Preventing crime waves

Philip Bond and Kathleen Hagerty*

September 2007

Abstract

We study the design of enforcement mechanisms when enforcement resources are chosen *ex ante* and are inelastic *ex post*. Multiple equilibria arise naturally. Our main result is that the relative costs and benefits of marginal deterrence differ dramatically across equilibria. In particular, marginal deterrence always generates a net benefit in the Pareto inferior ("crime wave") equilibrium; while under many conditions it produces a net cost in the Pareto superior equilibrium. We also explore the consequences of increasing per-capita enforcement resources, and show that such increases may worsen overall crime levels.

*The authors are from the University of Pennsylvania and the Northwestern University respectively. We are grateful for helpful comments from Itay Goldstein and Kathryn Spier, and from audiences at the Institute for Advance Study and the American Law and Economics Association meetings. Bond thanks the Institute for Advanced Study for hospitality and financial support (in conjunction with Deutsche Bank) over the academic year 2002-03. Any remaining errors are our own.

An important constraint faced by tax inspectors, regulators, police forces, etc. is that investigation resources are to a large extent fixed in the short-run. Human employees constitute the key resource in these organizations, and cannot be increased at short notice. Because of this, even if, for example, a tax inspector is in principle committed to auditing all returns claiming deductions above $100,000, if an unusually large number of returns fall into this category he is unable to audit them all.

For the most part, extant formal models of enforcement have ignored this constraint.[1] Instead, the typical assumption is that an enforcement authority can *ex ante* commit to investigate any number of individuals whose signal *ex post* fits the criteria for investigation (see, for example, Mookherjee and Png 1992, 1994, DeMarzo, Fishman and Hagerty 1998). In this paper we explicitly take the *ex post* inelasticity of enforcement resources into account. Our analysis focuses on a re-evaluation of the optimality of maximal versus marginal penalties given this constraint. Crime waves arise naturally when enforcement resources are inelastic *ex post*,[2] and we establish that the benefits of marginal penalties are much greater in crime waves than outside crime waves.

The optimality, or otherwise, of punishing proscribed behavior by using the severest sanction available — "maximal" penalties — is one of the oldest and most enduring questions in law and economics. As articulated by Becker (1968), there is a simple and compelling argument in favor of maximal penalties: given a non-maximal penalty, one can always achieve the same expected penalty by simultaneously reducing resources expended on detection and increasing the penalty. This argument is troubling because, of course, almost all real-world penalty schedules instead mandate different penalties for different actions. Arguably the leading counterargument to the optimality of maximal penalties relies on the need for marginal deterrence: in terms of a commonly given example, a penalty schedule needs to ensure not just that a potential offender prefers no crime to armed robbery, but also that he prefers armed robbery to murder.[3]

[1] Bassetto and Phelan (2006) are one recent exception.

[2] See the references in footnote 5 below.

[3] While this argument is widely associated with Stigler (1970), it has considerably older antecedents —

Although persuasive in many respects, the proposition that the need for marginal deterrence implies the optimality of non-maximal penalties is subject to at least two important caveats. First, as observed by Mookherjee and Png (1992), Shavell (1992), and Wilde (1992), it is often possible to monitor different actions at different rates (with the assumption that there is no limitation on the number of *ex post* investigations that can occur). Mookherjee and Png are the most specific in this regard, and point out that if the enforcement authority is able to observe a signal that is correlated to the action selected by the potential offender, it can vary its monitoring effort according to the signal observed. As such, the standard argument for maximal penalties still applies: the maximal penalty should be mandated for each action, with marginal deterrence across actions provided by varying the monitoring intensity.

Second, although marginal deterrence has the potential benefit of reducing the crimes of the worst offenders, it also has the cost of increasing the crimes of other individuals. In terms of the example above, reducing the penalty for armed robbery increases the attractiveness of this crime. This drawback of marginal deterrence is noted by Wilde (1992, page 334), and exists in Mookherjee and Png (1994). In their analysis, all crimes are monitored at the same rate regardless of the seriousness of the crime. This implies that marginal deterrence can only be accomplished through the penalty. They show that at the optimal penalty schedule, there are typically agents who cause a strictly positive amount of harm who would instead be fully deterred under the first best.[4] The harmful acts committed by these agents constitute the cost of marginal deterrence - were marginal deterrence not a concern, the expected penalty for these acts would be increased.

In this paper we show that in a model with limited *ex post* investigation resources, both issues can be simply and simultaneously addressed. First, this assumption immediately implies that an enforcement authority cannot reduce its expenditure by implementing marginal deterrence via varying monitoring intensities. Since monitoring resources must be

see Shavell's (1992) discussion.

[4]Specifically, in Mookherjee and Png (1994) the agent types in the lower interval of panel b of Figures 1, 3 and 4 choose harmful actions which could be deterred at low cost.

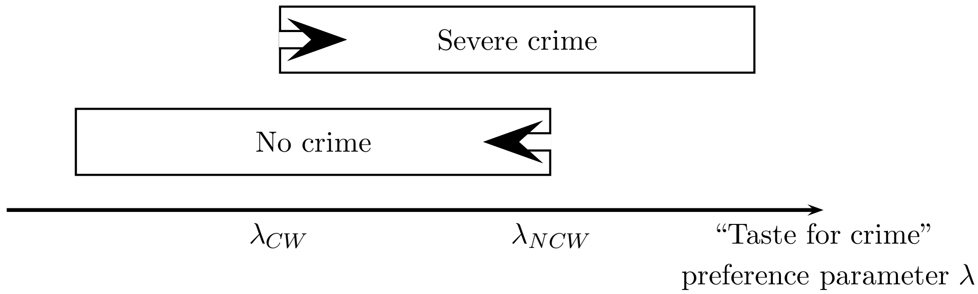

Figure 1: The diagram shows equilibrium crime levels when maximal penalties are used. No crime is an equilibrium for all sufficiently small realizations of the preference parameter λ. Severe crime is an equilibrium for all sufficiently large realizations. For moderate realizations of λ there is both a no crime equilibrium and a severe crime equilibrium. The arrows show the effect of introducing marginal deterrence.

decided ahead of time, there is nothing to be gained by reducing monitoring intensity *ex post*.

Second, as other authors have observed, the assumption that enforcement resources cannot be varied *ex post* leads to multiple equilibria in crime levels for standard congestion reasons.[5] When other individuals commit more crime, enforcement resources are stretched thin, and so the expected cost of crime falls.[6] For our purposes, the importance of multiple equilibria in crime levels is that they raise the question of which equilibrium the enforcement authority cares about. Our main result is that the balance of the costs and benefits of marginal deterrence (see above) differs dramatically across equilibria. Specifically, we show that when the high crime equilibrium is played — a situation we refer to as a *crime wave* — then there is always a level of marginal deterrence at which the benefits outweigh the costs. In contrast, when the low crime equilibrium is played, there are many parameter values for which the costs outweigh the benefits, and maximal penalties are optimal.

The intuition for this result is easiest to give using the simple diagram of Figure 1. In

[5]See Schrag and Scotchmer (1997), Tabarrok (1997), Fender (1999), and Jost (2001).

[6]That is, the crime decisions of different individuals are strategic complements when enforcement resources cannot be changed *ex post*.

our model, as in the papers referenced above, individuals are heterogeneous in their taste for crime, which we denote by λ. For the reasons discussed, for a given penalty schedule there are generally multiple equilibria in crime levels. More specifically, when extremal penalties are used both no crime and severe crime are equilibria for agents with a moderate taste for crime: in Figure 1 this is the range of taste parameters λ from λ_{CW} to λ_{NCW}.

Starting from the extreme of maximal penalties for moderate crime, the introduction of non-maximal penalties has two effects. On the one hand, severe crime is less likely to be an equilibrium outcome, since moderate crime is now a more attractive alternative. This is the familiar benefit of marginal deterrence noted by Stigler (1970). In terms of Figure 1, the value λ_{CW} increases. On the other hand, since moderate crime is now more attractive, no crime becomes harder to support as an equilibrium. This is the cost of marginal deterrence. In terms of Figure 1, the value λ_{NCW} decreases.

If one assumes that crime waves occur (that is, the highest crime equilibrium is played) then the equilibrium outcomes under maximal penalties are no crime when the taste parameter is less than λ_{CW}, and severe crime otherwise. So in this case, the benefits of marginal deterrence outweigh the costs: the relevant "boundary" in Figure 1 is the severe crime boundary λ_{CW}, which shifts right, while because of the equilibrium selection assumption the leftwards shift of the no crime boundary λ_{NCW} has no effect.

Conversely, if one instead assumes that the lowest crime equilibrium is played then the equilibrium outcomes under maximal penalties are no crime when the taste parameter is less than λ_{NCW}, and severe crime otherwise. In this case the costs of marginal deterrence outweigh the benefits, by a parallel argument. The relevant boundary in Figure 1 is now the no crime boundary λ_{NCW}, and this shifts leftwards.

