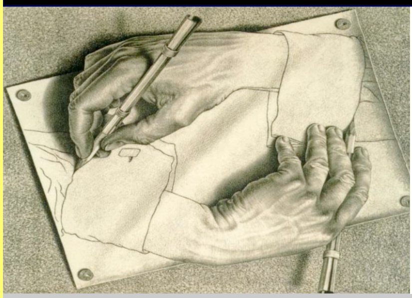
siena memos and papers on law and economics



MASSIMO D'ANTONI
ROBERTO GALBIATI

Simple
29/05

DETERRENCE AND INFORMATION: THE OPTIMAL USE OF MONETARY AND NONMONETARY SANCTIONS REVISITED

UNIVERSITY OF SIENA FACULTY OF ECONOMICS R.M. GOOD

Simple 29/05

MASSIMO D'ANTONI &ROBERTO GALBIATI

DETERRENCE AND INFORMATION: THE OPTIMAL

USE OF MONETARY AND NONMONETARY SANCTIONS REVISITED

Dipartimento di economia politica, Università di Siena

dantoni@unisi.it galbiati@unisi.it

DETERRENCE AND INFORMATION: THE OPTIMAL USE OF MONETARY AND NONMONETARY SANCTIONS REVISITED

MASSIMO D'ANTONI   ROBERTO GALBIATI

Department of Economics, University of Siena

Preliminary version — January 2005

Abstract

It is a widely accepted conclusion of the economic literature on deterrence that nonmonetary sanctions should be introduced only when monetary fines have been used up to their maximum extent. However, this conclusion is at odds with what is commonly observed e.g. in the case of deterrence of driving offenses, where nonmonetary sanctions such as licence suspension or vehicle withdrawal are extensively used, even though it would be possible to apply a monetary sanction of equivalent value.

In this paper it is shown that when the sanctioning policy conveys information about the riskiness of the sanctioned behavior, the conclusion about the optimal use of nonmonetary sanctions may be different. Namely, the use of nonmonetary sanctions can be optimal even though the monetary fine is not maximal. The argument is formalized in a model with rational though uninformed individuals, who know that the enforcer has superior information about the riskiness of action but are uncertain about the true objectives of the enforcer.

Keywords: optimal deterrence, nonmonetary sanctions, driving offenses

JEL classification: K42

1. Introduction

In order to deter individuals from engaging in harmful activities, monetary penalties are often supplemented by nonmonetary sanctions. Since the former are simply transfers, and therefore imply lower social costs than the latter, economists have tried to assess when and to what extent it is optimal to use nonmonetary sanctions. It is a widely accepted conclusion that these should be introduced only when fines have been used up to their maximum extent[1].

Nevertheless, there are cases in which sanctions are not designed according to this theoretical prescription. Consider for instance the deterrence of driving offenses: costly nonmonetary sanctions (e.g. licence withdrawal or vehicle forfeiture) are often used together with non-maximal fines. Indeed, even replacing the typical nonmonetary sanction (the suspension of the driving licence for some months) with a monetary sanction of equivalent value for the driver, fines would be much lower than driver's wealth which, according to the theory, should represent the upper limit of a monetary sanction[2]. If, on the other hand, the upper limit of the sanction were of a different nature, for example a social or ethical constraint to the penalty for a certain act[3] (and this would imply that both fines and nonmonetary sanctions for driving code violations would be subject to the same upper limit), it would be equally difficult to explain why the fine is not raised to this maximum level, and a more costly nonmonetary sanction is used. In any case, nonmonetary sanctions should play no role at such a low level of deterrence.

We believe that it is possible to explain such divergence between the theory and what can be observed in the real world without resorting to the argument of inefficient design of the policy[4].

[1] This conclusion has been formally stated by Polinsky and Shavell (1984), although we already find in Becker (1968, p. 193) that "social welfare is increased if fines are used *whenever feasible*". See also the surveys by Garoupa (1997) and Polinsky and Shavell (2000a).

[2] A recent empirical study supports the idea that fines could substitute license suspensions without violating any individual wealth constraint: Jorgensen and Wentel-Larsen (2002) estimate that the average willingness to pay of Norwegian car drivers for having back their license (suspended for six months) was about 600$ in 1997.

[3] If the penalty were defined according to the principle that the penalty should fit the crime, then neither a fine of thousands of euros nor imprisonment would be accepted as a sanction for double parking. The inclusion of notions of fairness in the theory of optimal sanctioning is discussed among others by Polinsky and Shavell (2000b) and Kaplow and Shavell (2001).

[4] An alternative argument to explain the use of nonmonetary sanctions may refer to incapacitation, as a sanction like driving licence suspension has the effect of keeping the driver from doing harm for some time. According to Shavell (1987) it is optimal to incapacitate an individual if the per period harm he can cause exceeds the per period social cost of incapacitation, and to keep him incapacitated as long as this condition holds. However, apart from licence suspension for drunken driving or similar cases in which the necessity to reduce an impending danger is obvious, this argument seems unable to provide by itself a rationale for nonmonetary sanctions in cases like that of

This paper argues that when a sanctioning policy is intended not only to deter individuals but also to inform them about the risk they incur in engaging in certain harmful activities, the conclusion about the optimal use of nonmonetary sanctions may differ from the standard one.

Public enforcement of law is motivated by the presence of externalities. However, we observe that a relevant characteristic of certain activities (such as driving or smoking) is that they are harmful to the individual engaging in them, and not only to others. As a consequence, the beliefs about the riskiness of those activities is an important determinant of behavior. In this situation, sanctions can affect the amount of harm produced not only by increasing the cost of harmful activities, but also by modifying the perception that individuals may have of the likely consequences of their activities.

The main result of the paper is that, when the sanctioning policy conveys information about the riskiness of the sanctioned behavior, the use of nonmonetary sanctions can be optimal even though the monetary sanction has not been used up to its possible maximum. This result, which is at variance with what is usually held, is obtained in a model with rational though uninformed individuals, and is based on the intuition that in some cases nonmonetary sanctions are more effective in conveying information about the riskiness of a certain harmful activity. We should note that in order to justify the fact that rational individuals cannot simply infer whether their actions are risky from the choices of the enforcer, we need to assume that they are uncertain about the objective function of the government. If the "credibility" of the enforcer becomes a concern, it is important for a benevolent enforcer to signal that the enforcement policy is aimed at deterring violations.

