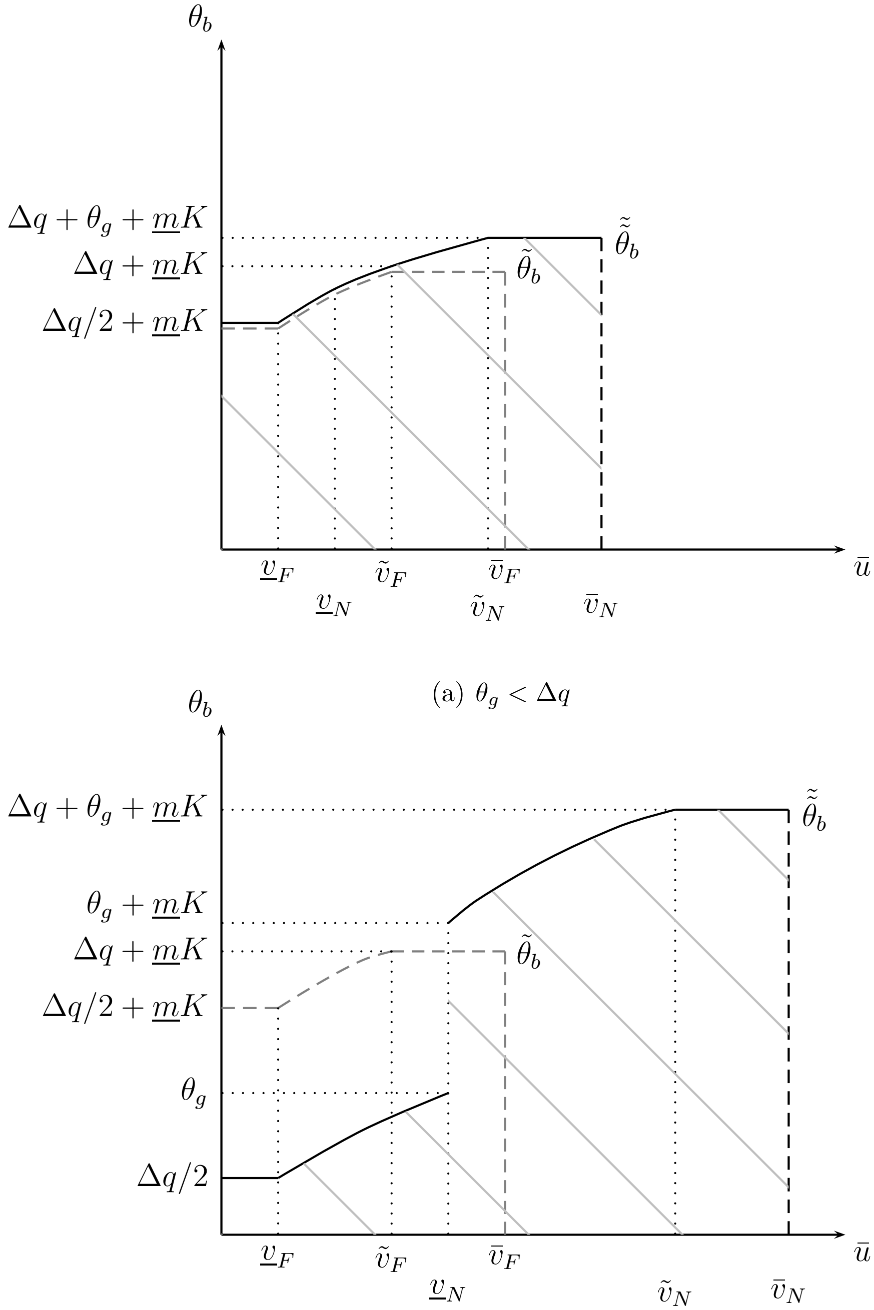
In order to determine the exact level of $\tilde{\tilde{\theta}}_b$, we then have to insert the optimal contracts in N and F into equation (8). For the sake of shortness, we will skip this exercise here. The interested reader may however find the detailed calculations in the Appendix. The results are also shown in Figure 3 which depicts the level of automatic deterrence in the nonprofit sector, $\tilde{\tilde{\theta}}_b$, 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, $\tilde{\theta}_b$, 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 ($\theta_g > \Delta q$) 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

$$u_{Nb}(d) = \underline{w} + \theta_b^2/(2a) \ ,$$



(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.

from choosing a negative effort in sector N, whereas they would get utility

$$u_{Fb}(e) = \underline{w} + \Delta q^2/(8a) ,$$

from choosing a positive effort in sector F. Therefore, all bad workers with $\theta_b > \Delta q/2$ will opt for sector N and 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 thus has provided 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.

