Emanuela Carbonara[♣] - Francesco Parisi[♦] - Georg von Wangenheim[♠]

# Unjust Laws and Illegal Norms[*]

Preliminary Version

Submission for annual Meeting of
**Verein für Socialpolitik 2009**

**Abstract:**

Due to a variety of circumstances, lawmakers occasionally create laws whose aims are perceived as outright unjust by the majority of the people. In other situations, the law may utilize improper means for the pursuit of a just goal. In all such cases, lawmaking processes generate rules that do not reflect the values of the underlying population. In these cases individuals may face legal commands or prohibitions that conflict with their sense of justice or fairness. Individuals can oppose unjust laws through protest. Social opposition to unjust laws may trigger social norms that can have countervailing effects on legal intervention. The dynamic effects of these phenomena are the object of this paper.

*JEL* Classifications*: K10, K42, D70, B52, Z13.*
*Keywords: Social* Norms*, Countervailing Effect, Expressive Function, Law Enforcement, Civil Disobedience.*

February 2009

[♣] University of Bologna, Department of Economics.
[♦] University of Minnesota Law School and University of Bologna, Department of Economics.
[♠] University of Kassel, Faculty of Law.
[*] We thank Robert Cooter, Tzili Dagan, Ben Depoorter, Yuval Feldman, Mark Grady, Claire Hill, Robert MacCoun, Richard McAdams, Brett McDonell, Gideon Parchomovsky, Hans-Bernd Schäfer and the participants at the Hebrew University and Bar-Ilan University Conference on Law and Social Norms, and the University of Minnesota Conference on Law and Empathy, and faculty workshops at the University of Hamburg and University of Bologna for insightful comments.

*"I think we should be men first, and subjects afterward. It is not desirable to cultivate a respect for the law, so much as for the right. The only obligation which I have a right to assume is to do at any time what I think right."*[1]

# 1. Introduction

Changing sanctions to enforce the law usually has the expected effects on the regulated action: higher sanctions induce less activity, lower sanctions allow for more of the activity. There are however, examples where the reverse happens.

During the high times of terrorism in West Germany in the 1970s, criminal law and criminal procedural law were made substantially stricter in all fields related to terrorism. Not only have foundation of, and membership in a terrorist organization been penalized, but also any "support of, or recruitment for" such organizations. All suspects in pre-trial custody and inmates found guilty of violating these provisions were or could be made subject to severe restrictions of their rights as prisoners. In the first years after enactment, the cases to which the new law has been applied were small in number but prominent in quality. The law was used to oppress in particular lesser forms of support of terrorism which could not be subsumed to any other criminal offense – sometimes consisting only of distributing flyers calling for better custody conditions of others. The effect, though, was a clear increase in the openly expressed opposition against these laws and at the same time an increase in the number of individuals actively criticizing custody conditions of terrorists and supporting them in more severe ways – all this under the encouragement of an increasing number of persons opposing to the strict stance the all state powers took on such activities.

In the early 2000's, the US Supreme Court handed down some rather liberal decisions on so-called partial birth abortions. The immediate effect was an upsurge in the discussion on the legality and legitimacy of this kind of abortions. As a consequence, the number of partial birth abortions declined.

I both examples, the sanction had not only the direct effect on the incentives to act but also an indirect effect transmitted by vivid discussions about social norms and by their effect on actual behavior.

In this paper, we discuss how the interaction of social norms expressed by protest and legal norms may entail reverse effects as described in the two examples. We first develop model of this interaction resulting in the possibility of having multiple equilibria of action and social norms (Section 2). The standard way to explain reverse effects in this class of models is to argue that by parametric changes, some of the multiple equilibria vanish and therefore big leaps in the endogenous variables occur, possible in a direction that may explain the reverse effects we presented in the examples. We argue that in our simple model of interaction between social norms and action, equilibria will vanish in the way we would need to explain reverse effects would require that a parametric shift in the size of the sanction affects the social-norm formation and the decisions on action in a counterintuitive way: the effect of the

[1] Henry David Thoreau, Civil Disobedience, in Walden and Other Writings by Henry David Thoreau 85, 86 (Joseph Wood Krutch ed., 1981) (1849).

change in the size of the sanction on social norms expressed by open protest against the sanction would have to far stronger than its effect on the individual decisions on action (Section 3). We therefore reject this explanation of reverse effects and proffer alternative explanations (Section 4), which not only are more plausible but also allow for policy recommendations on how to avoid such reverse effects (Section 5). Before we come to the conclusions (Section 7) we refine our model with respect to explanations of the existence of multiple equilibria (Section 6).

## 2. The Model

Imagine a simple framework, where there is one illegal action *a*, on which the law levies a sanction *S*. For instance, we might think of actions like smoking in public spaces, illegally copying recorded music, carrying out abortions, or supporting terrorists by advocating humane conditions of detainment. The sanction can vary in both its severity and frequency of application. For simplicity, we assume risk-neutrality and define *S* as the expected sanction in case of non-compliance. An expected sanction $S = 0$ means a law lacking an expected sanction, whereas *S* increases as the severity of the sanction or the level of its enforcement increase.[2] The primary effect of a sanction is to deter individuals from engaging in the relevant behavior. Sanctions increase the relative price of the sanctioned activity and possibly lead to a substitution towards other non-sanctioned activities.

The decision to perform the action not only depends on the legal sanction, but also on the value an individual attaches to the action and social sanctions. For expositional ease, we write the individual valuation of the action as $B - v_a$, where $B$ is a common value for all individuals and $v_a$ is the individuals' deviation from this common value.[3] We assume that the individual values of $v_a$ vary among individuals according to the distributions function $G(\cdot)$ with the corresponding density $g(\cdot)$.

