Black Sheep or Scapegoats?
Implementable Monitoring Policies under Unobservable Levels of Misbehavior*

Berno Buechel[†] & Gerd Muehlheusser[‡]

June 24, 2016

Abstract

An authority delegates a monitoring task to an agent. Thereby, it can only observe the number of detected offenders, but neither the monitoring intensity chosen by the agent nor the resulting level of misbehavior. We provide a necessary and sufficient condition for the implementability of monitoring policies. When several monitoring intensities lead to an observationally identical outcome, only the minimum of these is implementable, which can lead to under-enforcement. A comparative statics analysis reveals that increasing the punishment can undermine deterrence, since the maximal implementable monitoring intensity decreases. When the agent is strongly intrinsically motivated to curb crime, our results are mirrored and only high monitoring intensities are implementable. Then, higher monetary rewards for detections lead to a lower monitoring intensity and to a higher level of misbehavior.

JEL-Code: K42, D73

Keywords: Monitoring, Deterrence, Unobservable Misbehavior, Victimless Crime, Doping, Law & Economics

[†]University of St. Gallen, Institute of Economics (FGN), Varnbüelstrasse 19, 9000 St. Gallen, Switzerland; and Liechtenstein-Institute. Email: berno.buechel@unisg.ch. Web: www.berno.info

[‡]University of Hamburg, Department of Economics, Von-Melle-Park 5, D-20146 Hamburg, Germany. CESifo and IZA. Phone: +49 40 42838 5575. Email: gerd.muehlheusser@wiso.uni-hamburg.de

1 Introduction

In many contexts of delegated monitoring, looking only at detection statistics need not be informative about the quality of monitoring. For example, suppose a division head of a large company reports to cooperate headquarters a low number of violations against some corporate code of conduct for his division (e.g., compliance with certain ethical or safety standards). Then, it is not obvious what this information reveals about the true level of misbehavior in the division. A low number of detections could either result from a strict monitoring policy leading to few offenders only, most of which are detected (*black sheep*). Instead, the monitoring policy could be lax, leading to a large number of offenders, out of which only few (*scapegoats*) are discovered. As a further example, in sports competitions it is hard to judge for outsiders what a given number of detected dopers reveals about the seriousness and intensity of anti-doping measures by the respective agencies and the virulence of doping among athletes. Further examples include victimless crimes such as tax evasion, parking violations, prostitution, trafficking or drug dealing, where the number of detected offenders might not be too informative about the prevalence of an illegal activity.

The common feature of these examples is that an authority delegates the task of monitoring a population of individuals to an agent. Thereby, it is an outsider in the sense that it can neither observe the monitoring intensity chosen by the agent nor the resulting level of misbehavior.[1] In contrast, the potential offenders have a good assessment of the probability of being detected, which is a standard assumption in the economic literature on enforcement (see e.g., the survey by Polinsky and Shavell, 2007).[2] In this paper,

---

[1]The feature that several monitoring intensities lead to the same number of detections also applies to many further settings such as education or loan audits. However, as discussed in section 8 below, it is less clear in these contexts that the authority can be considered as an outsider who has to rely on the number of reported detections only.

[2]A similar informational structure is also considered in the model of Arlen (1994) in the context of corporate criminal liability. Moreover, as for the case of street prostitution, regular market participants might (correctly) perceive the actual threat of being arrested by the police (let alone conviction) to be
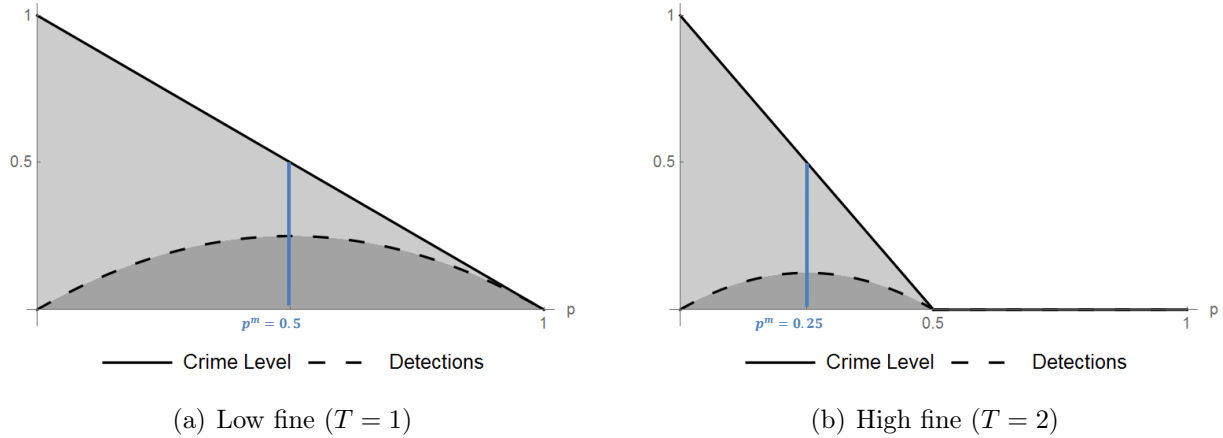
(a) Low fine $(T = 1)$ (b) High fine $(T = 2)$

Figure 1: Crime level and number of detections as a function of the detection probability p for different fine levels.

we develop a simple model which captures the interaction between the authority, the monitoring agent and potential offenders, and which builds on the previous literature on private law enforcement with a monopolistic enforcer (see e.g., Becker and Stigler, 1974; Landes and Posner, 1975; Polinsky, 1980; Besanko and Spulber, 1989; Garoupa and Klerman, 2002; Coşgel, Etkes, and Miceli, 2011).

Some key features of our model can be foreshadowed with the following simple example. The enforcement effort of the monitoring agent (the *inspector*) translates into p, the probability that an offense gets detected and punished. The number of offenses committed is decreasing in the expected punishment and, hence, decreasing in p for a given fine (T). For instance, the two panels of Figure 1 depict the number of offenses as a linear function $F(p, T) = 1 - p \cdot T$ for fine levels $T = 1$ and $T = 2$, respectively.[3] Figure 1 also depicts the number of detections $p \cdot F(p, T)$ for each case (dashed lines), which is hump-shaped and reaches its peak at $p^m(T = 1) = 0.5$ and $p^m(T = 2) = 0.25$, respectively.

The example is helpful in illustrating the main contributions of our paper compared to

_____

much smaller than might be presumed by outsiders (see e.g., Levitt and Venkatesh, 2007).

[3]As will become clear below, linear crime functions arise when the distribution of the gains from the crime is uniform.

the existing strands of literature on law enforcement and on delegated monitoring. First, we strengthen the argument already provided in Polinsky (1980) that high monitoring intensities might not be implementable because enforcers anticipate that strong deterrence reduces revenues. In particular, we provide a full characterization of implementable monitoring policies. Thus, we do not confine attention only to a particular monitoring policy which, for instance, maximizes social welfare (e.g., Polinsky, 1980) or some other objective function of the authority (Garoupa and Klerman, 2002). Intuitively, in Figure 1, for every monitoring intensity that corresponds to the decreasing part of the detection function there exists another one on the increasing part that gives rise to the same number of detected offenses, and is hence observationally identical for an outside observer. However, the inspector prefers the lower of the monitoring intensities as it involves less enforcement effort (thereby leading to more offenses), so that the largest monitoring intensity which can be implemented is p^m, where the number of detections is maximum. More generally, when several monitoring intensities give rise to the same number of detected offenses, then the agent can only be induced to choose the minimum of these.[4] For this reason, under quite general conditions (for example, with respect to the underlying distribution of individuals' gains from the offense or the agent's effort cost function), a large set of monitoring policies cannot be implemented by the authority, even if it had unlimited funds to reward the inspector. This puts a lower bound on the number of offenses which can be achieved in this setup and which can result in under-enforcement compared to an efficiency benchmark.

Second, to evaluate the scope of this result, we perform a comparative statics analysis both with respect to the distribution of gains from crime (where the previous literature has usually confined attention to the uniform case only) and the fine for detected offenders. Our results point to a novel trade-off between the severity of the punishment and the set of implementable monitoring intensities: on the one hand, when the punishment is relatively strong (for example, because there are not many individuals with sufficiently large gains

---

[4]In Ichino and Muehlheusser (2008), a low monitoring intensity results as an optimal choice of the inspector as this allows him to elicit private information from (potential) offenders.

from the offense, or because the fine is high), then the crime level tends to be low for any given monitoring intensity p. But on the other hand, also the set of implementable monitoring intensities is small. In Figure 1 this trade-off can be seen by comparing the two panels. The higher fine in panel (b) yields a lower number of offenses for any given monitoring intensity, but it also reduces the maximal implementable monitoring policy from $p^m(T = 1) = 0.5$ in panel (a) to $p^m(T = 2) = 0.25$ in panel (b). In this example, the two effects just offset each other so that the minimal number of offenses $F(p^m(T), T)$ is the same for both fine levels (i.e., $1 - 0.5 = 1 - 2 \cdot 0.25 = 0.5$), but we also show that the latter effect may dominate, so that deterrence may in fact *decrease* as the punishment becomes more severe.

This result is in contrast to the standard approach in the enforcement literature in the tradition of Becker (1968) where the two components of expected punishment – the probability of detection and the fine – can be set independently from each other. As a consequence, higher fines typically lead to more deterrence.[5] Moreover, the literature has provided various reasons against Becker's stark conclusion that fines should be set as large as possible. Examples include offenders who are risk averse or heterogeneous with respect to their wealth, offenders who engage in socially undesirable avoidance activities, costs of fine collection, or the requirement that the punishment should reflect the severity of the offense.[6] But also in these frameworks, higher fines would always lead to more deterrence. In contrast, our result suggests that in the context of delegated monitoring, even in the absence of all of these countervailing factors, optimal fines might not be

---

[5]One exception is the model of Nussim and Tabbach (2009), where in addition to the decision about their level of criminal behavior, offenders can also engage in avoidance activities. Further exceptions include frameworks of juror behavior (Andreoni, 1991; Feess and Wohlschlegel, 2009), inspection games (Tsebelis, 1990), and corruption (Kugler, Verdier, and Zenou, 2005).

[6]For a detailed discussion of these factors, see for example the survey by Polinsky and Shavell (2007) and the references cited therein.

too large because of the potentially detrimental effect on deterrence.[7] Importantly, the potentially inverse relationship between the severity of punishment and deterrence is not driven by behavioral biases or irrationality, e.g., on the side of the offenders. Finally, the novel trade-off identified in our framework might also contribute to the difficulty of the empirical literature in providing robust evidence in favor of Becker's deterrence hypothesis, apart from the well-known methodological issues (see e.g. Levitt, 1997; Di Tella and Schargrodsky, 2004; Levitt and Miles, 2007).