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 there is no "automatic" deterrence.

3.2 "Bad" Workers in the Profit-Oriented Sector

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. This is the case if θ_b is below $\tilde{\theta}_b$ as defined in (6). So, next we consider what happens if θ_b is higher than $\tilde{\theta}_b$ and how the optimal contracts then should be adjusted.

To do so, we assume that organizations in the profit-oriented sector do not take into account the policy of the mission-oriented sector, whereas the latter takes policies in the former sector as given. This is equivalent to assuming that sector N is small compared to sector F, i.e., the share of good workers in the population, x_g, is small relative to the share of regular workers, x_r.

3.2.1 Full Deterrence

If the principal wants to deter bad workers all together from being destructive, his maximization problem becomes[16]

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

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

subject to

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

where the last constraint is new. This deterrence constraint ensures that bad workers prefer to make a positive rather than a destructive effort. For $\theta_b > \tilde{\theta}_b$ the deterrence constraint becomes binding and we can hence 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

$$(LL) \qquad w_F \geq \underline{w} \ , \tag{9}$$

$$(PC) \qquad (\theta_b - m_F K)^2/(2a) + w_F \geq \bar{u}_j \ . \tag{10}$$

As before, the solution of this maximization problem gives rise to three different cases, depending on the reservation utility of the workers. We define $\underline{v}_{Fb}$ as the outside utility for which the modified participation constraint as given in (10) becomes binding. Furthermore, let us define $\tilde{v}_{Fb}$ as the level of outside utility at which the limited liability constraint ceases to be binding and $\bar{v}_{Fb}$ as the highest level of outside utility at which the for-profit firm still makes a nonnegative profit. Additionally, we have to define an upper bound for the level of negative intrinsic motivation $\bar{\theta}_b$: for $\theta_b > \bar{\theta}_b$, the monitoring level is equal to one and cannot increase further.

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

Proposition 2 *: For $\tilde{\theta}_b < \theta_b < \bar{\theta}_b$, the optimal contract with full deterrence $(m_F^{det}, t_F^{det}, w_F^{det})$ in sector F given a reservation payoff $\bar{u} \in [0, \bar{v}_{Fb}]$ has the following features:*

(a) *The optimal fixed wage is*

$$w_F^{det} = \begin{cases} \underline{w} & \text{if } \bar{u} \in [0, \tilde{v}_{Fb}] \\ \bar{u} - \frac{1}{2a}(\theta_b - m_F^{det} K)^2 & \text{if } \bar{u} \in [\tilde{v}_{Fb}, \bar{v}_{Fb}] \end{cases} \ ,$$

(b) *The optimal monitoring level* m_F^{det} *is such that the following conditions hold:*

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

(c) *The optimal bonus payment is*

$$t_F^{det} = \frac{\theta_b}{m_F^{det}} - K \ .$$

Note 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 Corollary 1. In particular, $\underline{v}_{Fb} > \underline{v}_F$, and $\tilde{v}_{Fb} > \tilde{v}_F$, but $\bar{v}_{Fb} < \bar{v}_F$. For a more detailed discussion, see the proof of Proposition 2 in the Appendix.

Furthermore, Proposition 2 implies that, in order to fully deter bad workers from bad actions, the principal in sector F has to raise his monitoring level relative to the benchmark case without destructive motivation, no matter what the outside utility of the agent, i.e. $m_F^{det} > m_F^*$. Besides, he has to raise the expected bonus payment for good effort, i.e., $m_F^{det}t_F^{det} \geq m_F^* t_F^*$. As a consequence, besides deterring bad workers from bad actions, this contract will also induce regular workers to choose a higher effort level.

3.2.2 No Full Deterrence

Alternatively, the principal may accept the possibility that destructive behavior may occur. Let β_F be the share of bad workers in sector F in this case.[17] Taking into account the agent's optimal effort choice, the principal's maximization problem then corresponds to

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

subject to the worker's limited liability and participation constraint as stated in (2) and (3).

The optimal contract then takes the following form:

[17] Note that β_F is actually endogenous. It is defined as the share of bad workers in sector F. If, for example, all regular and all bad workers opt for sector F, then the share of bad workers in this sector is $\beta_F = x_b/(x_r + x_b)$.

