

Formal sanctions in CPRs: why so often, but so low? Combining logics of rational choice using IAD-framework¹

H. Autto²

ABSTRACT

Many successful regimes managing common-pool resources have crafted formal rules, but the level of formal sanctions is usually low. These observations seem to constitute a contradiction of a sort. On the one hand, formal sanctions seem to be helpful. On the other hand, their deterrence effect is used only to a small extent.

These observations are explained by studying interaction effects between formal sanctions and other institutions that can sustain cooperation in a collective action game. We distinguish two positive mechanisms (direct mechanism, belief mechanism) between deterrence and collective action by combining logics of PD and coordination games in the IAD-framework using the concept of conformism. Despite conformism, deterrence is expected to contribute positively to the collective action, but exact effects depend on the distribution of conformism in the population. However, formal sanctions also interact with informal controls depending on the system characteristics. These effects may very well undermine or outweigh the potential for deterrence. Therefore, a formal agency does not face a simple optimization problem. Agencies prefer low sanctions as they rather complement informal controls than take a risk of supplanting them.

¹ Working paper (14.7.2011) for the IASC European Meeting 2011, Plovdiv, Bulgaria.

² Public Choice Research Centre, University of Turku. E-mail: hannu.autto@utu.fi

INTRODUCTION

Formal negative sanctions lie at the core of two classical suggestions to common-pool resource (CPR) -problem: state intervention and privatization (Ostrom 1990, 8-15). State intervention could change the incentives of resource users in the assumed Prisoner's Dilemma (herewith, PD) e.g. by making credible threats to sanction those users who extract over a quota. Similarly, formal sanctions also play a key role in enforcing private property rights against thefts and against frauds in trading. Sanctions, however, are not important only to external solutions, but also to robust self-governance.

Many robust self-governance systems have grafted rules that also assign sanctions for those who break them. This observation suggests that formal sanctions play some positive role in self-governance. From the point of view of classical decision theory, however, sanctions are often low (although graduated [Ostrom 1990, 90]). One could argue that once the sanctioning system has been established, more deterrence could be achieved with little extra effort merely by increasing sanctions. But this type of harsh regime is rarely encountered. One possible response would be to argue that formal sanctions do play a positive role in self-governance, but not necessarily through deterrence.³ Here I explore an explanation compatible with the deterrence hypothesis by identifying also negative mechanisms between formal sanctions and collective action as suggested by earlier literature. These negative mechanisms could set boundaries for the use of formal sanctions in collective action dilemmas.

In what follows I first identify two positive mechanisms (direct mechanism, belief mechanism) through which sanctions operate and study the interaction effects between deterrence and other institutions that favour robust self-governance using a model of collective action based on classical rational choice. More specifically, I study how conformism – a concept closely linked to ‘tit-for-tat’ and reciprocity – interacts with deterrence. The main point is that the positive effect of deterrence prevails, although changed, notwithstanding the number of equilibria in the collective action game.

Secondly, I point out negative causal mechanisms between formal sanctions and collective action and point out conditions under which they are likely to occur. Formal sanctions may lessen resource users' regulatory interest to each others' actions and, therefore, reduce peer monitoring that helps to maintain formal deterrence. They may also reduce other types of informal control. In addition, increased formal deterrence could lead to a loss of conformism because of motivation crowding-out. Depending on the characteristics of the system, an increase in formal sanctions may not success in increasing either the formal deterrence or the overall deterrence. Furthermore, in those cases where it does success in both, it may still be that the loss of conformism outweighs the deterrence effect. The problem is that of complex causality (Agrawal 2002), and there is clearly a need to study the relative importance of these mechanisms.

FORMAL SANCTIONS AND OTHER INSTITUTIONS SUSTAINING COOPERATION

Sometimes institutions create interesting interaction effects. Here I study interaction effects between formal sanctions and other institutions in the collective action. By “other institutions” I mean institutions that can sustain cooperative equilibria in a classical collective action game.

³ For example: Income from sanctions can be used to maintain monitoring, and the real deterrence is informal (e.g. the loss of reputation). This type of explanation would still need to explain why people are not deterred by more severe formal sanctions.

While discussing sanctions, Elinor Ostrom (1990, 94-100) links them to quasi-voluntary compliance (Levi 1988). Margaret Levi clarifies the concept of quasi-voluntary compliance in the context of taxation:

It is voluntary because taxpayers choose to pay. It is quasi-voluntary because the noncompliant are subject to coercion – if they are caught. [...] Taxpayers are strategic actors who will cooperate only when they can expect others to cooperate as well. The compliance of each depends on the compliance of the others. (Levi 1988, 52-53, emphasis in the original)

The most important point is that the compliance of this sort is contingent on others behaviour. Interestingly, Edna Ullman-Margalit's (1977) definition of conformism captures much of the same logic. According to Ullman-Margalit (1977, 93; emphasis in the original) "conformist [...] regards the majority of the situations in which he has to make a decision about action *as if* they were coordination problems. [...] the conformist perceives his preferences as conditional upon the actions of others [...]" More generally, the conformist is an imitator who does not want to stand out.