Third, we consider a model extension where we allow the inspector to be crime-sensitive in the sense that she directly benefits or suffers from criminal activity. In this respect, a disutility from crime can be naturally interpreted as resulting from intrinsic motivation to keep the crime level low.[8] We show first that when the agent's degree of intrinsic motivation is not too high, then, as in the baseline model, only relatively low monitoring intensities can be implemented (leading again to the same lower bound for the resulting crime level). In contrast, when the agent is strongly motivated to curb crime, a mirror results holds and only relatively high monitoring intensities can be implemented. In Figure 1, these implementable sets are given by the monitoring intensities which correspond to the increasing and decreasing part of the detection function, respectively. This gives rise to the possibility of over-enforcement, which has so far not been addressed in the existing literature on monopolistic enforcement (see e.g., Polinsky, 1980; Garoupa and Klerman, 2002; Coşgel, Etkes, and Miceli, 2011).[9] However, it might occur in cases when

---

[7]Note that this issue cannot be mitigated by replacing fines with imprisonment because our argument holds for any type of punishment.

[8]In contrast, the previous literature has only considered indirect ways to induce the agent to internalize the effect of his monitoring choice on the resulting crime level. For example, Garoupa and Klerman (2002, p.131) discuss penalties for the agent which are increasing in the number of offenses which, however, would require the latter to be observable and verifiable. Consequently, instead of imposing penalties, Coşgel, Etkes, and Miceli (2011) argue in favor of allowing the monitoring agent to also collect income taxes (which are the higher, the lower the level of crime).

[9]In Landes and Posner (1975) over-investment arises when there is a perfectly competitive market for private enforcement.

the harm from the offense is small or when the offender's cost of not committing it is large. One example in this respect would be minor parking offenses which are all deterred by an overly motivated agent. Moreover, intrinsic motivation to keep the number of offenses low yields an explanation for the phenomenon that there are (so many) inspectors who do monitor intensely, even if they would not suffer any material losses in case of shirking.

Fourth, depending on their degree of intrinsic motivation, agents react quite differently to incentive schemes, such as bounties for detected offenders, which are often analyzed in the literature (see e.g., Becker and Stigler, 1974; Landes and Posner, 1975; Polinsky, 1980; Besanko and Spulber, 1989; Garoupa and Klerman, 2002; Coşgel, Etkes, and Miceli, 2011). Intuitively, agents will generate more detections when they are rewarded for doing so independent of their intrinsic motivation. However, as the number of detections is generally non-monotonic in p (as depicted in Figure 1), it depends on whether the inspector's optimal choice is in the increasing or in the decreasing part of the detection function. For agents with weak intrinsic motivation the increasing part is relevant such that a higher per-detection reward indeed induces them to increase their monitoring effort. In contrast, optimal monitoring intensities of agents with strong intrinsic motivation are in the decreasing part of the detection function, so that they will *reduce* their effort, thereby also inducing a higher crime level. This latter result is similar to a crowding-out effect (see e.g., Deci, 1971; Deci, Koestner, and Ryan, 1999; Frey and Jegen, 2001; Gneezy, Meier, and Rey-Biel, 2011), but notice that here, monetary rewards do not directly impinge on the intrinsic motivation of the agent, for example in the sense of transforming an non-economic relationship into an economic one (Titmuss, 1970; Gneezy and Rustichini, 2000a,b). Rather, they simply introduce an incentive to generate more detections which, for agents with a strong intrinsic motivation, requires a lower monitoring effort. Our analysis suggests a beneficial role of intrinsic motivation in remedying the problem of under-enforcement, but it also reveals the crucial importance of distinguishing between different types of enforcers in order to avoid severely misguided incentives.

The remainder of the paper is organized as follows: We first set up the baseline framework in section 2, and then characterize implementable monitoring intensities (section 3),

followed by a discussion of some comparative statics properties (section 4). In section 5, we study the case of a crime-sensitive inspector. Section 6 considers linear reward schemes (bounties), while section 7 compares our results concerning implementability with an efficiency benchmark. Finally, section 8 discusses some implications from our analysis and concludes. All proofs are in the Appendix.

2   Model

There are three types of players: a population of *individuals* who are potential offenders, an *inspector* who monitors them, and an outside *governor* who incentivizes the inspector.

Individuals   There is a unit mass of individuals who differ with respect to their gains from committing an offense, g_i, which are distributed according to a twice continuously differentiable cumulative distribution function $G : \mathbb{R} \to [0, 1]$. Following the tradition of Becker (1968), for a given probability of detection $p \in [0, 1]$, and a (exogenous) penalty $T > 0$, each individual will commit the offense if and only if its gain g_i exceeds the expected costs $p \cdot T$. This yields a threshold $\overline{g} := p \cdot T$, such that all individuals satisfying $g_i > \overline{g}$ ($g_i \leq \overline{g}$) will (not) commit the offense, which leads to a fraction of offenders $F(p) := 1 - G(pT)$. We assume that the distribution of the gains from the offense has full support on $[0, T]$ such that $F(p)$ is strictly decreasing.[10]

Inspector   The inspector chooses the monitoring intensity $p \in [0, 1]$ which equals the probability each offender is detected with.[11] Monitoring is costly for the inspector, which

---

[10]Note that this assumption does not rule out the possibility that there exist individuals with $g_i < 0$ or $g_i > T$. It, however, excludes cases in which the number of offenses $F(p)$ reaches zero for monitoring intensities $p < 1$ (such as in panel (b) of Figure 1). These cases could also be included in the general model analysis, but this would only add notational inconvenience without qualitatively affecting the results.

[11]Alternatively, one could explicitly model the inspector's effort to affect the probability of detection through some (increasing) function. Under standard assumptions (e.g., Inada conditions), while adding notation, this approach would not affect our results qualitatively.

is captured by a strictly increasing cost function $C(p)$. Taking into account the optimal behavior of individuals as characterized above, a monitoring intensity p gives rise to a number of detected offenders $D(p) := p \cdot F(p)$. Denote by $\Delta \subseteq [0, 1]$ the image of $D(p)$, that is, $\Delta := \{d \mid d = D(p) \text{ for some } p \in [0, 1]\}$, and by p^m the smallest monitoring intensity for which the number of detections is maximal. In special cases, this occurs at the upper boundary ($p^m = 1$), while otherwise, p^m is characterized by the first-order condition $D'(p^m) = 0$.

We study a context where the inspector can only be rewarded on the basis of the number of detections $D(p)$, which is observable. Denoting the monetary reward by $R(D(p))$, the inspector's payoff is[12]

$$u(p) = R(D(p)) - C(p). \tag{1}$$

Governor The governor remunerates the inspector by setting a payment scheme $R(D(p))$, without being able to verify the inspector's actual behavior (p) or the crime level ($F(p)$).[13]

For the main part of the paper, it is not necessary to specify explicitly the preferences of the governor, for example, with regards to her distaste for crime. Rather, we assume that the governor aims at implementing some *desired* monitoring intensity $\hat{p} \in [0, 1]$. For instance, $\hat{p}$ could indeed be her privately optimal choice or, alternatively, it could arise from a social welfare function. This latter case will then be discussed in section 7 below.

3 Implementable Monitoring Policies

We now analyze under which circumstances the governor can successfully induce the inspector to choose $\hat{p}$, i.e. find payments R such that $\hat{p} \in \arg \max_p u(p)$. For any given

---

[12]Additive separability of rewards and costs is assumed for analytical convenience only. The assumption that the inspector's utility is not directly affected by the crime level $F(p)$ is relaxed in section 5 below.

[13]See Arlen (1994), Garoupa and Klerman (2002), and Coşgel, Etkes, and Miceli (2011) for similar assumptions concerning the governor's role as an "outsider" in the sense of lacking these crucial pieces of information.
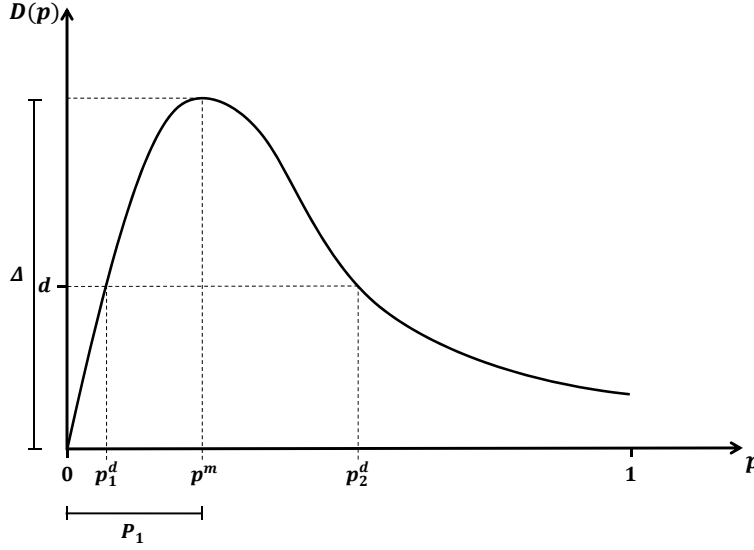
Figure 2: Number of detections as a function of monitoring intensity p

level of detections $d \in \Delta$, define an ordered set of monitoring intensities $(P^d, <)$ such that each $p_l^d \in P^d$ satisfies $D(p_l^d) = d$. Importantly, while the number of detected offenses is equal to d for all $p_l^d \in P^d$, the underlying crime level is decreasing in l, while the inspector's effort costs are increasing in l, that is, for all $l = 1, 2, \dots$ we have $F(p_l^d) \geq F(p_{l+1}^d)$ and $C(p_l^d) < C(p_{l+1}^d)$, respectively. Denote by P_1 the set containing all minimum monitoring intensities, that is, $P_1 = \{p \mid p = p_1^d \text{ for some } d \in \Delta\}$.

Theorem 1. *A desired monitoring policy $\hat{p}$ is implementable if and only if $\hat{p} \in P_1$. The resulting set of implementable monitoring policies P_1 satisfies $P_1 \subseteq [0, p^m]$ such that the induced number of offenses is at least $F(p^m)$.*

The result is illustrated in Figure 2. In special cases there is a unique monitoring intensity associated with a given number of (observable) detections.[14] Otherwise, when P^d is not a singleton, the inspector has a choice between several monitoring regimes in order to generate d detections. For example, he can either choose a low monitoring effort p_1^d (at

---

[14]In the example depicted in Figure 2, this is true for the global maximum $D(p^m)$ and when the number of detections is very small.

low cost) which leads to a relatively high number of offenders, out of which d are detected (*scapegoats*). Alternatively, the inspector can choose a higher effort $p_2^d > p_1^d$ (at higher cost) which leads to fewer offenses, but again to d detections (this time better referred to as *black sheep*). Since the governor can observe only the number of detections, but not the chosen monitoring intensity, these two effort choices are observationally identical from the governor's point of view. As the inspector's payment is the same for all $p_l^d \in P^d$, the inspector prefers to "deliver" any given number of detections $d \in \Delta$ at the lowest cost, and so his optimal choice is p_1^d. Consequently, only monitoring policies $p \in P_1$ can be implemented so that $\hat{p} \in P_1$ is a necessary condition for its implementation. As for sufficiency, all monitoring levels $\hat{p} \in P_1$ can be implemented by sufficiently rewarding the corresponding detection level $D(\hat{p})$, compared to all other detection levels $d \neq D(\hat{p})$.