Indeed, different kinds of sanctions can score differently on this regard. Consider the case that individuals may assign a positive probability to the event where the enforcer has a stake in the revenue accruing from monetary sanctions: if they are not sure that the enforcer is maximizing social welfare, individuals can interpret a high monetary sanction not as an optimal deterrence policy in the face of a very risky activity, but as motivated by the desire to raise revenue. In these circumstances, nonmonetary sanctions may have an advantage over monetary fines: the use of nonmonetary sanctions may be a more credible signal, compared to monetary sanctions, of the enforcer's commitment to deter harmful activities.

In order to illustrate our point we will refer specifically to the case of driving behavior: in the model, violators will be drivers not complying to driving laws, e.g. exceeding a speed limit. The case of driving offenses provides in our opinion a good example of the interplay between deterrence and information. Since a relevant part of the harm deriving from driving accidents is suffered by drivers themselves or by their passengers, it is often the case that more careful driving habits can be induced

---

driving offenses.

by influencing the drivers' perception of the risk of incautious behaviors. Clearly, education is a central aspect of government intervention. That drivers should be made aware of the risk they run by not driving with care is implicit in many campaigns financed by the government. These campaigns try to educate drivers by pointing out the risk of speed, alcohol and lack of care[5].

The paper joins a large stream of literature that qualifies Becker's (1968) theory of deterrence. In recent years some contributions resorted to information related arguments (Bebchuk and Kaplow, 1992; Ben-Shahar, 1997), focusing on imperfect information about the probability of detection. To our knowledge, the present contribution is the first to introduce imperfect information about the probability of undesired consequences for the wrongdoers themselves, and to emphasize the role played by nonmonetary sanctions as a signal of the riskiness of harmful activities.

The paper proceeds as follows. Section 2 introduces the basic hypotheses and the model setup. Section 3 discusses the case of monetary sanctions alone. Section 4 analyzes the optimal mix of monetary and nonmonetary sanctions and states the main results. Section 5 summarizes and offers concluding remarks.

2. Model setup and hypotheses

Consider a population of identical drivers who gain a benefit b from violating a legal rule (e.g. when the choice is whether to pass a speed limit or not, b is the cost of driving at low speed), where b is a random variable.

The violation of the rule will cause an accident with probability q. Let K and E be respectively the cost for the driver and the external cost of the accident. We consider that K includes also liability costs and any other negative effect of the accident that the drivers internalize; note that it is very likely that K is by far the larger component of social cost of accidents.

In order to deter violations, the enforcer will sanction the violator using monetary and nonmonetary sanctions. Let F be the monetary fine to be paid when caught, and H the money equivalent of the nonmonetary sanction.

The theory of optimal monetary sanctioning points to the wealth of individuals as an upper bound for monetary sanctions: in this case, we have $F \leq \bar{F}$, while H is not subject to any limit. However, we will allow for alternative possibilities, including social norms constraining the maximal sanction that can be imposed for a violation (in this case the upper bound will be on the total amount of the sanction $F+H$—see below).

Let π be the probability that a driver who violates is caught. Increasing π is

[5] Indeed we observe that, when regulating driving behavior, the government seems to start from the assumption that the judgement made by driver who violate road regulation is very far from the rational and informed choice assumed by economists.

costly for the enforcer: the cost of detection is described by $C(\pi)$, with $C' > 0$ and $C'' > 0$ (the marginal cost of additional increases in the detection probability increases in π).

Additionally, we consider that it is not possible to reduce too much the probability of detection. This assumption can be justified considering that if the probability is too low, individuals may not be able to observe it correctly, and an incorrect expectation about the probability of being caught is socially costly as erroneous beliefs lead individuals to make socially inefficient decisions—a point made by Ben-Shahar (1997). In order to maintain the model as simple as possible, we do not model explicitly the cost of reducing π, and we simply assume that there is a lower bound $\bar{\pi}$ for π (this simplification does not modify the conclusion with respect to the case of a finite cost of reducing the probability).

We make the assumption that the probability q can take two possible values: high or low (respectively q_H and q_L, with $q_L < q_H$). At the outset, drivers have a prior r that the probability of an accident is high (q_H). So, the prior probability they assign to the eventuality of an accident is $p_0 = q_L + r(q_H - q_L)$.

We assume that the enforcer is better informed than drivers. He knows the "true" probability q. This assumption seems plausible as the enforcer have access to more information than the drivers (e.g. accident statistics). It would be in the interest of the parties (enforcer and drivers) that this information is shared, but we assume that the enforcer cannot produce verifiable evidence about the value of q. The drivers can update their beliefs on the risk of an accident by observing the optimal strategy of the enforcer. If drivers knew the objective function of the enforcer, they would be able to infer the value of q from the sanctioning policy selected. However, we assume that the enforcer's objective function is not known with certainty by the drivers. Namely, we assume that drivers assign a positive probability to the case that the enforcer maximizes a function which is different from the social welfare. For example, the drivers may believe that with probability β the enforcer maximizes $U = W + \alpha R$, i.e. a weighted sum of social welfare W *and* fine revenue R.

The reason why the budget R enters the objective function may be that the enforcer (the government) earns a rent from managing public funds (the rent is assumed to be proportional to the funds). We will label respectively "benevolent" and "nonbenevolent" the two types of enforcer (with $\alpha = 0$ and $\alpha > 0$).

We are interested in the normative problem of determining the optimal strategy for an enforcer who maximizes social welfare. Namely, we try to characterize the optimal policy for a benevolent enforcer, taking account that the drivers assign a probability β to the case that the enforcer is nonbenevolent. It is worth emphasizing that drivers' expectations on the type of the enforcer play a role in this setting because the enforcer's behavior conveys information on the likelihood of an accident, hence on the final payoff of the drivers.