The remainder of the paper is as follows. Section 1 presents the model and some preliminary analysis. Section 2 characterizes the benchmark outcomes when maximal penalties are used. The only detail missing from the intuitive discussion above is that a shift to marginal deterrence moves the boundary values λ_{CW} and λ_{NCW} in ways that are more or less comparable. Section 3 shows just this, and so establishes our main result. Section 4 considers in more detail the optimal choice of the penalty for moderate crime when crime waves occur, and presents several comparative static results. Section 5 explores the consequences of increasing per capita enforcement resources. Finally, Section 6 concludes.

1 Model and preliminary results

There are two agents, labelled i and j. (All our main results hold for $N \geq 2$ agents — see the discussion on page 22.) Each agent chooses between three possible action levels: $a \in \{0, a_M, 1\}$, where $a_M \in (0,1)$. The social costs of these actions are 0, C_M and C_1 respectively, where $C_1 > C_M > 0$. We will often refer to actions $a = 0, a_M, 1$ respectively as *no crime*, *moderate crime* and *severe crime*.

An enforcement authority oversees the two agents with the aim of minimizing the social cost of their actions. However, the enforcement authority does not observe the actions of agents i and j directly. Instead, it observes only noisy signals of these actions, $a^i + \frac{\varepsilon^i}{h}$ and $a^j + \frac{\varepsilon^j}{h}$, where $h > 0$ is a constant measuring the precision of the signal. The error terms ε^i and ε^j are identically and independently distributed, with distribution and density functions F and f. The signals observed by the enforcement authority satisfy the monotone likelihood ratio property (MLRP); equivalently, the density function f is log-concave.[7]

The enforcement authority can impose penalties on the two agents, but must expend resources in order to do so. A central assumption in our analysis is that the enforcement authority's resources are fixed before agents choose their actions, and cannot be increased *ex post.* This appears to be a reasonable description of most real-world enforcement authorities: human employees typically constitute the key resource, and tax collection agencies, regulatory inspection agencies, police forces etc. cannot increase their staff at short notice.

[7] Let $\sigma^i = a^i + \frac{\varepsilon^i}{h}$ be the signal observed by the enforcement authority. MLRP is defined as

$$\frac{\partial}{\partial \sigma^i}\left(\frac{1}{\Pr(\sigma^i|a^i)}\frac{\partial \Pr(\sigma^i|a^i)}{\partial a^i}\right) = \frac{\partial}{\partial \sigma^i \partial a^i}\ln \Pr\left(\sigma^i|a^i\right) > 0.$$

Since $\Pr\left(\sigma^i|a^i\right) = f\left(\frac{\sigma^i - a^i}{h}\right)$, this condition is equivalent to log-concavity of the density function f. Many common distributions, including the normal distribution, satisfy log-concavity.

Throughout, we assume that the enforcement authority has resources to penalize just one of the two agents.[8]

Formally, based on the pair of signals $a^i + \frac{\varepsilon^i}{h}$ and $a^j + \frac{\varepsilon^j}{h}$ the enforcement authority chooses whether to investigate agent i or agent j. That is, an investigation policy is a mapping $\mu : \Re^2 \to \{i, j\}$.[9] We assume that the investigation policy is anonymous, in the sense that it is independent of the identity of the agent (if $\mu(x, x') = i$ then $\mu(x', x) = j$). Define $p(a^i, a^j)$ as the probability that agent i is investigated given actions a^i and a^j, i.e., $p(a^i, a^j) = \Pr\left(\mu\left(a^i + \frac{\varepsilon^i}{h}, a^j + \frac{\varepsilon^j}{h}\right) = i\right)$.

Investigating an agent allows the enforcement authority to perfectly observe the action chosen by that agent. We assume that penalty technology is such that only an agent who has been investigated can be penalized. The maximum feasible penalty is S. A general penalty specification is a triple $(s_0, s_M, s_1) \in [0, S]^3$. For simplicity, we assume that penalties impose no social cost (for example, they are wealth transfers).

The enforcement authority does not know how much agents benefit from the socially costly actions $a = a_M, 1$. Each agent's payoff to action a is given by λa, where λ is unobserved by the enforcement authority and is drawn from support $[0, \bar{\lambda}]$. For simplicity we assume that both agents share the same taste parameter λ. This assumption allows us to focus on symmetric equilibria throughout, and our results would be qualitatively unchanged by the introduction of a small amount of preference heterogeneity between the two agents. Economically, the assumption can be thought of as reflecting the widely held

[8]Of course, if society cares enough about preventing actions $a = a_M, 1$ it will endow the enforcement authority with resources to penalize both agents. In this case our environment corresponds to the single-agent problem analyzed by previous authors.

[9]Since the enforcement authority's resources are fixed in advance, it is natural to assume that it will always use these resources *ex post* by investigating one agent. If the enforcement authority derives any payoff from successfully penalizing an agent, then it will always choose to investigate one agent *ex post*. Qualitatively, we do not believe our results would change if instead the enforcement authority sometimes refrained from investigating anyone. On a technical level, the main issue is that our proof that there is always a symmetric pure-strategy equilibrium (Lemma 2) relies on the assumption that the enforcement authority always investigates one of the two agents.

notion that cultural norms against wrongdoing vary geographically, and may also change over time. Alternatively, the assumption can be motivated by variations in community enforced social sanctions; or by variations in the marginal utility of income.

No crime ($a = 0$) is an equilibrium if and only if

$$-s_0 p(0,0) \geq \lambda a_M - s_M p(a_M, 0) \tag{IC0-M}$$

$$-s_0 p(0,0) \geq \lambda - s_1 p(1,0) \tag{IC0-1}$$

Moderate crime ($a = a_M$) is an equilibrium if and only if

$$\lambda a_M - s_M p(a_M, a_M) \geq -s_0 p(0, a_M) \tag{ICM-0}$$

$$\lambda a_M - s_M p(a_M, a_M) \geq \lambda - s_1 p(1, a_M) \tag{ICM-1}$$

Severe crime ($a = 1$) is an equilibrium if and only if

$$\lambda - s_1 p(1,1) \geq -s_0 p(0,1) \tag{IC1-0}$$

$$\lambda - s_1 p(1,1) \geq \lambda a_M - s_M p(a_M, 1). \tag{IC1-M}$$

We assume that the social cost of actions $a = a_M, 1$ exceeds the direct private benefit, i.e., $C_1 - \bar{\lambda} > C_M - \bar{\lambda} a_M > 0$. So taking the enforcement authority's *ex post* investigative capacity as fixed at one investigation,[10] its objective is to choose an investigation policy μ and a penalty specification (s_0, s_M, s_1) so as to minimize the expected equilibrium crime level.

1.1 Preliminary results

The remainder of this section develops several preliminary results that allow us to simplify the equilibrium characterization. For use throughout, observe that for any pair of actions a^i and a^j,

$$p\left(a^i, a^j\right) = 1 - p\left(a^j, a^i\right). \tag{1}$$

[10] It would be straightforward to expand the problem to one in which the investigative capacity is also determined optimally.

Given that the investigation policy μ is anonymous, $p(a, a) = 1/2$ for any action a. That is, whenever both agents take the same action, each is investigated with probability 1/2.

Consider a pair of offsetting incentive conditions — (IC0-M) and (ICM-0), for example. Condition (IC0-M) says that agent i prefers no crime to moderate crime, conditional on agent j choosing no crime. Condition (ICM-0) says that agent i prefers moderate crime to no crime, conditional on agent j choosing the moderate crime. Since the probability that agent i is investigated depends on agent j's action, these two conditions are not symmetric. This raises the possibility that both (IC0-M) and (ICM-0) may simultaneously fail, a complication that would potentially lead to the non-existence of a pure-strategy symmetric equilibrium. However, (1) is enough to ensure that this possibility does *not* arise:

Lemma 1 *Suppose that both the penalty schedule and investigation policy are monotone, i.e., $s_1 \geq s_M \geq s_0$ and p is increasing in its first argument. Then at least one of (IC0-M) and (ICM-0) holds; at least one of (IC0-1) and (IC1-0) holds; and at least one of (ICM-1) and (IC1-M) holds.*

The existence of at least one pure-strategy symmetric equilibrium follows easily from Lemma 1:

Lemma 2 *Suppose that both the penalty schedule and investigation policy are monotone, i.e., $s_1 \geq s_M \geq s_0$ and p is increasing in its first argument. Then there exists at least one symmetric pure-strategy equilibrium.*

Having established that a pure-strategy equilibrium exists for any realization of λ, we turn now to simplifying the problem. Holding the investigation policy μ fixed, setting $s_0 = 0$ always increases the range of λ realizations for which no crime is an equilibrium; and setting $s_1 = S$ always increases the range of λ realizations for which severe crime is *not* an equilibrium. In a similar fashion, the investigation policy "investigate the agent with the higher signal" is the investigation policy that minimizes the range of the severe crime equilibrium and maximizes the range of the no crime equilibrium. Formally:

Lemma 3 *The probability that no crime is an equilibrium is maximized, and the probability that severe crime is an equilibrium is minimized, by choosing $s_0 = 0$, $s_1 = S$, and using the investigation policy "investigate the agent with the higher signal."*

Given Lemma 3, for the remainder of the paper we assume $s_0 = 0$, $s_1 = S$, and that the "investigate the agent with the higher signal" policy is used. Define $q\left(a^i - a^j\right) = p\left(a^i, a^j\right)$, since under this investigation policy only the *difference* in the actions of the two agents affects the investigation probability.