Proposition 3 *: For $\theta_b > \tilde{\theta}_b$, the optimal contract $(m_F^{nd}, t_F^{nd}, w_F^{nd})$ in sector F given a reservation payoff $\bar{u} \in [0, \bar{v}_F]$ has the following features:*

(a) *The optimal fixed wage is*

$$w_F^{nd} = \begin{cases} \underline{w} & \text{if } \bar{u} \in [0, \tilde{v}_F] \\ \bar{u} - \frac{1}{2a}\Delta q^2 & \text{if } \bar{u} \in [\tilde{v}_F, \bar{v}_F] \end{cases} ,$$

(b) *The monitoring level m_F^{nd} is such that $M'(m_F^{nd}) = \beta_F DK/a$.*

(c) *The optimal bonus payment is $t_F^{nd} = (m_F^* t_F^*)/m_F^{nd}$, where $m_F^* t_F^*$ as defined in 1.*

Note that the incentives to provide effort e are still the same as without bad workers, i.e., $m_F^* t_F^* = m_F^{nd} t_F^{nd}$ in each of the three cases. However, the transfer level t_F^{nd} is lower than without bad workers, whereas the monitoring level m_F^{nd} has increased. While the principal may not be able to prevent bad behavior, the expected reward for such behavior thus is lower and hence the level of destructive effort chosen by the workers is lower. Therefore, the expected damage from bad behavior goes down.

Whether the principal in sector F prefers full deterrence or whether he opts for no full deterrence depends on his respective expected profit. 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

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

As can be seen easily from the second function, the expected profit without full 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 $\pi_F^{det} > \pi_F^{nd}$, i.e., that the principal in sector F will prefer to fully deter bad workers. If, for instance, the number of regular workers in the population x_r is very high, this implies that the relative share of bad workers in sector F, β_F, is low and full deterrence hence is less attractive. Furthermore, the monitoring technology plays a role. If the marginal cost of an increased level of monitoring is high, then full deterrence may be too costly.

3.3 "Bad" Workers in the Mission-Oriented Sector

For low levels of "bad" motivation θ_b, the non-profit sector is protected from destructive behavior by the higher effort incentives offered in the profit-oriented sector, i.e., bad workers' utility from choosing a normal effort e in sector F is higher than their utility from choosing a destructive effort d in sector N. This is true as long as

$$w_F + \frac{1}{2a}(m_F t_F)^2 \geq w_N + \frac{1}{2a}(\theta_b - m_N K)^2 \ .$$

The principal in the mission-oriented sector therefore does not need to adapt his optimal wage policies (w_N^*, m_N^*, t_N^*) as defined in Corollary 2 as long as

$$\theta_b \leq \tilde{\tilde{\theta}}_b \equiv \sqrt{2a(w_F - w_N^*) + (m_F t_F)^2} + m_N^* K \ ,$$

given a contract (w_F, m_F, t_F) in sector F.

If the level of motivation of bad workers is higher, i.e., if $\theta_b > \tilde{\tilde{\theta}}_b$, then the principal in sector N will have to increase his monitoring level to deter bad workers from choosing a destructive effort. However, the principal cannot increase his monitoring level beyond $m_N = 1$. To account for this fact and make sure that $m_N \leq 1$, the following assumption is sufficient:

Assumption 3 $\theta_b \leq K + \Delta q/2$.

The effort incentives for good workers, however, need not be affected by this change in the intensity of monitoring: in order to induce good workers to provide effort, nothing more than the optimal incentives as described in Corollary 2 are needed.

Therefore the following proposition holds:

Proposition 4 *: For $\tilde{\tilde{\theta}}_b < \theta_b \leq K + \Delta q/2$, and given a contract (m_F, t_F, w_F) in sector F and a reservation payoff $\bar{u} \in [0, \bar{v}_F]$, 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 $w_N^{det} = w_N^*$ where w_N^* as defined in Corollary 2.*

(b) *The monitoring level is*

$$m_N^{det} = (\theta_b - \sqrt{2a(w_F - w_N) + (m_F t_F)^2})/K.$$

(c) *The bonus payment is* $t_N^{det} = m_N^* t_N^* / m_N^{det}$*, with* $m_N^* t_N^*$ *as defined in Corollary 2.*

Suppose there is full deterrence in sector F. For $\bar{u} \in [0, \tilde{v}_F]$, i.e., if the basic wage is the same in both sectors, N can achieve full deterrence of bad workers by choosing the same monitoring level in N as in F. If $\bar{u} > \tilde{v}_F$, then the basic wage in F is higher than in N, hence making work in N less attractive for bad workers. This gives some additional protection to sector N and hence allows principals in this sector to achieve full deterrence of bad workers with a monitoring level slightly lower than the one used in sector F, albeit still higher than the optimal monitoring level without bad workers.