In a next step, we analyze in more detail the set of implementable monitoring policies P_1. First, since Theorem 1 renders all $p > p^m$ non-implementable, the crime level is bounded from below by $F(p^m)$. Therefore, P_1's upper bound p^m becomes crucially important. The value of p^m is determined by the shape of the detection function $D(p)$, and its hump-shaped representation in Figure 2 is quite characteristic. To see this note that it always holds that $D(0) = 0$ and $D'(0) = F(0) > 0$ since without monitoring there are no detections (but a high crime level). Moreover, full monitoring typically also leads to few detections ($D(1) = 1 - G(T)$), because, given that every offense is detected, there are not many offenders (i.e., only those who are *undeterrable* as $g_i > T$). Thus, the maximal number of detections $D(p^m)$ is usually attained between these two extremes.

Second, even some $p < p^m$ might be non-implementable, in which case P_1 would be only a strict subset of the interval $[0, p^m]$. This case occurs when $D(p)$ is not monotonically increasing over this interval, such that there would exist several monitoring intensities leading to the same number of detections. By Theorem 1, only the minimum of these can be implemented. Otherwise, when $D(p)$ is monotonically increasing over $[0, p^m]$, we have $P_1 = [0, p^m]$ such that a monitoring intensity p is implementable if and only if $p \leq p^m$. The two properties discussed above – p^m interior and $D(p)$ monotonically increasing over $[0, p^m]$ – can be traced back to the distribution of gains from crime such that we get the

11

following Corollary of Theorem 1.

Corollary 1. *For the set of implementable monitoring policies P_1 the following holds:*

(i) *Let the number of undeterrable individuals be sufficiently small such that it satisfies the condition $1 - G(T) < T \cdot G'(T) (> 0)$. Then p^m is interior with the consequence that not all monitoring policies are implementable, i.e. $P_1 \subsetneq [0,1]$.*

(ii) *Let G be not too concave such that it satisfies the condition $G''(g) > -\frac{G'(g)}{g} (< 0)$ for $g \in [0,T]$. Then the number of detections $D(p)$ is monotonically increasing over $[0, p^m]$ with the consequence that a desired monitoring intensity $\hat{p}$ is implementable if and only if $\hat{p} \leq p^m$, i.e. $P_1 = [0, p^m]$.*

Corollary 1 provides two, arguably mild, conditions which are jointly sufficient for the detection function to be hump-shaped on its domain $[0,1]$. Both statements of Corollary 1 are derived from the detection function $D(p) = p \cdot (1 - G(pT))$. Writing the first-order condition $D'(p) = 0$ as

$$1 - G(pT) = p \cdot T \cdot G'(pT) \tag{2}$$

reveals the two underlying marginal effects: The term on the left-hand side captures the higher number of detections as the monitoring intensity increases (for a given crime level). The term on the right-hand side measures the marginal deterrence effect (for a given probability of detection). Condition (i) of Corollary 1 ensures that there exists an interior monitoring intensity p that satisfies Eq. (2), i.e., that balances the two marginal effects. This is achieved by simply requiring that for $p = 1$ the marginal detection effect (left-hand side) is smaller than the marginal deterrence effect (right-hand side), which is never true for $p = 0$. Since by Theorem 1 only $p \leq p^m$ are implementable, under Condition (i) there exist monitoring intensities which are not.[15] Condition (ii) of Corollary 1 ensures that the right-hand side of Eq. (2) is increasing in p, that is, the marginal deterrence increases as

---

[15]Polinsky (1980) provides further conditions which are sufficient for p^m interior such that monitoring policies close to 1 cannot be implemented.

monitoring becomes more and more intense.[16] Since the left-hand side of Eq. (2) is always decreasing in p, when condition (ii) holds, then there is at most one monitoring intensity p that solves Eq. (2). This implies that the slope of the detection function changes its sign at most once (that is, turn from increasing to decreasing since $D'(0) = F(0) > 0$). Therefore, under Condition (ii), $D(p)$ must be monotonically increasing for $p < p^m$ and, hence, all of these monitoring intensities are implementable.

Remark. *Theorem 1 can also be expressed in terms of the elasticity of crime* $\epsilon(p) := -F'(p)\frac{p}{F(p)}$. *Under Conditions* (i) *and* (ii) *of Corollary 1 it is readily derived that* $\epsilon(p) \geq 1$ *if and only if* $p \geq p^m$. *Thus, as a rule, inspectors cannot be induced to choose a monitoring regime in the elastic range of the crime function.*

4 Comparative Statics Analysis

The implications from Theorem 1 depend strongly on the model fundamentals, in particular the underlying distribution of the gains from misbehavior G and the fine T. In this section we use comparative statics analysis to assess the impact of these two factors.

4.1 Impact of the Distribution of Gains From Crime G

Different distributions of gains from crime G give rise to varying levels of misbehavior, detections and implementable monitoring intensities. We compare distributions which differ in the sense of first-order stochastic dominance (FOSD). To this end, consider two distributions G and $\tilde{G}$ where G is first-order stochastically dominated by $\tilde{G}$, i.e. $G(g) \geq \tilde{G}(g)$ for all g, with the interpretation that $\tilde{G}$ has more probability mass on high gains from crime. Denote by $\tilde{p}^m$ and $\tilde{P}_1$ the respective maximizer of the number of detections and the set of implementable monitoring policies resulting under $\tilde{G}$.

---

[16]In fact, Condition (ii) even implies that $D(p)$ is concave. Together with Condition (i), which guarantees that p^m is interior, this ensures that $D(p)$ is hump-shaped on $[0, 1]$, similar to its illustration in Figure 2.

Proposition 1. *Let G be first-order stochastically dominated by $\tilde{G}$.*

(i) *Then, for any given monitoring intensity p, the number of detections and the level of misbehavior is larger under $\tilde{G}$ than under G.*

(ii) *Let G and $\tilde{G}$ satisfy Conditions* (i) *and* (ii) *of Corollary 1. If the slopes of $\tilde{G}$ and G at $p = p^m$ are not too distinct, then the set of implementable monitoring intensities is larger under $\tilde{G}$ than under G. Formally, if*

$$\tilde{G}'(p^m T) - G'(p^m T) < \frac{G(p^m T) - \tilde{G}(p^m T)}{p^m \cdot T} (\geq 0),$$

then $\tilde{p}^m > p^m$ and hence $\tilde{P}_1 \supsetneq P_1$.

The two parts of Proposition 1 suggest that there is a trade-off in the sense that facing a population with a low tendency toward misbehavior (as exemplified by distribution G) is on the one hand beneficial, as the level of misbehavior is relatively low for any given monitoring intensity p (part (i)). But on the other hand, only a small set of (low) monitoring intensities is implementable, which in turn might still lead to relatively high levels of misbehavior (part (ii)).

How these two effects unfold is illustrated by Figure 3 for two classes of distributions (and for $T = 1$): Normal distributions $N(\mu, 0.5)$ and Power distributions with cumulative distribution functions (CDFs) $G(g) = g^\nu$ (defined for $g \in [0, 1]$). In each column, the three panels of Figure 3 depict the CDF, the detection function $D(p)$, and the crime function $F(p)$ for the different parameter values, where those distributions with higher parameter values of μ and ν, respectively, first-order stochastically dominate those with lower values.

Panels (c) and (d) illustrate the trade-off as emerging from Proposition 1 above: When there is large probability mass on low realizations of g (low values of μ and ν), the overall levels of misbehavior and detections are low. But also the value p^m where the (hump-shaped) detection function $D(p)$ reaches its peak is small, so that, from Corollary 1, the set of implementable monitoring intensities $P_1 = [0, p^m]$ is relatively small.
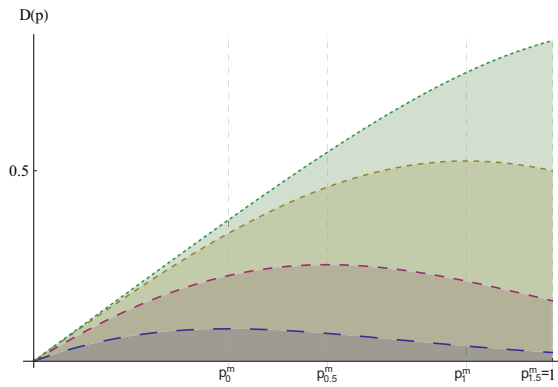
Shifting probability mass to higher realizations of g (i.e. increasing μ and ν, respectively) then leads to upward shifts of both $F(p)$ and $D(p)$ since deterrence is now weaker
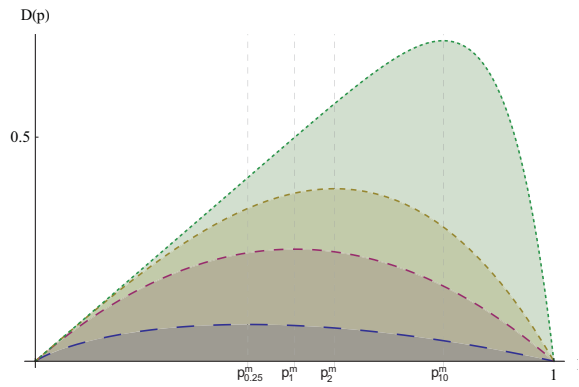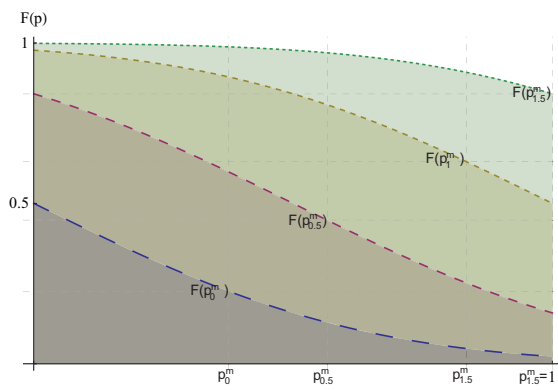
14
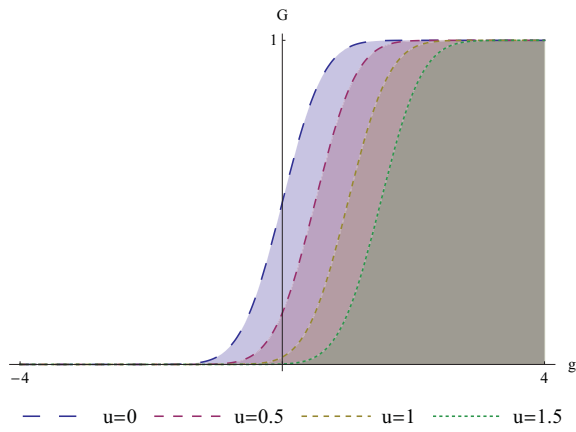
(a) Normal distribution: CDFs
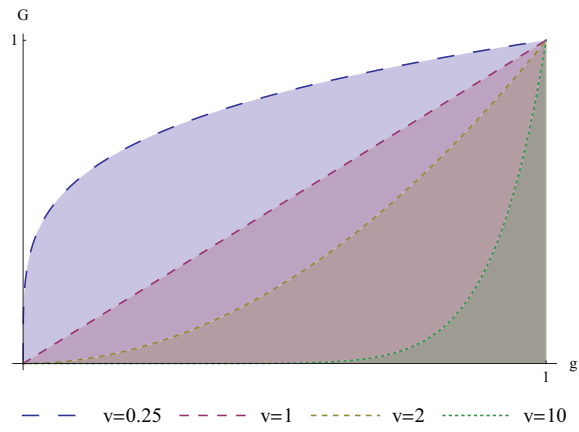
(b) Power distribution: CDFs