Social sanctions are assumed to be negatively related to currently expressed aggregate social norms (measured by $n_t$ – a variable to be explained in more detail shortly) and to the proportion $x_t$ of individuals currently performing action *a*, the latter because a high social prevalence of the action reduces social stigma. Analytically, this implies that all individuals with

$$B - v_a > S + \beta n_t - \alpha x_t \tag{1}$$

will perform the action, where $\alpha$ and $\beta$ are non-negative constants. Then the proportion of individuals who would be willing to perform the action is given by $G(B - S - \beta n_t + \alpha x_t)$. Individual decisions to act adapt over time to the current state. Thus, the change in the proportion of individuals performing the illegal action in each period is

$$\dot{x} = G(B - S - \beta n + \alpha x) - x \tag{2}$$

where the dot denotes the derivative with respect to time and we omitted the time index for notational simplicity.[4]

---

[2] A negative value of *S* would be a subsidy to the activity. We refrain from discussing this case in detail, but our results could be easily extended in this direction.

[3] We do not place any restrictions on $v_a$, so that it may be positive or negative and we can, for example, interpret *B* as the average benefit of the activity.

[4] In equation (2) the pace of this adaptation has been normalized by defining time periods so that only one individual gets to decide each time. We have then assumed continuous time. For a detailed discussion of this shift to continuous time see e.g. Weidlich and Haag (1983).

We further assume that besides the valuation $B - v_a$ relevant for the decision to act, each individual attaches a normative valuation to the activity, expressed by $v_n$, the size of the legal sanction the individual would deem appropriate for the action. The lower the normative valuation of the activity by an individual, the less she would like to see the action being performed by others and hence the larger her most preferred sanction $v_n$. Let the distribution of the individual normative valuation in the society be given by $F(\cdot)$, with density $f(\cdot)$.[5]

The individual values of $v_n$ and their distribution in society are initially opaque, but can be revealed to others through the expression of opinions and social reaction to action $a$ and to laws regulating and sanctioning *a*. Let $q_t$ be the proportion of individuals who disapprove of action $a$ to a degree that they protest against the law at time *t*, or otherwise openly object to the activity, because they deem the sanction for action *a* to be too low.

Similarly, let $p_t$ be the proportion of individuals who approve of action *a* to a degree that they protest against the law at time *t*, or otherwise openly support the activity, because they deem the sanction for action *a* to be too high. Taking these two forms of expressed opinions together, we define the expressed social norm on the action at time *t* as $n_t = q_t - p_t$, *i.e.*, by the expressed rejection of action $a$ beyond the legal norm, net of the expressed support of the action.

The open expression of valuations itself adapts to the current state of the society in the course of time. To understand its dynamic evolution, it should be recalled that individuals with a high valuation $v_n$ are more likely to disapprove of action $a$ and thus to oppose to the law for it being too lenient, whereas those with a low $v_n$ are more likely to approve of $a$ and thus to oppose the law for being too strict. However, the expression of opinions and protest are costly. Some of these costs are financial opportunity costs, like taking time off work or leisure to protest against an unjust law. Some others are psychological costs, like the cost of expressing opinions that run contrary to prevailing social norms either inferred from expressed opinions of others or from their behavior. The psychological costs of protest will be lower when an individual's preference regarding action $a$ and her possible dislike of the law are in line with the opinions of many other members of the society, inferred from observed protest and from observed action. Thus, an individual who would like a stricter law, will thus want to express her opinion openly only when her preference for stricter law is strong enough to offset the costs of expressing the opinion, i.e. when $v_n - S > c - \lambda n_t + \gamma x_t$, where $c$ represents the part of the costs of expressing one's opinion which is independent of other individuals' behavior and opinion expression, and $\lambda$ and $\gamma$ are non-negative constants. Similarly, an individual who would like to see the law to be more lenient, will want to express her opinion openly only when $S - v_n > c + \lambda n_t - \gamma x_t$.

Then, the proportion of individuals voicing their preference for a stricter law is $1 - F(S + c - \lambda n_t + \gamma x_t)$, whereas the proportion of those voicing a preference for a more lenient law is $F(S - c - \lambda n_t + \gamma x_t)$. The dynamic process of valuation expression may thus be represented by the two equations

$$\dot{q} = \sigma\left(1 - F\left(S + c - \lambda n + \gamma x\right) - q\right) \tag{3}$$

and

[5] It is highly likely that the values $v_a$ and $v_n$ are correlated. In general, we would expect $v_a$ and $v_n$ to be positively correlated: if an individual has a low $v_n$, she is also likely to perform *a*. Positive correlation allows individuals to infer normative valuations of others not only from their expressed opinions but also from their actions. Our model would however work also under the assumption of independent $v_a$ and $v_n$.

