The dynamics of neighbourhood watch and norm enforcement

Huck, S; Kosfeld, M
The dynamics of neighbourhood watch and norm enforcement

Abstract

We propose a dynamic model of neighbourhood watch schemes. While the state chooses punishment levels, apprehension of criminals depends on the watchfulness of citizens. We show that, contrary to standard intuition, crime levels can increase in punishments. This is because neighbourhood watch schemes can fall victim to their own success if recruitment of new members is driven by fear of crime - a finding that is in line with the empirical literature. We discuss the policy implications of this result and show how it extends to the more general problem of norm enforcement among interacting citizens.
The Dynamics of Neighbourhood Watch and Norm Enforcement*

Steffen Huck and Michael Kosfeld

Abstract

We propose a dynamic model of neighbourhood watch schemes. While the state chooses punishment levels, apprehension of criminals depends on the watchfulness of citizens. We show that, contrary to standard intuition, crime levels can increase in punishments. This is because neighbourhood watch schemes can fall victim to their own success if recruitment of new members is driven by fear of crime — a finding that is in line with the empirical literature. We discuss the policy implications of this result and show how it extends to the more general problem of norm enforcement among interacting citizens.

Pagehead Title: Dynamics of Neighbourhood Watch

JEL classifications: K42, C72

*Published in: The Economic Journal 117, 2007, 270-286. We are grateful to Armin Falk, Manfred Königstein, Norio Konno, Thomas Liggett, Larry Samuelson, Dolf Talman, Christian Zehnder, two referees, and two editors for very helpful comments and suggestions. We also wish to thank seminar participants at the University of Zurich, the Kennedy School of Government, and the University of Wisconsin-Madison for useful discussions on previous versions of the paper. The first author acknowledges financial support from the German Science Foundation (DFG) and the Economic and Social Research Council (UK). The second author’s acknowledges support from the European Commission and the Wissenschaftskolleg Berlin.
“[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 (2005)

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 to 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 popular over the last two decades. Neighbourhood watch programs have existed in the UK since the early 1980s. In 2000, the National Neighbourhood Watch Association estimated that over 155,000 programs operated throughout the UK (Sims, 2001). Results from the British Crime Survey show that the percentage of households in England and Wales that participate in a neighbourhood watch scheme almost doubled from 14 percent in 1988 to 27 percent (i.e., to about six million households) in 2000 (Dowds and Mayhew, 1992; Sims, 2001).¹

The key element for a successful neighbourhood watch program is its members’ voluntary active participation. Active participation, however, involves costs. For example, 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. The objective of this paper is to present a theoretical model for

¹Neighbourhood watch programs are similarly popular (and even more popular) in other countries. The 2000 National Crime Preventions Survey reported that 41 percent of the US population lived in areas that were covered by neighbourhood watch and that 58 percent of people who knew that they have a neighbourhood watch available participated in it (National Crime Prevention Council, 2001). A recent study for Australia estimates that 31 percent of Australian households in 2004 participated in a neighbourhood watch scheme (Johnson, 2005).
analyzing the conditions under which neighbourhood watch programs can be successfully sustained when recruitment of new members is an issue.

We study a local-interaction model of agents who participate in a neighbourhood watch program to protect themselves against burglars. Members of the program call the police if they see a burglar at their neighbours’ house. We assume that, if called, the police catch the burglar with positive probability and the state inflicts a punishment upon the burglar. If nobody calls the police, however, burglars are never caught and hence are never punished. Thus, successful deterrence in our model relies on both effective punishment by the state and active monitoring by neighbours. Simply put, neighbourhood watch and state punishment are complements. We assume that agents’ participation in neighbourhood watch is driven by two opposing dynamics. On the one hand, non-members are recruited by neighbours, who are members, to join the program; on the other hand, agents who are members leave the program for all sorts of exogenous reasons. The latter dynamic is captured by an exogenous drift term in our model. The former recruitment dynamic depends on the success of the neighbourhood watch program, i.e., may vary depending on whether or not burglaries occur in the neighbourhood.

