Institute for Empirical Research in Economics
University of Zurich

Working Paper Series
ISSN 1424-0459

Working Paper No. 199

The Dynamics of Neighbourhood Watch and Norm Enforcement

Steffen Huck and Michael Kosfeld

August 2004

The Dynamics of Neighbourhood Watch and Norm Enforcement

Steffen Huck
University College London[*]

Michael Kosfeld
University of Zurich[†]

July 2004

Abstract

We analyze the dynamics of neighbourhood watch programs in a local interaction framework. Agents can watch their neighbours' houses and thus deter burglars from breaking in. At the same time, agents also try to recruit their neighbours to join the neighbourhood watch program. The probability of an agent joining the neighbourhood watch program depends on the success of the program, i.e., whether burglaries continue to occur. We show that the punishment of burglars plays a dual role in this context. On the one hand, punishment deters burglaries if the level of punishment is sufficiently high. On the other hand, it also affects the probability of an agent joining the neighbourhood watch program. In particular, we show that if recruitment is harder when burglaries do not occur, a legal policy attempting to improve deterrence using more severe punishment is suboptimal. In a second part, we extend our model to the study of norm enforcement in public goods dilemmas and show that our results remain valid if agents can punish each other (instead of burglars) for not contributing to the public good. Our paper thus provides a first analysis of the evolution of "altruistic punishment" in large populations with local interaction.

Journal of Economic Literature Classification: C72, K42

Keywords: neighbourhood watch, norm enforcement, cooperation, punishment

[*]Department of Economics and ELSE, University College London, Gower Street, London WC1E 6BT, UK, eMail: s.huck@ucl.ac.uk

[†]Institute for Empirical Research in Economics, University of Zurich, Blümlisalpstrasse 10, 8006 Zurich, Switzerland, eMail: kosfeld@iew.unizh.ch

"[Punishment] does not serve, or serves only incidentally, to correct the guilty person or to scare off any possible imitators. Its real function is to maintain inviolate the cohesion of society by sustaining the common consciousness in all its vigour."

— Emile Durkheim (1893)

"It's an unfortunate fact that when a neighborhood crime crisis goes away, so does enthusiasm for Neighborhood Watch. Work to keep your Watch group a vital force for community well-being."

— Los Angeles Police Department (2004)

1 Introduction

When a burglar breaks into a house nobody is in a better position to report the crime to the police than a neighbour, the most likely person to observe suspicious movements or hear suspicious noises. If called, the police can intervene and the state can prosecute and punish. For all practical purposes, however, the state cannot substitute for the watchful eyes of neighbours. This is what has made neighbourhood watch programs so successful over the last two decades (see, for example, Sims 2001), despite the costs involved. Neighbours who want to report a possible crime have to interrupt what they are doing and might be ashamed if the alarm turns out to be false. This makes the issue of recruitment of new members vital for any neighbourhood watch program.

In this paper, we analyze the conditions under which neighbourhood watch programs can be successfully sustained in the long run, when recruitment is an issue. In particular, we examine how the legal framework (for example, punishments of burglars) interacts with the recruitment dynamics of a neighbourhood watch program. This interaction turns out to be quite subtle. In particular, we find that increasing punishment for burglary can actually *increase* the number of burglaries in the long run. The intuition for this finding is basically summarized in the above quote from the LAPD. Neighbourhood watch programs can fall victim to their own successes. If recruitment of new members is easier as long as there are some crimes, a neighbourhood watch can deteriorate when crime rates fall as a reaction to more severe punishments of criminals. In fact, the deterioration may lead to the complete dissolution of a program, at which point the crime rates will again increase.

While most of our paper focuses on neighbourhood watch programs where the danger comes from a third party (the burglar), we also show how our model can be extended to norm enforcement in public goods dilemmas. Interestingly, it turns out that the dynamics of cooperation and punishment follow similar paths to those identified for the neighbourhood watch case (these themes have recently attracted considerable attention in the experimental literature; see, e.g. Fehr and Gächter 2000, Masclet, Noussair, Tucker, and Villeval 2003). In particular, excessive punishment can destroy cooperation in the long run.[1]

Our paper thus carries a twofold message. First, our model emphasizes the dynamics that underlie active *participation* of individuals in society. Participation cannot be taken for granted and may depend on institutions such as the law in rather subtle ways. Second, we show that punishment is a two-edged sword in this context. On the one hand, punishment deters burglars from burglarizing and induces free riders to cooperate if sufficiently high. On the other hand, punishment also affects participation, i.e., individuals' contributions in combating deviant behaviour. The latter process can induce surprising non-monotonic relations, meaning that higher punishment can actually cause more deviant behaviour. The intuition for this is that a certain degree of deviant behaviour and consequential punishment has to exist in order to remind society's members constantly of the given problem. Our model thus provides an analytical foundation for an intuitive argument, formulated at the end of the 19th century by Emil Durkheim in his analysis of the cohesion of modern society (see the quote above), which still plays an important role in current social and economic policy.

Finally, on a purely methodological level we also introduce some new and quite powerful techniques from particle system theory (see Liggett 1985 for an excellent introduction) in order to analyze local recruitment and interaction dynamics between neighbours. Particle system theory has been used earlier, for example, in models of evolutionary game theory (Blume 1993, Kosfeld 2002) and social interaction (Glaeser, Sacerdote, and Scheinkman 1996). Our main results are based on new findings from this theory.

The paper is organized as follows. Section 2 introduces the model of neighbourhood watch and presents our main results. Section 3 extends the analysis to norm enforcement in public goods dilemmas. Finally, section 4 provides a discussion and concludes. All proofs are collected in an appendix.