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 (case Ib in the above), 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, there is no need for sector N to adapt its incentives otherwise, i.e., the optimal basic wage stays the same as before, and overall incentives will still be equal to $m_N^* t_N^*$. Even with full deterrence of bad workers, the profit in sector N therefore is still higher than in sector F.

The expected profit in N under full deterrence in both sectors is

$$\pi_N^{det} = q + (\Delta q - m_N^* t_N^*) \frac{m_N^* t_N^*}{a} - w_N^* - M(m_N^{det}) \; . \tag{11}$$

Under which circumstances will there be no full deterrence in sector N? If the principal in N chooses no full deterrence, his maximization problem is

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

subject to the limited liability and participation constraints of the workers as given by (2) and (3). Similar as in sector F, his optimal monitoring level then is $m_N^{nd} = \beta_N DK/a$, while the basic wage and the expected bonus payment stay the same as without bad workers, i.e., $w_N^{nd} = w_N^*$ and $t_N^{nd} = (m_N^* t_N^*)/m_N^{nd}$.

The expected profit in N without full deterrence therefore is

$$\begin{aligned} \pi_N^{nd} \;\; &= \;\; (1 - \beta_N)(\Delta q - m_N^* t_N^*) \frac{m_N^* t_N^*}{a} - \frac{\beta_N D \theta_b}{a} + \Big(\frac{\beta_N DK}{a}\Big)^2 \\ &\quad + q - w_N^* - M\Big(\frac{\beta_N DK}{a}\Big) \; . \end{aligned} \tag{12}$$

Comparing (12) and (11), we find that no full deterrence is the better strategy in sector N if

$$\frac{\beta_N}{a}\Big[(\Delta q - m_N^* t_N^*) m_N^* t_N^* + D\theta_b\Big] - (m_N^{nd})^2 < M(m_N^{det}) - M(m_N^{nd}) \ .$$

i.e., if the share of bad workers in sector N, β_N, is lower and if additional monitoring is very costly.

Proposition 5 *: If there is full deterrence in F, then full deterrence in N is optimal.*

The reasons for this statement are straightforward: We know that F will only prefer full deterrence if the expected damage from bad workers is large enough, i.e., in particular if β_F, the share of bad workers in F without full deterrence, is high.

Note that β_F is equal to the number of bad workers over all workers in sector F, i.e., it is equal to $x_b/(x_b + x_r)$ if all bad workers (share x_b in the overall population) and all regular (share x_r in the overall population) workers work in F. Similarly, $\beta_N = x_b/(x_g + x_b)$ if all bad workers choose to work in N, which also attracts all good workers (share x_g in the overall population).[18]

Since we assumed that $x_r > x_g$, $x_b/(x_b + x_r) < x_b/(x_b + x_g)$. Hence if full deterrence is optimal in F, then it must also be optimal in N since the otherwise expected damage in N is even higher than in F. Also, as we have seen above, the costs of full deterrence in N are lower than those in F.

If there is no full deterrence in F, whether N opts for full deterrence or not depends on the share of bad workers in the overall population, x_b, and the cost of increased monitoring. When the expected damage from bad workers is sufficiently low or if a high level of monitoring is too costly, there will be no full deterrence in N. However, N is more affected by the presence of bad workers since $x_g < x_r$ and hence may opt for full deterrence even if there is no full deterrence in F.

4 Ex Ante Control

The last section 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

[18]If all bad workers prefer sector N over F, then $\beta_N = 0$.

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 commonly observed especially in sectors where candidates, once hired, are difficult to fire, as for example civil servants,[19] or where the stakes are high, e.g., 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:[20] "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",[21] 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.[22] However, checking each applicant is costly, and therefore has to be seen in relation to the potential damage of hiring a bad

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

[20]See `https://www.cia.gov/careers/faq/index.html#a3`

[21]See `http://www.fbijobs.gov/61.asp#3`

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

worker.

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 their official police record ("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 record" ("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.[23]

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[24] 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,

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

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

and Owan (2003) have found that team incentives and mutual monitoring may indeed have positive effects on workers' effort.

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,[25] or providing joint housing for aid workers.[26] 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.

5 Conclusion

We have shown how the existence of "destructive" workers who derive satisfaction from actions that are detrimental to their employer or others may affect the optimal wage contracts offered. In particular, we discussed how this may affect nonprofit organizations that rely at least to some extent on the intrinsic motivation of their workers but may be unable to filter out workers with a "negative" motivation.