(c) Detections

(d) Detections

(e) Crime Level
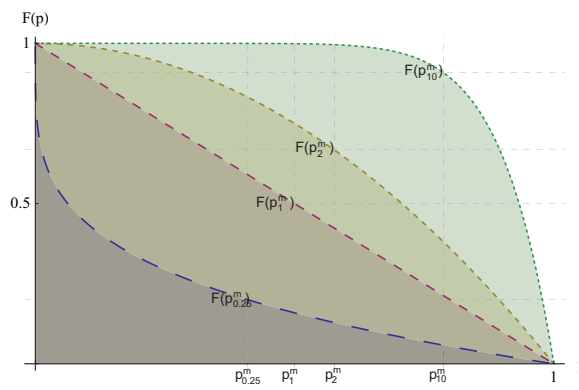
(f) Crime Level

On the left and right we depict (for $T = 1$), respectively, the case where the gains from crime are distributed according to Normal distributions $N(\mu, 0.5)$ for $\mu \in \{0, 0.5, 1, 1.5\}$, and Power distributions with CDF $G(g) = g^\nu$ defined on $[0, 1]$ for $\nu \in \{0.25, 1, 2, 10\}$. We also depict the respective values of p_μ^m and $F(p_\mu^m)$ (left), respectively, p_ν^m and $F(p_\nu^m)$ (right).

Figure 3: Illustrating Theorem 1 and Proposition 1 for different distributions

for any level of p. As a result, p^m and hence also the upper bound of the set P_1 increases.[17] When there are sufficiently many individuals with large gains from the offense, $D(p)$ may even become monotonically increasing on $[0,1]$ as in the uppermost graph in panel (c), so that $P_1 = [0,1]$, thereby coinciding with the choice set of the inspector. In this case, Condition (i) of Corollary 1 is violated, i.e. there are many undeterrable individuals, and there is no loss in terms of non-implementable monitoring intensities, so that Theorem 1 has no bite.

Interestingly, with respect to the resulting minimum crime level $F(p^m)$, the "benefit" from a larger choice set P_1 due to a larger p^m can outweigh the "cost" in the form of an upward shift of the crime function $F(p)$. For example, for the case of the Normal distribution with the monotonically increasing detection function ($\mu = 1.5$) just discussed, there are many criminals who are not deterred even when monitored with the maximum feasible intensity $p = p^m = 1$ (see panels (c) and (e)). Therefore, the minimum crime level is $F(1) = 0.84$. In contrast, for the Power distributions considered, full deterrence is in principle possible as $F(1) \equiv 0$ for all $\nu > 0$ (see panel (f)). However, only monitoring intensities $p \in [0, p^m]$ are implementable and, as can be seen from panel (d), $p^m < 1$ holds. Therefore, the minimum crime level $F(p^m)$ might well exceed the one resulting under the monotone case: For example, for $\nu = 10$ (upper graph in panel (f)), we get $p^m = 0.78 < 1$, which leads to $F(p^m) = 0.91 > 0.84$.

4.2 Varying the Fine T

To investigate the impact of changes of the exogenous fine T it is useful to now express explicitly, where appropriate, the dependency on T, i.e. we now write $F(p,T)$, $D(p,T)$ and $p^m(T)$, respectively. In analogy to the comparative statics concerning the distribution G (embodied in Proposition 1), changes in the fine T have two countervailing effects on

---

[17]For example, consider the case $\nu = 1$ in the Power distribution, which corresponds to the uniform distribution of gains over $[0,1]$. Since $p^m = \frac{1}{2}$, the crime level is at least $F(p^m) = \frac{1}{2}$. Thus, when the governor wants to deter more than half of the population (for instance, because the harm from the offense is large), then the required monitoring policy $\hat{p} > \frac{1}{2}$ cannot be implemented.

the set of implementable monitoring policies.

Proposition 2. *Consider two fines T and $\tilde{T}$ with $T < \tilde{T}$.*

(i) *For any given monitoring intensity $p > 0$, the number of detections and the level of misbehavior is smaller under $\tilde{T}$ than under T, i.e. $F(p, \tilde{T}) < F(p, T)$ and $D(p, \tilde{T}) < D(p, T)$.*

(ii) *Let G satisfy Conditions (i) and (ii) of Corollary 1. Then the set of implementable monitoring intensities is smaller under $\tilde{T}$ than under T. Formally, $\tilde{P}_1 \supsetneq P_1$ because $p^m(\tilde{T}) < p^m(T)$.*

Part (i) of Proposition 2 shows that an increase in T leads to a downward-shift of $F(p, T)$ and $D(p, T)$, thereby unambiguously lowering both the number of offenses and the number of detections. Intuitively, for every given p, a higher fine increases the expected penalty from committing the offense, and hence leads to fewer offenses and, as a result, also fewer detections.

However, analogously to the comparative statics concerning the distribution G, part (ii) points to a countervailing effect in the sense that it leads to a smaller set of implementable monitoring intensities $P_1 = [0, p^m(T)]$. This result is in contrast to the celebrated finding of Becker (1968) according to which any expected fine $p \cdot T$ should be implemented with T as large as possible, as increasing T is costless (in contrast to increasing p being costly). For the contexts considered in our paper, this suggests that there is a reduced benefit associated with raising T in the form of a shrinking set of implementable monitoring intensities. This trade-off is illustrated in Figure 4 for the case where the gains from crime are distributed according to a standard Normal distribution $N(0, 1)$ for different levels of T. As can be seen in Figure 4, low values of T give rise to a large set of implementing monitoring intensities (i.e., $p^m(T)$ is large), but also the resulting crime level (and the number of detections) is high. A higher value of T reduces the crime level, but it also reduces the set of monitoring intensities that can be implemented by the governor. Which effect dominates depends on whether $p^m(T) \cdot T$ is increasing or decreasing

in T, which in turn depends again on the underlying distribution G. Interestingly, in the example of the normal distribution depicted in Figure 4, the two effects just balance each other such that the minimum crime level $F(p^m(T), T)$ remains constant. For the case of power distributions this property can even be shown analytically: That is, as long as $p^m(T)$ is interior, $F(p^m(T), T)$ is independent of T.[18] This implies that, in contrast to standard arguments, the minimum crime level cannot be reduced by an increase of the fine T.

5   Crime-Sensitive Inspectors

One major implication of Theorem 1 is that inspectors cannot be induced to choose monitoring intensities beyond the ceiling p^m, which puts a lower bound $F(p^m)$ on the resulting crime level. We now investigate to which extent this result relies on the assumption maintained so far that inspectors only care about their remuneration and effort costs, but not about the level of crime itself (see Eq. (1)). In this section, we relax this assumption and allow for crime-sensitive inspectors, characterized by the more general utility function

$$u(p, \beta) = R(D(p)) - \beta F(p) - C(p), \tag{3}$$

where β measures how the inspector's utility is affected by the presence of crime. Thereby, disutility ($\beta > 0$) can be caused by intrinsic motivation to keep the crime level low, while benefits of crime ($\beta < 0$) can arise, for example, from accepting bribes.[19] The inspector's

---

[18]Indeed, consider $F(p, T) = 1 - (p \cdot T)^\nu$ for $p \cdot T \leq 1$. Using the first order condition, we get $p^m(T) = \frac{1}{(\nu+1)^{\frac{1}{\nu}} \cdot T}$, which is interior given that $T > (1 + \nu)^{-\frac{1}{\nu}}$. Finally, $F(p^m(T), T) = 1 - (p^m(T))^\nu \cdot T^\nu = \frac{\nu}{1+\nu}$.