In game-theoretic terms, quasi-voluntary compliance or conformism may change a single-equilibrium game (a PD) into a multiple equilibria game (coordination game, see Cooper 1999). But concepts such as 'quasi-voluntary compliance' or 'conformism' are not the only concepts with that power. Other well-known examples are 'trigger'- and 'tit-for-tat'-strategies, the norm of reciprocity and success-contingent benefits for cooperators.

General feasibility theorems⁴ say that in an infinitely repeated PD trigger strategies can support cooperative equilibria, i.e. make them possible outcomes (Rubinstein 1979; Medina 2007, 70-79). Trigger strategies are punishing strategies: a player punishes other players' earlier defections by defecting herself. An important sub-class of punishments, which also attains an elementary sense of fairness, is 'tit-for-tat'. In repeated PDs 'tit-for-tat' gives the other player the benefit of the doubt by cooperating in the first round and imitating the other player's earlier move in the subsequent rounds. The relative success of 'tit-for-tatters' might explain why some people seem to have internalized the norm of reciprocity: be nice, if you observe nice behaviour; be bad, if you observe bad behaviour. Yet another set of institutional statements that can support multiple equilibria is success-contingent benefits for cooperators (Medina 2007). If players expect that their nice behaviour is going to be rewarded if the collective action succeeds, they might find it better to behave nicely if the likelihood of success is sufficiently large. In general this likelihood depends on the behaviour of others.

Quasi-voluntary compliance, conformism, punishing strategies, reciprocity and success-contingent benefits for cooperators all share a simple common denominator, which can be expressed using the vocabulary of the institutional grammar (Ostrom 2005). The grammar identifies five parts that an institutional statement usually contains, thus making their comparison easier. One of these is the Condition giving the information about the conditions under which the statement applies. For example, the 'tit-for-tat' -strategy could be expressed as "one cooperates, if it is the first round or the other player cooperated in the previous round", and here the Condition refers to the sentence starting with "if". The common denominator is that the Condition of each of these statements refers to the behaviour of other players (usually, but not necessarily, of similar type than the *ego*). Success-contingent benefits do that via the likelihood of success, others more directly. More detailed comparison can be found in Table 1. The last entry also gives the definition of how formal

⁴ These are still better known as folk theorems.

sanctions are understood here: they are contingent only on *ego*'s own action, not on the behaviour of others.

<u>Syntax part</u>	<u>Attribute</u>	<u>Deontic</u>	<u>AIM</u>	<u>Conditions</u>	<u>Or Else</u>
Content of the part	to whom the statement applies	a modal verb	describes actions to which the deontic is assigned	defines when and where the statement applies	d. the institutionally assigned conseq. of not following
Tit-for-Tat	<i>Ego</i>		cooperates	first round; the other player cooperated in the previous round	
Reciprocity (direct)	<i>Ego</i>	must	cooperate	the player that <i>ego</i> faces cooperated previously	
Compliance (general)	<i>Ego</i>	must	take an action A	at least k % of others take an action A	
Success-contingent benefits	<i>Ego</i>	must	cooperate	the collective action is successful	<i>ego</i> receives no benefit
Conformism (general)	<i>Ego</i>	must	take an action A	at least k % of others take an action A	
Formal sanctions	<i>Ego</i>	must	cooperate	always	<i>ego</i> receives a negative sanction s

TABLE 1: Comparison of some well-known institutional statements

Interestingly, the existence of these types of statements opens up a second mechanism for formal sanctions to operate: *ego*'s behaviour is not only affected by possible sanctions directed to her, but also by possible sanctions directed to *others*. The first mechanism is here called the direct mechanism, and the second mechanism is here called the belief mechanism. It is rather obvious that the effect of sanctions is therefore mediated by other institutional statements. Studying these connections in a theoretical level hopefully offers a partial answer to the call of Arun Agrawal (2002) as he stresses the importance of finding interaction effects between variables that seem to foster robust self-governance in CPRs. We are also combining two rational choice approaches to collective action: single- and multiple equilibria models (Taylor & Ward 1982, Heckathorn 1993, Medina 2007).

A COLLECTIVE ACTION MODEL WITH DETERRENCE AND CONFORMISM

Next I study conformism in a more detailed manner using one-shot game, but it should be clear that the choice of the contingent statement is a question of convenience – not of content. By now there are both theoretical (e.g. 'tit-for-tat' in repeated games) and empirical (e.g. cheap talk effect in PD-experiments) evidence suggesting that many people have internalized the norm of conformism to some extent and that there may be a clear rationale for doing so. This implies that it is worthwhile to study the effect of formal deterrence in a population where some are conformists. I follow the

convention in Ostrom (2005), and represent the norm of conformism by delta-parameters (δ). The formalization of the general collective action game follows Medina (2007), but the model differs from those considered there.