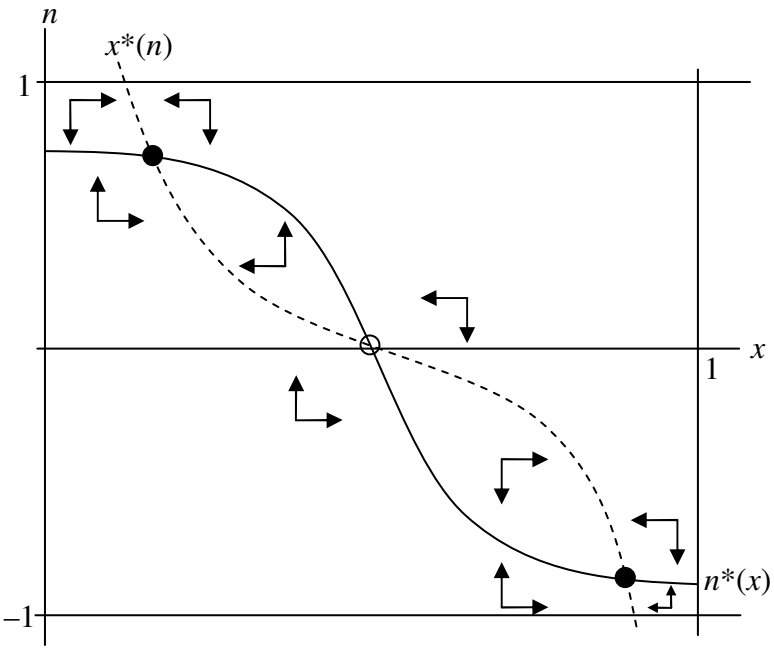


Figure 1: Equilibria, in the simplified model: Solid arrows denote direction of dynamics, filled and empty circles denote stable and unstable equilibria, respectively.

$$\dot{p} = \sigma\left(F\left(S - c - \lambda n + \gamma x\right) - p\right) \tag{4}$$

where $\sigma$ is the pace of adaptation of valuation expression relative to the pace of activity adaptation.[6]

From equations (3) and (4), the evolution of the social norm $n = q - p$ is given by

$$\dot{n} = \sigma\left(1 - F\left(S - c - \lambda n + \gamma x\right) - F\left(S + c - \lambda n + \gamma x\right) - n\right). \tag{5}$$

To simplify the graphical presentation of the rest of the argument, we note that the dynamic system consisting of equations (2) and (5) perfectly reflects the dynamic system consisting of equations (2) through (4), both with respect to the location and the stability of equilibria.[7]

By definition, the equilibria of the system are reached, when neither the proportion of individuals who perform the action nor the proportion individuals who express their opinion deviant from the law in either direction change, except for random disturbances, i.e. when $\dot{x} = \dot{n} = 0$. Not all equilibria defined in this way need to be stable, but only the stable equilibria will be relevant in the remaining analysis.

In order to present our argument in as simple a way as possible, we introduce some simplifications.[8] In particular, we assume that $\lambda = \alpha = 0$, so that equations (2) and (5) reduce to

$$\dot{x} = G\left(B - S - \beta n\right) - x \tag{2'}$$

and

[6] Coresponding to equation (2) we use the same length of the time period. As individuals may decide on their expression of opinions faster or slowlier than on their actions, we introduce a parameter $\sigma$ measuring how many individuals decide on the expression their opinion per time period.

[7] Proof in Appendix B.

[8] All simplifying assumptions we make in the remains of this Section do not strengthen our argument in any way and will be given up in Section 6.

$$\dot{n} = \sigma\left(1 - F\left(S - c + \gamma x\right) - F\left(S + c + \gamma x\right) - n\right). \tag{5'}$$

By posing $\dot{x} = 0$ we define the equilibrium proportion of active individuals as a function of the current social norm:

$$x^*(n) = G\left(B - S - \beta n\right) \tag{6}$$

and by posing $\dot{n} = 0$, we define the equilibrium social norm as a function of current proportion of active individuals:

$$n^*(x) = 1 - F\left(S - c + \gamma x\right) - F\left(S + c + \gamma x\right) \tag{7}$$

We further assume that both $f(\cdot)$ and $g(\cdot)$ are uni-modal and rather concentrated around the mode. Finally, we assume that the parameter $c$ is small. The latter two assumptions guarantee that both $x^*(n)$ and $n^*(x)$ are *S*-shaped functions. Figure 1 provides an example with three equilibria of which the one in the middle is unstable and the other two are stable, as one may easily infer from the solid arrows denoting the directions of the dynamics described by equations (2') and (5'). We should note that the equilibria need not be so nicely distributed across the entire range of *r* and *n*, but may rather be very concentrated for example in the upper left quarter of the figure, which would be the case if about one half of the population would never perform the action and never protest against the law as being too strict.

# 3. Feedback and Social Reaction: Rethinking the Effects of Legal Intervention

In this Section we will study the important interactions that the incentives created by sanctions and social norms have on deterrence when the adoption of a new law generates "social feedback." As it will be seen, when a social feedback is triggered by legal intervention, the reinforcement of a legal sanction can have countervailing effects on behavior. In some cases, this may undermine the goals pursued by the lawmaker. Our model goes beyond models of multiple equilibria being used in the literature to explain reverse effects of legal changes (Kuran 1989, Cooter 1998) as it explicitly separates the two endogenous variables, social norms *n* and behavior *x*.

We shall proceed considering the feedback between expressed opinions *p* and action *x* when a lawmaker implements a change in the severity of the sanction *S*.