[19]The literature has discussed other reasons, why inspectors might worry about the prevailing level of crime, for example, payments which are inversely related to the crime level (Garoupa and Klerman, 2002) or a lower tax revenue part of which accruing to the enforcer (Coşgel, Etkes, and Miceli, 2011). In a different model framework, Besanko and Spulber (1989) assume that an inspector can allocate a given budget between enforcement and perquisites, so that the marginal rate of substitution can also be interpreted as a measure of the inspector's concern about crime.
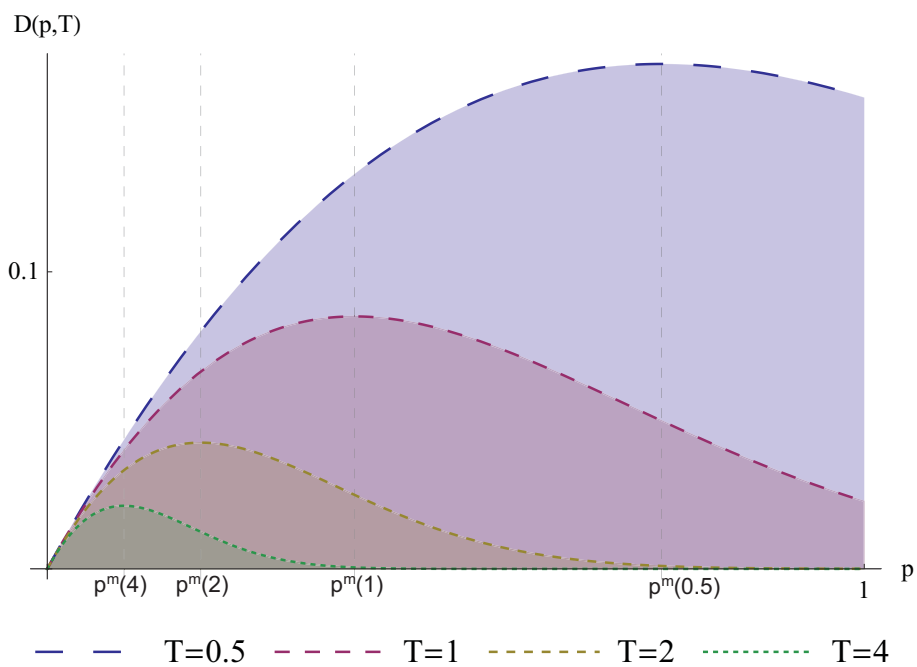
(a) Number of detections and implementable monitoring intensities
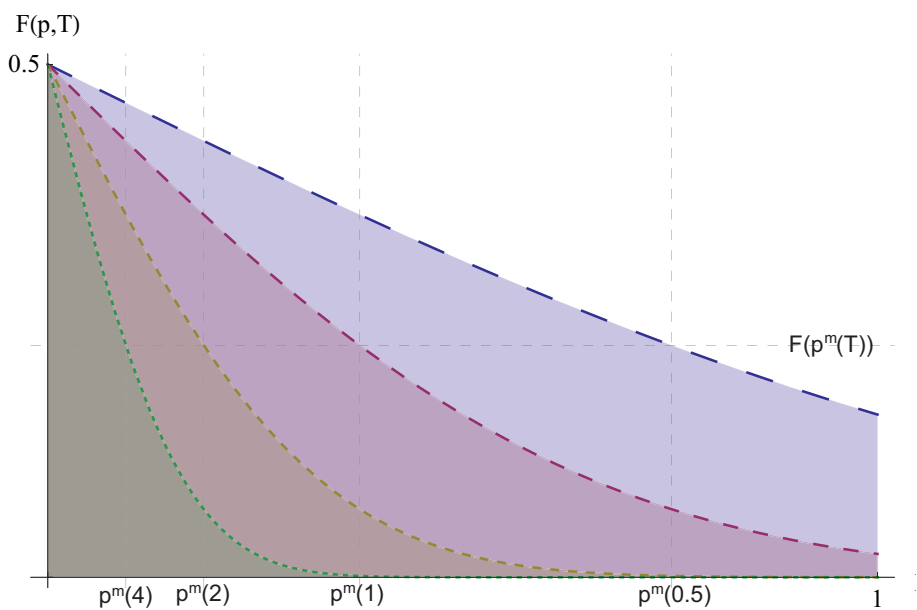
(b) Crime level

Gains from crime distributed uniformly on $[0,1]$ (special case of Power function $G(g) = g^\nu$ for $\nu = 1$) for $T = 0.5, 0.75, 1, 1.5$.

Figure 4: The impact of changes in the fine T

preferences from the basic model are nested as the special case $\beta = 0$ in this utility function.

As will be shown, inspectors with a sufficiently strong intrinsic motivation can be induced to implement monitoring intensities $(p > p^m)$, which are not implementable in the basic set-up. However, it turns out that for any type of inspector there is still a potentially large set of monitoring intensities which are not implementable.

To illustrate this point, we set $T = 1$ and consider quadratic costs of effort $C(p) = c \cdot p^2$ for some cost parameter $c > 0$. Moreover, for the sake of analytical tractability, we focus on the case where the gains from the offense are uniformly distributed on $[0, 1]$ which corresponds to the special case $\nu = 1$ for the Power distributions as considered in Figure 3 above, leading to $F(p) = 1 - p$. The resulting detection function $D(p) = p \cdot (1 - p)$ is then hump-shaped and symmetric around $p^m = \frac{1}{2}$. Also the conditions of Corollary 1 are satisfied, so that the set of implementable monitoring intensities is $P_1 = [0, \frac{1}{2}]$ for an inspector who is not crime-sensitive $(\beta = 0)$. The following result characterizes this set for crime-sensitive inspectors $(\beta \neq 0)$:

Proposition 3. *Let the distribution of gains be uniform on $[0, 1]$ and $T = 1$ such that $F(p) = 1 - p$ and $p^m = \frac{1}{2}$. Moreover, let the inspector's utility function be given by (3) with $C(p) = c \cdot p^2$. When the inspector's disutility from crime is sufficiently high (i.e. for $\beta > c$), then the set of implementable monitoring policies is $[\frac{1}{2}, 1]$. Otherwise (i.e. for $\beta < c$), it is given by $P_1 = [0, \frac{1}{2}]$.*

The underlying intuition for the proposition is simple: any reward offered for the desired number of detections $D(\hat{p})$ can also be gained with mimicking $\hat{p}$ by choosing $\tilde{p} = 1 - \hat{p}$ instead since $D(\tilde{p}) = D(\hat{p})$. It then depends on the inspector's sensitivity to crime β, in relation to the costs of monitoring, whether the higher or the lower monitoring intensity is preferred. In the knife-edge case of $\beta = c$ the inspector is indifferent between each pair of monitoring intensities that lead to the same number of detections.

Proposition 3 suggests that inspectors can be distinguished with respect to their β-type as follows: For "bad types" $(\beta < c)$, the intrinsic motivation for curbing crime is low,

20

and we are back to the setting underlying Theorem 1, rendering monitoring intensities $p > p^m$ non-implementable. The resulting crime level tends to be high (i.e. larger than $F(p^m)$) and detected offenders should hence be classified as scapegoats.

In contrast, for "good types" ($\beta > c$), the intrinsic motivation is sufficiently high such that they can be induced to choose high monitoring intensities. In this case, the number of offenses is low and the detected offenders should rather be classified as black sheep. Notice, however, that in this case, a mirror result of Theorem 1 applies in the sense that only monitoring intensities $p \geq p^m$ can be implemented, but not lower ones.

6   Linear Reward Schemes (Bounties)

The inspector's degree of intrinsic motivation also crucially determines his behavioral response to changes in the reward scheme $R(D(p))$. We illustrate this effect by considering simple linear payment schemes where for each detection, the inspector receives a predefined reward (or "bounty") r, i.e. $R(D(p)) = rD(p)$. This specification of payments has also been employed in the previous literature on delegated (private) enforcement (see e.g., Becker and Stigler, 1974; Landes and Posner, 1975; Polinsky, 1980; Besanko and Spulber, 1989). The inspector's optimal monitoring policy $p^*(r, \beta)$ then satisfies

$$p^*(r, \beta) \in \arg\max_{p \in [0,1]} rD(p) - \beta F(p) - C(p).$$

Keeping the parameterizations of the previous section, this leads to the following result:

Proposition 4. *Let the distribution of gains be uniform on $[0, 1]$ and $T = 1$ such that $F(p) = 1 - p$ and let the inspector's utility function be given by (3) with $C(p) = c \cdot p^2$. If $R(D(p)) = rD(p)$ with r not too small (i.e., $r \geq \max\{-\beta, \beta - 2c\}$), then*

(i) *the inspector's optimal monitoring policy is given by $p^*(r, \beta) = \frac{r+\beta}{2(r+c)}$, which is increasing in β.*

(ii) *an increase in the reward r leads bad (good) types to optimally choose a higher (lower) monitoring intensity, i.e. $\frac{\partial p^*(r,\beta)}{\partial r} > 0 \Leftrightarrow \beta < c$ and $\frac{\partial p^*(r,\beta)}{\partial r} < 0 \Leftrightarrow \beta > c$.*

21

As for part (i), more intrinsic motivation to curb crime (β) leads to higher monitoring effort.[20] Thereby $p^*(r, \beta) > p^m = \frac{1}{2}$ if and only if $\beta > c$ so that each type optimally chooses indeed a monitoring intensity from the set of implementable ones as characterized in Proposition 3.

Part (ii) of the proposition reveals that the response of the two types of inspectors to changes in the bounty r are quite different: For bad types ($\beta < c$), starting at the lower bound $r = -\beta$ we have $p^*(r, \beta) = 0$, and the monitoring intensity increases as r increases, approaching $p^*(r, \beta) = \frac{1}{2}$ from the left in the limiting case $r \to \infty$.[21] In this case, higher monetary rewards have the standard effect of increasing the monitoring intensity, generating more detections and thereby lowering the number of offenses.[22]

In contrast, for good types ($\beta > c$), $p^*(r, \beta)$ is *decreasing* in the reward r, starting at the lower bound $r = \beta - 2c$ where $p^*(r, \beta) = 1$ and approaching $p^*(r, \beta) = \frac{1}{2}$ from the right as r increases. Note that also these types respond with generating more detections when the marginal benefit r of doing so increases. However, they achieve more detections by *reducing* their monitoring intensity, which leads to more offenses. This result is in line with the literature on motivational crowding-out in the sense that the inspector's monitoring effort is the lower, the stronger the monetary incentives (see e.g. Deci, 1971; Deci, Koestner, and Ryan, 1999; Frey and Jegen, 2001; Gneezy, Meier, and Rey-Biel, 2011). In our framework, however, this effect is not due to the fact that such incentives directly reduce the inspector's degree of intrinsic motivation, for example, in the sense of turning an non-economic relationship into an economic one (see e.g., Titmuss, 1970;

---

[20]Also in the framework of Coşgel, Etkes, and Miceli (2011), the inspector chooses a higher effort when he benefits from a lower level of crime in the form of a higher taxation income.

[21]This is shown formally at the end of the proof of Proposition 4; see also Garoupa and Klerman (2002).

[22]In the context of corporate criminal liability considered in Arlen (1994), a firm is required to monitor its employees, and due to vicarious liability, it is liable for crimes committed by them. Liability is triggered (only) when a crime is detected by the firm so that the introduction of vicarious liability may in fact decrease the firm's optimal monitoring intensity. In our setup, vicarious liability can be interpreted as a negative bounty for each detection, the effects of which are characterized by Proposition 4.

Frey and Oberholzer-Gee, 1997; Gneezy and Rustichini, 2000b,a). Rather, for inspectors with a high degree of intrinsic motivation, generating more detections requires a lower monitoring effort, thereby also leading to more offenses.

The dependence of the inspector's response to monetary rewards on his degree of intrinsic motivation to keep the crime level low has further consequences: For example, in order to implement some desired monitoring intensity $\hat{p} > \frac{1}{2}$, the corresponding bounty is $r(\hat{p}, \beta) = \frac{\beta - 2\hat{p}c}{2\hat{p}-1}$, which is always positive when the inspector's intrinsic motivation is sufficiently strong (i.e., for $\beta > 2c$), but also $r(\hat{p}, \beta) < 0$ (i.e., a payment from the inspector to the governor for each detection) is possible. This occurs for good types with intermediate degrees of intrinsic motivation ($c < \beta < 2c$) and when $\hat{p} > p^*(0, \beta)$, that is, when $\hat{p}$ exceeds the monitoring intensity chosen by the inspector under intrinsic motivation alone ($r = 0$). In that case, the inspector needs to be punished by a negative bounty in order to increase his effort.