We choose not to model explicitly the political mechanism that can induce a government to act in a benevolent way. However, we make the assumption that the political system is effective enough that even a nonbenevolent government cannot overtly behave in a way that reveals its type. In other words, the individuals must not be able to infer from the behavior of the enforcer that he is nonbenevolent (though in some circumstances they may be able to tell apart a benevolent enforcer). To be more precise, we assume that

(1) the nonbenevolent government cannot choose a strategy which would never be chosen by a benevolent government; and
(2) when the risk is high, the nonbenevolent government cannot choose a low deterrence policy when the risk of accident is high; although he can overdeter violations when the risk is low.

The reason for assumption 2) is that individuals can detect that with underdeterrence the government has chosen the "wrong" policy by observing a high number of accidents; on the contrary, overdeterrence is difficult to detect, because although individuals see that there have been high sanctions and a small number of accidents, the counterfactual is missing, i.e. they do not know what the number of accident would have been with lower deterrence.

We will show that the nonbenevolent enforcer may indeed find it convenient to pretend that $q = q_H$ (i.e. that the risk of an accident is high) even when it is $q = q_L$. This means that high sanctions do not necessarily mean that $q = q_H$, as they may well be compatible with $q = q_L$ *and* a nonbenevolent enforcer.

3. The optimal use of monetary sanctions alone

We consider first the case that the only instrument available to deter undesired behaviors is a monetary sanction. We will introduce nonmonetary sanctions in the framework in the next section.

In order to analyze the optimal policy it is convenient to refer directly to the expected sanction $S = \pi F$. $\overline{S} = \overline{\pi}\overline{F}$ will denote the minimum expected sanction which can be obtained by setting the sanction F at its maximum level $\overline{F}$. We indicate by $c(S)$ the minimum cost of detection necessary to implement an expected sanction S: it is $c(S) = C(\overline{\pi}) = c_0$ for $S \leq \overline{S}$ and $c(S) = C(S/\overline{F})$ with $c', c'' > 0$ for $S > \overline{S}$ (where the enforcer will be able to increase or decrease S only by changing π).

Given S, the individual will choose to violate the rule if

$$b - S - pK > 0 \tag{1}$$

where p is the probability that drivers assign to the occurrence of an accident (not necessarily equal to q). The social benefit from a violation of the driving code net of loss and enforcement cost is

$$b - q(K + E) - c(S); \tag{2}$$

the enforcer selects S to maximize

$$\int_{S+pK}^{\infty} [b - q(E + K)]dG(b) - c(S) \tag{3}$$

where G is the cumulative distribution of b, or $G(x) = \text{Prob}\{b \leq x\}$, with density g.

3.1. *Perfect information*

We start our analysis by considering the case of perfectly informed drivers who know the "true" probability of incurring an accident when violating a legal rule (i.e. $p = q_i$, with $i = L, H$). In order to align individual incentives and social costs/benefits, the optimal expected sanction S^i when the probability is q_i satisfies

$$S^i = q_i E - c'(S^i)/g(S^i + q_i K); \tag{4}$$

for $c'(S^i) = 0$ (e.g. because $S^i < \bar{S}$) the expected sanction $S^i = q_i E$ corresponds to the Pigouvian corrective tax for the externality, while for $c'(S^i) > 0$ it is $S^i < q_i E$.

3.2. *Asymmetric information*

When information is asymmetric, drivers update their expectation on the probability of accident by observing the sanctioning policy and considering what a benevolent and a nonbenevolent enforcer would do. In the equilibrium each type of enforcer maximizes his payoff given the expected optimal response by the drivers; and drivers choose on the basis of their beliefs, which are formed consistently with the observed behavior of the enforcers.

Consider the case that the benevolent enforcer fixes the expected sanctions at the perfect information levels. This strategy will result in an efficient outcome only if the drivers can infer the "true" risk of accident from the observed sanctions.

A driver observing a low sanction will infer that the probability is low, since by our assumptions it is not possible for the enforcer, even when nonbenevolent, to choose a low sanction when the risk is high. Therefore, when the risk is low there is no difference with respect to the perfect information case, and the optimal sanction will be S^L.

However, a driver observing $S > S^L$ cannot infer that the risk is high, since the nonbenevolent enforcer could find it convenient to fix a high sanction even when the risk is low, if by doing so he can increase his fine revenue. In other words, drivers are not able to distinguish between the case of a truly high risk and that of low risk with nonbenevolent enforcer.

We indicate by $p(S)$ the probability that drivers assign to incurring an accident as a function of the sanctioning policy implemented by the enforcer. This will be referred to as the "posterior" belief, which for $S > S^L$ and in case drivers cannot tell apart the case of high risk from that of a nonbenevolent enforcer, will be[6]

$$p(S) = q_L + \frac{r}{r+(1-r)\beta}(q_H - q_L) \equiv \tilde{p}; \tag{5}$$

comparing it to the expression for the prior p_0 above, we have for $\beta > 0$ and $0 < r < 1$

$$q_L < p_0 < \tilde{p} < q_H \tag{6}$$

while we have $\tilde{p} = q_H$ for $\beta = 0$, $\tilde{p} = p_0$ for $\beta = 1$. Drivers, observing a sanction $S > S^L$, update their belief and assign a higher probability to the case that the accident is more likely to occur, but due the possibility that the enforcer is nonbenevolent they do not assign to this case a probability equal to one. The higher is β, the lower is the distance between the prior and the posterior probabilities p_0 and $\tilde{p}$).

For a given sanction $S > S^L$, the number of drivers violating the rule is $1 - G(p(S)K + S)$. When the risk of accident is q_L, by setting a sanction equal to $S > S^L$ the non benevolent enforcer gets a payoff equal to

$$U(S) \equiv \int_{S+p(S)K}^{\infty} [b - q_L(K+E)]dG(b) + \alpha S[1 - G(p(S)K+S)] - (1+\alpha)c(S) \tag{7}$$

Therefore, he will disguise if S is such that $U(S) > U_L$ where U_L is the payoff obtained by setting S^L. In figure 1 we represent function $U(S)$ (which can be safely assumed to be $\cap$-shaped[7]) and the interval of values where $U(S) > U_L$, which we assume to be nonempty. The curve above $U(S)$ in the figure represents the payoff of the nonbenevolent enforcer should the drivers know that the true probability of accident is low. We call respectively S_m and S_M the lower and upper bounds of this interval.