First of all, we showed that 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: we showed that if bad workers join sector N then only to follow their destructive instincts and not because they want to provide a positive effort.

In order to reduce the negative impact of bad workers, both the profit- and the mission-oriented sector have to increase their monitoring levels. We showed

[25]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.

[26]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.

that to achieve full deterrence of bad workers, the profit-oriented sector may even have to increase effort incentives beyond the optimal level as described in Section 2, i.e., not only monitoring has to be increased, but also the expected return for effort $m_F t_F$ is higher. By contrast, 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 m_N increases, the bonus payment t_N decreases such that the overall effort incentives are still at their optimal level $m_N^* t_N^*$. The mission-oriented sector therefore still may enjoy a certain cost advantage, since it is cheaper to get already motivated workers to provide effort.

However, increased monitoring of workers may be difficult and expensive under many circumstances, thus requiring firms to make a better ex ante candidate selection.

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.

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.[27] Taking into account such effects would mean that the more the mission-oriented sector N increases monitoring in order to prevent damage from bad workers, the lower would be the intrinsic motivation of good workers. N would therefore also have to increase his monetary 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

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

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 therefore be not too much affected.

6 Appendix

6.1 Proof of Proposition 1

The principal in sector i who wants to hire an agent j faces the following maximization problem:

$$\max_{w_{ij},t_{ij},m_i} \pi_{ij} = q + (\Delta q - m_i t_{ij})(m_i t_{ij} + \theta_{ij})\frac{1}{a} - w_{ij} - M(m_i) \ ,$$

subject to

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

Let λ_{LL} and λ_{PC} be the respective Lagrange multipliers of the two constraints for the modified optimization problem stated above. The resulting Lagrangian is

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

and the corresponding first-order conditions are

$$
\begin{aligned}
\frac{\partial L}{\partial w_{ij}} &= -1+\lambda_{LL}+\lambda_{PC}\leq 0\ , && (13)\\
\frac{\partial L}{\partial t_{ij}} &= \frac{m_i}{a}[\Delta q-2m_it_{ij}-\theta_{ij}+\lambda_{PC}(m_it_{ij}+\theta_{ij})]\leq 0\ , && (14)\\
\frac{\partial L}{\partial m_i} &= \frac{t_{ij}}{a}[\Delta q-2m_it_{ij}-\theta_{ij}+\lambda_{PC}(m_it_{ij}+\theta_{ij})]-M'(m_i)\leq 0\ , && (15)\\
\frac{\partial L}{\partial \lambda_{LL}} &= w_{ij}-\underline{w}\geq 0\ , && (16)\\
\frac{\partial L}{\partial \lambda_{PC}} &= w_{ij}+(m_it_{ij}+\theta_{ij})^2/(2a)-\bar{u}_j\geq 0\ , && (17)\\
0 &= \lambda_{LL}(w_{ij}-\underline{w})\ , && (18)\\
0 &= \lambda_{PC}(w_{ij}+(m_it_{ij}+\theta_{ij})^2/(2a)-\bar{u}_j)\ , && (19)
\end{aligned}
$$

From (13) 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 $\lambda_{LL}=\lambda_{PC}=0$, (13) implies that the profit of the principal could be increased by reducing w_{ij} to its minimum legal level $\underline{w}$, a contradiction with $\lambda_{LL}=0$.

Furthermore, if (14) is binding, then (15) 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 $w_{ij}=\underline{w}$. 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 (14) it follows that $m_it_{ij}=\max\{0,(\Delta q-\theta_{ij})/2\}$ is optimal.

The principal's payoff then is

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

and the agent's payoff is

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

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

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

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 $w_{ij} = \underline{w}$. If the (PC) is also binding ($\lambda_{PC} > 0$), then from (17) follows that $m_i t_{ij} = \sqrt{2a(\bar{u}_j - \underline{w})} - \theta_{ij}$ is optimal. For this to be a solution, it is necessary that $m_i t_{ij} \geq 0$ which is equivalent to $\bar{u}_j \geq \underline{w} + \frac{\theta_{ij}^2}{2a}$. The agent's payoff is by construction

$$u_{ij} = \bar{u}_j \ .$$

The principal's payoff is

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

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 (PC) constraint is binding ($\lambda_{PC} > 0$), then $w_{ij} = \bar{u}_j - (m_i t_{ij} + \theta_{ij})^2/(2a)$.