Consider N players, each arbitrary player i having a possibility to either cooperate C_i or defect D_i . The payoff from the successful collective action is denoted by b_i , the cost of cooperation by l_i , and sanctions for defection by s_i . Sanctions refer to overall deterrence meaning that it captures both the severity of formal- and informal sanctions and the likelihood monitoring. Possible payoffs for a player i are then as follows:

- i cooperates, collective action succeeds, k-majority cooperates: $w_{1i} = b_i - l_i + \delta_i$
- i cooperates, collective action succeeds, no k-majority: $w_{2i} = b_i - l_i$
- i cooperates, collective action succeeds, k-majority defects: $w_{3i} = b_i - l_i - \delta_i$
- i cooperates, collective action fails, k-majority cooperates: $w_{4i} = -l_i + \delta_i$
- i cooperates, collective action fails, no k-majority: $w_{5i} = -l_i$
- i cooperates, collective action fails, k-majority cooperates: $w_{6i} = -l_i - \delta_i$

- i defects, collective action succeeds, k-majority cooperates: $w_{7i} = b_i - s_i - \delta_i$
- i defects, collective action succeeds, no k-majority: $w_{8i} = b_i - s_i$
- i defects, collective action succeeds, k-majority defects: $w_{9i} = b_i - s_i + \delta_i$
- i defects, collective action fails, k-majority cooperates: $w_{10i} = -\delta_i - s_i$
- i defects, collective action fails, no k-majority: $w_{11i} = -s_i$
- i defects, collective action fails, k-majority defects: $w_{12i} = \delta_i - s_i$

Denote the likelihood of a successful collective action by $\Pr(\text{Succ})$ and the likelihood of a failure by $\Pr(\text{Fail})$. I assume $1 - \Pr(\text{Succ}) = \Pr(\text{Fail})$. Denote the likelihood of the event that k-majority cooperates by $\Pr(\text{Maj } C)$, the event that k-majority defects by $\Pr(\text{Maj } D)$, and the event that there are no k-majority by $\Pr(\text{No } \text{Maj})$. Naturally, $\Pr(\text{No } \text{Maj}) = 1 - \Pr(\text{Maj } C) - \Pr(\text{Maj } D)$. All the probabilities depend on the strategy vector, and more specifically, on the rate of cooperating players, $0 < \gamma < 1$. Denote conditional probabilities as in $\Pr(\text{Succ}|C_i, \gamma)$, which refers to the probability of success given that γN others are cooperating, and player i cooperates. Recalling that the k-majority is defined as a proportion of others (*ego* not included), it holds that e.g. $\Pr(\text{Maj } C|C_i, \gamma) = \Pr(\text{Maj } C|D_i, \gamma)$. It is convenient to write $\Pr(\gamma < 1 - k)$ instead of awkward $\Pr(\text{Maj } D|C_i, \gamma)$, $\Pr(1 - k < \gamma < k)$ instead of $\Pr(\text{No } \text{Maj}|C_i, \gamma)$, and $\Pr(k < \gamma)$ instead of $\Pr(\text{Maj } C|C_i, \gamma)$.

Let the function $v_i(C_i, \gamma)$ denote the payoff for player i after her cooperation given a certain rate of cooperating players, and $v_i(D_i, \gamma)$ the corresponding payoff after her defection. Any arbitrary player i then cooperates, iff

$$v_i(C_i, \gamma) > v_i(D_i, \gamma), \quad (1a)$$

where

$$v_i(C_i, \gamma) = w_{1i} \Pr(\text{Succ}|C_i, \gamma) \Pr(k < \gamma) + w_{2i} \Pr(\text{Succ}|C_i, \gamma) \Pr(1 - k < \gamma < k) + w_{3i} \Pr(\text{Succ}|C_i, \gamma) \Pr(\gamma < k) + w_{4i} \Pr(\text{Fail}|C_i, \gamma) \Pr(k < \gamma) + w_{5i} \Pr(\text{Fail}|C_i, \gamma) \Pr(1 - k < \gamma < k) + w_{6i} \Pr(\text{Fail}|C_i, \gamma) \Pr(\gamma < k)$$

and

$$v_i(D_i, \gamma) = w_{7i} \Pr(\text{Succ}|D_i, \gamma) \Pr(k < \gamma) + w_{8i} \Pr(\text{Succ}|D_i, \gamma) \Pr(1 - k < \gamma < k) + w_{9i} \Pr(\text{Succ}|D_i, \gamma) \Pr(\gamma < k) + w_{10i} \Pr(\text{Fail}|D_i, \gamma) \Pr(k < \gamma) + w_{11i} \Pr(\text{Fail}|D_i, \gamma) \Pr(1 - k < \gamma < k) + w_{12i} \Pr(\text{Fail}|D_i, \gamma) \Pr(\gamma < k)$$

It is assumed conventionally that no single contribution has a significant effect to the possibility of success, therefore, $\Pr(\text{Succ}|C_i, \gamma) \approx \Pr(\text{Succ}|D_i, \gamma)$. Condition (1a) then simplifies to

$$\Pr(\gamma > k) - \Pr(\gamma < 1 - k) > \frac{l_i - s_i}{2\delta_i}. \quad (1b)$$

Note that the decision in this model does not depend on the likelihood of success (for that to be the case, there should be success-contingent benefits). I visualize how the decision making problem of a single player depends on parameters and her beliefs in Table 2. Recall that l_i here captures the initial temptation and s_i the deterrence.