Since outside the interval the nonbenevolent enforcer will have no incentive to disguise and deviate from the efficient strategy, we have that, for $S > S^L$,

$$p(S) = \begin{cases} \tilde{p} & S_m < S < S_M \\ q_H & \text{otherwise.} \end{cases} \tag{8}$$

[6] This is a simple application of Bayes' rule where r is the prior probability that $q = q_H$. Observing S the drivers update their prior about the event that the probability of an accident is high and they form their posterior about the probability of an accident.

[7] The first and third addendum on the right hand side of (7) are concave functions; the term $S[1-G]$ is concave if g does not decrease too rapidly as S increases (specifically, it must be $g'S/g > -2$); moreover, $U(S)$ is certainly increasing for low values and decreasing for high values of S.

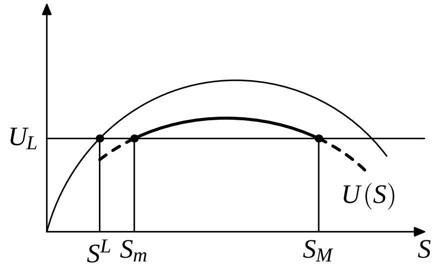


Figure 1

Therefore, there is no difference at all between the cases of symmetric and asymmetric information if $S^H \leq S_m$ or $S^H \geq S_M$. In order to focus on the case in which asymmetric information affects the outcome, we assume that $S_m < S^H < S_M$.

Given the best response of the nonbenevolent enforcer and its effect on the drivers' expectation, what is the optimal strategy for a benevolent enforcer? To induce an efficient level of violation when $q = q_H$, he will set S in order to maximize

$$\int_{S+p(S)K}^{\infty} [b - q_H(E+K)]dG(b) - c(S). \tag{9}$$

To simplify further the analysis we will restrict our attention to the case in which $c'(S)/g(S+pK)$ is monotonically increasing in S (this is indeed verified if the density function g is decreasing or is not increasing too rapidly, which is certainly true, when S is sufficiently high, for all single peaked density functions). Yet, since $p(S)$ introduces some discontinuities in the objective function, the problem has several local maxima that must be considered as possible optimal strategies for the benevolent enforcer. The following proposition characterizes these possible optimal strategies

Proposition 1. *The objective function*

1) *increases as $S \leq S_m$ and jumps downward at S_m, hence S_m is a local maximum; at S_m there is underdeterrence;*
2) *may or it may not admit a maximum S^* internal to the open interval (S_m, S_M); if such a maximum exists, it is $S^* > S^H$ for plausible values of the parameters, and in S^* there are more violations than in S^H;*
3) *has a discontinuity at $S = S_M$, and is decreasing for $S > S_M$: this implies two possibilities, depending respectively on whether the function jumps downward or upward at S_M. Namely:*
   - 3a) *the function admits no maximum at S_M: however, if the function is increasing for all $S < S_M$, it is optimal to set S as close as possible to (but less than) S_M (we denote this solution as S_M^-); in this case there will be underdeterrence;*

3b) S_M is a local maximum: at S_M there will be overdeterrence.

PROOF. See Appendix

Therefore, depending on the parameters of the model, and in particular on the shape of functions $c(S)$ and G—i.e. on how fast the cost of detection increases and the number of violation decreases as S is raised—the problem may admit an optimum with underdeterrence at S_m, S^* or S_M^-, or an optimum with overdeterrence at S_M.

In any case, the solution implies a lower social welfare than we have in the perfect information case.

3.3. A graphical representation

In order to illustrate the content of Proposition 1, we represent the different cases graphically, under the simplifying assumptions that $c'(S) = 0$ throughout the whole interval of relevant values of S.

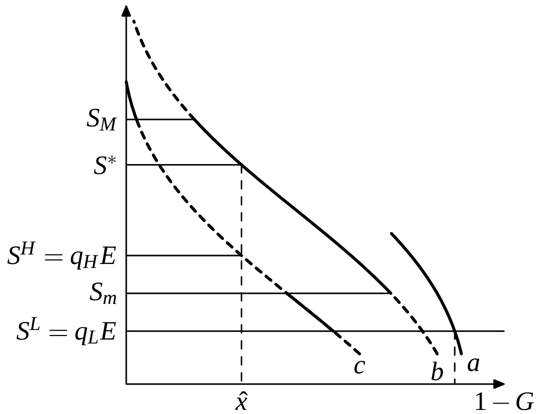


Figure 2

We plot three curves, representing the frequency of violation $1-G(S+pK)$ as a function of S for p equal respectively to q_L, $\tilde{p}$ and q_H (a higher perceived risk shifts the curve downwards, inducing a lower number of violations for given S). We put the sanction on the vertical axis, to emphasize the analogy with demand curves: on the horizontal axis we have the "consumption" of violations as a function of their "price".

Figure 2 represents the case in which a maximum S^* ("internal" solution) exists. The sanction is such that drivers cannot infer the "true" value of q; they use their posterior $\tilde{p}$ to take their decisions, so that the relevant demand curve is the middle one. With posterior probability less than the true probability, the enforcer must

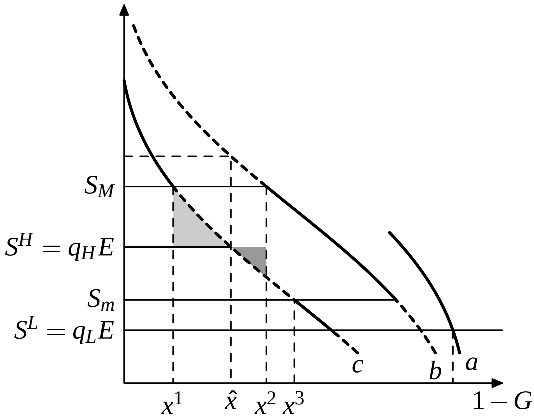


Figure 3

increase the sanction above S^H in order to obtain the same level of deterrence that we have in the perfect information case (frequency of violation $\hat{x}$). Having assumed that $c'(S) = 0$, (22) simplifies to

$$S^* = [q_H - \tilde{p}]K + q_H E. \tag{10}$$

Note that, under this simplifying assumption, the fact that $S^* > S^H$ implies no additional cost for the enforcer, while in the general and more realistic case that $c'(S) > 0$ there is an additional cost equal to $c(S^*) - c(S^H)$. Moreover, with $c'(S) = 0$ the solution S^* always dominates S_M.