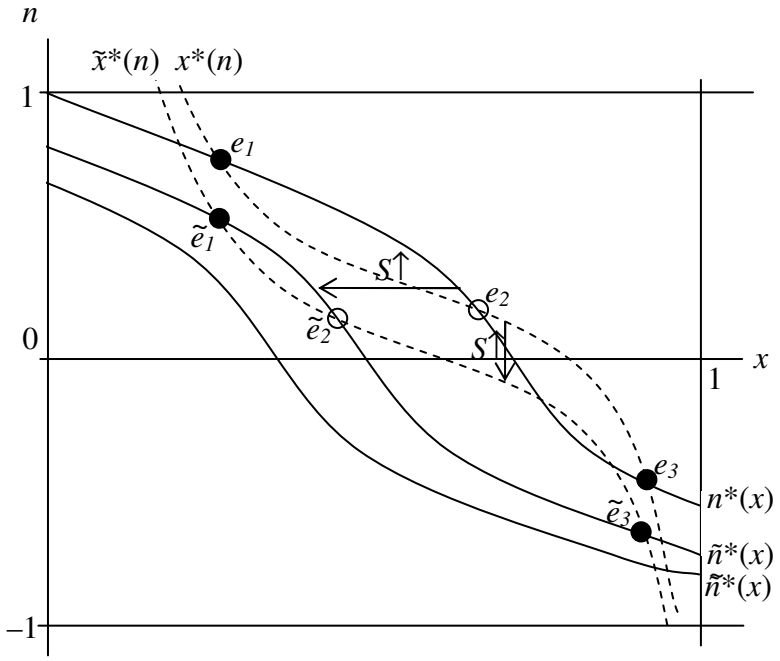


Figure 2: Shifting the equilibrium functions by an increase in *S*: Arrows indicate the shift of the equilibrium functions.

The standard argument, of which we will borrow some elements, would explain our introductory example on terrorism as follows. Suppose the social system is in the equilibrium described by the left-most intersection of the original partial equilibrium functions $x^*(n)$ and $n^*(x)$ in Figure 2 (marked as $e_1$). The increase in the sanction *S* shifts the graph of the $x^*(n)$-function downwards and the graph of the $n^*(x)$-function to the left, so that the new partial equilibrium functions are now $\tilde{x}^*(n)$ and $\tilde{n}^*(x)$. Would the $n^*(x)$-function shift much more than the $x^*(n)$-function (even more than already depicted in Figure 2), the equilibria labeled $\tilde{e}_1$ and $\tilde{e}_2$ would first converge and then disappear. As a consequence, the social system would approach the now unique equilibrium described by the intersection of the two partial equilibrium functions located at high values of $x$ and low values of $n$ (marked as $\tilde{e}_3$). Hence, proportion of individuals performing the action would have substantially increased (and the social norm turned very much against the law) as a reaction to the (possibly small) *reduction* of the sanction — which is exactly the reverse effect we described in the introduction. A similar explanation could be given for our other example, simply with reversed signs.

The problem with this kind of explanation, which only becomes apparent by our clear separation of social norms and the level of activity, is that the $n^*(x)$-function has to shift much more than the $x^*(n)$-function. In words, the reaction of the evolution of social norms to a change in the sanction must be substantially stronger than the corresponding reaction of the evolution of the level of action in society. To be very clear: it is not the equilibrium values of social norms which have to react strongly, but the entire dynamics of social norm formation. This condition typically will not be satisfied, since one would more plausibly expect that

action reacts in a stronger way to changes in sanctions than norms do. We therefore present alternative and — to our impression: far more plausible — explanations for the reverse effects of legal sanctions on actual behavior as we described them in the introduction.

# 4. More Plausible Explanations of Reverse Effects

So far we have assumed that legal intervention has effects on both behavior and protest, and that the dynamics of both these variables react at the same time to changes in the law. The direct implication of this assumption in our model is that the schedules $n^*(x)$ and $x^*(n)$ shift simultaneously. Sudden large reactions to minor legal changes were only possible when the number of equilibria declined.

In this section, we discuss three effects which may entail sudden large reactions to minor legal changes even when no equilibrium disappears. Both effects are based on the countervailing effects of social protest and are observed when the initial enforcement of a new legal sanction has a more immediate impact on protest than on the level of compliance (action). According to their theoretical basis, we call these effects the "announcement effect" and the "social acceptance effect." The first of these effects may be explained with the very simple version of the model we used so far, while the other requires the introduction of slight variations.

## *a. Delaying the Entry into Force of Law: The Announcement Effect*

Changes in the law are usually announced some time before a new law takes effect, at least as far as statutory law is concerned. The idea of the announcement effect is the following: legal sanctions that are announced but are not yet applicable are unable to deter current behavior, but could nevertheless immediately trigger a reaction of social norms. When the entry into force of an enacted law is delayed, social-norm reaction is likely to be triggered before the law is able to produce its incentive effects. When the law eventually comes into force, it will be applied to a modified environment. Anticipatory protest may guide the social system into a different equilibrium from that we could expect when incentives and social reaction materialize simultaneously.

In terms of our model and the example of support of terrorism, the announcement of a more severe law would shift the relevant equilibrium schedule leftwards, from $n^*(x)$ to $\tilde{n}^*(x)$, some time before the corresponding downward shift of $x^*(n)$ to $\tilde{x}^*(n)$ follows. Suppose the social system is well described by the equilibrium labeled $e_1$ in Figure 2 before any legal change, i.e. the action — "support" of terrorism — is rather small and there are clear social norms in favor of strict laws. Now suppose that an increase in the sanction is announced but not yet enforced. While that is the case, support of terrorism is not yet deterred by sanctions ($x^*(n)$ still describes the levels of $x$ towards which the system is drawn for given values of $n$), but it already is affected by the observed social reaction, which exhibits a quick increase in the protest against the stricter law due to the shift in the dynamics of social norm formation ($n^*(x)$ shifts to $\tilde{n}^*(x)$). As depicted in the figure, this may temporarily reduce the number of intersections of the two partial equilibrium functions to one: for the time when only the announcement affects social norm formation but no higher legal sanction deters the activity yet, the intersection of $\tilde{n}^*(x)$ and $x^*(n)$ close to $e_3$ and $\tilde{e}_3$ marks the unique equilibrium of the system. Expressed opinions will become less and less supportive of the high sanction and as a consequence, social norms will more and more encourage "support" of terrorism. Actively "supporting" terrorism will become a more frequent action. If this is the case for a