		<i>i</i> believes that it is more likely that <i>k</i> -majority	
		<i>cooperates</i>	<i>defects</i>
Temptation vs. deterrence	<i>temptation is higher</i>	depends on δ_i	defects always
	<i>deterrence is higher</i>	cooperates always	depends on δ_i

TABLE 2: Determinants of player *i*'s decision

I consider a population that is homogenous in respect to l_i and s_i so that the temptation is higher, that is, $l_i > s_i$, but heterogeneous in respect to $\delta_i \in [0, \infty)$. This means that some people have internalized the norm of conformism to a greater extent than others. Irrespective of her beliefs on what the *k*-majority will do, player *i* defects, if

$$\delta_i < \frac{l_i - s_i}{2}. \quad (2)$$

This means that she has a dominant strategy. For convenience denote $\delta^* = \frac{l_i - s_i}{2}$. As it is possible to eliminate all the strictly dominated strategies, it is also clear that all the others know that she is not going to cooperate. As I assume that the distribution f of delta parameters in the population is common knowledge among the players, it is possible to calculate the amount of players who are going to defect unconditionally. I call these players 'Prisoners' and refer to their number with m . The rest $n = N - m$ players are 'Conformists'. Figure 1 illustrates.

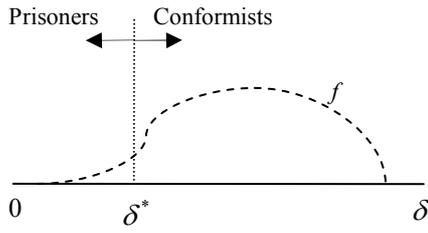


FIGURE 1: Prisoners and Conformists

Now we are ready to consider how the role of sanctions depends on the spread and strength of conformism in the population. Consider first two initial populations of Prisoners. For the first population let $\delta_1 = \delta_2 \dots = \delta_N = 0$. This population cooperates only, if the deterrence is higher than the temptation, and defects otherwise. In either case we have a single equilibrium, either mutual cooperation or mutual defection, and the sanctions operate only via direct mechanism. This implies that (increased) deterrence is a necessary and sufficient condition for any cooperation to occur.

For the second initial population of Prisoners let $\delta_1, \delta_2, \dots, \delta_N < \delta^*$ and, at least for some subset J of players, $\delta_j > 0$. The only equilibrium for this population is mutual defection, until the sanctions are increased to a level where a sufficient number of players no longer have a dominant strategy to defect. This requires k -majority, that is, when conformists think that a sufficient number of other players do not have a dominant strategy to defect. From that on at least two strategy vectors are Nash equilibria: the final outcome depends on conformists' beliefs on others' likely behaviour. Before this limit is reached, deterrence is a necessary, but not a sufficient, condition for any cooperation to occur. After this limit is reached, further deterrence is no longer a strictly necessary condition for any cooperation occurring, but it does continue to increase the expected amount of cooperation. This is studied in a more general population below using an exogenously given belief condition.

Consider again just some distribution f of delta parameters in the population. I model the beliefs here simply by assuming that players assign a probability p to the event that a single other conformist is going to cooperate. This probability captures the collective mood among the conformist and I leave it exogenous, which suffices to my purposes for now.

With these assumption the expected number of cooperators in the system follows a binomial distribution $B(N - m, p)$. Note that the number of trials, and therefore the expected number of cooperators, decreases in the number of Prisoners, m . Assuming that f is continuous, higher deterrence can decrease the number of Prisoners in the population making it more likely that some k -majority is cooperating. The general message is that sanctions looming above *others* send a signal to conformists about their peers' likely behaviour. This causes a conformist to adjust her beliefs about the likely actions of the majority as new equilibria become possible. The rest of this chapter elaborates the quantification put forth here.

$B(n, p)$ can be approximated using a normal distribution $g : N(np, p(1 - p))$. After standardization

we get the following approximations: $\Pr(\gamma < k) = \Phi\left(\frac{(1 - k)N - (N - m)p}{\sqrt{(N - m)p(1 - p)}}\right)$ and

$\Pr(\gamma > k) = 1 - \Phi\left(\frac{kN - (N-m)p}{\sqrt{(N-m)p(1-p)}}\right)$. The number of Prisoners, m , is given by $F\left(\delta_i < \frac{l_i - s_i}{2}\right)N$

(but preferably by its floor). For convenience denote $\delta^* = \frac{l_i - s_i}{2}$. The decision making problem of a single player, the condition (1b), can then be re-written as

$$1 - \Phi\left(\frac{N[k - (1 - F(\delta_i < \delta^*))p]}{\sqrt{Np(1-p)(1 - F(\delta_i < \delta^*))}}\right) - \Phi\left(\frac{N[1 - k - (1 - F(\delta_i < \delta^*))p]}{\sqrt{Np(1-p)(1 - F(\delta_i < \delta^*))}}\right) > \frac{l_i - s_i}{2\delta_i}. \quad (3)$$