Figure 3 represents the case in which the payoff function is always increasing for $S < S_M$. In this case the sanction that would be necessary to induce the same deterrence as S^H is higher than S_M. As discussed in Proposition 1, there are two possibilities. One is that the enforcer selects S_M, so that drivers infer that the risk of accident is high ($p(S_M) = q_H$): the relevant "demand" curve is the lowest one, and the enforcer will overdeter, since the number of violations will be suboptimal (equal to x_1 in the figure). Alternatively, the enforcer can select S_M^-, so that the quantity of violations is x_2. The choice between these two policies depends on the relative efficiency costs of oversanctioning (x_1) and undersanctioning (x_2), represented by the shaded triangles in the figure.

In figure 3, it is also illustrated the choice of S_m, which induced a level x_3 of violations (there is underdeterrence, since when drivers posterior is correct the optimal sanction should be S^H).

4. The optimal mix of monetary and nonmonetary sanctions

In this section we show how the use of nonmonetary sanctions can increase efficiency for reasons different from those emphasized in the existing literature.

As is well known, an important difference between monetary and nonmonetary sanctions is that the latter imply a social cost, while the former are simply a transfer, and therefore the cost for those sanctioned is exactly offset by a gain of those who receive the fine revenue. Another aspect, which is relevant in our model, is that the nonbenevolent enforcer has a direct interest in increasing the fine revenue, while the gain from increasing a nonmonetary sanction is lower (if not zero). We will see that, if we take into account both aspects, the standard conclusion about the superiority of monetary sanctions is not warranted anymore.

We consider here the case that nonmonetary sanction are not costly to impose for the enforcer; we have in mind sanctions different from imprisonment, such as vehicle forfeiture or driving licence withdrawal. Moreover, we allow for some benefits accruing from the nonmonetary sanctions; for example when drivers are required to attend a course in order to have their driving licence back after a withdrawal, it may be the case that the course increases the ability of the drivers, and affects the future risk of accidents: in this case, the benefit for society corresponds to the reduction in future externalities. In addition, a sanction as driving licence withdrawal may prevent the individual from doing harm for a certain period (this is the incapacitation argument we mentioned in footnote 4). However, since these gains are generally less than the loss for those sanctioned[8], we consider that a nonmonetary sanction whose amount is H implies a social loss of δH, with $\delta > 0$ (if $\delta = 0$, the nonmonetary sanction would be equivalent to a monetary transfer).

If we suppose that nonmonetary sanctions are not limited by wealth or other constraints, any level of expected monetary sanction $N = \pi H$ can be reached for a given probability of detection π simply by increasing the amount of the sanction H. Therefore, nonmonetary sanctions do not affect the cost of detection as long as S is nonzero (so that $\pi > 0$). Below, we will consider as a separate case the possibility that H is bounded too.

The objective function of the enforcer when both monetary and nonmonetary sanctions are available is

$$\int_{\pi(H+F)+pK}^{\infty} [b - q(E+K) - \delta\pi H] dG(b) - C(\pi) \tag{11}$$

4.1. Nonmonetary sanctions and perfect information.

The standard conclusion on the optimal mix of monetary/nonmonetary sanctions is that a nonmonetary sanction should be used only to supplement a monetary one when the latter has been used to its maximum extent. In our context, this can be expressed as follows:

[8]If the gains from the nonmonetary sanction were higher than the benefit for the driver, the imposition of the sanction would be optimal irrespective of its deterrence effect. A case that is very unlikely to be verified in most circumstances, and that for this reason we rule out from our analysis.

Proposition 2. *When $p = \underline{q}$ (perfect information), it is never optimal that at the same time $H > 0$ and $F < \bar{F}$.*

PROOF. We simply show that, for given π, if $F < \bar{F}$ and $H > 0$, it is always possible to increase (11) by increasing F. Consider the effect of an increase in F exactly offset in terms of deterrence by a decrease in H, or $dH = -dF$: since $dH + dF = 0$, the change in social welfare is $[1 - G(\pi(H+F) + qK)]\delta\pi dF > 0$. □

We emphasize that this result is not warranted anymore when a change in F and H affects the probability p, as it is the case when information is asymmetric.

With perfect information, we can rewrite (11) as

$$\int_{N+S+qK}^{\infty} [b - q(E+K) - \delta N] dG(b) - c(S) \tag{12}$$

and maximize it with respect to S and N. The necessary conditions for maximization are

$$-[qE - S - (1-\delta)N] g(N+S+qK) - c'(S) = 0. \tag{13}$$

and

$$-[qE - S - (1-\delta)N] g(N+S+qK) - \delta[1 - G(N+S+qK)] \leq 0 \tag{14}$$

where the last expression must be an equality for $N > 0$. It follows that

$$c'(S) \leq \delta[1 - G(N+S+qK)] \tag{15}$$

with the equality when $N > 0$. In words: the enforcer will not use nonmonetary sanctions as long as the marginal cost of of detection $c'(S)$ is lower than the marginal cost of nonmonetary sanction to those sanctioned.

4.2. *Nonmonetary sanctions and asymmetric information.*

When information is asymmetric with respect to both the risk of accidents and the objective function of the enforcer, we must take into account that monetary and nonmonetary sanction may have different signaling power, which can be exploited by the enforcer in order to increase efficiency.

The way in which the introduction of nonmonetary sanction affects the optimal solution depends on which of the cases discussed in Proposition 1 is the relevant one.