2 Neighbourhood Watch

We analyze a community that faces a threat from burglars. Burglars can only be detected if a neighbour sees and reports them to the police. (For convenience, we assume that burglars never try to rob a house when the owner is at home.) However, reporting suspicious behavior is costly and there are no immediate rewards for doing so. In particular, people might feel ashamed if the alarm turns out to be false. This is why the community sets up a neighbourhood watch program. Essentially, members of a neighbourhood watch program engage in two activities: they call the

[1]Recent papers by Bohnet, Frey, and Huck (2001) and Rege (2004) have analyzed a similar crowding out effect of government policy.

police when they see something suspicious and they try to recruit their neighbours to join the program. Thus, recruiting means basically convincing a neighbour that calling the police is the right thing to do if he observes something suspicious. Insofar, we will say that agents will call the police if and only if they have joined the program.

More specifically, we consider a population of agents located on the one-dimensional set of integers $\mathbb{Z}$. Each agent is identified by his location and denoted by $x, y, z \in \mathbb{Z}$.[2] Agents have two neighbours located to their left and right. Thus, for each agent $x \in \mathbb{Z}$, the set of neighbours is equal to $\{x-1, x+1\}$. There are two alternatives for each agent: either being a member of the neighbourhood watch program or not. $\mathcal{M}_t \subset \mathbb{Z}$ denotes the set of members at time t.

The number of neighbours who are members of the neighbourhood watch program determines both whether the police are called and whether the burglars are actually captured. We assume that a burglar who breaks into agent x's house is caught with probability $\alpha_1 > 0$ if only one neighbour is a member of the neighbourhood watch program, and with probability $\alpha_2 \geq \alpha_1$ if both neighbours are members. If neither neighbour is a member of the neighbourhood watch, nobody calls the police and therefore the probability of a burglar being caught is zero. A burglar whom the police capture receives a punishment $p > 0$ that the state sets. If we normalize a burglar's utility from robbing and not robbing an agent's house to one and zero respectively, and assume that burglars are risk-neutral, it follows that burglaries will be deterred if $\alpha p > 1$, while burglaries occur if $\alpha p < 1$, where $\alpha \in \{\alpha_1, \alpha_2\}$ depends on the number of members from the neighbourhood watch program in agent x's neighbourhood. More precisely, there exist two thresholds $\underline{p} = \frac{1}{\alpha_2} \leq \overline{p} = \frac{1}{\alpha_1}$, such that if p is small ($p < \underline{p}$), a burglary occurs regardless of how many neighbours watch agent x's house. If p is intermediate ($\underline{p} < p < \overline{p}$), a burglary occurs if only one neighbour is a member but is deterred if both neighbours are members. If p is large ($p > \overline{p}$), a burglary is deterred if at least one neighbour is a member of the program. Obviously, if $\alpha_1 = \alpha_2$, i.e., the probability of getting caught is independent of how many neighbours keep an eye on x's house, $\underline{p}$ and $\overline{p}$ coincide and hence the intermediate case vanishes.

The recruitment of new members for the neighbourhood watch program is modelled by a continuous-time Markov process. Our main assumptions are the following:

1. *(Drift)* There is a constant positive probability for any agent $x \in \mathcal{M}_t$ to leave the set $\mathcal{M}_t$. This drift captures some general laziness and the tendency to leave voluntary organizations at some later point in time for all sorts of exogenous reasons.

2. *(Recruitment)* A neighbour must convince a perspective member to join $\mathcal{M}_t$. Hence, if neither neighbour of x is a member, the probability of x becoming a member of $\mathcal{M}_t$ is zero.

[2]Thus, the population of agents considered in our model is infinite. The analysis of an infinite model is legitimate because *(i)* the behavior of the infinite model captures important features of large finite population models at large finite times and *(ii)* the analysis of these features is simpler in an infinite than in a finite model (cf. Liggett 1999, pp71).

If, on the other hand, at least one neighbour is in $\mathcal{M}_t$, there is a strictly positive probability of x joining $\mathcal{M}_t$, as well.

3. *(Crime Crisis)* The likelihood of becoming a member of $\mathcal{M}_t$ may depend on the success of the neighbourhood watch program. It may be easier to convince someone to join the program when it has been very successful and there is little crime (*no crime crises*). Alternatively, people might find it more compelling to become a member of $\mathcal{M}_t$ if lots of burglaries occur (*crime crises*).

The transition probabilities of the Markov process are given by individual Poisson rates $m(x, \mathcal{M}_t)$ with $x \in \mathbb{Z}$ and $\mathcal{M}_t \subset \mathbb{Z}$. These rates determine the probability that agent x will change his membership status within an infinitesimally short period of time, given the state of the process $\mathcal{M}_t$.[3] Let $n_t(x) \in \{0, 1, 2\}$ denote the number of agent x's neighbours who are members of the neighbourhood watch program at time t, i.e., $n_t(x) = |\mathcal{M}_t \cap \{x-1, x+1\}|$. Then rates are defined as follows:

$$m(x, \mathcal{M}_t) = \begin{cases} \epsilon & \text{if} \quad x \in \mathcal{M}_t, \\ f(n_t(x), p, \alpha_1, \alpha_2) & \text{if} \quad x \notin \mathcal{M}_t. \end{cases} \tag{1}$$