If the (LL) constraint is not binding ($\lambda_{LL} = 0$), then $w_{ij} > \underline{w}$. This implies in (13) an interior solution so that $\lambda_{PC} = 1$. We deduce that if $m_i = \underline{m} > 0$, by (14), 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 $\bar{u}_j - (\Delta q + \theta_{ij})^2/(2a) > \underline{w} > 0$. That is, Case III is only relevant for agents with a reservation utility above

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

The principal's payoff then is

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

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

$$u_{ij} = \bar{u}_j \ .$$

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

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

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 $\tilde{v}(\theta_{ij})$.

This means that Case III is relevant when the agent's reservation utility is $\bar{u}_j \in [\tilde{v}(\theta_{ij}), \bar{v}(\theta_{ij})]$, that case II is relevant when the agent's reservation utility is $\bar{u}_j \in [\underline{v}(\theta_{ij}), \tilde{v}(\theta_{ij})]$, and that 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, $q - \underline{w} - M(\underline{m}) > 0$ is a sufficient assumption. It also ensures that $\underline{v}(\theta_{ij}) \leq \tilde{v}(\theta_{ij}) \leq \bar{v}(\theta_{ij})$. QED

6.2 Calculating Automatic Deterrence in N

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

Let us first consider the case where $\theta_g > \Delta q$. Depending on the level of reservation utility of the agents, Figure 4 indicates which of the cases derived in Corollaries 1 and 2 is relevant in each sector and summarizes the resulting utility levels $u_{Nb}(d)$ and $u_{Fb}(e)$ that can be achieved by bad workers. We then have to compare each possible combination of utility levels in order to determine the relevant level of automatic deterrence. For instance, Case Ib

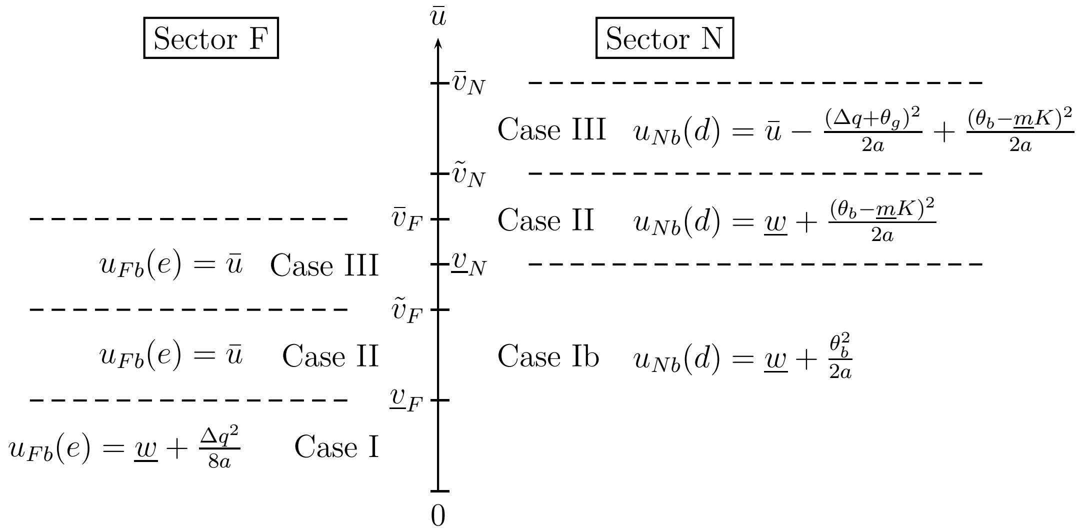


Figure 4: Bad workers' utility from positive effort in F (i.e., $u_{Fb}(e)$) and negative effort in N (i.e., $u_{Nb}(d)$) for $\theta_g > \Delta q$.

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 (7), we find that $\tilde{\tilde{\theta}}_b = \Delta q/2$ if $\bar{u} < \underline{v}_F$ and $\tilde{\tilde{\theta}}_b = \sqrt{2a(\bar{u} - \underline{w})}$ 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 $\theta_g < \Delta q$.

6.3 Proof of Proposition 2

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:

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

The corresponding first-order conditions are

$$\frac{\partial L}{\partial w_F} = -1 + \lambda_{LL} + \lambda_{PC} = 0 \, , \tag{20}$$

$$\frac{\partial L}{\partial m_F} = \frac{K}{a}[2\theta_b - 2m_F K - \Delta q] - M'(m_F) + \lambda_{PC}\frac{K}{a}(\theta_b - m_F K) \, . \tag{21}$$

Furthermore it has to hold that

$$0 = \lambda_{LL}(w_F - \underline{w}) \tag{22}$$

$$0 = \lambda_{PC}(w_F + (\theta_b - m_F K)^2/(2a) - \bar{u}_j) \ . \tag{23}$$

As before, we get three cases:

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

If the limited liability constraint is binding, $\lambda_{LL} > 0$, and given condition (22) it follows immediately that the optimal basic wage in case I $w^I = \underline{w}$. If the (PC) is not binding, then $\lambda_{PC} = 0$. Hence, from condition (21) it follows that the optimal monitoring level m_F^I has to be such that

$$2\theta_b - \Delta q = \frac{a}{K} M'(m_F) + 2m_F K \ .$$

Case II: (LL) and (PC) binding

If both conditions are binding, then $\lambda_{LL} > 0$ and $\lambda_{PC} > 0$. Again, by condition (22) we therefore have that the optimal wage in case II $w^{II} = \underline{w}$. Furthermore, condition (23) is fulfilled iff

$$m_F^{II} = \frac{\theta_b - \sqrt{2a(\bar{u} - \underline{w})}}{K} \ .$$

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

Since the limited liability constraint is not binding, $\lambda_{LL} = 0$ and hence by (20) $\lambda_{PC} = 1$. Inserting this in (21), we get that the monitoring level in case III m_F^{III} has to be such that the following holds:

$$\theta_b - \Delta q = \frac{a}{K} M'(m_F) + m_F K \ .$$

Furthermore, since the participation constraint is binding the optimal basic wage is

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

In all three cases $k = I, II, III$ the optimal transfer level t_F^k is calculated as

$$t_F^k = \frac{\theta_b - m_F^k K}{\underline{m}} \ .$$

This is due to the fact that the deterrence constraint $m_F t_F = \theta_b - m_F K$ is binding.

Having calculated these three solutions, the question is, when each of them is relevant, i.e., we have to calculate the critical values of the agent's outside utility delimiting the above three cases $\underline{v}_{Fb}$, $\tilde{v}_{Fb}$ and $\bar{v}_{Fb}$.

Let us start with $\underline{v}_{Fb}$ which is defined as the outside utility for which the participation constraint of the agents becomes binding. That is

$$(\theta_b - m_F^I K)^2/(2a) + w_F^I = \bar{u}_j \ ,$$

has to hold, and hence $\underline{v}_{Fb}$ is defined as

$$\underline{v}_{Fb} \equiv (\theta_b - m_F^I K)^2/(2a) + \underline{w} \ .$$

Recall that $\underline{v}_F = (m_F^* t_F^*)^2/(2a) + \underline{w}$ and that $(\theta_b - m_F^I K) > m_F^* t_F^*$ since $\theta_b > \tilde{\theta}_b$. Therefore $\underline{v}_{Fb} > \underline{v}_F$.

Next, let us consider $\tilde{v}_{Fb}$, which defines the border between case II and III. Case III is only relevant if $\bar{u}_j - (\theta_b - m_F^{III} K)^2/(2a) > \underline{w} > 0$. That is, Case III is only relevant for agents with a reservation utility above

$$\tilde{v}_{Fb} \equiv \frac{1}{2a}(\theta_b - m_F^{III} K)^2 + \underline{w} \ .$$

For outside values above this one, case III holds. Note that the limited liability constraint is trivially fulfilled if $\bar{u} > \tilde{v}_{Fb}$. Again, since we consider only cases where $\theta_b > \tilde{\theta}_b$ and hence $(\theta_b - m_F^{III} K) > m_F^* t_F^*$, we get that $\tilde{v}_{Fb} > \tilde{v}_F$.

Finally, $\bar{v}_{Fb}$ is defined as the outside utility of the agent for which the principal's profit in case III becomes zero, i.e., for which $\pi_F^{III} = 0$. That is:

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

where m_F^{III} is such that

$$\theta_b - \Delta q = \frac{a}{K} M'(m_F) + m_F K \ .$$

The derivative of $\bar{v}_{Fb}$ with respect to θ_b is given as

$$\begin{aligned} \frac{d\bar{v}_{Fb}}{d\theta_b} \ &= \ \frac{\partial \bar{v}_{Fb}}{\partial \theta_b} + \frac{\partial \bar{v}_{Fb}}{\partial m_F^{III}} \cdot \frac{\partial m_F^{III}}{\partial \theta_b} \\ &= \ \frac{1}{a}(\Delta q - \theta_b + m_F^{III}) \\ &- \ \left[\frac{K}{a}(\Delta q - \theta_b + m_F^{III} K) + M'(m_F^{III})\right] \frac{1}{K + M''(m_F^{III})} \ . \end{aligned}$$

This expression is smaller than zero if

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

We assumed that $M'(m) > 0$ and $M''(m) > 0$. Since we consider only cases where $\theta_b > \tilde{\theta}_b$ 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 $\bar{v}_{Fb}$ is decreasing in θ_b.