One can also relax the assumption that the fine is fixed at $T = 1$ and explore how changes in T affect the monitoring intensity chosen by the inspector.[23] Then the per-detection reward necessary to induce a given implementable monitoring intensity is given by $r(\hat{p}, \beta, T) = \frac{2c\hat{p} - \beta T}{1 - 2T\hat{p}}$, when the inspector's choice is interior. For the case of purely monetary incentives ($\beta = 0$) this is increasing in T so that higher fines induce higher per-detection rewards. With crime-sensitive inspectors ($\beta \neq 0$), however, the sign of the effect is ambiguous. Taking the derivative with respect to T reveals that it delicately depends on the parametrization of the model whether rewarding the inspector for detections becomes more or less costly for the governor as the fine increases.

All in all, our findings suggest that payment schemes based on the number of detections such as bounties are a delicate instrument, which is highly sensitive to the inspector's concern to keep the crime level low. In particular, ignoring an inspector's intrinsic motivation

---

[23]A special case of a linear reward occurs when the inspector keeps the collected fines, i.e., $r = T$. If interior, the optimal monitoring intensity is then given by $p(\beta, T) = \frac{(1+\beta)T}{2T^2+2c}$, which is hump-shaped in T where the maximum is attained at $T = \sqrt{c}$. Hence, for fines beyond this threshold, a higher fine (i.e., a higher reward per detection) would induce the inspector to choose a lower monitoring intensity.

might lead to severely misguided incentives.

7 Social Welfare

So far the analysis was focused on the implementability of a desired monitoring intensity ($\hat{p}$) which was assumed to be exogenous given. In this section we relax this assumption and derive $\hat{p}$ endogenously as the solution to a maximization problem of the governor who can be considered as a social planner. This will provide us with an efficiency benchmark against which we then compare our results on implementability. In this respect, the underlying social welfare function contains the following elements: First, let the society's harm from crime be $h \cdot F(p)$ where $h > 0$ is a constant harm from each offense. Furthermore, from the inspector's payoff function (see Eq. (3)), we include the inspector's enforcement costs $C(p)$ and his disutility from crime $\beta \cdot F(p)$, while the remuneration $R(D(p))$ is a transfer payment between the governor and the inspector which cancels. Moreover, also the offenders' gains from the offense are typically included (see e.g., Polinsky and Shavell, 2007), but this is not an uncontroversial issue (see e.g., Stigler, 1970). For this reason we use a general specification which nests both possibilities as special cases, so that social welfare is given by

$$SW(p) := -(h + \beta) \cdot F(p) - C(p) + \pi \int_{pT}^{\infty} g \cdot G'(q) dq, \tag{4}$$

where the parameter $\pi \in [0,1]$ reflects the weight that the social planner puts on the offenders' gains from the offense.[24]

Throughout we continue with the uniform-quadratic specification from the previous sections, i.e., the gains from the offense are uniformly distributed on $[0,1]$ and $T = 1$ (so that $F(p) = 1 - p$ and $p^m = \frac{1}{2}$) and $C(p) = c \cdot p^2$. The desired monitoring intensity $\hat{p}$ can

---

[24]The special case $\pi = 0$ is consistent with an alternative interpretation. If the inspector has an outside option of value zero and the governor chooses the reward to satisfy the inspector's participation constraint, then $R(D(p)) = \beta F(p) - C(p)$, such that the governor's payoff coincides with social welfare defined in Eq. (4), when setting $\pi = 0$.

thus be defined as maximizing the following social welfare function:

$$\hat{p} = \arg max_{p \in [0,1]} SW(p) = (1 - p)\left(\frac{\pi}{2}(1 + p) - h - \beta\right) - c \cdot p^2. \tag{5}$$

It is easily seen that social welfare is always strictly concave in p and, given that $h + \beta > 0$, also strictly increasing at $p = 0$ such that the welfare optimum is either attained at the boundary ($\hat{p} = 1$), which here leads to full deterrence, or it is interior.[25] In the latter case it solves the first-order condition $SW'(\hat{p}) = 0$ and is given by

$$\hat{p} = \frac{h + \beta}{\pi + 2c}. \tag{6}$$

Hence, the efficient (interior) monitoring intensity increases as offenses become more detrimental for the society (h), when the inspector becomes more crime-sensitive (β), as the gains from the offense receive less weight in the social welfare function (π) and when the cost of monitoring decreases (c).

In a next step, we compare the efficiency benchmark with our previous findings concerning implementability (in particular Theorem 1 and Proposition 3). This leads to the following result.

Proposition 5. *Let social welfare be given by Eq. (5) so that the resulting socially optimal monitoring intensity is $\hat{p}$.*

(i) *When the inspector's crime-sensitivity is low ($\beta < c$, bad type), then $\hat{p}$ is implementable if and only if the harm from the offenses is sufficiently small, i.e. $h \le c - \beta + \frac{\pi}{2}$. Otherwise, under-enforcement arises.*

(ii) *When the inspector's crime-sensitivity is high ($\beta > c$, good type), then $\hat{p}$ is implementable if and only if the harm from the offenses is sufficiently large, i.e. $h \ge c - \beta + \frac{\pi}{2}$. Otherwise, over-enforcement arises.*

---

[25]The socially optimal monitoring intensity is interior for sufficiently large costs of monitoring. More precisely, given that $h + \beta > 0$, $\hat{p}$ is interior if and only if $c > \frac{h + \beta - \pi}{2}$.

The proposition can be illustrated using the taxonomy presented in Table 1. Thereby, p^* denotes the (privately) optimal monitoring intensity chosen by the inspector for a given reward scheme $R(D(p))$, i.e., p^* maximizes the inspector's payoff as given in Eq. (3).[26]

As for part (i) of the Proposition, bad types exhibit a low intrinsic motivation to curb crime such that only p^* from the set $\in [0, p^m]$ can be implemented (see Proposition 3). If the social harm from crime is sufficiently low (cell 1 of Table 1), the welfare-maximizing monitoring intensity $\hat{p}$ also belongs to this set and hence becomes implementable. For example, it can be shown that by setting a bounty $\hat{r} = \frac{ch - \beta\frac{\pi}{2}}{\frac{\pi}{2} - h - \beta + c}$, the governor indeed induces the inspector to choose $p^* = \hat{p}$.[27] For larger harm from crime (cell 2 of Table 1), $p^* \leq p^m < \hat{p}$ holds, which implies that under-enforcement is inevitable.

Analogously, in part (ii) of the Proposition the inspector's intrinsic motivation is high, so that only high monitoring intensities $p^* \in [p^m, 1]$ can be implemented (see Proposition 3). Hence, the welfare-maximizing monitoring intensity also belongs to this set only when the harm from crime is sufficiently high (cell 4) such that $\hat{p} > p^m$ holds. For low harm from crime (cell 3), over-enforcement arises since $\hat{p} < p^m \leq p^*$. Note that the conditions for over-enforcement require $h < \frac{\pi}{2}$ (since we must have $\beta + h - \frac{\pi}{2} < c < \beta$), so that over-enforcement can only occur when also the offenders' gains from the offense enter the social welfare function ($\pi > 0$).

In terms of this taxonomy, one could argue that the previous literature (see e.g., Polinsky, 1980; Garoupa and Klerman, 2002; Coşgel, Etkes, and Miceli, 2011) has confined

---

[26]Note there typically exist several reward schemes which lead to the same p^*. For example, in the proofs of Theorem 1 and Proposition 3 we consider a simple, discontinuous reward scheme. But it can be shown that, except for p^m (where the detection function $D(p)$ reaches its hump), all implementable monitoring policies can also be reached with a linear reward scheme as discussed in section 6.

[27]Importantly, in our model there are no restrictions on the feasible transfers between the governor and the inspector (for example, due to limited liability). This implies that in our setup there is no issue with additional distortions due to a rent versus efficiency trade-off, which arises in many agency models (see e.g., Laffont and Martimort, 2002), and which would affect the possible implementability of $\hat{p}$ even "within cells" of Table 1.

| | harm from crime | |
	low: $h < c - \beta + \frac{\pi}{2}$	high: $h > c - \beta + \frac{\pi}{2}$
(i) low crime-sensitivity: $\beta < c$	(1) $p^*, \hat{p} \leq p^m$ implementable	(2) $p^* \leq p^m < \hat{p}$ under-enforcement
(ii) high crime-sensitivity: $\beta > c$	(3) $\hat{p} < p^m \leq p^*$ over-enforcement	(4) $p^*, \hat{p} \geq p^m$ implementable

Table 1: Privately optimal monitoring intensity p^* versus (socially) efficient monitoring intensity $\hat{p}$

attention to the two upper cells in Table 1, while we also analyze the remaining two cases.[28] We consider both of these novel cases to be empirically relevant. Intrinsic motivation might be one explanation why in many situations where desired monitoring intensity is high (cell 4) there is no issue of implementability, although an inspector's shirking would not reduce his payments. And intrinsic motivation can cause over-enforcement when the governor's desired monitoring level is low (cell 3), e.g. in the case of minor offenses which are fully deterred by an overly motivated inspector.

Generally, our analysis points to some degree of "congruity" of inspector's crime sensitivity (β) and the social harm from crime (h), which is required in order to make the efficient monitoring intensity implementable. This is satisfied on the diagonal of Table 1. Hence, when the harm is large, inspectors with strong intrinsic motivation are socially desirable in order to overcome the under-enforcement issue. In contrast, when the harm from crime is low, inspectors with low intrinsic motivation seem better suited since they do not over-enforce.[29]

---

[28]Indeed, Propositions 1 and 5 in Polinsky (1980) refer to cell (2); Garoupa and Klerman (2002) show that there is a linear reward scheme ("bounty") that implements a desired monitoring policy in cell (1), but not in cell (2); and Coşgel, Etkes, and Miceli (2011) introduce an inspector's crime-sensitivity by the right to collect taxes, but they do not explore the possibility that the disutility from crime is strong enough such that the lower cells (3) and (4) become relevant.