The parameter $\epsilon > 0$ captures the constant drift away from membership due to different exogenous reasons, of which laziness might be one. The function f models the recruitment dynamics at the local level. As described above, it depends on the number of members in agent x's neighbourhood and whether burglaries occur or not (the latter depending on the punishment level p and the probabilities α_1 and α_2). We assume that f takes value $\chi > 0$ if at least one neighbour is already a member and burglaries occur *(crime crisis)* and value $\pi > 0$ if at least one neighbour is a member but burglaries do not occur *(no crime crisis)*. If neither neighbour is a member, the agent cannot be recruited and f is therefore zero:

$$f(n_t(x), p, \alpha_1, \alpha_2) = \begin{cases} 0 & \text{if } n_t(x) = 0, \\ \chi & \text{if } n_t(x) \geq 1 \text{ and burglaries occur}, \\ \pi & \text{if } n_t(x) \geq 1 \text{ and no burglaries occur}. \end{cases} \tag{2}$$

Recall from above that burglaries occur only if $p < \underline{p}$, or if $\underline{p} < p < \overline{p}$ and $n_t(x) \leq 1$. In all other cases, i.e., if $\underline{p} < p < \overline{p}$ and $n_t = 2$, or if $p > \overline{p}$ and $n_t \geq 1$, no burglaries happen because the neighbourhood watch together with a sufficiently high punishment level successfully deters burglars.

[3]The Markov process we consider is a so-called interacting particle system. See Liggett (1985) for an introduction. Intuitively, just as in the case of a standard Poisson process, rates (that take values in $[0, \infty]$) and probabilities (that take values in $[0, 1]$) form a monotone relation: the higher the rate, the higher the probability of a type change within a short period of time. For example, a Poisson rate equal to infinity implies an instantaneous type change, i.e., the probability of a type change equals one. On the other hand, a Poisson rate equal to zero corresponds to a situation where the probability of a type change is zero, as well.

Our goal is to analyze the evolutionary dynamics of $\{\mathcal{M}_t\}_{t\geq 0}$. In particular, we are interested in the effects of an exogenous increase in the punishment level p on the stability of neighbourhood watch. To analyze the dynamics we need, of course, some assumptions about the initial state — basically we have to ensure that there are a sufficient number of initial program members to initiate the recruitment process at all. More technically, we take the following approach. Suppose that each agent is programmed at time zero with probability q of watching his neighbours' house and is not programmed to do so with remaining probability $1-q$. For example, the government and the local police may start a large policy campaign in favour of neighbourhood watch. Suppose that q is strictly positive (however small). The following questions then arise: first, can $\mathcal{M}_t$ be non-empty at every time $t > 0$ and second, can it have sufficient density asymptotically.[4] The following three results provide an answer to the first question. All proofs are collected in the appendix.

Proposition 1 *For every $\epsilon > 0$ there exists a critical value $s(\epsilon) < \infty$ such that:*
(A) if $\max\{\pi, \chi\} < s(\epsilon)$, independently of p the unique limit of $\mathcal{M}_t$ is the empty set,
(B) if $\min\{\pi, \chi\} > s(\epsilon)$, independently of p with probability one $\mathcal{M}_t$ is non-empty for every $t > 0$.
Upper and lower bounds for $s(\epsilon)$ are given by $1.224\epsilon \leq s(\epsilon) \leq 2.17\epsilon$.

Proposition 1 gives a precise condition on transition rates for ensuring non-emptiness of $\mathcal{M}_t$ for all $t > 0$. Simply said, both π and χ must have sufficient size. More precisely, they must be larger than a finite critical value $s(\epsilon)$. If both rates are smaller, the neighbourhood watch program will disintegrate. The value of $s(\epsilon)$ depends purely on the drift rate ϵ and, as bounds on $s(\epsilon)$ indicate, this dependency is monotonic: the larger ϵ, the larger is $s(\epsilon)$. Thus, the result in Proposition 1 can also be formulated the other way round. If the rates π and χ are fixed, then a positive value exists such that $\mathcal{M}_t$ is non-empty for all $t > 0$ if the drift rate is smaller than this value. In consequence, it is not necessary to let the drift rate go to zero to ensure asymptotic non-emptiness of the set $\mathcal{M}_t$.[5] The program can permanently lose some members and yet survive.

Note that the optimal policy for deterring burglaries in case *(B)* is to choose a punishment level $p > \overline{p}$. In case *(A)*, any level of p is obviously ineffective and hence irrelevant.

The following two Propositions deal with the cases where π and χ lie on opposite sides of the critical value $s(\epsilon)$, i.e., either $\chi < s(\epsilon) < \pi$ (Proposition 2) or $\pi < s(\epsilon) < \chi$ (Proposition 3). In view of the preceding result, we say that the neighbourhood watch *breaks down* if, as in case *(A)*, the unique limit of $\mathcal{M}_t$ is the empty set, whereas the neighbourhood watch *survives* if, as in case *(B)*, the set $\mathcal{M}_t$ is always non-empty.

[4]Note that once $\mathcal{M}_t$ is empty, it will remain empty forever as the empty set forms a trap for the process.

[5]Our model differs with this respect from standard mutation models in evolutionary theory that assume that mutations eventually disappear (e.g., Kandori, Mailath, and Rob 1993, Young 1993).

Proposition 2 *Let $\epsilon > 0$ and suppose $\chi < s(\epsilon) < \pi \leq 2\chi$. There exists a critical value $\tilde{s}(\theta, \epsilon) \leq s(\epsilon)$ that depends on the ratio $\theta = \frac{\pi}{\chi}$ such that:*
(C) the neighbourhood watch program breaks down for $p < \underline{p}$ and survives for $p > \underline{p}$ if $\chi > \tilde{s}(\theta, \epsilon)$,
(D) the neighbourhood watch program breaks down for $p < \overline{p}$ and survives for $p > \overline{p}$ if $\chi < \tilde{s}(\theta, \epsilon)$.