Our results show that if recruitment of new members is sufficiently strong (in particular in response to the program’s success, i.e., the absence of burglaries), neighbourhood watch can be sustained and crime rates are bounded above whenever punishment is effective. Moreover, crime rates fall very quickly and converge to zero in the limit, as the rate of recruitment of new members rises. This is the key positive message from our analysis, which suggests that neighbourhood watch can provide a powerful instrument for successful community crime prevention. There is a caveat, however. If recruitment is easy in times of a crime crisis but difficult in the absence of burglaries, neighbourhood watch can be sustained if and only if a certain, non-zero, burglary rate remains. The intuition is that as soon as burglaries disappear completely, new members are less likely to join the program and hence the neighbourhood watch will be slowly hollowed out due to the constant outflow of members. Since neighbours’ monitoring is essential in order to achieve effective deterrence of burglars, burglary rates will eventually rise again and reach maximum levels in response to the decrease in participation. Thus, full deterrence in the short run leads to zero
deterrence in the long run. We show that the optimal punishment policy in this situation is to set punishment at an intermediate level that does not deter burglaries completely but maintains a certain crime rate in order to keep participation in neighbourhood watch active.

At first glance, this second result might seem counterintuitive: neighbourhood watch programs can fall victim to their own success. Yet, the result captures an important and very practical issue, namely the problem of maintaining a watch program when the primary aim (the reduction of crime) has been achieved. Lindsay and McGillis (1986), for example, document this problem in their assessment of the Seattle Community Crime Prevention Project, one of the first neighbourhood watch programmes in the US. The statement from the LAPD at the beginning of the paper provides a recent illustration of this phenomenon. In fact, the notion that a certain crime rate might be necessary to maintain active participation in neighbourhood watch programs has also been emphasized in empirical work on community crime prevention (Rosenbaum, 1987; Hope, 1988).

Hope (1988), for example, finds that individual support for neighbourhood watch correlates with anxieties about becoming the victim of burglary. Using data from the 1984 British Crime Survey, Hope analyzes the likelihood of a person’s support for neighbourhood watch based on personal characteristics, attitudes and experiences. He finds that people who worry about becoming the victim of burglary are more likely to support neighbourhood watch than people who do not do so. Furthermore, support for neighbourhood watch seems to be increasing in the frequency of burglaries in the neighbourhood. Hope concludes that “a certain level of ‘fear arousal’ may be necessary to stimulate a decision to join Neighbourhood Watch” (Hope, 1988, p. 154). This is exactly what our result predicts. The result is also in line with Rosenbaum (1987) who argues, based on early evidence from neighbourhood watch programs in the US, that “moderate increases in perceived vulnerability may be necessary to induce behaviour change directed at minimizing the risk of victimization.” (Rosenbaum, 1987, p. 129).

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
“altruistic 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, 2002; Masclet et al., 2003). In particular, we show that high levels of punishment can hollow out players’ enthusiasm for norm enforcement and hence destroy cooperation in the long run.\footnote{Rege (2004) analyzes a similar crowding out effect of government policy.}

In sum, our paper 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: higher punishment can cause more deviant behaviour. The main 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.\footnote{In this sense, our model provides an analytical foundation of a famous proposition put forward by Emil Durkheim in his 19th century analysis of the cohesion of modern society (see the quote at the beginning of the paper).}

Finally, on a 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 was used earlier, for example, in models of evolutionary game theory (Blume, 1993; Kosfeld, 2002) and social interaction (Glaeser et al., 1996). Our main results are based on new findings from this theory.

The paper is organized as follows. Section 1 introduces the model of neighbourhood watch and presents our main results. Section 2 extends the analysis to norm enforcement in public goods dilemmas. Finally, section 3 provides a discussion and concludes. All proofs are collected in an appendix.
1 Neighbourhood Watch

We analyze a community that faces a threat from burglars. Burglars can only be detected by watching neighbours who may call the police. (For convenience, we assume that burglars never try to rob a house when the owner is at home.) But reporting suspicious behavior is costly and there are no immediate rewards for doing so. For example, 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 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 they observe 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 $\mathbb{Z}$. An agent is identified by his location and denoted by $x \in \mathbb{Z}$. 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 an agent’s house and not robbing it all 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 $\bar{p} = \frac{1}{\alpha_2} \leq \underline{p} = \frac{1}{\alpha_1}$, 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 ($\underline{p} < p < \bar{p}$), a burglary occurs if only one neighbour is a member but is deterred if both neighbours are members. If $p$ is large ($p > \bar{p}$), 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 $\bar{p}$ 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. 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*).