[29]The optimal alignment of preferences in principal-agent settings with delegation is also discussed in contexts other than law enforcement. For instance, Bubb and Warren (2014) study a principal's optimal

It is also interesting to investigate how a change of the fine T would affect these results. While a full analysis is beyond the scope of the paper, it is clear that the following effects need to be taken into account. First, as shown in Proposition 2, the monitoring intensity for which the number of detections is maximized becomes smaller as the fine increases (i.e., $p^m(T)$ is decreasing in T), which shrinks the set of implementable monitoring intensities for inspectors with low intrinsic motivation (bad types). Second, a higher fine changes the threshold for the classification as either bad or good types (now the condition is $\beta \lessgtr c/T^2$ instead of $\beta \lessgtr c$). Third, the fine also affects the welfare-maximizing (interior) monitoring intensity $\hat{p} = \frac{T(h+\beta)}{\pi T^2 + 2c}$, which is increasing in T as long as the weight of the gains from the offenses in the social welfare function (π) is not too large. Taken together, whether a higher fine increase the scope of implementability of $\hat{p}$ or instead fosters under-respective over-enforcement is generally ambiguous, and depends on the underlying distributional assumptions as well as on the other model parameters. For any given fine, however, one can distinguish cases analogously to Table 1 and obtain the corresponding implementability results.

8 Discussion

In this paper, we contribute to the literature on delegated monitoring where only the number of detections, but neither the underlying monitoring intensity nor the level of misbehavior can be observed by the delegating authority (governor). This literature has pointed to a problem of under-enforcement in the sense that the first-best monitoring intensity need not be implementable. The reason is that when several monitoring intensities are observationally identical (i.e., give rise to the same number of detections), then

---

delegation decision, when agents differ with respect to how strongly they care about the social costs and benefits from a regulatory policy such as environmental protection. They show that it is typically not optimal for the principal to select an unbiased agent (i.e., one with the same preferences as the principal with respect to the policy). The reason is that biased agents might have a stronger incentive to engage in information acquisition about new regulatory opportunities.

the inspector to which the monitoring task is delegated can typically only be induced to choose the minimum of these. This also imposes a lower bound on the resulting crime level.

We first generalize the results from the previous literature and characterize the full set of implementable monitoring policies. We then perform a comparative statics analysis to study how the set of implementable monitoring intensities varies with changes in the distribution of gains from the offense and in the fine. It turns out that this set is small for low benefits from crime or high fines. In particular, the largest monitoring policy that is implementable decreases with the fine such that deterrence need not be increasing in the fine.

We then consider a model extension where the utility of inspectors is directly influenced by the number of offenses in the population. In this respect, we believe that the intrinsic motivation to keep the level of crime low is an empirically relevant (and commonly neglected) factor. We first show that low levels of intrinsic motivation do not qualitatively affect the findings from the baseline model, and still only relatively low monitoring intensities can be implemented. However, when the degree of intrinsic motivation is sufficiently large, a mirror result prevails and the set of implementable policies is bounded from below instead. Consequently, either the desired monitoring policy is large enough to be implementable or over-enforcement occurs, that is, for any payment scheme the inspector will choose a monitoring intensity which is larger than the one which the governor wants to implement.

Our results also suggest that when inspectors are crime-sensitive (e.g., due to intrinsic motivation), then incentive schemes such as bounties as often discussed in the related literature might be an even more delicate instrument than was previously thought. The reason is that it crucially depends on the inspector's degree of intrinsic motivation whether such extrinsic incentives tend to reinforce the monitoring incentives generated by intrinsic motivation or crowd them out.

Our analysis, hence, points to the potentially crucial role of the degree of intrinsic motivation in the context of delegated monitoring which, however, will typically be of-

ten unobservable for governors. Therefore, an interesting model extension would be a screening framework where the governor can offer different pairs of monitoring intensities $p(\hat{\beta})$ and payments $R(\hat{\beta})$, depending on inspector's reported type $\hat{\beta}$ (which does not necessarily coincide with his true type β). In contrast to standard screening models, the resulting design problem for the governor becomes significantly more intricate because of additional incentive constraints. This is due to the fundamental property of this setting that multiple monitoring intensities give rise to the same number of detections, which increases the scope of mimicking. A full analysis of such an extended screening framework is, however, beyond the scope of the present paper. From a more practical point of view, one possibility to learn about an inspector's intrinsic motivation over time would be to exploit the different comparative statics properties of agents with different degrees of intrinsic motivation. For example, if it is possible to manipulate the costs of monitoring, then good types can be separated from bad types since the number of detections increases in only one of the two cases.

Finally, our analysis also sheds light on the question which contexts of delegated monitoring are more prone to issues such as under- or over-enforcement and the ensuing consequences. In this respect, note that our framework does not only apply to the arena of law & economics, but in principle also to other settings of monitoring such as enforcing safety and ethical standards in the manufacturing industry, or hygienic standards in the food industry. Another relevant context is education, where a school authority delegates the task of educating pupils to schools and teachers. In doing so, it is an outsider in the sense that it can typically only observe which grades are awarded in a school. However, it does not observe whether, for instance, good grades are due to the fact that the school, its teachers and the pupils are all hard-working or whether they are the result of a (tacit) agreement among these parties to take things easy. As a result, looking at grades only might not be too informative about the level of education of pupils, or the quality of schools. In this context, however, by featuring state- or nationwide tests which all pupils must take, it seems that authorities have successfully implemented institutional changes to ameliorate the problem that schools of different quality might be observationally identi-

cal.[30] Even though such or similar measures might not always be available in the contexts to which our framework applies, our analysis nevertheless clearly points to the beneficial role of institutional changes which help authorities overcome their "outsider" status.

---

[30]In Germany, for example, many states have recently introduced mandatory state-wide tests in German, English, and mathematics.

Appendix: Proofs

A   Proof of Theorem 1

For $\hat{p} \notin P_1$ there is a $\tilde{p} \in P_1$ such that $D(\tilde{p}) = D(\hat{p})$ by definition of P_1. Noting, that $R(D(\tilde{p})) = R(D(\hat{p}))$ while $C(\tilde{p}) < C(\hat{p})$, it follows that $u(\tilde{p}) > u(\hat{p})$, which shows that $\hat{p}$ cannot be implemented. Now, suppose $\hat{p} \in P_1$. Let $R(d) = C(p_1^d) + \varepsilon$ if $d = \hat{d} := D(\hat{p})$ and $R(d) = 0$ otherwise. Then $u(\hat{p}) = \varepsilon > 0$, while for ε small enough we have $u(p) \leq 0$ for all $p \neq \hat{p}$ because other monitoring intensities $\tilde{p}$ that lead to the same number of detections (i.e. $\tilde{p} \in P^{\hat{d}}$) are associated with higher costs and all other choices (i.e. $p \notin P^{\hat{d}}$) do not lead to any reward.

For the second statement note that G continuous renders F continuous and thus D as well. Since $D(p)$ starts with $D(0) = 0$ and reaches its global maximum for the first time at p^m, $D(p)$ attains every value of its image Δ in the interval $[0, p^m]$. Thus, for any $d \in \Delta$, $p_1^d \in [0, p^m]$. $\qquad\square$

B   Proof of Proposition 1

Consider two distributions G and $\tilde{G}$ where G is first-order stochastically dominated by $\tilde{G}$, i.e. $G(g) \geq \tilde{G}(g)$ for all g. Denote by $\tilde{F}(p)$ and $\tilde{D}(p)$ the respective functions resulting under $\tilde{G}$.

(i) For all $p \in [0, 1]$, we have

$$F(p) = 1 - G(pT) \;\leq\; 1 - \tilde{G}(pT) = \tilde{F}(p) \quad \text{and}$$
$$D(p) = p \cdot (1 - G(pT)) \;\leq\; p \cdot (1 - \tilde{G}(pT)) = \tilde{D}(p).$$

(ii) Under Condition (i) of Corollary 1, p^m and $\tilde{p}^m$ are characterized by the first-order conditions $1 - G(pT) = p \cdot T \cdot G'(pT)$ and $1 - \tilde{G}(pT) = p \cdot T \cdot \tilde{G}'(pT)$. By Condition (ii) of Corollary 1, the right-hand sides of both equations are increasing in p. The left-hand sides of both equations are decreasing in p such that p^m and $\tilde{p}^m$ are the
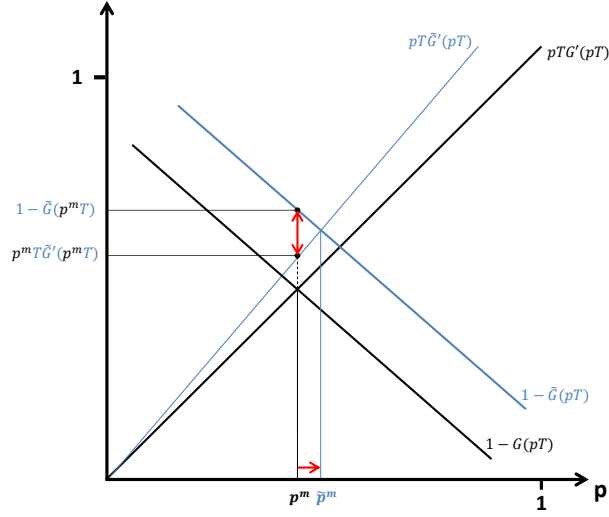
Figure 5: The marginal deterrence effect and the marginal detection effect for two distributions G and $\tilde{G}$

only intersections. This is illustrated in Figure 5 (while FOSD always implies that $1 - \tilde{G}(pT) \geq 1 - G(pT)$, the relation of the two right-hand sides is ambiguous). Observe that $\tilde{p}^m \geq p^m$ if and only if $1 - \tilde{G}(p^m T) \geq p^m \cdot T \cdot \tilde{G}'(p^m T)$.

The assumption $\tilde{G}'(p^m T) - G'(p^m T) < \frac{G(p^m T) - \tilde{G}(p^m T)}{p^m \cdot T}$ is equivalent to

$$1 - \tilde{G}(p^m T) > p^m \cdot T \cdot \tilde{G}'(p^m T) + \underbrace{(1 - G(p^m T)) - p^m \cdot T \cdot G'(p^m T)}_{=0} \qquad (7)$$

which yields the result.

$\square$

C   Proof of Proposition 2

To show the assertions for two fines T and $\tilde{T}$, we determine the slopes of the relevant functions.

(i) For all $p \in (0, 1]$, we have

$$\frac{\partial F(p, T)}{\partial T} = -p \cdot G'(pT) < 0 \quad \text{and} \quad \frac{\partial D(p, T)}{\partial T} = -p^2 \cdot G'(pT) < 0.$$

(ii) By Condition (i) of Corollary 1, $p^m(T)$ is interior. Using the implicit function theorem, we have

$$\frac{\partial p^m(T)}{\partial T} = -\frac{\frac{\partial^2 D(p,T)}{\partial p \partial T}}{\frac{\partial^2 D(p,T)}{\partial^2 p}} = -\frac{2p^m(T) \cdot G'(p^m(T)T) + p^m(T)^2 \cdot T \cdot G''(p^m(T)T)}{2T \cdot G'(p^m(T)T) + p^m(T) \cdot T^2 \cdot G''(p^m(T)T)}.$$

By Condition (ii) of Corollary 1, $D(p,T)$ is concave and thus the denominator is negative (this also follows from the second order condition for $p^m(T)$ to maximize $D(p,T)$). Observe that the negative sign of the denominator is equivalent to $G''(p^m(T)T) < \frac{-2G'(p^m(T)T)}{p^m(T) \cdot T}$ which is also equivalent to the numerator being negative. Taken together the expression is negative since the fraction after the minus sign is positive.