Proposition 2 shows that an increase in p can only lead to survival of the neighbourhood watch program. If χ is larger than a certain threshold $\tilde{s}(\theta, \epsilon)$, survival will already occur at $p > \underline{p}$; if not, survival occurs at $p > \overline{p}$. Therefore, an increase in p is an optimal policy for a legislator. The intuition behind Proposition 2 is that *both* deterrence *and* recruitment effects are raised. Figure 1 illustrates the four possible cases that can arise if $\chi \leq \pi$.

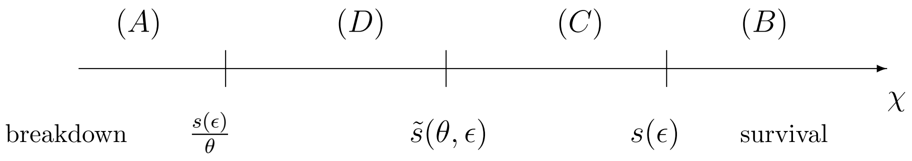

Figure 1: $\chi < s(\epsilon) < \pi \leq 2\chi$

While an increase in p is always optimal in Propositions 1 and 2, the next result shows that this is no longer true if the asymmetry between π and χ is reversed, i.e., when recruitment probabilities decrease as burglaries disappear and parameters do not lie on the same side of the critical value. Unfortunately, the evolution of $\{\mathcal{M}_t\}_{t\geq 0}$ is harder to analyze in this case. The reason is that the process is no longer attractive, the latter requiring that the probability of an agent joining $\mathcal{M}_t$ is non-decreasing in the number of neighbours who are already in $\mathcal{M}_t$. In consequence, we do not have a picture as complete as before.[6]

Proposition 3 *Let $\epsilon > 0$ and suppose $\pi < s(\epsilon) < \chi$.*
(E) The neighbourhood watch program survives for $p < \underline{p}$ and breaks down for $p > \overline{p}$.

Proposition 3 reveals an important result in our model which supports the intuition of the Los Angeles police department stated at the beginning of the paper. As before, consider a legislator who may influence the level of punishment p. Suppose that the initial situation is such that $p < \underline{p}$. Since χ is large enough, the neighbourhood watch survives. However, burglaries happen because p is too low to deter effectively. A plausible strategy is to increase p. Indeed, if $p > \underline{p}$ burglaries are successfully deterred whenever both neighbours of x are members of the program. However, exactly these situations also result in a lower recruitment effect on agent x (namely π).

[6]Precisely, in case (E) we can neither guarantee survival nor breakdown if $\underline{p} < p < \overline{p}$. Note, however, that this case does not exist if $\alpha_1 = \alpha_2$.

In consequence, we can no longer guarantee survival. If p is further increased to $p > \overline{p}$ to make deterrence even more effective, the neighbourhood watch eventually breaks down. In this case the "standard policy" of achieving better deterrence by more severe punishment is suboptimal. Figure 2 summarizes the scenario if $\pi < \chi$.

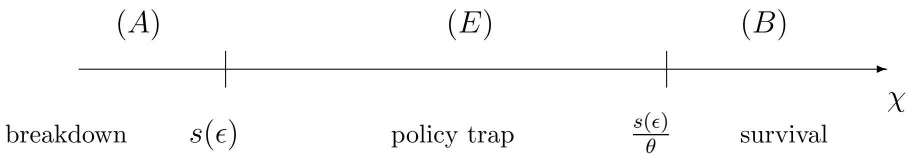

Figure 2: $\pi < s(\epsilon) < \chi$

Given that we can ensure survival of the process, a key question concerns the asymptotic frequency of the neighbourhood watch program members in the overall population. An answer to this question is important because only a sufficiently large fraction of members will be able to create a reasonable deterrence on the global level. For simplicity, we consider only the boundary cases where exactly one recruitment effect is at work.[7] That is, either $p < \underline{p}$, so the recruitment rate equals χ if $n \geq 1$ and zero otherwise, or $p > \overline{p}$, in which case the rate equals π if $n \geq 1$ and zero otherwise. The latter is of particular interest for a policy maker, since it can induce the highest degree of deterrence. Recall from Proposition 1 that the threshold guaranteeing survival is $\min\{\pi, \chi\} \geq 2.17\epsilon$.

Proposition 4 *Let* $\epsilon > 0$ *and suppose* $\min\{\pi, \chi\} \geq 2.17\epsilon$. *Then the asymptotic probability of an agent* $x \in \mathbb{Z}$ *being a member of* $\mathcal{M}_t$ *is bounded below,*

$$\lim_{t \to \infty} Prob(x \in \mathcal{M}_t) \geq \frac{1}{\psi}, \tag{3}$$

where ψ *is given by the equation*

$$\psi = \lambda - \sqrt{\lambda^2 - 2\lambda - \frac{\lambda - 1}{\lambda + 1}}, \tag{4}$$

and λ *equals the relative recruitment rate* $\frac{\chi}{\epsilon}$ *if* $p < \underline{p}$ *and* $\frac{\pi}{\epsilon}$ *if* $p > \overline{p}$.

Remark. The left hand side in equation (3) is indeed independent of x since the limiting distribution of $\mathcal{M}_t$ is translation invariant (cf. Konno 1994).

Note first that as the relative recruitment rate λ approaches infinity, i.e., either π and χ go to infinity or ϵ goes to zero, ψ converges to one. Thus, the asymptotic probability of agents joining the program converges to one as well: every agent is a member. Secondly, this probability is already quite large for relatively small values of λ. Figure 3 illustrates $\frac{1}{\psi}$ for different values of λ.

[7] These are the only cases if $\alpha_1 = \alpha_2$.