Lemma 4 *The investigation probability function $q : [-1, 1] \to [0, 1]$ is:*
(I) Symmetric about 0: $q(a) - q(0) = q(0) - q(-a)$.
(II) Increasing (decreasing) in signal precision for positive (negative) values.
(III) Concave (convex) over positive (negative) values.

For use in the remainder of the paper, the equilibrium conditions simplify to: no crime $(a = 0)$ is an equilibrium if and only if

$$0 \geq \lambda a_M - s_M q(a_M) \tag{IC0-M}$$

$$0 \geq \lambda - Sq(1)\,; \tag{IC0-1}$$

moderate crime $(a = a_M)$ is an equilibrium if and only if

$$\lambda a_M - s_M q(0) \geq 0 \tag{ICM-0}$$

$$\lambda a_M - s_M q(0) \geq \lambda - Sq(1 - a_M)\,; \tag{ICM-1}$$

and severe crime $(a = 1)$ is an equilibrium if and only if

$$\lambda - Sq(0) \geq 0 \tag{IC1-0}$$

$$\lambda - Sq(0) \geq \lambda a_M - s_M q(a_M - 1)\,. \tag{IC1-M}$$

An immediate Corollary to part (II) of Lemma 4 is:

Corollary 1 *For any penalty specification s_M, an increase in signal precision h unambiguously improves the distribution of crime.*

Proof of Corollary 1: From Lemma 4, an increase in h increases $q(a_M)$, $q(1-a_M)$ and $q(1)$, decreases $q(a_M-1)$, and leaves $q(0)$ unchanged. As such, no crime ($a=0$) is now an equilibrium for more realizations of λ, while severe crime ($a=1$) is an equilibrium for fewer realizations. ∎

1.2 Extremal vs marginal penalties

Observe that in light of Lemma 3, the enforcement authority's problem has reduced to choosing the penalty s_M to impose on an agent who chooses action a_M. This is the focus of the paper. We refer to the choice $s_M = s_1 = S$ as an *extremal penalty*, and to any choice $s_M < s_1 = S$ as a *marginal penalty.*[11]

1.3 Assumptions

We assume that the supremum of the support of the taste parameter λ satisfies $\bar{\lambda} > S\frac{q(1-a_M)}{1-a_M}$. This guarantees that severe crime is the only equilibrium when the taste parameter is sufficiently high, regardless of the choice of penalty s_M.[12] Moreover, to make our analysis as transparent as possible, we assume that signal precision is sufficiently poor such that

$$q(0) > q'(0). \tag{2}$$

This assumption ensures that moderate crime is never an equilibrium under extremal penalties. However, we stress that it is not essential for our main results, which hold under much weaker conditions. In particular, our results hold whenever a_M is close to 1/2, regardless of whether or not (2) is satisfied.[13]

[11]Clearly one could also term the choice $s_M = 0$ as an extremal penalty. This is purely an issue of terminology.

[12]It is immediate that (ICM-1) does not hold at $\bar{\lambda}$. By the concavity of q over positive values, (IC0-1) does not hold either: $q(a) \geq aq(1) + (1-a)q(0) > aq(1)$ for any $a \in (0,1)$, and so $S\frac{q(1-a_M)}{1-a_M} > Sq(1)$. Hence $a=1$ is the only equilibrium at $\lambda = \bar{\lambda}$.

[13]See footnotes 17 and 18.

2 A benchmark: equilibrium outcomes under extremal penalties

Our main object of enquiry is when extremal penalties are, and are not, optimal. Accordingly, we begin our analysis by characterizing the equilibrium outcomes when extremal penalties are used. No crime ($a = 0$) is an equilibrium when (IC0-M) and (IC0-1) hold, i.e., if $\lambda \leq \frac{Sq(a_M)}{a_M}$ and $\lambda \leq Sq(1)$. Since q is concave over positive values,

$$q(a_M) \geq a_M q(1) + (1 - a_M) q(0) > a_M q(1). \quad (3)$$

As such, for extremal penalties (IC0-M) holds whenever (IC0-1) does, and so no crime is an equilibrium if and only if $\lambda \leq Sq(1)$.

At the other extreme, severe crime is an equilibrium when (IC1-0) and (IC1-M) hold, i.e., $\lambda \geq Sq(0)$ and $\lambda \geq S\frac{q(0)-q(a_M-1)}{1-a_M}$. Under assumption (2), for extremal penalties (IC1-M) holds whenever (IC1-0) does.[14]

Assumption (2) says that the investigation probability does not change very quickly as an agent changes his action choice. This means that any agent with preferences for crime that are strong enough for him to prefer moderate crime to no crime would derive an even higher payoff from severe crime, given that under extremal penalties s_M and s_1 coincide. As such, under extremal penalties moderate crime is never an equilibrium. Formally, moderate crime ($a = a_M$) is an equilibrium if (ICM-0) and (ICM-1) hold, i.e., if $\lambda \geq \frac{Sq(0)}{a_M}$ and $\lambda \leq S\frac{q(1-a_M)-q(0)}{1-a_M}$, and assumption (2) implies that both inequalities cannot be satisfied at once.[15]

[14]We must show that $q(0) \geq \frac{q(0)-q(a_M-1)}{1-a_M}$, or equivalently, $q(a_M - 1) - a_M q(0) \geq 0$. Convexity of q over negative values implies $q(0) - q'(0)(1 - a_M) \leq q(a_M - 1)$. From (2) it follows that $q(0) a_M = q(0) - q(0)(1 - a_M) < q(a_M - 1)$.

[15]Specifically, assumption (2) implies $\frac{q(0)}{a_M} > \frac{q(1-a_M)-q(0)}{1-a_M}$, or equivalently, $a_M q(1 - a_M) < q(0)$, as follows. Concavity of q over positive values and assumption (2) together imply

$$q(1 - a_M) \leq q(0) + (1 - a_M) q'(0) < (2 - a_M) q(0).$$

Since $a_M (2 - a_M) \leq 1$, the result follows.

Proposition 1 *When extremal penalties are used, no crime ($a = 0$) is an equilibrium if $\lambda \leq Sq(1)$; severe crime ($a = 1$) is an equilibrium if $\lambda \geq Sq(0)$; moderate crime ($a = a_M$) is never an equilibrium.*

As an immediate implication:

Corollary 2 *Both no crime and severe crime are equilibria when $\lambda \in [Sq(0), Sq(1)]$.*

The source of multiple equilibria is, or course, the enforcement authority's limited resources *ex post*. This means that the action choices of the two agents are strategic complements: when agent j chooses a more socially costly action, it reduces the investigation probability faced by agent i, and so increases the utility gain to agent i of choosing a more socially costly action.

Given the existence of multiple equilibria, an equilibrium selection rule is needed to select the optimal choice of the penalty s_M. Our main results characterize how the equilibrium selection rule affects the optimal choice. We consider two selection rules: (CW) A *crime wave* selection rule, in which whenever multiple equilibria exist we assume the one with the highest crime level is played; and (NCW), a *no crime wave selection rule*, in which whenever multiple equilibria exist we assume the one with the lowest crime level is played. We often refer to these equilibrium selection rules in terms of whether or not crime waves occur.

3 The costs and benefits of marginal penalties

From Proposition 1, when extremal penalties are used the only equilibria are no crime ($a = 0$) and severe crime ($a = 1$). As discussed in the introduction, the benefit of introducing marginal penalties is that doing so shrinks the range of taste parameters λ for which severe crime is an equilibrium. In terms of Figure 1, the point λ_{CW} increases. This benefit is obtained only if there is some taste parameter λ for which (IC1-0) is satisfied (so that severe crime is an equilibrium under extremal penalties $s_M = S$), and for which (IC1-M) fails. That is, the benefits of marginal penalties are obtained only if s_M is set so that for some λ,

$$\overbrace{Sq(0) \leq}^{\text{(IC1-0) holds}} \lambda < \overbrace{\frac{Sq(0) - s_M q(a_M - 1)}{1 - a_M}}^{\text{(IC1-M) fails}}. \tag{4}$$

Define s_M^B as the penalty for moderate crime at which benefits are first experienced. That is, s_M^B is the highest value of s_M such that (4) is satisfied for some taste parameter λ. Evaluating,

$$s_M^B = S a_M \frac{q(0)}{q(a_M - 1)}.$$

Likewise, the cost of introducing marginal penalties is that doing so shrinks the range of taste parameters λ for which no crime is an equilibrium. In terms of Figure 1, the point λ_{NCW} decreases. This cost is experienced only if there is some taste parameter λ for which (IC0-1) is satisfied (so that no crime is an equilibrium under extremal penalties $s_M = S$), and for which (IC0-M) fails. That is, the costs of marginal penalties are incurred only if s_M is set so that for some λ,

$$\overbrace{\frac{s_M q(a_M)}{a_M}}^{\text{(IC0-M) fails}} < \lambda \overbrace{\leq Sq(1)}^{\text{(IC0-1) holds}} \tag{5}$$