Proposition 3. *When information about the risk of an accident and about the objectives of the enforcer is asymmetrically distributed between the enforcer and drivers, the nonmonetary sanction N which optimally supplements a monetary sanction S is as follows:*

1) *when the optimal monetary sanction is* S_M *it is* $N = 0$;
2) *when the optimal monetary sanction is* $S^* \in (S_m, S_M)$, $N > 0$ *only if the marginal cost of an increase in the nonmonetary sanction is less than the marginal increase in the cost of detection (as in the perfect information case)*

$$c'(S^*) > \delta[1 - G(S^* + \tilde{p}K)] \qquad (16)$$

3) *if the optimal monetary sanction is* S_m *or if it is* S_M^-, *it can be* $N > 0$ *even if the marginal cost of the nonmonetary sanction is higher than the marginal increase in the cost of detection, or*

$$c'(S) < \delta[1 - G(S + pK)] \qquad (17)$$

PROOF. See Appendix

Result 3) of the Proposition shows that the scope of nonmonetary sanctions is widened in the case of asymmetric information. The reason is that nonmonetary sanction, though more costly, make it possible to increase deterrence while controlling at the same time drivers' beliefs. The case in which S_m is optimal is of particular interest: here nonmonetary sanctions allow the enforcer to increase deterrence above S_m without making the signal "noisy", as it would have been the case had the enforcer set a monetary sanction $S > S_m$[9].

The conclusion that nonmonetary sanctions are not necessarily dominated by the monetary sanction when information is asymmetric can be made even more clear when we consider the case that $S < \bar{S}$. We defined $\bar{S}$ as the minimum monetary sanction such that the fine is maximum ($F = \bar{F}$) and satisfying $\pi \geq \bar{\pi}$, where $\bar{\pi}$ is the lower bound on the probability of detection. A reduction of the expected sanction below the level $\bar{S} = \bar{\pi}\bar{F}$ is possible only by setting $F < \bar{F}$. It follows that for $S < \bar{S}$ it is $c'(S) = 0$.

Consider now the case that the optimal sanction is S_m. If $S_m < \bar{S}$, it will be $F = S_m/\bar{\pi} < \bar{F}$ in the optimum, even if $N > 0$. This can be contrasted to what is stated in Proposition 2 for the case of perfect information.

4.3. *Upper limit to the total sanction*

Now we consider the case that the upper limit on sanctions is not related to the wealth of individuals, but reflects some other external constraint. For instance, the government may want to respond to some view of fairness shared by the individuals,

[9] Similar considerations can be advanced for the case in which the optimal sanction is S_M^-, even though in that case the role of nonmonetary sanction is to add noise to the signal received by drivers, concealing the true value of the risk of accident.

who believe that the sanction should "fit the crime"[10]. Clearly, in this case it is not plausible that the upper bound applies to the monetary sanction alone. Rather, it will be a limit on the *total* amount of the sanction, monetary and nonmonetary.

The existence of an upper limit to total sanction can be expressed by the constraint $F + H \leq \bar{F}$. Under this assumption, there is no role at all for nonmonetary sanctions in the perfect information case. Indeed we have

Proposition 4. *Under conditions of perfect information ($p = q$), when there is an upper limit on the total sanction—monetary and nonmonetary—it is always inefficient to set $N > 0$.*

PROOF. It is sufficient to show that, for any value of H and F such that $H + F \leq \bar{F}$ and $H > 0$, it is always advantageous to substitute a monetary sanction for the nonmonetary one by increasing F and decreasing H. Considering dH and dF such that $dH + dF = 0$ the change in the payoff (11) is $[1 - G(\pi(H+F) + qK)]\delta\pi dF > 0$. It follows that the payoff increases as H is reduced. □

Once again, asymmetric information brings us to different conclusions. Let us start from a situation where S_m is optimal when only monetary sanctions are used. To see that the use of nonmonetary sanctions can improve on the payoff obtained in S_m, we analyze the effect of a marginal increase of N at $S = S_m$ and $N = 0$. Considering that the constraint $F + H = \bar{F}$ must be always binding in an optimum[11], so that we have $C(\pi) = c(S + N)$, the marginal change in the payoff is

$$[S_m - qE]g(S_m + qK) - \delta[1 - G(S_m + qK)] - c'(S_m) \tag{18}$$

since there is nothing preventing this expression from being positive, it may well be optimal to increase N above zero.

We summarize our conclusions of this and the previous section about the use of nonmonetary sanctions in the following

Proposition 5. *When information about the risk of an accident and about the objectives of the enforcer is asymmetrically distributed between the enforcer and drivers,*

[10]Polinsky and Shavell (2000b) consider the case of a government which responds to fairness oriented preferences of the citizens: "When the issue of fairness is added to the analysis, however, the usual solution is generally not optimal because a very high sanction will be seen as unfair, or more precisely, will result in the lowering of individuals' fairness-related utility. With respect to double parking, even a sanction of $100 might be considered unfair, let alone a sanction of $10,000". Our analysis differ in that for simplicity we take the upper limit as a constraint in the enforcer's objective function, rather than introducing a concern for fairness in the objective function itself.

[11]If this were not the case, by choosing $dF > 0$ and $d\pi < 0$ such that $d\pi(F + H) + \pi dF = 0$, it would be possible to decrease π while leaving H and total deterrence $N + S = \pi(F + H)$ unchanged. Clearly, this increases efficiency.

so that the sanctioning policy conveys information on the riskiness of the sanctioned behavior, the use of nonmonetary sanctions can be optimal even when the monetary sanction has not been used to its maximum possible extent.

In other words, Proposition 2 and Proposition 4 do not carry over to the case of asymmetric information.

5. Concluding remarks

In this paper, we have incorporated educational and informative concerns into the standard theory of deterrence by considering a situation in which a certain act may be harmful both to the individual committing it and to others. In this case, the belief about the riskiness of the activity is an important determinant of behavior and therefore of the amount of harm produced. We have proposed a model in which the enforcer's sanctioning policies can affect this belief. In this framework, the effectiveness of sanctions depends not only on their direct deterrence effect, but also on how credible it is as a signal of a high risk of harm.

We have shown that in these circumstances a benevolent government, namely a government which maximizes the expected social cost from the harmful activity, can benefit from the use of nonmonetary sanctions even when monetary sanctions have not been used to their maximum extent.

There are circumstances, such as those described in our model, in which monetary and nonmonetary sanctions are not perfect substitutes not only because they imply different costs, but also in that they provide different signals to individuals. In particular, nonmonetary sanctions are more credible in transmitting information about the riskiness of possible individuals' actions.