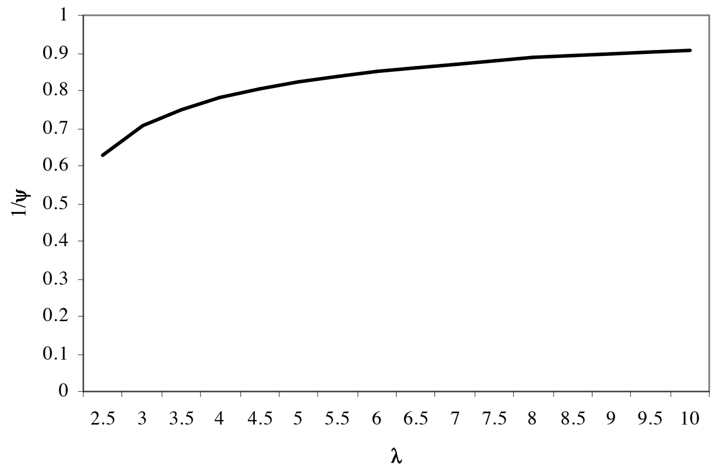

Figure 3: Asymptotic density of neighbourhood watch

For example, if recruitment rates are four times larger than the drift rate, the asymptotic probability of any agent being a member of the program is larger than 0.75. In other words, there are three members in the population for every non-member. If recruitment rates are five times larger, only every fifth agent does not join the program. Even if recruitment rates only exceed the drift by a factor of 2.5, asymptotically more than 60% of the population are members of the neighbourhood watch program. This shows that, once the program attains survival, the fraction of members will be large; and it will be large even for relatively small values of λ.

3 Norm Enforcement

We considered a situation of neighbourhood watch in the previous section where agents keep an eye on their neighbour's house in order to deter burglaries. In this section, we show how the model can be extended to study the dynamics of cooperation and punishment in local public goods games.

Following the work of Fehr and Gächter (2000, 2002), the experimental analysis of so-called "altruistic punishment" has attracted considerable attention in recent years (see, e.g., Falk, Fehr, and Fischbacher 2001, Anderson and Putterman 2003, Carpenter 2003a,b, Masclet *et al.* 2003). These studies have provided solid evidence for the fact that many individuals are willing to punish free-riding behaviour in social dilemma games, even if punishment is costly and players face a one-shot interaction. Falk *et al.* (2001) show that non-strategic factors (such as norms of fairness or spite), rather than strategic concerns (as, e.g., reputation or future payoff calculations) largely are the major drivers of the motivation to punish. While the authors are "unable to detect a significant impact of strategic forces on sanctioning behavior, non-strategic sanctions are large and significant" (Falk *et al.* 2001, p4). Given this empirical evidence, an important question is how altruistic punishment and norm enforcement might evolve in a large society where some agents are willing to punish free-riding behaviour while others are not, and repeated game effects are absent. In particular, the question is whether non-strategic motives will suffice for the survival of altruistic punishment, or whether norm enforcement will eventually die out if players do not take future payoffs into account. Fortunately, we can directly apply our model from the previous section to give an answer to this question.

Suppose that neighbours play the following simple public goods games. Each agent $x \in \mathbb{Z}$ has one unit of an endowment, which he can invest in a local public good or keep for himself. Investment in the public good decreases one's own payoff by one and at the same time also increases each neighbour's payoff by one. Thus, investing is costly on an individual basis but beneficial overall. Denoting the investment decision of agent x by $i_x \in \{0, 1\}$, the payoff of agent x is equal to

$$\Pi_x = 1 - i_x + \sum_{y \in \{x-1, x+1\}} i_y \tag{5}$$

Assuming that agents maximize their individual payoffs, no one will invest in the public good. In consequence, each agent earns a payoff of 1 while agents could earn a payoff of 2 if everyone invested in the public good. But now suppose that agents can try to punish any neighbours who did not invest in the public good, for example, by "naming and shaming" them. Agents who have been named and shamed might be ostracized, in which case they experience a utility loss p.

Following the experimental evidence, we assume that punishment of defectors is a matter of agents' types rather than a consequence of future payoff calculations, i.e., there are some agents who punish (P-types) and others who don't. This gives rise to a setup very similar to that in the previous section. A defector's payoff is reduced by p with probability $\alpha_1(\alpha_2)$ if one (two) neighbour(s) are P-types. In consequence, agents will cooperate if p is sufficiently high and at least one neighbour "names and shames".[8] In particular, cooperation is optimal if $p > \overline{p} = \frac{1}{\alpha_1}$ and at least one neighbour is a P-type or if $\overline{p} > p > \underline{p} = \frac{1}{\alpha_2}$ and both neighbours are P-types. Investment in the public good is not optimal in all other cases.

As above, agents can change their types, where imitation of role models (in combination with a constant drift) drives the dynamics of type change.[9] This means we assume agent x will become a P-type with positive probability if at least one neighbour is a P-type. Furthermore, agent x's likelihood of becoming a P-type depends on whether his neighbours' types successfully deter him or

[8]We assume that also P-types maximize their monetary income in the public goods game for the following two reasons. First, we do not want to make our lives too easy. If there was a link between punishing and cooperation, there would always be weakly more cooperators across any given population compared to the present model. Second, experimental evidence also shows that punishers do not cooperate all the time (Falk et *al.* 2001).

[9]See Offerman, Potters, and Sonnemans (2002) who provide experimental evidence for imitation of "exemplary behaviour".

not. In the first case, x experiences only the threat of being named and shamed (which successfully deters him). In the second case, he actually experiences ostracism after having defected. Both experiences may change x's type. In both cases, x may also become a P-type, as his neighbours may serve as a role model. In analogy to our previous analysis, we can assume that a rate χ drives the likelihood of agent x becoming a norm-enforcing P-type himself in the latter situation, where punishment is actually executed, while a rate π drives it in the former situation, where there is only the (successful) threat of punishment.