Define s_M^C as the penalty for marginal crime at which costs are first experienced. That is, s_M^C is the highest value of s_M such that (5) is satisfied for some taste parameter λ. Evaluating,

$$s_M^C = S a_M \frac{q(1)}{q(a_M)}.$$

Now that we have identified the critical values s_M^B and s_M^C at which marginal penalties first generate benefits and costs, we are ready to state our first main result:

Proposition 2 *Suppose that crime waves occur. Then as marginal penalties are introduced by reducing s_M away from S, equilibrium crime levels are at first unaffected and then reduced.*

Proof of Proposition 2: The key to establishing this result is to show that $s_M^B > s_M^C$, i.e., that the benefits are experienced before the costs. This is sufficient to establish the result because if $s_M \geq \max\left\{s_M^B, s_M^C\right\}$ then the equilibria are the same as for $s_M = S$; while as s_M is reduced from s_M^B to s_M^C, the range of taste parameters for which severe crime is an equilibrium shrinks, and the range of taste parameters for which no crime is an equilibrium remains unchanged. In terms of Figure 1, λ_{CW} is increased while λ_{NCW} remains unchanged. Because of the equilibrium selection rule, the reduction in the severe crime equilibrium range represents an unambiguous drop in the crime level.

The fact that $s_M^B > s_M^C$ follows from the convexity of q over negative values and its concavity over positive values. A rough intuitive argument is as follows. The benefit of marginal penalties derives from inducing an agent to switch from severe crime to moderate crime when the other agent is committing the severe crime. The associated fall in the investigation probability is from $q(0)$ to $q(a_M - 1)$. Because of the convexity of q over negative values this reduction in the investigation probability is relatively large, and so the penalty s_M does not need to be set very low to induce the change. Conversely, the cost of marginal penalties derives from inducing an agent to switch from no crime to moderate crime when the other agent is committing no crime. No crime is an equilibrium only if each agent dislikes incurring the penalty S with probability $q(1)$. Because of the concavity of q over positive values, the investigation probability $q(a_M)$ is relatively close to $q(1)$. As such, an agent who prefers no crime to severe crime with an expected penalty $Sq(1)$ likewise prefers no crime to moderate crime with an expected penalty of $s_M q(a_M)$ — unless s_M is very low.

Formally, the inequality $s_M^B > s_M^C$ is equivalent to $\frac{q(0)}{q(a_M-1)} > \frac{q(1)}{q(a_M)}$. As such, we must establish

$$q(1)\, q(a_M - 1) - q(0)\, q(a_M) < 0. \tag{6}$$

The lefthand side of (6) is strictly convex in a_M (since q is convex over negative values and concave over positive values), and is clearly equal to 0 when $a_M = 1$. Moreover, by the symmetry property of q (see Lemma 4) it is readily established that $q(1)\, q(-1) < q(0)\, q(0)$, and so the lefthand side of (6) is negative at $a_M = 0$. Consequently inequality (6) is indeed satisfied for any $a_M < 1$. ∎

It is important to note that Proposition 2 says only that when marginal penalties first have an impact on crime levels, that impact is positive. As will be clear from Proposition 7 below, it does not follow that the penalty s_M should be reduced all the way to 0.

We turn now to the opposite case to Proposition 2 in which crime waves do not occur — that is, the equilibrium selection rule is to pick the lowest crime equilibrium. As we will see, for many parameter values the introduction of marginal penalties actually increases the crime level in this case. Again, Figure 1 gives the basic intuition. Marginal penalties increase λ_{CW} but decrease λ_{NCW}, so that under the assumption that the lowest crime equilibrium is played the latter effect is the one that counts: the range of taste parameters for which no crime is an equilibrium shrinks, and so crime rises. This conclusion is avoided only if s_M^B is *much* higher than s_M^C, in which case as s_M is reduced from S the severe crime equilibrium range starts to shrink long before the no crime equilibrium range does.

To formalize this argument, we start by characterizing the equilibrium outcomes under the lowest crime equilibrium selection rule:

Proposition 3 *Suppose crime waves do not occur. Then the equilibrium is no crime ($a = 0$) whenever (IC0-1) and (IC0-M) both hold; the equilibrium is moderate crime ($a = a_M$) whenever at least one of (IC0-1) and (IC0-M) fail but (ICM-1) holds; and the equilibrium is severe crime ($a = 1$) otherwise.*

Proposition 3 allows us to characterize the equilibrium outcome using just three of the incentive constraints, namely (IC0-1), (IC0-M), and (ICM-1). Figures 2 and 3 graphically present the equilibrium outcome as a function of the extent of marginal penalties, s_M, and the realization of the taste parameter, λ. To understand the graphs, note first that each of (IC0-1), (IC0-M) and (ICM-1) can be rewritten as upper bounds on the taste-for-crime parameter λ:

$$\lambda \leq Sq(1) \tag{IC0-1}$$

$$\lambda \leq \frac{s_M q(a_M)}{a_M} \tag{IC0-M}$$

$$\lambda \leq \frac{Sq(1-a_M) - s_M q(0)}{1-a_M} \tag{ICM-1}$$

By the concavity of q over positive values, $q(a)/a > q(1)$ for any $a \in (0,1)$.[16] As such, the (IC0-1) line certainly falls below the maximum values of the (IC0-1) and (IC0-M) lines.

The (IC0-1) line may cross the (IC0-M) line at either a higher or lower value of s_M than at which it crosses the (ICM-1) line. Figures 2 and 3 correspond to the two cases.

Economically, the intersection of the (IC0-1) and (IC0-M) lines determines the highest penalty s_M such that, for some λ, no crime is an equilibrium under extremal penalties, but is not an equilibrium when the moderate crime penalty is instead s_M. That is, the intersection of (IC0-1) and (IC0-M) occurs at the penalty level where the costs of marginal penalties are first experienced, namely s_M^C.

In a similar fashion, the intersection of the (IC0-1) and (ICM-1) lines determines the highest penalty s_M such that, for some λ, severe crime is the only equilibrium under extremal penalties, but moderate crime is an equilibrium when the moderate crime penalty is instead s_M. The intersection of the (IC0-1) and (ICM-1) lines occurs at $s_M = \frac{S(q(1-a_M)-(1-a_M)q(1))}{q(0)}$.

Consequently, when

$$s_M^C > \frac{S(q(1-a_M) - (1-a_M)q(1))}{q(0)} \tag{7}$$

a reduction of s_M away from S induces costs (the collapse of the no crime equilibrium) before the benefits (the creation of a moderate crime equilibrium) are experienced.[17] This is the

[16] See footnote 12.

[17] This conclusion requires that S exceed the value of s_M at which the (IC0-M) and (ICM-1) intersect, i.e.,

$$S > \frac{Sa_M q(1-a_M)}{(1-a_M)q(a_M) + a_M q(0)}.$$

This certainly holds whenever assumption (2) does: by footnote 15, $q(0) > a_M q(1-a_M)$, and so the denominator of the righthand side exceeds $a_M q(1-a_M)$. However, even if assumption (2) is not satisfied this inequality trivially holds in the neighborhood of $a_M = 1/2$.

situation depicted in Figure 2. In this case, any choice of penalty $s_M > \frac{S(q(1-a_M)-(1-a_M)q(1))}{q(0)}$ clearly leads to a worse outcome than do extremal penalties. As such, extremal penalties are at least locally optimal. (Whether the optimal penalty is $s_M = S$, or lies in $\left[0, \frac{S(q(1-a_M)-(1-a_M)q(1))}{q(0)}\right]$, clearly depends on the distribution of the taste parameter λ and on the relative social costs of severe and moderate crime.)

We have thus established our main result, namely that marginal penalties help when crime waves occur, but increase crime when crime waves do not occur:

Proposition 4 *Suppose condition (7) holds. When crime waves occur extremal penalties are suboptimal, while when crime waves do not occur extremal penalties are locally optimal.*

To see when condition (7) holds, it is useful to substitute in for s_M^C and rearrange the inequality to

$$\frac{a_M q(0)}{q(a_M)}(q(1)-q(a_M)) > q(1-a_M) - a_M q(0) - (1-a_M)q(1).$$

The righthand side would be zero if the investigation probability function q were linear. As such, condition (7) is satisfied whenever the function q has sufficiently little curvature. In general, the function q has less curvature when the signal precision is low. Formally:

Proposition 5 *Whenever the signal precision h is low enough condition (7) holds and extremal penalties are locally optimal absent crime waves.*

Remark: It is also readily verified that (7) holds for all a_M sufficiently small.