The left-hand side of (3) is a scalar between -1 and 1 . Denote it by y . If $y < 0$, that is, if it is more likely that k -majority is defecting rather than cooperating, all players are defecting. This conclusion was already drawn in the table 2. However, if $y > 0$, some people may be cooperative. We can finally write the rate of defectors in the system as

$$Defecting(\%) = \begin{cases} F\left(\frac{l_i - s_i}{2y(s_i)}\right) & , y > 0 \\ 0 & , y < 0 \end{cases}. \quad (4)$$

For convenience, consider two numerical examples that highlight the role of deterrence and mechanisms through which they operate.

Example 1:

Consider a system of 100 players, $N = 100$. Let $l_i = 2$ for temptation and $s_i = 1$ for deterrence, so that $\delta^* = 0.5$. Assume that the delta parameter is uniformly distributed from 0 to 5, that is, $f : U(0,5)$. The percentage of Prisoners in the system is given by $F(\delta^* = 0.5) = 0.1$ as illustrated in the Figure 2:

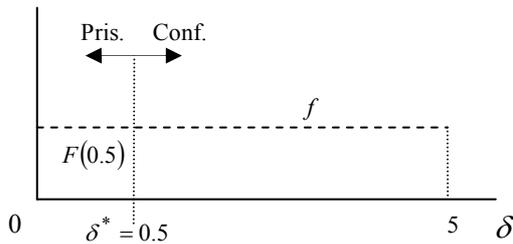


FIGURE 2: Example 1

This means that 10 % of the players have a dominant strategy to defect. Set $k = 0.75$ and assume people are pretty confident that the conformists are more likely to cooperate than defect by setting $p = 0.8$. Then $y \approx 0.21$ meaning that it is approximately 21 percentage units more likely that 0.75 - majority will coordinate to cooperation rather than to defection. Finally we derive the defection rate using (4) and attain 0.47. About 47 % of the players are now going to defect, which implies that a fair amount of Conformist is following the lead of Prisoners. Note that this model captures an important property of real-life coordination problems: people do play diverse strategies at the

equilibrium. The reason is the exogenously given belief condition. Some form of belief updating can very well make conformists' strategies converge to either full (conformist) cooperation or full defection.

Example 2:

Consider now the same system with the following difference: the sanctioning agency is able to increase the deterrence so that $s_i = 1.5$. Then $\delta^* = 0.25$ and the percentage of Prisoners in the system is $F(\delta^* = 0.25) = 0.05$ as illustrated in Figure 3:

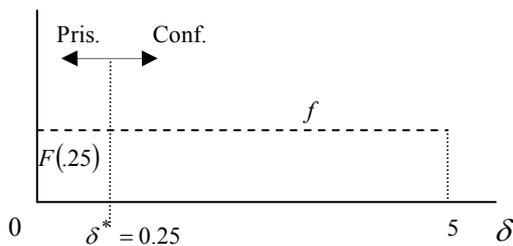


FIGURE 3: Example 2

This means that only 5 % of the players now have a dominant strategy to defect. This affects conformists' beliefs about the likely decision of the majority. Now $y = 0.6$, meaning that it is 60 % more likely that the majority will coordinate to cooperation rather than to defection. In addition to increasing this probability, sanctions also affect directly to the necessary threshold value of δ_i in (4). Finally we derive the defection rate and attain 0.08. About 8 % of the population is now defecting as some Conformists still follow the lead of the Prisoners.

To conclude: Notwithstanding the distribution of conformists in the population formal sanctions are expected to contribute positively to the collective action. But both the mechanism and the degree of necessity of formal sanctions depends on the distribution of conformism in the population.⁵

WHY FORMAL SANCTIONS CAN NOT SIMPLY SOLVE IT?

According to the earlier discussion formal sanctions contribute positively to collective action irrespective of the amount of conformism in the population assuming that they increase the deterrence. This suggestion seems to be in line with the observation that many successful self-governance systems have adapted formal sanctions although cooperation is possible without them. However, it leaves another question unanswered: why the level of formal sanctions is moderate at best? More deterrence could be achieved with little costs by merely increasing sanctions. But harsh and successful regimes do not seem to be a rule. I assume that the observed state of affairs reflects some type of optimality or resiliency. In what follows I try to offer a partial answer by identifying negative mechanisms between high sanctions and collective action as suggested by the earlier literature.

⁵ More generally, deterrence is expected to contribute positively to the collective action notwithstanding the number of equilibria in the collective action game. Cooperative equilibria may also follow, if the assumption that a single contribution does not have a significant effect on the outcome is dropped. (see also Medina 2007.)

For deterrence following is necessary: a positive possibility of getting caught and a positive probability of a negative consequence after being caught. Increasing probabilities is expensive to an agency. It usually involves hiring more people as monitors, expansion of bureaucracy, and improving monitoring techniques. An important part of the incentive structure in CPR-cases is that (due to externalities) individuals have a regulatory interest to others' actions (Heckathorn 1988). This can give rise to diverse types of informal control depending on how individuals' action space is defined. Informal controls interact with formal sanctions.