This setup of norm enforcement is identical to the previous model of neighbourhood watch, with the only difference that the neighbours themselves (instead of burglars) can behave badly. While neighbourhood watch in our earlier analysis is supposed to deter crime from the outside, punishment is supposed to induce cooperation among neighbours in this case. The previous results can be transferred one-to-one to the present setup. As a consequence, we find that increasing levels of ostracism can actually cause more defection in the long run. The intuition is as before. As long as the utility loss from ostracism is low, agents will defect and they will be ostracized for not contributing to the public good. If the level of ostracism is raised, at some point ($p > \overline{p}$) defection no longer pays if at least one neighbour names and shames. Agents now cooperate and in consequence also ostracism no longer takes place. If, however, the actual experience of ostracism induces agents to name and shame, the disappearance of ostracism will cause the number of P-types to decline steadily. This will continue until naming and shaming eventually disappears completely, at which point defection will take over and will become the norm.

4 Conclusion

We have studied the (local) dynamics of neighbourhood watch programs and norm enforcement, and how public policy setting fines for criminal or deviant behavior can drive the rise and fall of such programs. Surprisingly, we found that increasing punishment can have adverse consequences. If the survival of neighbourhood watch programs depends on successful recruiting, deterrence can be *too* effective. As crime rates fall, recruitment for neighbourhood watch programs can become harder and, ultimately, programs might be destroyed by their own success. Once they are destroyed, crime rates can and will pick up again. This suggests that optimal policy might aim at a tolerable low crime rate rather than total prevention of crime. Of course, should neighbourhood watch programs actually become victims of their own success, the state has also the option of campaigning for neighbourhood watch programs or subsidizing them. Such campaigns would be equivalent to a second drift term in our model, which would transform non-members into (founding) members of (new) programs. Obviously, such campaigns would, if vigorous enough, offset the deterioration caused by the "bad" drift term.

We also show that the dynamics of cooperation and punishment in public goods provision schemes, that has recently attracted considerable attention in the literature, can follow similar paths. In particular, we show that self-enforcing systems of cooperation and punishment can survive if a tendency to punish deviant behaviour is "learned" from punishing neighbours. However, as in the neighbourhood watch model, we find that punishment can be *too* severe. A successful dynamic system of cooperation and punishment needs a constant influx of new agents who are willing to spend resources on punishment. And this might require punishment sometimes be actually carried out. Again, mild levels of antisocial behaviour can actually help sustain a system of norm enforcement. Our analysis therefore suggests in some sense, that policy targets combatting antisocial or criminal behaviour should not be too ambitious. Furthermore, our analysis also spells out a rather pronounced warning that the standard law-and-order approach, assuming that higher punishments will always reduce crime, may not necessarily work.

Two avenues for further research seem obvious. Theoretically, an analysis of the dynamics in more complex spaces and, in particular, in endogenously formed neighbourhoods (that is in models where agents can move) would be of interest. However, neighbourhood watch and any forms of local social control are also in desperate need for more empirical work. Searching Google for "neighborhood watch" in edu-domains or "neighbourhood watch" in ac.uk-domains yields dozens and dozens of neighbourhood watch programs which universities participate in, but virtually no data and research. This is particularly surprising, as neighbourhood watch programs are a rather interesting object from the perspective of the broader social capital, trust, and reciprocity literature.

Appendix: Proofs

Proposition 1 to 4 are implications of results for a particular class of stochastic processes: the *contact process*, the *threshold contact process*, and the *θ-contact process* (Liggett 1991, Konno 1994). The contact process is a continuous-time Markov process with state space $\{A|A \subset \mathbb{Z}\}$. Sites in A are regarded as infected, whereas the other sites are regarded as being healthy. The contact process has transition rates where each infected site independently recovers at rate 1, and a healthy site becomes infected at rate λ times the number of neighbours that are infected, with $\lambda > 0$. The transition rates of the threshold contact process are such that an infected site recovers at rate 1, while a healthy site is infected at rate $\lambda > 0$ if *at least one* neighbour is infected. The θ-contact process forms a generalization of the two processes: each infected site recovers at rate 1; a healthy site is infected at rate λ if one neighbour is infected, and it is infected at rate $\theta\lambda$ if both neighbours are infected, where $1 \leq \theta \leq 2$. Obviously, the θ-contact process coincides with the basic contact process if $\theta = 2$ and coincides with the threshold contact process if $\theta = 1$.

The model of neighbourhood watch in this paper represents a particular combination of the threshold and the θ-contact process. To see this, simply regard each infected site as an agent being a member of $\mathcal{M}_t$. By dividing Poisson rates by ϵ we can transform our model into an *equivalent* contact-process model that has rate 1 for an infected site to recover, and rates $\frac{\chi}{\epsilon}$ and $\frac{\pi}{\epsilon}$ for a healthy site to become infected (or, in our terminology to become a member of the neighbourhood watch program). The two models are equivalent in the sense that their asymptotic behavior is the same. The division by the positive number ϵ only affects the time scale.

Now, if only one kind of recruitment effect is at work (either π or χ), our model is equivalent to the threshold contact process. If both effects are at work and $\chi \leq \pi \leq 2\chi$, the model is equivalent to the θ-contact process with $\theta = \frac{\pi}{\chi}$. In fact, if both effects are at work and $\pi < \chi$, our model is also a θ-contact process; however this time $\theta < 1$. Unfortunately, not so much is know in this case as the process is not attractive, which requires that the probability of a site becoming infected does not decrease in the number of neighbouring sites that are infected.