Formally, the transition probabilities of the Markov process are given by individual Poisson rates $m(x, \mathcal{M}_t)$ with $x \in \mathcal{Z}$ and $\mathcal{M}_t \subset \mathcal{Z}$. 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 $\mathcal{M}_t$.\footnote{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} 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 $t$. Then rates are defined as follows:
\[
m(x, \mathcal{M}_t) = \begin{cases} 
\epsilon & \text{if } x \in \mathcal{M}_t, \\
f(n_t(x), p, \alpha_1, \alpha_2) & \text{if } x \notin \mathcal{M}_t.
\end{cases}
\] (1)

The parameter \( \epsilon > 0 \) captures the constant drift away from membership due to different exogenous reasons, of which laziness might be one. Another interpretation of \( \epsilon \) is that it captures agents moving away and being replaced by new neighbours who, by default, are not yet member of the scheme. 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 \( \alpha_1 \) and \( \alpha_2 \)).

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:

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

Recall from above that burglaries occur only if \( p < \underline{p}, \) if \( \underline{p} < p < \overline{p} \) and \( n_t(x) \leq 1, \) or if \( p > \overline{p} \) and \( n_t(x) = 0. \) 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.

Our goal is to analyze the evolutionary dynamics of the process \( \{\mathcal{M}_t\}_{t \geq 0} \). 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 \( q \) to watch his neighbours’ house and is programmed not to do so with remaining probability \( 1 - q. \)
the government and the local police may start a large policy campaign in favour of neighbour-
hood 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.\(^5\) The following three results provide an answer to the first question. All proofs
are collected in an appendix.

**Proposition 1** For every positive drift rate $\epsilon$ there exists a finite critical value $s(\epsilon)$ 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 recruitment rates for ensuring non-emptiness of $\mathcal{M}_t$
for all $t > 0$. Simply said, both $\pi$ and $\chi$ 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 $\epsilon$ and, as bounds
on $s(\epsilon)$ indicate, this dependency is monotonic: the larger $\epsilon$, the larger is $s(\epsilon)$. Thus, the result
in Proposition 1 can also be formulated the other way round. If the recruitment rates $\pi$ and $\chi$
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 $\mathcal{M}_t$.\(^6\) The program can permanently lose some members and
yet survive.

The following two Propositions deal with the cases where $\pi$ and $\chi$ 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.

\(^5\)Note that once $\mathcal{M}_t$ is empty, it will remain empty forever as the empty set forms a trap for the process.

\(^6\)Our model differs with this respect from standard mutation models in evolutionary theory that assume that
mutations eventually disappear (e.g., Kandori et al., 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 $p < p$ and survives for $p > \overline{p}$ 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 says that if $\chi < s(\epsilon) < \pi$, an increase in $p$ is beneficial since it can only lead to survival of the neighbourhood watch program. Note that $\chi$ is the smaller of the two recruitment rates and, by assumption, below the critical value $s(\epsilon)$. If $\chi$ is not too small, i.e., larger than a certain threshold $\tilde{s}(\theta, \epsilon)$, the programm will yet survive for $p > \overline{p}$; if not, survival occurs at $p > \overline{p}$. In the first case, both recruitment rates $\chi$ and $\pi$ are at work; in the second case, only $\pi$ is at work (and hence $\chi$ is irrelevant).

While high levels of punishment are beneficial in Proposition 2, the next result shows that this is no longer true if the asymmetry between $\pi$ and $\chi$ is reversed, i.e., when recruitment rates decrease as burglaries disappear and parameters do not lie on the same side of the critical value. Unfortunately, the evolution of $\{M_t\}_{t \geq 0}$ 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 $M_t$ is non-decreasing in the number of neighbours who are already in $M_t$. In consequence, we do not have a picture as complete as before.\(^7\)

Proposition 3 Let $\epsilon > 0$ and suppose $\pi < s(\epsilon) < \chi$.

(E) The neighbourhood watch program survives for $p < \overline{p}$ and breaks down for $p > \overline{p}$.