Note first that extreme sanctions are not an option, if there is a positive probability that either the agency or an individual makes a mistake (innocent person is convicted, individual mistakenly defects) and the individual is able to take an exit-option or revolt. Both conditions apply very generally. If it is possible for them, individuals vote against extreme and non-perfect agencies with their feet or by revolt. One could give a simple condition for this theoretical upper limit, but it suffices here to note that such an anchor point exists. Next I identify negative mechanisms from formal sanctions to collective action.

Passive monitoring

If individuals may report others' deviations from the rule (passive monitoring), the regulatory interest fosters formal deterrence. The agency can now partly rely on peer monitoring. However, if individuals are interdependent so that negative sanctions to one also have negative consequences to others (Heckathorn 1988; 1990), higher sanctions may lead to less passive monitoring. As individuals can not anymore prevent their peer's deviation, they may be better off by not reporting the deviation. Moreover, their threats to do so may cease to be credible if they are themselves hurt in the process. Therefore, increasing formal deterrence by higher sanctions may also decrease formal deterrence due to a loss of passive peer monitoring.

Active monitoring

If individuals may choose to engage in active monitoring, that is, spend resources to gain information about their peers' behaviour, other mechanisms emerge.⁶ Again, regulatory interest rationalizes active monitoring in some cases. However, official guards or high sanctions may lead to fewer active peer monitors. The reason for this relies on the nature of catch-and-evade games, which do not necessarily have equilibria in pure strategies. Mixed Nash equilibria strategies have the interesting property of being (at least partly) determined by the payoffs of other players. Increasing sanctions does not only affect the behaviour of the potential criminal, but it also affects the behaviour of monitoring agents.

This point was forcefully made by George Tsebelis (1989; 1990a; 1990b; 1991) in a series of papers with a help of two-player games. In two-player games this effect is most striking, because a player's strategy in a mixed Nash equilibrium is completely determined by the payoffs of the other player. A vivid discussion followed Tsebelis' argument, which can not be covered here in detail.⁷ In one line of responses, Weissing and Ostrom (1991a; 1991b; 1993) considered irrigation systems with both peer monitors and official guards. Their results largely confirm that this type of "perverse" and somewhat counterintuitive feedback loops do occur also in larger games, although they do not

⁶ The difference between passive and active monitoring is that in active monitoring the individual chooses to spend resources on monitoring. She does not merely obtain information during her daily routines. Both types of monitoring, of course, occur in real world.

⁷ See Bianco et al. (1990), Hirshleifer and Rasmusen (1990), Mayer (1991), Rapoport (1990) and Weissing and Ostrom (1991a; 1991b; 1993).

always alone determine the effect (for conditions, see Weissing and Ostrom 1991b) and they are not always dominant factors (Weissing and Ostrom 1991a, 253-254). Interestingly, increasing formal sanctions in hope for increasing formal deterrence may lead to fewer active peer monitors and, therefore, to a decrease in formal deterrence irrespective of whether sanctions have spill-over effects or not. Rather complex game theoretic arguments behind these effects can be simplified for current purposes by saying that the policy decreases peer monitors' regulatory interest: formal sanctions and official guards are already taking care of the deterrence lessening the gains of peer monitoring. Of course if circumstances are such that passive peer monitoring does not work, we should not expect any active peer monitoring occurring either (if one is not going to report, why should one spend resources for monitoring in the first place).

Informal control

If individuals may choose to contribute to informal control affecting others' possibilities to take actions, the formal deterrence and the informal control co-exist. The informal control might operate *via* persuasion that affects others' preferences, *via* selective incentives as also formal deterrence does or *via* control of opportunity structures, i.e. limiting players' action space (Heckathorn 1990, 369). When also the informal control is an option, increasing the formal deterrence might lead to a decrease in peers' regulatory interest (following the earlier argument) and, therefore, to a decrease in the informal control. The agency might merely substitute the informal control by the formal deterrence.

Sanctions with spill-over effects are again an exception (Heckathorn 1990; 1993). The spill-over actually increases peers' regulatory interest: if one is sanctioned, others hurt as well. This type of complementary control by collective sanctions can be very efficient as examples from army boot camps and tribal societies illustrate (Heckathorn 1990). It is interesting to note how the effects of spill-over depend on how players' action space is defined: Spill-over effects may decrease passive monitoring (that occurs after the deviation), but may foster informal control (that is aimed to prevent the deviation by a complementary control system).⁸ The latter effect is yet contingent on the agency's ability to monitor individuals (without relying on their help) and their ability to exercise informal control on one another. If agency can not monitor them efficiently enough and they can not impose informal control on one another, individuals might find it beneficial to rebel against the agency's control attempts. If some of them are going to deviate in any case, suffering the mere externality is a lesser evil (Heckathorn 1988, 553-556).