However, If $\theta \geq 1$ the θ-contact process is attractive. A well-known consequence (Liggett 1985, Theorem 2.3, Chapter III) is that the so-called *upper invariant measure*

$$\overline{\nu} = \lim_{t \to \infty} \delta_{\mathbb{Z}} S(t) \tag{6}$$

exists. Here $\delta_{\mathbb{Z}}$ denotes the Dirac measure that puts probability one on the state where every site is infected and $S(t)$ denotes the semigroup (the continuous-time analogue to the transition matrix) of the process. Of course, it may well be that $\overline{\nu} = \delta_{\emptyset}$, the latter denoting the Dirac measure where with probability one no site is infected. A main result, however, is that this is not the case if λ is large enough.

Precisely, for the threshold contact process there exists a critical value λ_c such that $\overline{\nu} = \delta_{\emptyset}$ if $\lambda < \lambda_c$ and $\overline{\nu} \neq \delta_{\emptyset}$ if $\lambda > \lambda_c$. Moreover, if $\lambda > \lambda_c$, it holds that $\overline{\nu}(\emptyset) = 0$, so in this case with probability one the set of infected sites is non-empty. In consequence, it is said that the process *survives* if $\lambda > \lambda_c$ and that it *dies out* if $\lambda < \lambda_c$. If the process survives, we obtain convergence to $\overline{\nu}$ from any translation invariant distribution putting mass zero on $\emptyset$. In particular, there is convergence starting from the Bernoulli product measure with strictly positive infection probability $q > 0$. Liggett (1991) and Konno (1994) provide lower and upper bounds for the critical value λ_c of the threshold contact process showing that

$$1.224 \leq \lambda_c \leq 2.17. \tag{7}$$

The θ-contact process has a critical value $\lambda_c(\theta)$ as well, such that survival occurs if $\lambda > \lambda_c(\theta)$ and the process dies out if $\lambda < \lambda_c(\theta)$. For $1 \leq \theta \leq 2$, it can be shown that $\lambda_c(\theta)$ is strictly decreasing in θ and that $\theta\lambda_c(\theta)$ is strictly increasing in θ (Durrett and Griffeath 1983). Obviously, $\lambda_c(1) = \lambda_c$. This implies that $\frac{\lambda_c}{\theta} < \lambda_c(\theta) < \lambda_c$.

Via multiplication with $\epsilon > 0$ we obtain the equivalent critical values for our model. For example, λ_c translates into $s(\epsilon) = \epsilon\lambda_c$. Similarly, the critical value $\tilde{s}(\theta, \epsilon)$ is obtained through $\tilde{s}(\theta, \epsilon) = \epsilon\lambda_c(\theta)$. This provides us with enough information to prove our results.

Proof of Proposition 1: *Case (A)*: Suppose

$$m = \max\{\pi, \chi\} \quad < \quad s(\epsilon) \tag{8}$$

$$\Leftrightarrow \frac{m}{\epsilon} \quad < \quad \lambda_c. \tag{9}$$

Then the threshold contact process with infection rate $\lambda = \frac{m}{\epsilon}$ dies out. Equivalently, the system of neighbourhood watch with single recruitment rate m breaks down. By definition of m, and using a standard dominance argument, the original neighbourhood watch program with rates π and χ must break down as well. *Case (B)*: Suppose

$$m = \min\{\pi, \chi\} \quad > \quad s(\epsilon) \tag{10}$$

$$\Leftrightarrow \frac{m}{\epsilon} \quad > \quad \lambda_c. \tag{11}$$

In this case the threshold contact process with infection rate $\lambda = \frac{m}{\epsilon}$ survives. Equivalently, neighhourhood watch with single recruitment rate m survives. Again, by the same dominance argument, the same holds for the original process The bounds for $s(\epsilon)$ follow immediately from those for λ_c. □

Proof of Proposition 2: Suppose $\chi < s(\epsilon) < \pi \leq 2\chi$. If $p < \underline{p}$, burglaries occur and the recruitment rate is always equal to χ. Hence, the process of neighbourhood watch is equivalent to the threshold contact process with infection rate $\lambda = \frac{\chi}{\epsilon}$. If $p > \overline{p}$, no burglaries occur and the recruitment rate is always equal to π. Thus, in this case the process is equivalent to the threshold contact process with infection rate $\tilde{\lambda} = \frac{\pi}{\epsilon}$. Since $\chi < s(\epsilon) < \pi$, or equivalently, $\lambda < \lambda_c < \tilde{\lambda}$, the threshold contact process with infection rate λ dies out but the threshold contact process with infection rate $\tilde{\lambda}$ survives. We are thus left with the situation where $\underline{p} < p < \overline{p}$. Let $\theta = \frac{\pi}{\chi}$. *Case (C)*: If $\chi > \tilde{s}(\theta, \epsilon)$, the θ-contact process with infection rate $\lambda = \frac{\chi}{\epsilon}$ survives. Equivalently, the neighbourhood watch program with recruitment rates χ and π survives. *Case (D)*: If $\chi < \tilde{s}(\theta, \epsilon)$, the θ-contact process with infection rate $\lambda = \frac{\chi}{\epsilon}$ dies out and therefore also the neighbourhood watch program with recruitment rates χ and π breaks down. □

Proof of Proposition 3: If $p < \underline{p}$, burglaries occur and the recruitment rate is always equal to χ. Hence, the process of neighbourhood watch is equivalent to the threshold contact process with infection rate $\lambda = \frac{\chi}{\epsilon}$. If $p > \overline{p}$, no burglaries occur and the recruitment rate is always equal to π. Thus, in this case neighbourhood watch is equivalent to the threshold contact process with infection rate $\tilde{\lambda} = \frac{\pi}{\epsilon}$. Since $\pi < s(\epsilon) < \chi$, or equivalently, $\tilde{\lambda} < \lambda_c < \lambda$, the threshold contact process with infection rate λ survives, but the threshold contact process with infection rate $\tilde{\lambda}$ dies out. □