Proposition 3 reveals an important result in our model. If $\pi < s(\epsilon) < \chi$, an increase in $p$ is detrimental to the stability of an existing neighbourhood watch program. While the program survives as long as the punishment of burglars is low it breaks down if punishment is high. The intuition for the break-down at high punishment levels $p > \overline{p}$ is the following. Consider a string of agents who are members of the neighbourhood watch program at some time $t > 0$. Gradually, due to drift some members become non-members. For the program to be sustained it is essential that these non-members become members again, i.e., are re-recruited by adjacent agents who are still

\(^7\)Precisely, in case (E) we can neither guarantee survival nor breakdown if $p < p < \overline{p}$. 

members. However, since punishment is high no burglaries occur and therefore the recruitment rate is low. So most of these agents do not become members. Meanwhile the drift pushes other members into non-membership, including some of the neighbours of the above agents. Once both neighbours of an agent are non-members the agent is no longer recruited. The program is gradually hollowed out, and since recruitment of new (and old) members is too weak to compensate for the constant outflow, the program breaks down.

Figure 1 summarizes our findings on the dynamics of neighbourhood watch. Proposition 1 shows that if both recruitment rates are low (A), the program breaks down, while if both rates are high (B), the program survives. In both cases, the dynamics are independent of the punishment the state sets. If recruitment is relatively easy when no burglaries occur but is hard in times of a crime crises (C and D), neighbourhood watch survives if and only if the level of punishment is sufficiently high (Proposition 2). However, if recruitment is easy when there is a crime crisis but is hard when no burglaries occur (E), punishment has an adverse effect on the dynamic stability of the program. While the program survives if punishment is low, it breaks down if punishment is high (Proposition 3).

![Fig. 1. The Dynamics of Neighbourhood Watch](image-url)
Given the dynamics of neighbourhood watch we can now solve for the optimal punishment the state can set. For this we have to bear in mind that punishment has two effects in our model. On the one hand, it directly determines the incentive for burglars to burglarize (deterrence effect). On the other hand, it also determines indirectly the rate of recruitment of new members (recruitment effect) and hence the dynamic stability of the program. The direct effect is straightforward: higher punishment always leads to greater deterrence of burglars. The indirect effect, however, can go both ways. If it is easier to recruit members when crime rates are low ($\pi > \chi$), higher punishment leads to less burglaries and hence increases the rate of recruitment. If, instead, agents are less likely to join the neighbourhood watch scheme if no or only few burglaries occur ($\pi < \chi$), an increase of $p$ leads to greater deterrence but reduces the rate of recruitment. Let us consider the different situations as illustrated in Figure 1. In case (A), any punishment is obviously ineffective and hence irrelevant. Since recruitment rates are too low, the neighbourhood watch program breaks down independent of $p$. In case (B), the program survives independent of $p$ and hence the optimal policy is to choose a high punishment $p > \bar{p}$ since this generates the greatest degree of deterrence. This policy is also optimal in case (C) and (D), where the direct deterrence effect and the indirect recruitment effect go into the same direction. Although survival in case (C) occurs already if $p > p_1$, high punishment $p > \bar{p}$ is optimal because it yields the lower crime rate. In case (D), high punishment is necessary both for optimal deterrence and survival of the neighbourhood watch program. The most complex situation is case (E). In this case, a punishment increase has a positive direct effect on deterrence but has a negative indirect effect on recruitment. This suggests a trade-off policy with regard to optimal punishment. In fact, if punishment is low ($p < \bar{p}$) the direct effect dominates the indirect effect. The program survives but punishment is insufficient to deter any burglary. If punishment is high ($p > \bar{p}$), however, the indirect effect dominates the direct effect. Increasing punishments beyond $\bar{p}$ initially reduces crime rates (wherever at least one neighbour is a member of the neighbourhood watch program), but since recruitment rates are now too low the program eventually breaks down. The optimal policy in case (E) is to set punishment at an intermediate level $\underline{p} < p < \bar{p}$. This leads to a lesser degree of deterrence (and thus to a higher crime rate) than high punishment but circumvents the guaranteed breakdown of the program. In
fact, it is exactly the incidence of some crime that keeps the recruitment of program members active. Proposition 4 summarizes our findings with regard to the optimal punishment policy.

**Proposition 4 (Optimal punishment)** The optimal policy is to choose high punishment \( p > \bar{p} \) if the recruitment of new members for the neighbourhood watch program is relatively easy in the absence of crime, i.e., if \( \pi > s(\epsilon) \). If recruitment is easy in times of a crime crises, i.e., \( \chi > s(\epsilon) \), but is hard when the crime crisis goes away, i.e., \( \pi < s(\epsilon) \), the optimal punishment lies at an intermediate level \( \underline{p} < p < \bar{p} \).