Motivation crowding-out

Richard T. Titmuss (1970) claims that monetary incentives lessen an individual's sense of obligation, that is, they kill the intrinsic motivation to cooperate. Bruno S. Frey and Felix Oberholzer-Gee (1996), for example, find supportive empirical evidence for the claim while studying willingness of communities to host locally unwanted projects ("Not in My Backyard" - projects, e.g. nuclear facilities). They conclude that there is a need to reconsider the use of monetary incentives in all areas where intrinsic motivation is likely to be important (*ibid.*, 753). From the point of view of rational choice theory, the presence of external incentives may make it impossible for an individual to indulge in altruistic feelings (*ibid.*, 747). But also another rational choice explanation can be suggested that generalizes the type of norms to which the claim applies. The change in players' preferences, a norm like conformism, is usually maintained by members of the community in social interaction. This persuasion is merely one mechanism suggested by

⁸ Another line of research (Kitts 2006; 2008) considers rival incentives. These studies, however, concentrate on positive incentives and their generalization to negative incentives is not straightforward (Kitts 2006, 253).

Heckathorn (1990) that enables peers to exercise informal control on one another. In the last chapter we saw how conformism may turn a PD into a multiple equilibria game. If formal sanctions lead to a loss of conformism, they also reduce the number of cooperative equilibria in the game (that can be sustained by some initial belief conditions) and effectively push the original game towards PD, thus making themselves more necessary than they actually were in the first place.

Summary

The possible negative consequences of increasing formal sanctions can then be summarized in following questions:

- 1) Does the increase of formal sanctions lead to a regime failure?
- 2) Does the increase of formal sanctions actually increase the formal deterrence?
- 3) Does the increase of formal sanctions actually increase the overall deterrence?
- 4) Does the increase of formal sanctions lead to a loss of conformism?

The answer to these questions and their relative importance depends in rather complicated manner on whether people are able to exercise informal control on one another and with what costs, to which extent the agency must rely on informal monitoring, and whether sanctions have spill-over effects or not. In addition, the deterrence effect depends on the spread and strength of conformism in the population and the degree of motivation crowding-out. Broadly put, the straightforward deterrence hypothesis is likely to apply in cases of atomized populations in which players do not develop, or the situation does not invoke conformism (e.g. drop-in communities, communities with heterogeneous roles), in which informal control is either expensive or impossible for individuals, and in which weak interdependency between individuals implies weak spill-over effects.

These characteristics do not offer a good description of traditional cases of governed CPRs. Usually people are strongly interdependent, they communicate and interact regularly which gives them a chance to exercise informal control and to engage in mutual monitoring, and they face the same situation repeatedly. Informal norms are created and they play a crucial role in determining the social outcome. In other words, the situation is favourable for negative mechanisms between formal sanctions and collective action. Agencies that operate with very limited budgets are not willing to risk the services that peers are providing for free. It is then perhaps not surprising that formal sanctions tend to be very far from the theoretical maximum, and that agencies rather complement informal control with some amount of formal sanctions than take a risk of supplanting it.

CONCLUSIONS

In this paper I have discussed the role of formal sanctions in collective action in general and in cases of CPRs in particular. Although many successful regimes use formal sanctions, the level of sanctions is moderate at best. From the point of view of rational choice theory this is somewhat puzzling: the potential for deterrence is created, but it is not utilized to a great extent. I suggest an answer by studying both positive and negative mechanisms between formal sanctions and collective action. It becomes crucial to study the interaction effects between formal sanctions and other types of institutional statements that may sustain cooperation in a collective action game.

There are two mechanisms through which deterrence contributes positively to the collective action. When sanctions affect through direct mechanism, the deterrence affects the behaviour of the player who is under deterrence. But many institutional statements that can sustain cooperation in some form of collective action game (including trigger- and 'tit-for-tat' -strategies, compliance,

reciprocity and conformism) give rise to belief mechanism. Because these statements make a player's choice contingent on the behaviour of others, the deterrence directed to a player may also affect the behaviour of others. I studied both mechanisms using a collective action game with deterrence and a norm of conformism and concluded that although the magnitude of the deterrence effect and the degree of its necessity for collective action are contingent on the distribution of conformism in the population, the qualitative effect is not. To some extent this explains why formal sanctions are adapted so often, although cooperation may very well occur without deterrence.

But not all deterrence is formal deterrence. Reviewing and synthesising the earlier literature it is possible to identify several possible negative mechanisms between formal sanctions and collective action. When CPR-users are able to exercise some type of informal control on each others' actions, high formal sanctions may be counterproductive as they decrease peers' regulatory interest. This may lessen peer monitoring and, therefore, undermine formal deterrence, or it may lessen informal deterrence and, therefore, undermine the overall deterrence. Interestingly, an increase in formal sanctions may also lead to a loss of conformism. The irony of the last effect is that by it formal sanctions actually decrease the number of cooperative equilibria in the original game, thus making themselves more necessary than they were in the first place. Many traditional CPR-environments satisfy the necessary assumptions for these negative mechanisms. I argue that the reason for relatively low sanctions in successful CPR-cases is that formal agencies rather complement informal control with some amount of formal sanctions than take a risk of supplanting it.