We have referred specifically to the case of driving offenses. The hypotheses at the basis of our model seem to fit well this case: most of the harm from accidents is very often suffered by drivers' themselves; governments seem to design their policies assuming that drivers are not able to assess correctly the risk of driving (indeed, information campaigns are often an important dimension of these policies); sometimes monetary fines are interpreted by drivers as a way to raise revenue rather than an instrument to deter harmful activities. Our model provides an explanation of the wide use of nonmonetary sanctions which takes into account all these characteristics.

Appendix

Proof of Proposition1

First, we calculate the derivative of the objective function (9) for given belief p by the individuals:

$$-[S+pK-q_H(E+K)]g(S+pK)-c'(S) \tag{19}$$

This is monotonically decreasing in S if $c'(S)/g(S+pK)$ is increasing in S, as we assumed. Moreover, by definition of S^H, it is equal to zero in S^H for $p=q_H$.

1) For $S \leq S_m$ it is $p=q_H$, while for $S>S_m$ it is $p=\tilde{p}<q_H$. The change in the objective function as S increases above S_m is

$$\int_{S_m+\tilde{p}K}^{S_m+q_HK} bg(b)db - q_H(E+K)[G(S_m+q_HK)-G(S_m+\tilde{p}K)]<0 \tag{20}$$

(the sign of the inequality follows from $S_m<S^H\leq q_HE$). Therefore, the objective function jumps downward above S_m. There is underdeterrence, since when drivers' posterior is correct the optimal sanction is $S^H>S_m$.

2) In the interval (S_m,S_M) we have $p=\tilde{p}$. A maximum internal to the interval will satisfy

$$-[S+\tilde{p}K-q_H(E+K)]g(S+\tilde{p}K)-c'(S)=0. \tag{21}$$

In this case we have

$$S^*=[q_H-\tilde{p}]K+q_HE-c'(S^*)/g(S^*+\tilde{p}K). \tag{22}$$

It can be shown that $S^*>S^H$ (the optimal sanction is higher than in the perfect information case) provided that g' is low enough. More specifically, by differentiating the first order condition we have that a sufficient condition for $dS/d\tilde{p}<0$ is that $g'<g^2/c'$, so that this is certainly verified for $g'<0$ (which is the case for a single peaked density function when S^H is large enough) or for $c'=0$.

Moreover, the amount of violation is clearly higher than in the perfect information case, since from the first order condition we have that $S^*<S^H+(q_H-\tilde{p})K$, so that $G(S^H+q_HK)>G(S^*+\tilde{p}K)$.

3) Since for $S\geq S_M$ drivers learn the true value q_H, at S_M the objective function is discontinuous; the change at S_M is

$$-\int_{S_M+\tilde{p}K}^{S_M+q_HK} bg(b)db+q_H(E+K)[G(S_M+q_HpK)-G(S_M+\tilde{p}K)] \tag{23}$$

this can be either positive or negative depending on the parameters of the model[12].

[12] Note that (23) corresponds to the difference between the two shaded triangles in figure 3 (the one on the left minus the one on the right).

If there is no $S < S_M$ satisfying (21) because the derivative is always positive, then the objective function admits no local maximum in the open interval (S_m, S_M). However, when (23) is negative, S_M is the supremum of the function, and the "optimal" strategy for the enforcer is to set the sanction as high as possible compatible with $S < S_M$ (note that this implies that $S > S^H$).

The level of violations will be higher than in the perfect information case because, for (19) to be increasing for $S < S^M$, it must be $S_M < S^H + (q_H - \tilde{p})K$, so that $G(S^H + q_H K) > G(S_M + \tilde{p}K)$.

If (23) is positive, then at S_M the function jumps upward; moreover, since $p = q_H$, the function is the same as in the perfect information case: from $S_M > S^H$ follows that S_M is a local maximum and that at S_M there is overdeterrence. □

Proof of Proposition3

1) Since in S_M we have overdeterrence, there is no gain from supplementing a monetary sanction which is higher than optimal.

2) The necessary condition simply follows from differentiation of the objective function with respect to N

$$\int_{N+S+\tilde{p}K}^{\infty} [b - q(E+K) - \delta N] dG(b) - c(S); \tag{24}$$

using the first order condition for the maximization with respect to S,

$$-[qE + (q - \tilde{p})K - S - (1-\delta)N] g(N + S + \tilde{p}K) - c'(S) = 0 \tag{25}$$

we have that at $S = S^*$ and $N = 0$, social welfare increases only if condition (16) is satisfied.

3) If S_m is the optimal monetary sanction, the objective function is the same as in the case of perfect information, i.e. (24). Therefore, $N > 0$ increases social welfare at $S = S_m$ and $N = 0$ if

$$[S_m - q_H E] g(S_m + q_H K) - \delta [1 - G(S_m + q_H K)] > 0 \tag{26}$$

since at S_m we have that in general

$$[S_m - q_H E] g(S + q_H K) - c'(S_m) > 0 \tag{27}$$

it may well be that it is optimal to have $N > 0$ even when (17) is satisfied.

It is possible to proceed in a similar way for the case in which the optimal nonmonetary sanction is S_M^-. □

References

Bebchuk, L. A., Kaplow, L., 1992. "Optimal sanctions when individuals are imperfectly informed about the probability of apprehension". *Journal of Legal Studies*, vol. 21, pp. 365–70.

Becker, G. S., 1968. "Crime and punishment: an economic approach". *Journal of Political Economy*, vol. 76, pp. 169–217.

Ben-Shahar, O., 1997. "Playing without a rulebook: optimal enforcement when individuals learn the penalty only by committing the crime". *International Review of Law and Economics*, vol. 17, pp. 409–21.

Garoupa, N., 1997. "The theory of optimal law enforcement". *Journal of Economic Surveys*, vol. 11, pp. 267–95.

Jorgensen, F., Wentel-Larsen, T., 2002. "Car drivers' willingness to pay for not loosing their driving license". *Transportation*, vol. 29, pp. 271–86.

Kaplow, L., Shavell, S., 2001. "Any non-welfarist method of policy assessment violates the Pareto principle". *Journal of Political Economy*, vol. 109, pp. 281–6.