4 Comparative statics under crime waves

In the previous section we showed that when crime waves occur (that is, when the equilibrium selection rule is to choose the highest crime equilibrium), some use of marginal penalties is always optimal. As we showed, this contrasts sharply to the situation in which crime waves do not occur. There remains the question of the level of marginal penalties that should be employed when crime waves occur. This the subject of the current section.

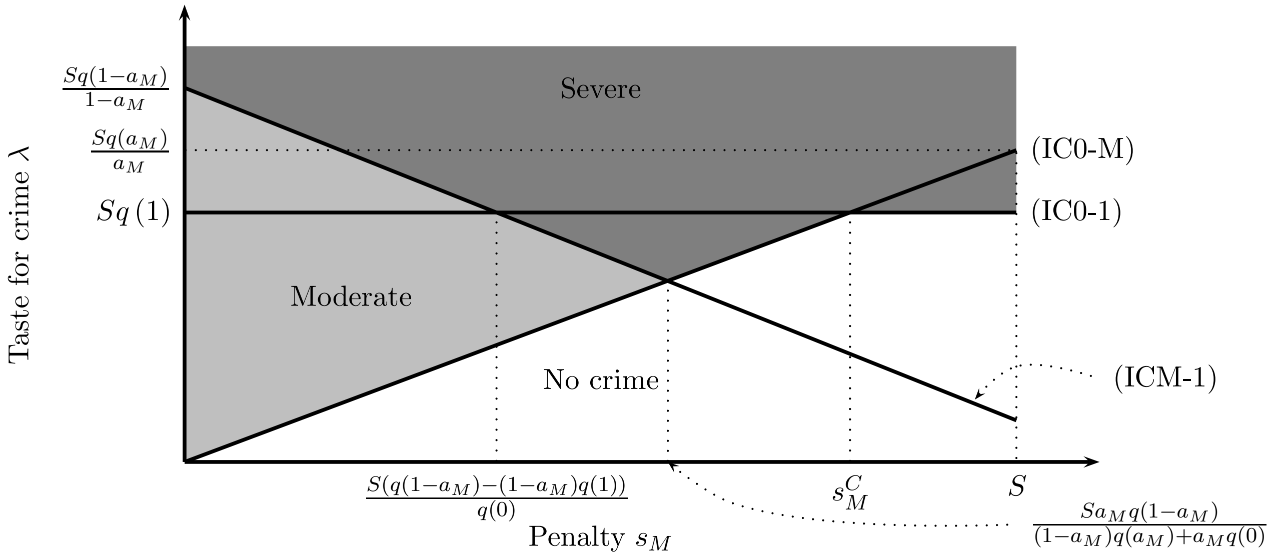

Figure 2: Marginal penalties increase crime when crime waves do not occur and $\frac{S(q(1-a_M)-(1-a_M)q(1))}{q(0)} < s_M^C$. The graph shows the lowest crime equilibrium for each moderate crime penalty s_M and taste-for-crime parameter λ.

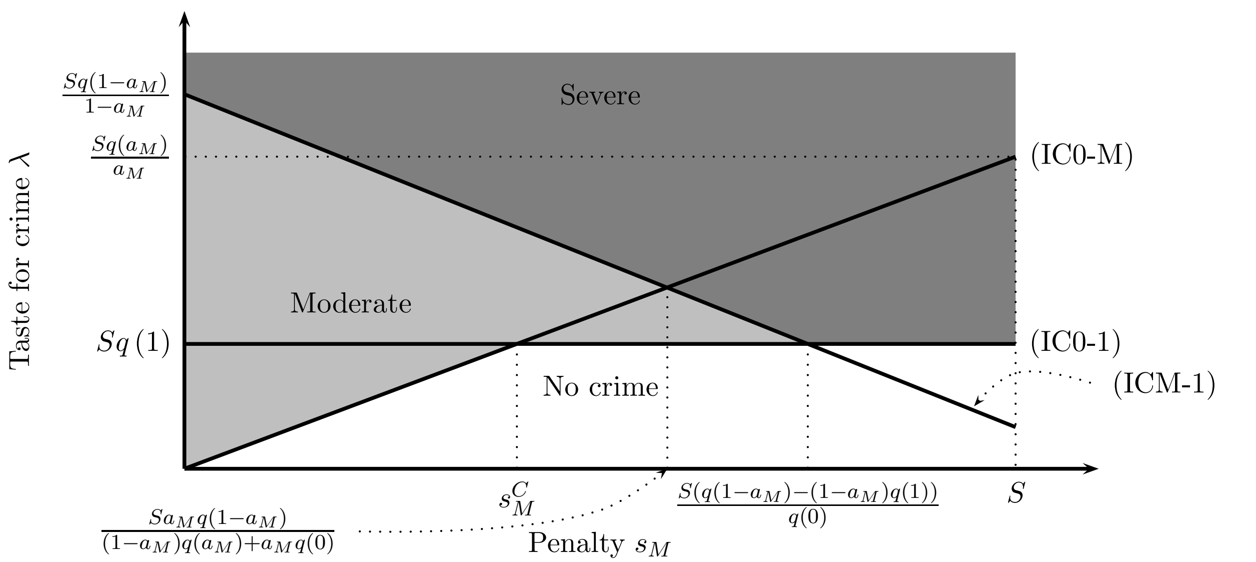

Figure 3: Marginal penalties reduce crime when crime waves do not occur and $\frac{S(q(1-a_M)-(1-a_M)q(1))}{q(0)} > s_M^C$. The graph shows the lowest crime equilibrium for each moderate crime penalty s_M and taste-for-crime parameter λ.

As discussed, marginal penalties generate a benefit only if they destroy a severe crime equilibrium that exists under extremal penalties. This occurs if s_M is set so that for some λ at which (IC1-0) holds (so that severe crime is an equilibrium under extremal penalties), (IC1-M) fails. As established, this requires that s_M be lowered at least as far as s_M^B.

When crime waves occur, marginal penalties have a cost only if for some λ, no crime is the only equilibrium under extremal penalties, but moderate crime is an equilibrium when the moderate crime penalty is instead s_M. That is, marginal penalties have a cost only if for some λ and s_M (IC1-0) fails (so that no crime is the equilibrium outcome under extremal penalties) while (ICM-0) holds (so that moderate crime is potentially an equilibrium). Algebraically this condition is

$$\overbrace{\frac{s_M q(0)}{a_M} \leq \lambda}^{\text{(ICM-0) holds}} \overbrace{< Sq(0)}^{\text{(IC1-0) fails}},$$

from which it is clear that marginal penalties have a cost under crime waves only if $s_M < Sa_M$. Observe that Sa_M is strictly less than s_M^C, the penalty at which the range of the no crime equilibria starts to shrink.

These observations permit us to characterize the equilibrium under crime waves using just three of the six equilibrium conditions, in a way analogous to Proposition 3:

Proposition 6 *Suppose that crime waves occur. Then the equilibrium is severe crime ($a = 1$) whenever (IC1-0) and (IC1-M) both hold; the equilibrium is moderate crime ($a = a_M$) whenever (IC1-M) fails but (ICM-0) holds; and the equilibrium is no crime ($a = 0$) otherwise.*

Given Proposition 6, we can graphically represent the equilibrium outcomes under crime waves in a way analogous to Figures 2 and 3. As before, it is useful to rewrite the incentive

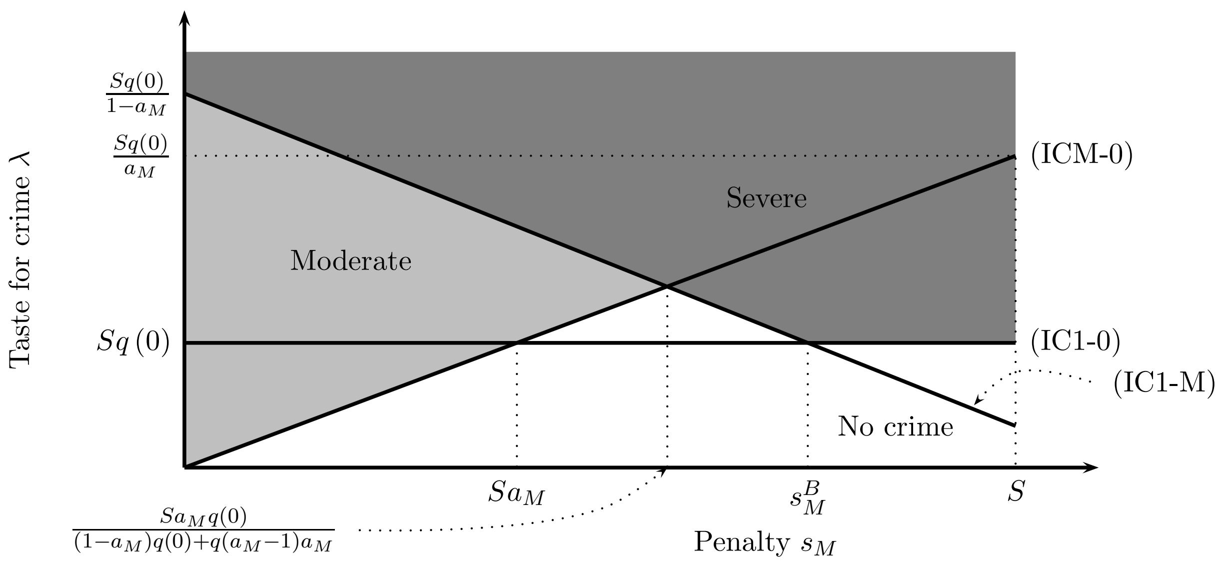

Figure 4: Equilibrium outcomes when crime waves occur. The graph shows the highest crime equilibrium for each moderate crime penalty s_M and taste-for-crime parameter λ.