Proof of Proposition 4: The result follows from Katori and Konno (1993), who prove that the density of the upper invariant measure of the θ-contact process has the following lower bound. Let $1 \leq \theta \leq 2$. Define

$$\xi = \frac{2-\theta}{\lambda} + (\theta - 1), \tag{12}$$

$$\eta = (2-\theta) + \theta\lambda, \tag{13}$$

and

$$\psi = \frac{\eta}{1+\xi}\left(1 - \sqrt{1 - \frac{2(1+\xi)}{\eta} - \frac{1-\xi^2}{\eta^2}}\right). \tag{14}$$

Then

$$\overline{\nu}(A|\, x \in A) \geq \frac{1}{\psi} \tag{15}$$

for any $x \in \mathbb{Z}$ and $\lambda \geq \lambda_U(\theta)$, where $\lambda_U(\theta)$ is the upper bound for the critical value of the θ-contact process, which is given by the largest root of the cubic equation

$$\theta\lambda^3 - (3\theta - 2)\lambda^2 - 3(2-\theta)\lambda + (2-\theta) = 0. \tag{16}$$

As Proposition 4 considers the case of the threshold contact process only, i.e., $\theta = 1$, the above equations become much simpler. First, $\xi = \frac{1}{\lambda}$ and $\eta = 1 + \lambda$. Putting this into equation (14) and using some basic algebra leads to

$$\psi = \lambda - \sqrt{\lambda^2 - 2\lambda - \frac{\lambda - 1}{\lambda + 1}}. \tag{17}$$

If $\theta = 1$, $\lambda_U(\theta)$ coincides with the upper bound for the critical value of the threshold contact process, which is 2.17, thereby concluding the proof. □

References

Anderson, C.M. and Putterman, L. (2003). 'Do Non-strategic Sanctions Obey the Law of Demand? The Demand for Punishment in the Voluntary Contribution Mechanism', Brown University Economics Working Paper No. 03-15.

Carpenter, J.P. (2003a). 'The Demand for Punishment', mimeo, Middlebury College.

Carpenter, J.P. (2003b). 'Punishing Free-Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods', mimeo, Middlebury College.

Blume, L.E. (1993). 'The Statistical Mechanics of Strategic Interaction', *Games and Economic Behavior*, vol. 5, pp. 387-424.

Bohnet, I., Frey, B.S., and Huck, S. (2001) 'More Order with less Law: On Contract Enforcement, Trust, and Crowding', *American Political Science Review*, vol. 95, pp. 131-144.

Durkheim, E. (1893). *De la Division du Travail Social*, Alcan, Paris. (Translated by Halls, W.D., *The Division of Labor in Society*, 1984, The Free Press, New York.)

Durrett, R. and Griffeath, D. (1983). 'Supercritical Contact Processes on $\mathbb{Z}$', *Annals of Probability*, vol. 11, pp. 1-15.

Falk, A., Fehr, E., and Fischbacher, U. (2001). 'Driving Forces of Informal Sanctions', Working Paper No. 59, Institute for Empirical Research in Economic, University of Zurich.

Fehr, E. and Gächter, S. (2000). 'Cooperation and Punishment in Public Good Experiments', *American Economic Review*, vol. 90, pp. 980-994.

Fehr, E. and Gächter, S. (2002). 'Altruistic Punishment in Humans', *Nature*, vol. 415, pp. 980-994.

Glaeser, E.L., Sacerdote, B. and Scheinkman, J.A. (1996). 'Crime and Social Interactions', *Quarterly Journal of Economics*, vol. 111, pp. 507-548.

Kandori, M., Mailath, G.J., and Rob, R. (1993). 'Learning, Mutation and Long Run Equilibria in Games', *Econometrica*, vol. 61, pp. 29-56.

Katori, M. and Konno, N. (1993). 'Bounds for the critical line of the θ-contact process with $1 \leq \theta \leq 2$', *Journal of Physics A*, vol. 26, pp. 6597-6614.

Konno, N. (1994). *Phase Transitions of Interacting Particle Systems*, World Scientific, Singapore.

Kosfeld, M. (2002). 'Stochastic Strategy Adjustment in Coordination Games', *Economic Theory*, vol. 20, pp. 321-339.

Liggett, T.M. (1985). *Interacting Particle Systems*, Springer, New York.

Liggett, T.M. (1991). 'The Periodic Threshold Contact Process,' in R. Durrett and H. Kesten (Eds.), *Random Walks, Brownian Motion and Interacting Particle Systems*, Birkhäuser, Boston.

Liggett, T.M. (1999). *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*, Springer, New York.

Masclet, D., Noussair, Ch., Tucker, S. and Villeval, M.-C. (2003). 'Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism', *American Economic Review*, vol. 93, pp. 366-380.

Offerman, T., Potters, J., and Sonnemans, J. (2002). 'Imitation and Belief Learning in an Oligopoly Experiment', *Review of Economic Studies*, vol. 69, pp. 973-997.

Rege, M. (2004). 'Social Norms and Private Provisions of Public Goods', *Journal of Public Economic Theory*, vol. 6, pp. 65-77.

Sims, L. (2001). 'Neighbourhood Watch: Findings from the 2000 British Crime Survey', *Home Office Findings* 150.

Young, H.P. (1993). 'The Evolution of Conventions', *Econometrica*, vol. 61, pp. 57-84.