Polinsky, A. M., Shavell, S., 1984. "The optimal use of fines and imprisonment". *Journal of Public Economics*, vol. 24, pp. 89–99.

Polinsky, A. M., Shavell, S., 2000a. "The economic theory of public enforcement of law". *Journal of Economic Literature*, vol. 38, pp. 45–76.

Polinsky, A. M., Shavell, S., 2000b. "The fairness of sanctions: some implications for optimal enforcement policy". *American Law and Economics Review*, vol. 2, pp. 223–37.

Shavell, S., 1987. "A model of optimal incapacitation". *American Economic Review (Papers and Proceedings)*, vol. 77, pp. 107–10.

SIMPLE

http://www.unisi.it/lawandeconomics/simple.html

ELENCO DEI CONTRIBUTI RICEVUTI FINO AL NUMERO ODIERNO

2005		
29	Massimo D'Antoni, Roberto Galbiati	Deterrence And Information: The Optimal Use Of Monetary And Nonmonetary Sanctions Revisited
2004		
28	Antonio Nicita, Ugo Pagano	Institutional complementarities, corporate governance and financial-technological equilibria
27	Laura Ammannati	Tutela della concorrenza e accesso al mercato dei servizi pubblici locali dell'energia: il caso del gas
26	Elisabetta Cervone	EU conduct of business rules and the liberalisation ethos.The challenging case of investment research
25	Antonio Nicita	Hold-up, competition and vertical integration:another look at fisher body/general motors?
24	Antonio Nicita, Ernesto Savaglio	Minimal liberty and the 'coasian liberal': boundaries and complementarities between the state and the market
23	Antonio Nicita, Ugo Pagano	Law and Economics in retrospect
22	Antonio Nicita, Matteo Rizzolli	Much Ado About The Cathedral: Property rules and liability rules when rights are incomplete
21	Pier Luigi Parcu	The dominant position and its abuse: suggestions from game theory
20	Emilio Asaro	Collusion and auctions: a survey
2003		
19	Francesco Farina	Game-theoretic solutions to endogenous contractual incompleteness
18	Simone Sepe	La cartolarizzazione: aspetti giuridici ed economici
17	Francesco Denozza	Consumer transaction costs at the intersection of antitrust and intellectual property
16	Thomas S. Ulen	Firmly grounded: economics in the future of the law
15	Luca Fiorito	Transaction cost economics before Coase: L.K. Frank on industrial integration
14	Roberto Galbiati	Durable goods monopolies: some insights on time inconsistency
13	Antonio Nicita	Concorrenza e sanità: il 'pendolo delle riforme' in Europa
12	Laura Ammannati, Roberto Galbiati, Giorgio Nizi	Regolazione e concorrenza nei servizi pubblici locali
11	Marcello Clarich	Servizi pubblici e diritto europeo della concorrenza: l'esperienza italiana e tedesca a confronto
10	Giuseppe Bellantuono	Contratti incompleti e norme sociali
9	Donatella Porrini	Financial internalization of environmental damage: legal framework for economic solutions
8	Ruggero Paladini	Law & Economics ed economia pubblica
7	Maurizio Franzini, Antonio Nicita	Externality, efficiency and fairness: understanding inertia in coase theorem
6	Lorenzo Sacconi	Incomplete contracts and corporate ethics: a game theoretical model under fuzzy information
5	Antonio Nicita	Esternalità, transazioni e ambiente. Una rivisitazione del teorema di Coase
4	Arman Khachatrian	Theory of the firm. A critical survey
3	Giovanni Ramello	Diritto d'autore e tutela della concorrenza nell'esperienza antitrust europea e italiana
2	Antonio Nicita, Ugo Pagano	Le istituzioni economiche del capitalismo alla prova della globalizzazione
1	Laura Ammmannati	Diritto e Mercato. Una rilettura delle loro attuali relazioni alla luce della nozione di 'transaction' di Commons

simple

Siena Memos and Papers on Law & Economics

***La collana di quaderni SIMPLE** nasce nell'ambito della ricerca collegata alla Scuola di Dottorato di Ricerca in Diritto ed Economia (sezione Law and Economics), nata dalla collaborazione tra l'Università di Siena, l'Università di Trento, il Liuc Castellanza e la Scuola Superiore in Economia e Finanza. Il dottorato L&E si propone di formare studiosi e dirigenti di istituzioni economico-sociali nazionali ed internazionali e di imprese per le quali sia richiesta una preparazione interdisciplinare con particolare attenzione all'area giuridico-economica.*

Il percorso didattico, al quale contribuiscono docenti provenienti da diverse università italiane ed estere, nonché esperti di istituzioni pubbliche e private, si propone di costruire un bagaglio di conoscenze e strumenti comuni nelle aree di diritto ed economia, partendo da basi disciplinari differenziate, come quelle offerte nella facoltà di economia, giurisprudenza e scienze politiche. In particolare, l'attività didattica e di ricerca offerta dal dottorato è rappresentata dalle seguenti macro-aree:Contratti, Diritti di Proprietà, Responsabilità; Organizzazioni economiche e Corporate Governance; Concorrenza, Regolazione e Antitrust; Istituzioni, Organizzazioni e Scelte pubbliche; Regolazione e Organizzazioni internazionali. Nell'ambito di questo percorso formativo sono previsti inoltre moduli di formazione di base per economisti e giuristi e moduli/seminari specialistici, tra i quali: diritti di proprietà intellettuale, finanza d'impresa, sistemi finanziari, analisi dell'impatto della regolamentazione, etica e impresa, diritto penale ed economia del crimine, tutela ambientale e tutela del consumatore. L' iter formativo si divide in tre segmenti: uno di frequenza obbligatoria delle lezioni della durata di un anno e mezzo, un secondo di eventuale formazione all'estero e/o di stage presso istituzioni o enti di ricerca, di durata variabile da un semestre ad un anno, l'ultimo, focalizzato alla preparazione della tesi. L'obiettivo formativo è diretto a condurre lo studente ad una preparazione interdisciplinare orientata ad una successiva specializzazione tematica.

http://www.unisi.it/lawandeconomics