constraints (IC1-0), (IC1-M) and (ICM-0) as bounds on λ:

$$\lambda \geq Sq(0) \tag{IC1-0}$$

$$\lambda \geq \frac{Sq(0) - s_M q(a_M - 1)}{1 - a_M} \tag{IC1-M}$$

$$\lambda \geq \frac{s_M q(0)}{a_M}. \tag{ICM-0}$$

From the text preceding Proposition 6, the lines (IC1-0) and (IC1-M) intersect at s_M^B, while the lines (IC1-0) and (ICM-0) intersect at $s_M = Sa_M$. Since clearly $s_M^B > Sa_M$ the (IC1-0) line lies below the intersection of (IC1-M) and (ICM-0) — see Figure 4. As earlier established in Proposition 2, when crime waves occur the introduction of marginal penalties always reduces the crime level.[18]

[18] The conclusion that marginal benefits initially help under crime waves requires

$$S > \frac{Sa_M q(0)}{(1-a_M)q(0) + q(a_M - 1)a_M}.$$

Assumption (2) implies that $s_M^B = \frac{Sa_M q(0)}{q(a_M - 1)} < S$, and so this inequality is certainly satisfied (see footnote 14). However, even if assumption (2) is not satisfied, this inequality trivially holds provided a_M is not much greater than 1/2.

Let g denote the probability density of the taste parameter λ. From Figure 4 one can easily calculate social welfare for each penalty level s_M, as follows. First, note that the penalty for moderate crime should be set below the intersection point of (ICM-0) and (IC1-M), i.e., below $\frac{Sa_M q(0)}{(1-a_M)q(0)+q(a_M-1)a_M}$, since above this point marginal penalities generate only benefits, and no cost. For penalties s_M below this level, social welfare under crime waves is given by

$$SW \equiv \int_{s_M q(0)/a_M}^{\frac{Sq(0)-s_M q(a_M-1)}{1-a_M}} (\lambda a_M - C_M)\, g(\lambda)\, d\lambda + \int_{\frac{Sq(0)-s_M q(a_M-1)}{1-a_M}}^{\bar{\lambda}} (\lambda - C_1)\, g(\lambda)\, d\lambda,$$

where the two terms represent the expected net social cost of moderate and severe crime respectively.[19] From this expression it is straightforward to show:

Proposition 7 *Suppose that crime waves occur. The optimal penalty for moderate crime lies in the interval* $\left[0, \frac{Sa_M q(0)}{(1-a_M)q(0)+q(a_M-1)a_M}\right]$, *and is decreasing in* C_1 *and increasing in* C_M.

The intuition behind Proposition 7 is clear. When the social cost of severe crime is high, society should do everything it can to curtail severe crime. Setting s_M low achieves this objective because it induces agents to switch from the severe crime to the moderate crime equilibrium. Of course, setting s_M low also engenders more moderate crime, but when C_M is low relative to C_1 this is a price worth paying.

The second comparative static we consider is the effect of signal precision h on the optimal choice of s_M. Signal precision affects social welfare via the investigation probability $q(a_M - 1)$, which is decreasing in precision h: when signal precision is high, the investigation probability faced by an agent who takes action a_M while the other takes action $a = 1$ is low. The investigation probability $q(a_M - 1)$ measures the strength of incentives provided by the investigation policy for agents to abstain from severe crime: if agent j chooses severe crime, agent i is punished with probability $q(0) = 1/2$ if he also chooses severe crime, but with probability $q(a_M - 1)$ if instead he chooses moderate crime. As such, when $q(a_M - 1)$ is

[19] One can also calculate social welfare without including the private benefits λa_M and λ from actions $a = a_M, 1$. Doing so has no qualitative effect on our results.

low, the investigation policy alone delivers considerable "marginality" in expected penalties even when s_M is close to the maximum penalty S. If instead $q(a_M - 1)$ is high, the investigation policy delivers little marginality, and it is worthwhile setting s_M much lower than S in order to increase the difference in expected penalties for severe and moderate crime.

The effect of $q(a_M - 1)$, and hence of signal precision h, on social welfare depends in part on the density g of the taste parameter λ. To abstract from these effects, we consider the special case in which λ is uniformally distributed:[20]

Proposition 8 *Suppose that crime waves occur and the taste parameter is distributed uniformally. The optimal penalty for moderate crime is decreasing in the investigation probability* $q(a_M - 1)$ *and so is increasing in signal precision* h.

5 Varying per capita enforcement resources

Thus far we have assumed that the enforcement agency has the resources to conduct one investigation for each two potential offenders. We have focused on this case purely for expositional convenience: our main results apply equally when instead the enforcement authority oversees $N \geq 2$ agents, and has the resources to investigate and penalize just one of them. Details are available from the authors' webpages. In this section, we analyze the effect of changes in N, that is, of changes in per capita enforcement resources.

With N agents, the probability of investigation faced by an agent choosing action $\tilde{a}$ while the other $N-1$ agents choose action a is given by $q_N(\tilde{a} - a)$, for some increasing function q. Clearly as the number of agents increases, the probability of investigation $q_N(x)$ decreases for any value of x. Moreover, it is possible to show that the ratio $q_N(x)/q_N(0)$ is increasing (respectively, decreasing) in the number of agents N if $x > 0$ (respectively, $x < 0$). That is, as N increases the probability of investigation decreases faster for an agent who commits a lesser crime, holding the actions of other agents fixed.

[20] More generally, Proposition 8 would hold whenever the derivative of the density function g' is sufficiently small.

We can use this result to consider the effect of changes in per capita enforcement resources on crime levels. Suppose that enforcement resources increase, i.e., N is reduced. This generates some unambiguously positive effects. For example, it increases the range of λ realizations for which (IC0-1) and (IC0-M) hold, implying that no crime ($a = 0$) is an equilibrium more often. However, for $s_M > 0$ a decrease in N has a tendency to make (ICM-1) less likely to hold, since as N decreases the ratio $q_N (1 - a_M) / q_N (0)$ decreases. Economically, a decrease in N increases $q_N (0)$ by a larger amount than $q_N (1 - a_M)$, and this makes it harder to prevent an agent deviating from moderate crime to severe crime. In this case, the increase in resources actually has a negative effect on the distribution of crime.

The above discussion gives the main intuition, but is somewhat loose in that (ICM-1) involves the terms $q_N (1 - a_M)$ and $q_N (0)$ separately, as opposed to in ratio. The following result gives one reasonably concise set of sufficient conditions for a decrease in N to have the negative effect described:[21]

Proposition 9 *Suppose that crime waves do not occur, and $a_M = 1/2$. Then for any moderate crime penalty $s_M \in (0, S/(1 + 1/N))$ and any number of agents $N > 2$, there exists a signal precision $\bar{h}$ such that whenever $h \geq \bar{h}$, an increase in per capita enforcement (i.e., a reduction in N) increases the probability of severe crime (though it also increases the probability of no crime).*

In general, an enforcement agency that experiences an increase to its budget can choose to spend these extra funds in one of two ways. On the one hand, the enforcement agency can seek to increase the precision of its pre-investigation information, that is, of the information upon which it decides whom to investigate (i.e., increase h). On the other hand, the enforcement agency can expand its capacity to investigate and penalize agents (i.e., decrease N). Between them, Corollary 1 and Proposition 9 imply that the former is often the more

[21] We emphasise that the effect described holds much more generally than the sufficient conditions of Proposition 9.

attractive alternative, since it leads to an unambiguous decrease in crime, while the latter has the potential to actually increase the crime level under some circumstances.

To some extent this prediction is consistent with the recent rise in "community policing." While the term encompasses a variety of distinct ideas, one important element is that police officers should spend more time patrolling streets by foot, and less time in patrol cars and responding to emergency calls.[22] This, it is often argued, will engender much better relations between police officers and the communities they oversee.[23] In terms of our analysis, this aspect of community policing can be thought of as corresponding to an increase in signal accuracy achieved at the cost of a decrease in investigation resources (increase in N). Enabling better communication between a community and its police is akin to increasing the amount of information a police department has on which to base its more formal investigations. However, taking police out of patrol cars reduces their ability to arrive promptly at a crime scene and possibly apprehend a criminal immediately.