Suppose now that survival of the neighbourhood watch program is ensured. A key question is then the asymptotic frequency of program members in the overall population and the resulting asymptotic burglary rate. 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 case of high punishment \( p > \bar{p} \), i.e., where the recruitment rate equals \( \pi \) if \( n \geq 1 \) and zero otherwise. This case is also 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 \( \pi > 2.17\epsilon \).

**Proposition 5 (Asymptotic density)** Let \( \epsilon > 0 \) and suppose \( \pi > 2.17\epsilon \). The asymptotic probability of an agent \( x \in \mathcal{Z} \) being a member of \( \mathcal{M}_t \) is bounded below,

\[
\lim_{t \to \infty} \text{Prob}(x \in \mathcal{M}_t) \geq \frac{1}{\psi},
\]

where \( \psi \) is given by the equation

\[
\psi = \lambda - \frac{\sqrt{\lambda^2 - 2\lambda - \frac{\lambda - 1}{\lambda + 1}}}{\lambda + 1},
\]

and \( \lambda \) equals the relative recruitment rate \( \frac{\pi}{\epsilon} \).

Note first that as the relative recruitment rate \( \lambda \) approaches infinity, i.e., either \( \pi \) goes to infinity or \( \epsilon \) goes to zero, \( \psi \) 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 $\lambda$. Figure 2 illustrates the lower bound $\frac{1}{\psi}$ for different values of $\lambda$.

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 percent 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 $\lambda$.

A consequence of the high density of program members is that once neighbourhood watch can be sustained the resulting crime rate will be low. Recall that burglaries are successfully deterred at site $x$ if at least one neighbour of $x$ is a member of the program (since punishment is high). This means that $1 - \frac{1}{\psi}$ serves as an upper bound for the asymptotic probability of a burglary at site $x$. Because $\frac{1}{\psi}$ goes to one as the relative recruitment rate $\lambda$ goes to infinity, the asymptotic crime rate converges to zero. Moreover, the number of burglaries is already quite low if the rate of recruitment is only little higher than necessary (Figure 2).
2 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 et al., 2001; Masclet et al., 2003; Anderson and Putterman, 2006; Carpenter, forthcoming a,b). 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) 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, p. 4). 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. 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
\[ \Pi_x = 1 - i_x + \sum_{y \in \{x-1,x+1\}} i_y \] (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 \( p > \bar{p} = \frac{1}{\alpha_1} \) and at least one neighbour is a \( P \)-type or if \( \bar{p} > p > p = \frac{1}{\alpha_2} \) 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 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 \( \chi \) 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 \( \pi \) drives it in the former situation, where there is

---

\(^8\)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 cooperating, 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 et al. (2002) who provide experimental evidence for imitation of “exemplary behaviour”. 

15
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, here punishment is supposed to induce cooperation among the neighbours themselves. 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 \( p > \bar{p} \) 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.

3 Conclusion

We have studied the dynamics of neighbourhood watch programs and how public policy (by setting fines for criminal behavior) can drive the rise and fall of such programs. Quite 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 become victims of their own success. Once programs have dissolved, 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, in addition to setting fines, the state might also have the option of campaigning for neighbourhood watch programs or subsidizing them. Such campaigns would formally 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. Alternatively, central campaigning could also re-in-
stall neighbourhood watch programs once crime rates start rising again. In this case, our model
suggests that high punishment levels would give rise to cycles of high and low crime, while inter-
mediate levels of punishment would keep crime rates within certain bounds. Optimally, of course,
the best strategy would be to sustain neighbourhood watch by maintaining participation rates at
a sufficiently high level, also when crime rates are low. Since the weak point of watch programs
seems to be its single-issue focus (i.e., the reduction of crime), a promising solution might be to
create community programs that involve multiple forms of cooperation between neighbours, such
as, the organization of street parties or the setting-up of a shuttle system to bring neighbouring
children to school (see also Rosenbaum, 1987 on this issue).