sufficient amount of time, the social system will get so close to this temporarily unique equilibrium that it falls into the region of attraction of the equilibrium marked by $\tilde{e}_3$ once the higher legal sanction becomes effective and thus $\tilde{x}^*(n)$ replaces $x^*(n)$ as attractor of $x$ for given levels of $n$. The social system will then not return to the low-activity equilibrium $\tilde{e}_1$ but rather evolve towards $\tilde{e}_3$. As a consequence of the increase in the sanction, the frequency of the illegal activity in society has increased substantially social norms have turned to far more support of the illegal activity.

### *b. Getting Used to Unjust Law only later: The Outrage Effect*

Laws may be perceived as excessively strict or even unjust when enacted but later be tolerated and eventually accepted over time. Initial outrage over the more severe sanction may eventually give way to acceptance and perhaps even internalization. So, initially sharp changes of the social norm dynamics may fade away partially or even completely. Nevertheless, the initial sharp reaction of social norms may move the society into a different condition. When the social norm dynamics returns to its original form, it will – together with the individual action dynamics – start from a modified situation and be attracted by a different equilibrium. Like the announcement effect, this transitory shift of the social norms dynamics – which we call the outrage effect – may move the social system from one equilibrium into the region of attraction of another.

Referring to our model this effect is easily described by replacing *S* in the inequalities describing the decisions to express one's opinion in one way or other by a term $S + \Delta S_t = S + \sum_{\tau=0}^{t-1} (S_{\tau+1} - S_\tau)(1+r)^{t-\tau}$ to the size of the (changed) sanction, where $\tau$ is the period in which the sanction is changed and *r* is a depreciation rate for the relevance of past legal changes. Then equation (5') becomes

$$\dot{n} = \sigma\left(1 - F\left(S + \Delta S_o e^{-r(t-\tau)} - c + \gamma x\right) - F\left(S + \Delta S_o e^{-r(t-\tau)} + c + \gamma x\right) - n\right), \qquad (5'')$$

where $\Delta S_o$ is the size of the change in the legal sanction, *e* is the basis of the natural logarithm and $t - \tau$ is the time that has elapsed since the legal change. In terms of Figure 2, then the outrage effect is a temporary severe shift of the $n^*(x)$-function, for example to $\tilde{\tilde{n}}^*(x)$ which partially withers away as time elapses, while $x^*(n)$ only shifts to $\tilde{x}^*(n)$. Then, as was the case for the announcement effect, all but one equilibrium (described by the intersection of $\tilde{\tilde{n}}^*(x)$ and $\tilde{x}^*(n)$ close to $\tilde{e}_3$) temporarily disappear, and the system tends towards this equilibrium. If this situation holds for a sufficiently long time, i.e. if outrage is strong long enough, then the system will be in the region of attraction of $\tilde{e}_3$ even when acceptance replaces outrage and hence the dynamics of the system is described by $\tilde{n}^*(x)$ and $\tilde{x}^*(n)$. The system will not return to a low-activity equilibrium with social norms in line with the law (described in the figure by $\tilde{e}_1$), which now exists again, but will remain close to the equilibrium described by $\tilde{e}_3$ and, absent further disturbances, will tend toward this equilibrium.

What happened with respect to "support" of terrorism during the late 70s in West Germany could be explained by this phenomenon: many people got really upset by the extreme increase in the sanction for "supporting" terrorism and the corresponding decline of custody conditions

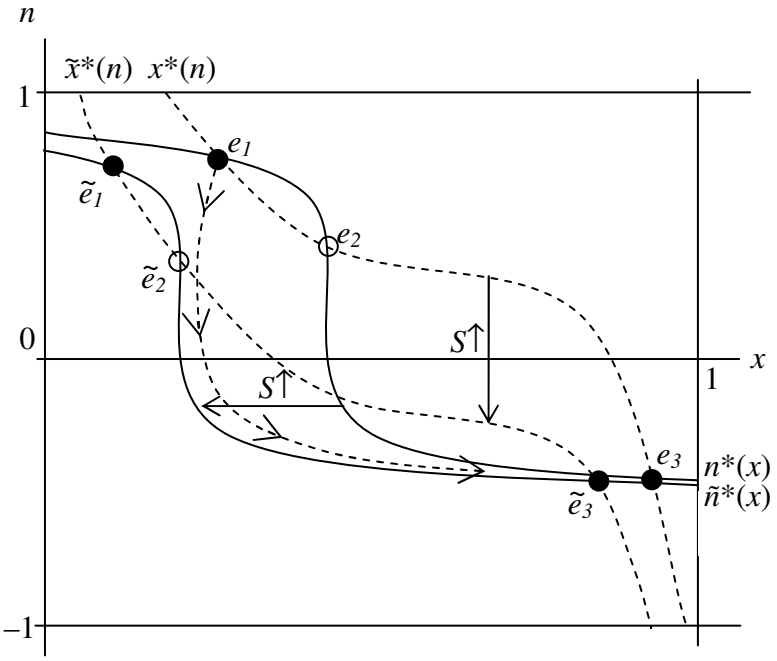


Figure 3: Fast-Reaction-of-Protest Effect: dashed arrow shows evolution of the system from $e_1$ to $\tilde{e}_3$ when protest reacts much faster than action.

of individuals charged with acts of "supporting" terrorism. As a consequence, active support of jailed terrorist increased, which made protest against the stricter law even more attractive for more individuals. While the discussion became less intense after some time (i.e. the outcry effect $\Delta S_o e^{-r(t-\tau)}$ vanished over time) still a large number of people rejected the new laws and many still "supported" terrorists.