6 Conclusion

Our analysis identifies a new answer to the old question of why non-maximal penalties are used to punish moderate actions. In environments in which multiple equilibria arise naturally, marginal penalties are much more attractive in the Pareto inferior crime wave equilibrium. Specifically, although marginal penalties have both costs and benefits (as identified by prior authors — see the introduction), the net benefit is strictly positive in the crime wave equilibrium. In contrast, for a wide range of parameter values marginal penalties have a net cost in the non-crime wave equilibrium.

The economic intuition for this difference is as follows. The potential benefit of marginal

[22] Perhaps the aspect of community policing to have attracted most comment is the "broken windows" theory, which emphasizes the importance of the eliminating small crimes (or at least their effects) for the control of more serious crimes. Although logically distinct from increasing the number of police officers on foot-patrol, in practice the two ideas are closely related.

[23] For example, in their much-cited article, Wilson and Kelling (1982) emphasize the awkwardness faced by an individual who wants to communicate with a police officer seated in a patrol car.

penalties is that severe crime is eliminated for some realizations of the taste parameter λ. The potential cost is that the no crime equilibrium is eliminated for other realizations of λ. In both instances, it is realizations of λ close to the "marginal" type that are impacted first. The key to our result is the marginal type has less taste for crime when crime waves are an issue (specifically, the marginal types are $Sq(0)$ and $Sq(1)$ under crime waves and no crime waves). Since it is easier to remove severe crime for low types than high types, the benefits of marginal penalties are reached earlier under crime waves. Conversely, high types switch more readily away from the no crime equilibrium, and so the costs of marginal deterrence are experienced earlier outside crime waves.

References

Marco Bassetto and Christopher Phelan. Tax riots. *Working Paper*, 2006.

Gary Becker. Crime and punishment: An economic approach. *Journal of Political Economy*, 76:169–217, March-April 1968.

Peter M DeMarzo, Michael J Fishman, and Kathleen M Hagerty. The optimal enforcement of insider trading regulations. *Journal of Political Economy*, 106(3):602–632, June 1998.

John Fender. A general equilibrium model of crime and punishment. *Journal of Behavior and Organization*, 39:437–453, 1999.

Peter J Jost. Crime, coordination, and punishment: an equilibrium analysis. *International Review of Law and Economics*, 21(1):23–46, March 2001.

Dilip Mookherjee and Ivan Png. Monitoring vis-a-vis investigation in enforcement of law. *American Economic Review*, 82(3):556–65, June 1992.

Dilip Mookherjee and Ivan Png. Marginal deterrence in enforcement of law. *Journal of Political Economy*, 102(5):1039–66, October 1994.

Joel Schrag and Suzanne Scotchmer. The self-reinforcing nature of crime. *International Review of Law and Economics*, 17(3):325–35, September 1997.

Steven Shavell. A note on marginal deterrence. *International Review of Law and Economics*, 12:345–355, 1992.

George J Stigler. The optimum enforcement of laws. *Journal of Political Economy*, 78(3): 526–536, May-June 1970.

Alexander Tabarrok. A simple model of crime waves, riots, and revolutions. *Atlantic Economic Journal*, 25(3):274–88, September 1997.

Louis L Wilde. Criminal choice, nonmonetary sanctions, and marginal deterrence: a normative analysis. *Internationl Review of Law and Economics*, 12:333–344, 1992.

James Q Wilson and George L Kelling. Broken windows. *Atlantic Monthly*, pages 29–38, March 1982.

A Proofs omitted from main text

Proof of Lemma 1: We will show that at least one of (ICM-1) and (IC1-M) holds — the other two claims follow by parallel arguments. Suppose that (ICM-1) does not hold, i.e.,

$$\lambda\left(1-a_M\right) > s_1 p\left(1, a_M\right) - s_M p\left(a_M, a_M\right).$$

To show that (IC1-M) holds it suffices to show

$$s_1 p\left(1, a_M\right) - s_M p\left(a_M, a_M\right) \geq s_1 p\left(1, 1\right) - s_M p\left(a_M, 1\right),$$

or equivalently,

$$s_1\left(p\left(1, a_M\right) - p\left(1, 1\right)\right) \geq s_M\left(p\left(a_M, a_M\right) - p\left(a_M, 1\right)\right). \tag{8}$$

From (1), $p(a_M, a_M) - p(a_M, 1) = p(1, a_M) + p(a_M, a_M) - 1$. This expression is non-negative since p is increasing in its first argument and $p(a_M, a_M) = 1/2$, and equals $p(1, a_M) - p(1, 1)$ since $p(1, 1) = 1/2$ also. The result then follows from $s_1 \geq s_M$. ∎

Proof of Lemma 2: Suppose that contrary to the claimed result no pure strategy equilibrium exists. Since $a = 0$ is not an equilibrium, at least one of (IC0-M) and (IC0-1) must fail to hold. Observe that by (1), p is increasing in its first argument if and only if it is decreasing in its second argument.

First, suppose that (IC0-M) fails to hold. From Lemma 1, (ICM-0) holds, and so (ICM-1) fails to hold (else $a = a_M$ is an equilibrium); and so (IC1-M) holds, and (IC1-0) fails to hold (else $a = 1$ is an equilibrium). So

$$\begin{aligned} -s_0 p(0, 1) > \lambda - s_1 p(1, 1) \geq \lambda - s_1 p(1, a_M) \\ > \lambda a_M - s_M p(a_M, a_M) \geq \lambda a_M - s_M p(a_M, 0) > -s_0 p(0, 0), \end{aligned}$$

a contradiction. (The strict inequalities follow from the failure of (IC1-0), (ICM-1) and (IC0-M) respectively; the weak inequalities and contradiction follow from the fact that p is decreasing in its second argument.)

Second, suppose that (IC0-1) fails to hold. Then (IC1-0) holds, and so (IC1-M) must fail to hold (else $a = 1$ is an equilibrium). This in turn implies that (ICM-1) holds, and (ICM-0) fails to hold (else $a = a_M$ is an equilibrium). So

$$-s_0 p(0, a_M) > \lambda a_M - s_M p(a_M, a_M) \geq \lambda - s_1 p(1, a_M) \geq \lambda - s_1 p(1, 0) > -s_0 p(0, 0),$$

a contradiction. (The inequalities follow from the failure of (ICM-0), (ICM-1), p decreasing in its second argument, and the failure of (IC0-1).) ∎

Proof of Lemma 3: The statement about the choice of penalties is established in the main text. For the statement regarding the investigation policy, it suffices to show that if any other investigation policy is used then shifting to the "investigate the agent with the higher signal" policy strictly increases $p(a_M, 0)$ and $p(1, 0)$, and strictly decreases $p(0, a_M)$ and $p(0, 1)$. The proof is as follows. For an arbitrary investigation policy μ and action

$a > 0$,

$$p\left(a,0\right) = \int\int \mu\left(a + \frac{\varepsilon^i}{h}, \frac{\varepsilon^j}{h}\right) f\left(\varepsilon^i\right) f\left(\varepsilon^j\right) d\varepsilon^i d\varepsilon^j.$$

Changing variables,

$$p\left(a,0\right) = \int\int \mu\left(\frac{\varepsilon^i}{h}, \frac{\varepsilon^j}{h}\right) \frac{f\left(\varepsilon^i - ah\right)}{f\left(\varepsilon^i\right)} f\left(\varepsilon^i\right) f\left(\varepsilon^j\right) d\varepsilon^i d\varepsilon^j.$$

Anonymity of μ implies that

$$\begin{aligned} p\left(a,0\right) = 1 - p\left(0,a\right) &= \int\int \left(1 - \mu\left(\frac{\varepsilon^i}{h}, a + \frac{\varepsilon^j}{h}\right)\right) f\left(\varepsilon^i\right) f\left(\varepsilon^j\right) d\varepsilon^i d\varepsilon^j \\ &= 1 - \int\int \mu\left(\frac{\varepsilon^i}{h}, \frac{\varepsilon^j}{h}\right) \frac{f\left(\varepsilon^j - ah\right)}{f\left(\varepsilon^j\right)} f\left(\varepsilon^i\right) f\left(\varepsilon^j\right) d\varepsilon^i d\varepsilon^j. \end{aligned}$$

Thus

$$p\left(a,0\right) = \frac{1}{2} + \frac{1}{2}\int\int \mu\left(\frac{\varepsilon^i}{h}, \frac{\varepsilon^j}{h}\right)\left(\frac{f\left(\varepsilon^i - ah\right)}{f\left(\varepsilon^i\right)} - \frac{f\left(\varepsilon^j - ah\right)}{f\left(\varepsilon^j\right)}\right) f\left(\varepsilon^i\right) f\left(\varepsilon^j\right) d\varepsilon^i d\varepsilon^j.$$