REFERENCES

- Agrawal, Arun. 2002. Common Resources and Institutional Stability. In Ostrom, Elinor et al. (ed.): *Drama of the Commons*. Washington: National Academy Press.
- Bianco, William and Peter Ordeshook and George Tsebelis. 1990. Crime and Punishment: Are One-Shot, Two-Person Games Enough? *The American Political Science Review*. Vol. 84, No. 2.
- Cooper, Russell. 1999. *Coordination Games. Complementaries and Macroeconomics*. USA: Cambridge University Press.
- Frey, Bruno S. and Felix Oberholzer-Gee. 1997. The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out. *The American Economic Review* 87, 4.
- Gordon, H. 1954. The Economic Theory of a Common-Property Resource: The Fishery. *Journal of Political Economy* 62: 124-142.
- Hardin, G. 1968. The Tragedy of the Commons. *Science* 162: 1243-1248.
- Heckathorn, D. 1988. Collective Sanctions and the Creation of Prisoner's Dilemma Norms. *American Journal of Sociology* 94, 3.
- Heckathorn, D. 1990. Collective Sanctions and Compliance Norms: A Formal Theory of Group-Mediated Social Control. *American Sociological Review* 55, 3.
- Heckathorn, D. 1993. Collective Action and Group Heterogeneity: Voluntary Provision Versus Selective Incentives. *American Sociological Review* 58, 3.
- Heckathorn, D. 1996. The Dynamics and Dilemmas of Collective Action. *American Sociological Review* 61: 250-277.
- Hirshleifer, D. and E. Rasmusen. 1989. Co-operation in a Repeated Prisoner's Dilemma with Ostracism. *Journal of Economic Behavior and Organization* 12: 87-106.
- Hirshleifer, J. & E. Rasmusen. 1990. Are Equilibria Strategies Unaffected by Incentives? The Police Game. UCLA Department of Economics, Working Paper No. 595. August.
- Kitts, James. 2006. Collective Action, Rival Incentives, and the Emergence of Antisocial Norms. *American Sociological Review* 71.
- Kitts, James. 2008. Dynamics and Stability of Collective Action Norms. *Journal of Mathematical Sociology* 32.
- Levi, Margaret. 1988. *Of Rule and Revenue*. Berkeley and Los Angeles: University of California Press.
- Mayer, Thomas. 1991. Comment on Tsebelis 'Penalty has no Impact on Crime. *Rationality and Society*. Vol. 3, No. 2.

- Medina, Lus Fernando. 2007. *A Unified Theory of Collective Action and Social Change*. USA: University of Michigan Press.
- Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions For Collective Action*. New York: Cambridge University Press.
- Ostrom, Elinor. 2005. *Understanding Institutional Diversity*. Princeton: Princeton University Press.
- Rapoport, Anatol. 1990. Comments on Tsebelis. *Rationality and Society*. Vol. 2, No. 4.
- Rubinstein, Ariel. 1979. Equilibrium in Supergames with the Overtaking Criterion. *Journal of Economic Theory* 21: 1-9.
- Taylor, M. and H. Ward. 1987. Chickens, Whales, and Lumpy Goods: Alternative Models of Public Good Provision. *Political Studies* 30: 350-370.
- Titmuss, Richard M. 1970. *The Gift Relationship*. London: Allen and Unwin.
- Tsebelis, G. 1989. The Abuse of Probability in Political Analysis: the Robinson Crusoe Fallacy. *American Political Science Review* 83: 77-91.
- Tsebelis, G. 1990a. Are sanctions effective? A Game-Theoretic Analysis. *Journal of Conflict Resolution*, vol. 34, No. 1.
- Tsebelis, G. 1990b. Penalty has no Impact on Crime: A Game-Analytic Analysis. *Rationality and Society*. Vol. 2, No. 3.
- Tsebelis, G. 1991. The Effect of Fines on Regulated Industries: Game Theory vs. Decision Theory. *Journal of Theoretical Politics*, Vol. 3. No. 1.
- Ullman-Margalit, Edna. 1977. *The Emergence of Norms*. Oxford: Oxford University Press.
- Weissing, F. & E. Ostrom. 1991. Irrigation institutions and the games irrigators play: rule enforcement without guards. In R. Selten (ed.) *Game Equilibrium Models 2 – Methods, Morals and Markets*. Springer-Verlag, Heidelberg.
- Weissing, F. & E. Ostrom 1991b. Crime and Punishment: Further Reflections on the Counterintuitive Results of Mixed Equilibria Games. *Journal of Theoretical Politics*. Vol. 3, No. 3.
- Weissing, F. & E. Ostrom. 1993. Irrigation Institutions and the Games Irrigators Play: Rule Enforcement on Government- and Farmer-Managed Systems. In Scharpf, F, W (ed.): *Games in Hierarchies and Networks. Analytical and Empirical Approaches to the Study of Governance Institutions*. Campus Verlag, Westview Press.