We also show in our paper that the dynamics of norm enforcement in public goods provision
schemes, which 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 that punishment be sometimes
actually carried out. Again, mild levels of antisocial behaviour can actually help sustain a system
of norm enforcement. In some sense our analysis therefore suggests that policy targets combating
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. Similar perverse effects of “more law” have recently been documented in both
laboratory and field experiments. Bohnet et al. (2001) show in a laboratory experiment how better
contract enforcement can actually increase the frequency of contract breach. Gneezy and Rusti-
chini (2000) demonstrate in a field experiment how the introduction of a monetary punishment
worsens behaviour. Our paper analyses one possible channel that can cause such counterintuitive
incentive effects.
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. So far, only a limited number of studies that analyze the effectiveness of neighbourhood watch exist. All of these studies are based on data from time periods before the mid 1990s. A recent meta-analysis of this work by Bennett et al. (2005) finds that “across all studies combined, neighbourhood watch is associated with a reduction in residential burglaries.” (Bennett et al., 2005, p. 2) However, this results needs to be qualified, due to the strong variation in outcomes across individual studies. On balance, the authors conclude that “the results (...) are encouraging, but more research needs to be done to help explain why these variations exist.” (ibid, p. 3). Clearly, a better empirical foundation of our understanding of neighbourhood watch programs would be of great interest from the perspective of the broader social capital, trust, and reciprocity literature.

Appendix: Proofs

Proposition 1 to 5 are implications of results for a particular class of stochastic processes: the contact process, the threshold contact process, and the \( \theta \)-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 \( \lambda \) 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 \( \theta \)-contact process forms a generalization of the two processes: each infected site recovers at rate 1; a healthy site is infected at rate \( \lambda \) 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 \( \theta \)-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 $\theta$-contact process. To see this, simply regard each infected site as an agent being a member of $M_t$. By dividing Poisson rates by $\epsilon$ 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 $\epsilon$ only affects the time scale.

Now, if only one kind of recruitment rate is at work (either $\pi$ or $\chi$), 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 $\theta$-contact process with $\theta = \frac{\pi}{\chi}$. In fact, if both effects are at work and $\pi < \chi$, our model is also a $\theta$-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 $\theta$-contact process is attractive. A well-known consequence (Liggett, 1985, Chapter III, Theorem 2.3) is that the so-called upper invariant measure

$$\nu = \lim_{t \to \infty} \delta_\mathfrak{z} S(t)$$

exists. Here $\delta_\mathfrak{z}$ 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 $\nu = \delta_\emptyset$, 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 $\lambda$ is large enough.

Precisely, for the threshold contact process there exists a critical value $\lambda_c$ such that $\nu = \delta_\emptyset$ if $\lambda < \lambda_c$ and $\nu \neq \delta_\emptyset$ if $\lambda > \lambda_c$. Moreover, if $\lambda > \lambda_c$, it holds that $\nu(\emptyset) = 0$, 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 $\nu$ 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 $\lambda_c$. 

19
of the threshold contact process showing that

$$1.224 \leq \lambda_c \leq 2.17. \quad (7)$$

The $\theta$-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 $\theta$ and that $\theta \lambda_c(\theta)$ is strictly increasing in $\theta$ (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, $\lambda_c$ 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

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

$$\iff \frac{m}{\epsilon} < \lambda_c. \quad (8)$$

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 $\pi$ and $\chi$ must break down as well. Case (B): Suppose

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

$$\iff \frac{m}{\epsilon} > \lambda_c. \quad (9)$$

In this case the threshold contact process with infection rate $\lambda = \frac{m}{\epsilon}$ survives. Equivalently, neighbourhood 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 $\lambda_c$. \hfill \square

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

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

**Proof of Proposition 5:** Note first that the left hand side in equation (3) is indeed independent of \( x \) since the limiting distribution of \( M_t \) is translation invariant (cf. Konno, 1994). The result then follows from Katori and Konno (1993), who prove that the density of the upper invariant measure of the \( \theta \)-contact process has the following lower bound. Let \( 1 \leq \theta \leq 2 \). Define

\[
\xi = \frac{2 - \theta}{\lambda} + (\theta - 1),
\]

(12)

\[
\eta = (2 - \theta) + \theta \lambda,
\]

(13)

and

\[
\psi = \frac{\eta}{1 + \xi} \left( 1 - \sqrt{1 - \frac{2(1 + \xi)}{\eta} \left( 1 - \frac{1 - \xi^2}{\eta^2} \right)} \right).
\]

(14)

Then

\[
\nu(A \mid x \in A) \geq \frac{1}{\psi}
\]

(15)

21
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 $\theta$-contact process, which is given by the largest root of the cubic equation

$$\theta \lambda^3 - (3\theta - 2)\lambda^2 - 3(2 - \theta)\lambda + (2 - \theta) = 0. \quad (16)$$

As Proposition 5 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

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

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.

University College London

University of Zurich

References