[We might want to make more use of the previous text on the "acceptance" or "habituation" effect in this section.]

## *c. Fast-Reaction-of-Protest Effect*

The last effect which may explain reverse effects of sanctions on behavior is what we call the "Fast-Reaction-of-Protest Effect". This effect occurs when social norms are extremely sensitive to increases in the sanction for the level of activity prevalent before the legal change and social norms expressed by open protest adapt very quickly to new situations. The extreme sensitivity of social norms may be due to a highly concentrated distribution of the normative valuations of the regulated activity, but also follow from the alternative explanations for multiple equilibria to be proffered in section 6. We will refer to the high concentration of the normative distribution here, but the interested reader may be willing to simply replace this explanation by any of the other. What may happen is the following: Suppose the vast majority of individuals who actually support openly the sanction (and even want to see a stricter sanction) are only a small bit beyond the limit where they would acquiesce because their costs of open opinion expression would be too high. Now imagine a substantial increase in the sanction. Now this large number of individuals who were just willing to express their preference for a higher sanction will now tend to express the opposite opinion: in their minds, the sanction has become too high. At the same time, many individuals will tend to give up the

activity due to the higher deterrence by the new sanction. If the adaptation in expressed opinion occurs much faster than the adaptation in action, the new norms may offset the legal deterrence before action has reduced by a substantial degree. As a consequence, the countervailing effect of social norms may more than outweigh the increase in deterrence in the long run.

Figure 3 exemplifies the effect in terms of our model. Both partial equilibrium functions are *S*-shaped in a rather strong way. An increase in the sanction shifts the partial equilibrium function of social norms $n^*(x)$ leftward and the partial equilibrium function of behavior $x^*(n)$ downwards. All three equilibria continue to exist. The initial combination social norms and activity after the legal change is given by the intersection marked as $e_1$. The dynamics starting from this point will be directed to the South-West, since current situation is to the North of the partial equilibrium function of social norms and to the East of the partial equilibrium function of actions. However, the movement to the West will be rather small, given the assumed fast reaction of social norms. If the speed differential is large enough, the phase describing the dynamics of the system (marked by the dashed arrow in the figure) will intersect the partial equilibrium function $\tilde{x}^*(n)$ before it intersects $\tilde{n}^*(x)$. Once this happens, the system will be driven to the South-East, towards the new equilibrium $\tilde{e}_3$. The result of the (not so small) increase in the sanction is an increase in the activity. The corresponding shift of social norms against too high sanctions is much stringer than one would get from a movement from equilibrium $e_1$ to equilibrium $\tilde{e}_1$, which one would expect from an analysis disregarding the differential speed of reactions of protest and action.

[I refrain from writing about the terrorism example as long as we have not decided to keep the effect in the paper.]

# 5. Remedies to Overcome Reverse Effects

The announcement effect is due to the immediate upsurge of protest when the law is announced, before it is actually implemented. In order to reduce the impact of this effect on violations, it is important that protest is kept at a low level during the transitory phase. There are several strategies that the lawmakers could implement to avoid the announcement effect. Among them, four are worthy of mention: (i) shortening the lag between enactment and entry into force of the law; (ii) proceeding with piecemeal legislation; (iii) sunset clauses; and (iv) executive legal intervention. [9]

By shortening the enforcement time lag the lawmaker can reduce the duration of the transitional period. This leaves the social dynamics with less time to get into the basin of attraction of the high-activity-level equilibrium. For example, if in Figure 4 the higher sanction becomes effective when the system has evolved only to point $C_1$, the law would come into force when the evolved reality is still in the basin of attraction of the low-activity-level equilibrium and is thus most likely to evolve towards it. The normal dynamics of legal intervention studied in Section 5 would not be disturbed by a short-term announcement effect.

[9] We could of course think of several other strategies. For example, limitation of the freedom of speech during the transitory phase would clearly reduce the social reaction, and possibly limit or eliminate the announcement effect. But it would violate modern constitutional or democratic principles.

Piecemeal legislation may also solve the problem by avoiding the transitional increase of protest and disappearance of the low-activity-level equilibrium. In an initial phase, the lawmaker could implement the law with low sanctions to let both the high-activity-level equilibrium and the low-activity-level equilibrium persist. The announcement effect of a moderate law would then be sufficiently contained and would not disturb the normal dynamics of legal intervention studied in Sections 4 and 5. The first phase would then be characterized by a small sanction and a small reduction in the activity level. When the sanction introduced in phase one becomes effective, the social system will evolve towards the low-activity-level equilibrium as desired, because it will not have left its basin of attraction regardless of the duration of the transitional phase.[10] Once the social system has evolved sufficiently close to the new low-activity-level equilibrium, the second step of the piecemeal legislation can be undertaken, with an increase in the sanction. The stricter law would trigger an increase in protest, but this reaction would take place in an environment characterized by higher levels of compliance, as obtained in the previous phase. In this way, by introducing small consecutive increments in the strictness of the law, piecemeal legislation could more effectively achieve the chosen policy objective and avoid undesirable announcement effects.