$\square$

D Proof of Proposition 3

For $C(p) = c \cdot p^2$ and $F(p) = 1 - p$, the inspector's optimization problem is: $\max_p R(p(1 - p)) - \beta \cdot (1 - p) - c \cdot p^2$. Consider some desired monitoring intensity $\hat{p}$ and let $\tilde{p} := 1 - \hat{p}$. Noting that $R(\hat{p}(1 - \hat{p})) = R(\tilde{p}(1 - \tilde{p}))$, we get $u(\hat{p}) \geq u(\tilde{p})$ if and only if $\beta \cdot (2\hat{p} - 1) \geq c \cdot (2\hat{p} - 1)$.

Suppose $\beta > c$. Then $u(\hat{p}) > u(\tilde{p})$ if and only if $\hat{p} > \frac{1}{2}$ such that any desired intensity $\hat{p} < \frac{1}{2}$ is strictly worse than $\tilde{p} = 1 - \hat{p}$. Now, let $R(d) = x > \beta + c$ if $d = \hat{d} := D(\hat{p})$ and $R(d) = 0$ otherwise. Then choosing x large enough, e.g. $x = \beta + c$, yields $u(\hat{p}) = x - \beta \cdot (1 - \hat{p}) - c \cdot \hat{p}^2 > 0 - \beta \cdot (1 - p) - c \cdot p^2$ for any $p \neq \hat{p}, \tilde{p}$. The case $\beta < c$ is fully analogous. $\square$

E Proof of Proposition 4

Suppose that the crime function is $F(p) = 1 - p$ and the inspector's utility is given by Eq. (3). If every detection is rewarded with some bounty r, the inspector's optimization

problem becomes:

$$\max_{p \in [0,1]} r \cdot p(1-p) - \beta \cdot (1-p) - c \cdot p^2.$$

An interior solution is characterized by the first-order condition $p^*(r, \beta) = \frac{r+\beta}{2(r+c)}$, which is the first part of the proposition. The second part follows directly from

$$\frac{\partial p^*(r, \beta)}{\partial r} = \frac{c - \beta}{2(r + c)^2}$$

given, again, that the solution is interior.

An interior solution is obtained if $r \geq \max\{-c, -\beta, \beta - 2c\}$, where the first condition is equivalent to concavity of the maximization problem, while the second and the third condition assure that the function is increasing at the boundary $p = 0$ and decreasing at the boundary $p = 1$. To show that $r \geq \max\{-\beta, \beta-2c\}$ is sufficient for an interior solution $p^*(r, \beta)$ and for the sake of completeness, we formally provide the optimal monitoring policies for all possible types.

1. For $\beta < c$ (bad types), we have $-\beta > -c > \beta - 2c$. Thus, the optimal behavior is:

$$p^*(r, \beta) = \begin{cases} \frac{r+\beta}{2(r+c)}, & \text{if } r \geq -\beta \\ 0, & \text{if } r < -\beta \end{cases}.$$

For $r = -\beta$ the inspector just chooses $p^* = 0$ and for growing r, $p^* = \frac{1}{2}$ is approached from the left.

2. For $\beta > c$ (good types), we have $\beta - 2c > -c > -\beta$. Thus, the optimal behavior is:

$$p^*(r, \beta) = \begin{cases} \frac{r+\beta}{2(r+c)}, & \text{if } r \geq \beta - 2c \\ 1, & \text{if } r < \beta - 2c \end{cases}.$$

For $r = \beta - 2c$ the inspector just chooses $p^* = 1$ and for growing r, $p^* = \frac{1}{2}$ is approached from the right.

3. For the "non-generic case" $\beta = c$, we have $-\beta = -c = \beta - 2c$. Then, the optimal behavior is:

$$p^*(r, \beta) = \begin{cases} \frac{r+\beta}{2(r+c)} = \frac{1}{2}, & \text{if } r > -\beta \\ \{0, 1\}, & \text{if } r < -\beta \\ [0, 1], & \text{if } r = -\beta \end{cases}.$$

For $r = -\beta$ the inspector is indifferent between all choices, for larger r he maximizes the number of detections, for smaller r he minimizes them.

□

F Proof of Proposition 5

We first establish that the monitoring intensity $\hat{p}$ that maximizes Eq. (4) satisfies

$$\hat{p} \geq p^m = \frac{1}{2} \iff h \geq c - \beta + \frac{\pi}{2}. \tag{8}$$

If maximizing Eq. (4) yields an interior solution, then the condition above follows from a simple rearrangement of the first order condition provided by Eq. (6). Otherwise, maximizing Eq. (4) yields either the boundary solution $\hat{p} = 0$ for $h + \beta < 0$, which implies $h < c - \beta + \frac{\pi}{2}$; or it yields the boundary solution $\hat{p} = 1$ for $h + \beta \geq 0$ and $c < \frac{h+\beta-\pi}{2}$, which implies $h > 2c - \beta + \pi > c - \beta + \frac{\pi}{2}$. This shows the first assertion.

(i) By Proposition 3, $\beta < c$ implies that any p is implementable if and only if $p \in P_1 = [0, p^m]$. Hence, $\hat{p}$ is implementable if and only if $\hat{p} \leq p^m$, which is satisfied if and only if $h \leq c - \beta + \frac{\pi}{2}$. Otherwise, we have $h > c - \beta + \frac{\pi}{2}$, which implies $\hat{p} > p^m$, while only $p \leq p^m$ are implementable.

(ii) By Proposition 3, $\beta > c$ implies that any p is implementable if and only if $p \in [p^m, 1]$. Hence, $\hat{p}$ is implementable if and only if $\hat{p} \geq p^m$, which is satisfied if and only if $h \geq c - \beta + \frac{\pi}{2}$. Otherwise, we have $h < c - \beta + \frac{\pi}{2}$, which implies $\hat{p} < p^m$, while only $p \geq p^m$ are implementable.

□

References

ANDREONI, J. (1991): "Reasonable doubt and the optimal magnitude of fines: should the penalty fit the crime?," *RAND Journal of Economics*, 22(3), 385–395.

ARLEN, J. (1994): "The potentially perverse effects of corporate criminal liability," *Journal of Legal Studies*, 23(2), 833–867.

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

BECKER, G. S., AND G. J. STIGLER (1974): "Law enforcement, malfeasance, and compensation of enforcers," *Journal of Legal Studies*, 3, 1–18.

BESANKO, D., AND D. F. SPULBER (1989): "Delegated law enforcement and noncooperative behavior," *Journal of Law, Economics & Organization*, 5, 25–52.

BUBB, R., AND P. L. WARREN (2014): "Optimal agency bias and regulatory review," *Journal of Legal Studies*, 43(1), 95–135.

COŞGEL, M. M., H. ETKES, AND T. J. MICELI (2011): "Private law enforcement, fine sharing, and tax collection: Theory and historical evidence," *Journal of Economic Behavior & Organization*, 80, 546–552.

DECI, E. (1971): "Effects of externally mediated rewards on intrinsic motivation," *Journal of Personality and Social Psychology*, 18(1), 105–115.

DECI, E., R. KOESTNER, AND R. RYAN (1999): "A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation," *Psychological Bulletin*, 125(6), 627–668.

DI TELLA, R., AND E. SCHARGRODSKY (2004): "Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack," *American Economic Review*, 94(1), 115–133.

FEESS, E., AND A. WOHLSCHLEGEL (2009): "Why higher punishment may reduce deterrence," *Economics Letters*, 104(2), 69–71.

FREY, B., AND R. JEGEN (2001): "Motivation crowding theory," *Journal of Economic Surveys*, 15(5), 589–611.

FREY, B. S., AND F. OBERHOLZER-GEE (1997): "The cost of price incentives: An empirical analysis of motivation crowding-out," *American Economic Review*, 87(4), 746–755.

GAROUPA, N., AND D. KLERMAN (2002): "Optimal law enforcement with a rent-seeking government," *American Law & Economics Review*, 4(1), 116–140.

GNEEZY, U., S. MEIER, AND P. REY-BIEL (2011): "When and why incentives (don't) work to modify behavior," *Journal of Economic Perspectives*, 21(4), 1–21.

GNEEZY, U., AND A. RUSTICHINI (2000a): "A fine is a price," *Journal of Legal Studies*, 29(1), 1–17.

GNEEZY, U., AND A. RUSTICHINI (2000b): "Pay enough or don't pay at all," *Quarterly Journal of Economics*, 115(2), 791–810.

ICHINO, A., AND G. MUEHLHEUSSER (2008): "How often should you open the door? Optimal monitoring to screen heterogeneous agents," *Journal of Economic Behavior & Organization*, 67, 820–831.

KUGLER, M., T. VERDIER, AND Y. ZENOU (2005): "Organized crime, corruption and punishment," *Journal of Public Economics*, 89(9), 1639–1663.

LAFFONT, J.-J., AND D. MARTIMORT (2002): *The theory of incentives: The principal-agent problem.* Princeton University Press, Princeton.

LANDES, W. M., AND R. A. POSNER (1975): "The private enforcement of law," *Journal of Legal Studies*, 4, 1–46.

LEVITT, S. D. (1997): "Using electoral cycles in police hiring to estimate the effect of police on crime," *American Economic Review*, 87(3), 270–290.

LEVITT, S. D., AND T. J. MILES (2007): "Empirical study of criminal punishment," in *Handbook of Law and Economics*, ed. by A. M. Polinsky, and S. Shavell, vol. 1, pp. 455–495. Elsevier.

LEVITT, S. D., AND S. A. VENKATESH (2007): "An empirical analysis of street-level prostitution," *University of Chicago, mimeo.*

NUSSIM, J., AND A. D. TABBACH (2009): "Deterrence and avoidance," *International Review of Law and Economics*, 29(4), 314–323.

POLINSKY, A. M. (1980): "Private versus public enforcement of fines," *Journal of Legal Studies*, 9, 105–127.

POLINSKY, A. M., AND S. SHAVELL (2007): "The theory of public enforcement of law," in *Handbook of Law and Economics*, ed. by A. M. Polinsky, and S. Shavell, vol. 1, pp. 403–454. Elsevier.

STIGLER, G. J. (1970): "The optimum enforcement of laws," *Journal of Political Economy*, 78(3), 526–36.

TITMUSS, R. (1970): *The gift relationship: From human blood to social policy.* London; Allen and Unwin.

TSEBELIS, G. (1990): "Penalty has no impact on crime: A game-theoretic analysis," *Rationality and Society*, 2(3), 255–286.