Log-concavity of f implies that $\frac{f\left(\varepsilon^i - ah\right)}{f\left(\varepsilon^i\right)} \geq \frac{f\left(\varepsilon^j - ah\right)}{f\left(\varepsilon^j\right)}$ if and only if $\varepsilon^j \leq \varepsilon^i$. As such, the investigation policy μ that maximizes $p\left(a,0\right)$ sets $\mu\left(x^i, x^j\right) = 1$ whenever $x^i > x^j$ and $\mu\left(x^i, x^j\right) = 0$ whenever $x^i < x^j$ — that is, the "investigate the agent with the higher signal" policy. ∎

Proof of Lemma 4: Part (I) follows from rewriting (1) in terms of q, which gives $q\left(a\right) + q\left(-a\right) = 1 = q\left(0\right) + q\left(0\right)$ for any $a \in [0,1]$. For Part (II), we use the explicit expression for the function q:

$$q\left(a\right) = \int \Pr\left(a + \frac{\varepsilon^i}{h} \geq \frac{\varepsilon^j}{h}\right) f\left(\varepsilon^j\right) d\varepsilon^j = \int \left(1 - F\left(\varepsilon - ha\right)\right) f\left(\varepsilon\right) d\varepsilon.$$

Clearly $q\left(a\right)$ is increasing (decreasing) in precision h when a is positive (negative). For Part (III), we first evaluate the derivatives of q:

$$\begin{aligned} q'\left(a\right) &= h\int f\left(\varepsilon - ha\right) f\left(\varepsilon\right) d\varepsilon > 0 \\ q''\left(a\right) &= -h^2 \int f'\left(\varepsilon - ha\right) f\left(\varepsilon\right) d\varepsilon. \end{aligned}$$

Integration by parts implies

$$q''(a) = h^2 \int f(\varepsilon - ha) f'(\varepsilon) \, d\varepsilon.$$

Since f is log-concave, f'/f is decreasing. As such, $f'(\varepsilon - ha)/f(\varepsilon - ha) > f'(\varepsilon)/f(\varepsilon)$ for any $a > 0$, which in turn implies

$$\int f(\varepsilon - ha) f'(\varepsilon) \, d\varepsilon < \int f'(\varepsilon - ha) f(\varepsilon) \, d\varepsilon = -q''(a).$$

Thus we have established that $q''(a) < -q''(a)$, and so $q(a)$ must be concave over positive values of a. The symmetry property of Part (I) implies that q is likewise convex over negative values. ∎

Proof of Proposition 3: The statement regarding no crime is immediate. Clearly moderate crime is only an equilibrium when (ICM-1) holds. As such, it remains only to show that (ICM-0) holds whenever (ICM-1) holds and at least one of (IC0-1) and (IC0-M) fails. There are two cases. The first (and easier) case is when (IC0-M) fails, for then (ICM-0) holds by Lemma 1. The second case is when (IC0-1) fails. Clearly (ICM-1) can hold while (IC0-1) fails only if for some λ

$$\overbrace{Sq(1) <}^{\text{(IC0-1) fails}} \lambda \overbrace{\leq \frac{Sq(1 - a_M) - s_M q(0)}{1 - a_M}}^{\text{(ICM-1) holds}}.$$

This requires that $s_M q(0) < Sq(1 - a_M) - (1 - a_M) Sq(1)$, and so

$$\frac{s_M q(0)}{a_M} < Sq(1) - S\frac{q(1) - q(1 - a_M)}{a_M} < Sq(1).$$

As such, whenever λ is high enough that (IC0-1) fails, it is also high enough that (ICM-0) holds. ∎

Proof of Proposition 5: Condition (7) is satisfied if and only if

$$Q(a_M) \equiv (1 - a_M) q(a_M) q(1) + a_M q(0) q(1) - q(a_M) q(1 - a_M) > 0.$$

Evaluating the derivatives of Q gives

$$
\begin{aligned}
Q'(a_M) &= q(0)\,q(1) + q(a_M)\left(q'(1-a_M) - q(1)\right) + q'(a_M)\left((1-a_M)\,q(1) - q(1-a_M)\right) \\
Q''(a_M) &= 2q'(a_M)\left(q'(1-a_M) - q(1)\right) \\
&\quad + q''(a_M)\left((1-a_M)\,q(1) - q(1-a_M)\right) - q(a_M)\,q''(1-a_M).
\end{aligned}
$$

Observe that $Q(0) = Q(1) = 0$. As such, it suffices to show that $Q''(a_M) < 0$ for all a_M. From the proof of Lemma 4,

$$
\frac{q''(a)}{q'(a)} = -h\frac{\int f'(\varepsilon - ha)\,f(\varepsilon)\,d\varepsilon}{\int f(\varepsilon - ha)\,f(\varepsilon)\,d\varepsilon}.
$$

At $h = 0$, $\int f(\varepsilon - ha)\,f(\varepsilon)\,d\varepsilon > 0$ while $\int f'(\varepsilon - ha)\,f(\varepsilon)\,d\varepsilon = 0$.[24] It follows that $\frac{Q''(a_M)}{q'(a_M)} \to -2q(1)$ as $h \to 0$.[25] This completes the proof. ■

Proof of Proposition 6: As a preliminary, observe that (IC1-0) holds whenever both (IC1-M) and (ICM-0) do. To see this, suppose to the contrary that for some penalty s_M and taste parameter λ (IC1-M) and (ICM-0) both hold but (IC1-0) fails. On the one hand, (ICM-0) can hold while (IC1-0) fails only if $s_M < Sa_M$ (see the text immediately prior to the proof). On the other hand, by reversing the inequalities in (4) it is clear that (IC1-M) can hold while (IC1-0) fails only if $s_M \geq s_M^B$. Since $s_M^B > Sa_M$ this delivers the required contradiction.

The Proposition's statement regarding severe crime is immediate. The statement regarding moderate crime is immediate given that (ICM-1) holds whenever (IC1-M) fails (see Lemma 1). For the remaining case, suppose that neither (IC1-0) and (IC1-M) both hold, nor that (IC1-M) fails but (ICM-0) holds. There are two remaining possibilities. If (IC1-M) and (ICM-0) both fail, the only remaining equilibrium candidate is no crime. Alternatively, if (IC1-M) holds then (IC1-0) must fail. The preliminary above then implies that (ICM-0) fails. Again, the only remaining equilibrium candidate is no crime. ■

Proof of Proposition 7: Differentiating the social welfare SW with respect to s_M gives:

$$
\frac{\partial SW}{\partial s_M} = -\left(\lambda_M a_M - C_M\right)\frac{q(0)}{a_M}g(\lambda_M) - \left(C_1 - C_M - \lambda_1(1-a_M)\right)\frac{q(a_M-1)}{1-a_M}g(\lambda_1),
$$

[24] Observe that $\int f'(\varepsilon)\,f(\varepsilon)\,d\varepsilon = \left[\frac{1}{2}f(\varepsilon)^2\right]_{-\infty}^{\infty} = 0$.

[25] To handle the final term, note that $\frac{q''(1-a_M)}{q'(a_M)} = \frac{q'(1-a_M)}{q'(a_M)}\frac{q''(1-a_M)}{q'(1-a_M)}$ and $\frac{q'(1-a_M)}{q'(a_M)} \to 1$ as $h \to 0$.

where $\lambda_M = \frac{s_M q(0)}{a_M}$ and $\lambda_1 = \frac{Sq(0) - s_M q(a_M - 1)}{1 - a_M}$. Substituting in,

$$\frac{\partial SW}{\partial s_M} = (C_M - s_M q(0)) \frac{q(0)}{a_M} g(\lambda_M) - (C_1 - C_M - Sq(0) + s_M q(a_M - 1)) \frac{q(a_M - 1)}{1 - a_M} g(\lambda_1).$$

Since the derivative is increasing in C_M and decreasing in C_1, the result follows. ∎

Proof of Proposition 8: From the proof of Proposition 7, the derivative $\partial SW / \partial s_M$ is decreasing in $q(a_M - 1)$, which implies the result. ∎

Proof of Proposition 9: We know that for any $x > 0$ and any N, $q_N(x) \to 1$ as signal precision $h \to \infty$. Consequently, since $a_M = 1/2$ the penalty level

$$\frac{S a_M q_N (1 - a_M)}{(1 - a_M) q_N (a_M) + a_M q_N (0)} \to \frac{S}{1 + \frac{1}{N}} \text{ as signal precision } h \to \infty.$$

As such, whenever precision is high enough the penalty s_M lies to the left of the intersection of the (IC0-1) and (IC0-M) lines. Moreover, we claim that at s_M the (ICM-1) line falls as N is reduced. From Figures 2 and 3 the result follows from this claim. Thus we must show that

$$Sq_N (1 - a_M) - s_M q_N (0) > Sq_{N-1} (1 - a_M) - s_M q_{N-1} (0).$$

Rearranging, and substituting in $q_N(0) = 1/N$,

$$\frac{s_M}{N(N-1)} > S (q_{N-1} (1 - a_M) - q_N (1 - a_M)).$$

The righthand side converges to 0 as signal precision $h \to \infty$, while the lefthand side is clearly independent of h. This completes the proof. ∎