How high is $\bar{v}_{Fb}$ relative to $\bar{v}_F$? Recall that

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

Hence $\bar{v}_{Fb} > \bar{v}_F$ if

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

As we have seen above, $\bar{v}_{Fb}$ is decreasing in θ_b, and the lowest value of θ_b for which case III is actually relevant is $\theta_b = \tilde{\theta}_b = \Delta q + \underline{m}K$. If we plug this into the above inequality, after some simplification we find that $\bar{v}_{Fb} > \bar{v}_F$ if

$$-M(m_F^{III}) + M(\underline{m}) - \frac{K^2}{2a}(m_F^{III} - \underline{m})^2 > 0\ .$$

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 by contradiction that $\bar{v}_{Fb} < \bar{v}_F$ must hold.

6.4 Proof of Proposition 3

The solution of the principal's maximization problem when there are bad workers follows the analysis in the benchmark model when there are only good and regular workers. As we have seen, the principal's maximization problem without full deterrence corresponds to

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

subject to the following constraints

$$\begin{array}{ll} (LL) & w_F \geq \underline{w} \ , \\ (PC) & (m_F t_F)^2/(2a) + w_F \geq \bar{u}_j \ . \end{array}$$

Let again λ_{LL} and λ_{PC} be the respective Lagrange multipliers of the two constraints for the modified optimization problem stated above. The resulting Lagrangian is

$$\begin{aligned} \max_{w_F, m_F, t_F, \lambda_{LL}, \lambda_{PC}} \quad & L(w_F, m_F, t_F, \lambda_{LL}, \lambda_{PC}) = q - w_F - M(m_F) \\ + \quad & (1-\beta_F)(\Delta q - m_F t_F)\frac{m_F t_F}{a} - \beta_F D(\theta_b - m_F K)\frac{1}{a} \\ + \quad & \lambda_{LL}(w_F - \underline{w}) + \lambda_{PC}(w_F + (m_F t_F)^2/(2a) - \bar{u}_j) \ , \end{aligned}$$

and the corresponding first-order conditions are

$$\frac{\partial L}{\partial w_F} = -1 + \lambda_{LL} + \lambda_{PC} = 0 \ , \tag{24}$$

$$\frac{\partial L}{\partial t_F} = \frac{m_F}{a}[(1-\beta_F)(\Delta q - 2m_F t_F) - \lambda_{PC} m_F t_F] = 0 \ , \tag{25}$$

$$\begin{aligned} \frac{\partial L}{\partial m_F} &= \frac{t_F}{a}[(1-\beta_F)(\Delta q - 2m_F t_F) - \lambda_{PC} m_F t_F] \\ & \quad -M'(m_F) + \frac{\beta_F D K}{a} = 0 \ . \end{aligned} \tag{26}$$

Furthermore, the following has to be true:

$$0 = \lambda_{LL}(w_F - \underline{w}) \ , \tag{27}$$

$$0 = \lambda_{PC}(w_F + (m_F t_F)^2/(2a) - \bar{u}_j) \ . \tag{28}$$

Equation (25), i.e. ,the first-order condition with respect to t_F, is fulfilled if the expression in square brackets is equal to zero. This implies that (26), the first-order condition with respect to m_F, simplifies to

$$-M'(m_F) + \frac{\beta_F D K}{a} = 0 \ ,$$

and hence the optimal level of monitoring without full deterrence of bad workers is m_F^{nd} is such that $M'(m_F^{nd}) = \beta_F DK/a$.

Note that no other change in extrinsic incentives is needed in order to account for the presence of bad workers. In particular, effort incentives for regular workers can stay at the same level. The further solution of the problem hence runs along the same lines as the proof of Proposition 1, except that the optimal monetary transfer level t_F^{nd} is adapted such that the overall incentives are still the same, i.e., that $m_F^* t_F^* = m_F^{nd} t_F^{nd}$.

References

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.

——— (2009): "Individual and Corporate Social Responsibility," Discussion paper, IZA Discussion Paper 4570.

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.

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.

Francois, P. (2000): "'Public service motivation' as an argument for government provision," *Journal of Public Economics*, 78, 275–299.

——— (2002): "Not-for-profit Provision of Public Services," *CMPO Working Paper Series*, No. 03/060.

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," Working Paper 5535, National Bureau of Economic Research.

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 No. 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.

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.

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.