Another possible method through which lawmakers could mitigate the announcement effect when the enactment of a strict law is unavoidable is by introducing a "sunset clause." A sunset clause operates as a sort of termination date beyond which the law will expire and the sanction will go back to its original level. This potentially reduces protest and prevents the realization of the announcement effect. This was the tactic used by Congress when the U.S. Patriot Act was passed on October 24, 2001, just a few weeks after the terrorist attack on the World Trade Center in New York. The Act contains some substantial reductions of human rights, justified by the need to protect one of the fundamental freedoms protected by the Universal Declaration of Human Rights, the freedom from fear.[11] Although the Patriot's Act was approved almost with no dissent in the aftermath of September 11, the law was set to expire in the year 2005. This was due to strong criticism that the Act excessively weakened the protection of civil liberties. After its "sunset", the bill was reauthorized (with no substantial change) in March 2006 and is now set to expire on December 31st, 2009. As we will see in the next section, the use of sunset clauses is also useful when people "adjust" to the new law over time.

Finally, one last strategy would be to avoid publicizing future legal changes while implementing them in a rapid fashion. A law enacted with no prior notice and rendered immediately effective would simultaneously trigger incentive effects and social reaction, generating a dynamic similar to the one we studied in Sections 4 and 5 above. There are several examples of rapid legal changes in both common law and civil law countries, notwithstanding the complex procedures to pass laws in the latter.

One example is the recent (November 2007) passing of a bill in Italy containing provisions to improve security and reduce crime after a woman was robbed, raped, and killed by a Romanian illegal immigrant. The main provisions in the bill are directed at immigrants, who can be repatriated not only if they reside in the country illegally but also if they are unable to prove that they have a job and enough income to make a living. The bill was passed using an expedited procedure to defeat protests claiming that the government was using immigrants as scapegoats and unfairly punishing the poor. According to the Italian procedure,

---

[10] One should be aware that during the transitional phase random effects may more easily push the system into the basin of attraction of the high-activity-level equilibrium. Therefore, the small steps undertaken by the piecemeal legislation should not drive the transitory equilibrium too close to the maximum of the graph of $p(x)$ , since then this equilibrium's basin of attraction is very narrow.

[11] The Act allows measures that in practice imply invasions of privacy, "scapegoating" immigrants and refugees, the creation of "extra-legal zones", most prominently at the naval base at Guantánamo Bay in Cuba. Finally, it implies a reduced American human-rights presence through the rest of the globe (see Koh in The Economist, 30th October 2003).

the bill will have to be confirmed by the Parliament. In the meantime, however, it is enforceable. The announcement effect has been avoided.

# 6. Alternative Explanations for Multiplicity of Equilibria

In Sections 2 through 5 we have made a number of simplifying assumptions in order to get equations (2') and (5') and easily analyze them. In this section, we will discuss the effects of giving up the simplifications. It will turn out that our results do not hinge on any of the simplifications. On the contrary: relaxing them strengthens our argument.

## *a. Frequency Dependencies*

The most radical simplification was to set $\alpha = \lambda = 0$ in equations (2') and (5'). As a consequence, it was easy to derive the *S*-shaped form of the functions $x^*(n)$ and $n^*(x)$ by simple reference to uni-modal density functions of $v_a$ and $v_n$. Allowing for $\alpha > 0$ and $\lambda > 0$ makes the analysis somewhat more complicated but stronger. Consider the partial equilibrium of social norms first. Equating the time derivative in equation (2) to zero, may then yield up to three solutions in $x$ for any given $n$. However, studying the inverse of $x^*(n)$ defined by

$$n(x^*) = \left(B - S + \alpha x^* - G^{-1}(x^*)\right)/\beta, \qquad (6')$$

where $G^{-1}(\cdot)$ stands for the inverse of $G(\cdot)$, we can see that this inverse remains a function (exactly one $n(x^*)$ for every $x^*$) and that this function is strictly decreasing unless $\alpha\, g\left(B - S - \beta n(x^*) + \alpha x^*\right) > 1$, since the slope of $n(x^*)$ is given by

$$\frac{dn(x^*)}{dx^*} = \frac{\alpha\, g\left(B - S - \beta n(x^*) + \alpha x^*\right) - 1}{\beta\, g\left(B - S - \beta n(x^*) + \alpha x^*\right)},$$

as the interested reader may easily verify by taking the total derivative of equation (2) after setting $\dot{x} = 0$. Saying that $\alpha\, g\left(B - S - \beta n(x^*) + \alpha x^*\right) > 1$ holds true is nothing else than arguing that either the density function $g(\cdot)$ takes high values in the relevant range (high concentration of probability mass of $v_a$ as assumed in Sections 2 through 5) or that $\alpha$ is sufficiently large, i.e. that there is a strong direct feedback of the frequency $x$ of the action on the proportion of individuals willing to perform the action. If either of these conditions is satisfied, $n(x^*)$ has an upward sloping part and thus $x^*(n)$ (treated as a correspondence, not a function any more) is not only *S*-shaped, but has a backward bending part in the middle. This, however, only facilitates making our arguments as multiple equilibria become more likely and all effects we discuss may occur even when $n^*(x)$ is not so strongly *S*-shaped as we assumed it in the figures.

Similarly, if $\lambda\left(f(S - c - \lambda n + \gamma x) + f(S + c - \lambda n + \gamma x)\right) > 1$, then $n^*(x)$ has a backwards bending part, implying that multiplicity of equilibria is more likely and our arguments may become valid even if $x^*(n)$ is not so strongly *S*-shaped as we assumed it in the figures. Interpreting this condition means that our argument is particularly strong, if either the

probability mass of $v_n$ is highly concentrated (as assumed in Sections 2 through 5) or social norms exhibit a strong positive feedback on their own formation (large $\lambda$).

One should note that if both $x^*(n)$ and $n^*(x)$ have backwards bending parts, more equilibria become possible. Out of the nine possible equilibria, four are stable — all equilibria given by the intersection of the backward bending part of either of the correspondences with the other correspondence are unstable. Our arguments may easily be extended to this case. However, due to the steeper form of the partial equilibrium correspondence $x^*(n)$ off its backward-bending part, the reverse effect of legal sanctions will be smaller.

### *b. Larger Costs of Protest*

A second important restriction we placed on our model to make the analysis more simple was the restriction of the part of the costs of expressing one's opinion which is independent of other individuals' behavior and opinion expression, i.e. the parameter $c$ to a "sufficiently small" level. Allowing for larger values of $c$ will induce the sum $f(S-c-\lambda n+\gamma x)+f(S+c-\lambda n+\gamma x)$ to have two local maxima in $x$ and $n$ so that $F(S-c-\lambda n+\gamma x)+F(S+c-\lambda n+\gamma x)$ is double-*S*-shaped in $x$ and $n$. This double-*S*-shape transfers to the partial equilibrium function or correspondence $n^*(x)$ if $\lambda$ is small and turns into two backwards bending parts of $n^*(x)$ if $\lambda$ is large enough. In both cases, the number of equilibria may increase, with the same effects as we discussed them for high frequency dependencies. As there, our argument becomes stronger from relaxing our restrictive assumptions.

### *c. Multi-Modal Distributions of Valuations*

We also assumed that the densities $f(v_n)$ and $g(v_a)$ are uni-modal. Giving up this assumption entails multiple increasing and decreasing parts of the distribution functions $F(\cdot)$. and $G(\cdot)$ The gentle reader will already suspect the result: we get double- or multiple-*S*-shaped functions or correspondences $x^*(n)$ and $n^*(x)$ possibly with backward bending parts. As before, this allows for more equilibria and the effects on which we based our argument become more likely, though they may occur on a smaller scale.

## 7. Conclusions

We have developed a model of the interaction between social norm formation and decisions on performing an illegal action. This model, exhibiting multiple equilibria, allowed us to explain reverse effects of changes in the size of legal sanctions as we observe them in our introductory examples of support of terrorism, music downloads and partial birth abortion. The existing literature explains such reverse effects only by the disappearance of equilibria due to the legal change. We have argued that such disappearance of equilibria requires that the dynamics (not the level!) of the endogenous variables (social norms and the level of the illegal activity) react to the legal change in an implausible way: The social norm dynamics has to react in a far stronger way than the dynamics of behavior, although sanctions affect the

latter more directly. Only our clear distinction between social norms and levels of activities made it possible to discover this implausibility.

In contrast to the literature, we therefore offer explanations which do not rely on the disappearance of equilibria. We explained three effects of the interaction of social norm formation and adaptation of behavior to explain reverse effects: the Announcement Effect based on the idea that increases in legal sanctions which are announced some time before they come into force may affect social norms long before they affect the action they penalize; the Outrage Effect based on the idea that social norm formation and protest may react in a particularly strong way shortly after the legal change, and come to its new equilibrium level only after some time of adaptation; and finally the Fast-Reaction-of-Protest Effect based on the idea that adaptations of opinions may take place at a much faster pace than adaptations of behavior. We do not claim that any of these effects always or frequently interferes with legislation, but we argue that they are better explanations for the existing cases were reverse effects of legislation are observable.

Also in contrast to the existing literature relying on the disappearance of equilibria as explanations for reverse effects, we are able to offer policy recommendations to avoid the reverse effects. Most prominent, because helpful against all three sources of reverse effects, is piecemeal legislation. Giving the social system time to adapt to small changes in the circumstances avoids strong reactions of social norms before the adaptation of behavior follows, mitigates temporary upsurges of protest and avoids leaving the social system in a situation which is so far from the equilibria that it is unclear to which equilibrium will tend.

## References (preliminary)


Bicchieri, Cristina (2006). *The Grammar of Society: the Nature and Dynamics of Social Norms*. New York: Cambridge University Press.

Bikhchandani, Sushil, David Hirschleifer and Ivo Welch (1992). "A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades." *Journal of Political Economy* 58: 211-221.

Bar-Gill, Oren and Alon Harel (2001). "Crime Rates and Expected Sanctions: The Economics of Deterrence Revisited." *Journal of Legal Studies* 30: 485-501.

Binmore, Ken and Larry Samuelson (1994). "An Economist's Perspective on the Evolution of Norms." *Journal of Institutional and Theoretical Economics* 150: 45- 63.

Bohnet, Iris and Robert Cooter (2001). "Expressive Law: Framing or Equilibrium Selection?" Berkeley Olin Program in Law & Economics, 31 Working Paper Series, 2001.

Bohnet, Iris, Bruno S. Frey and Steffen Huck (2001). "More Order with Less Law: On Contract Enforcement, Trust and Crowding." *American Political Science Review* 95: 131-144.

Borland, John (2003). "Lawsuits Yield Mixed Results." CNET News. Com, Dec. 4, 2003.

Cooter, Robert (1998). "Expressive Law and Economics." *Journal of Legal Studies* 27: 585-608.

Cooter, Robert (2000). "Do Good Laws Make Good Citizens? An Economic Analysis of Internalized Norms" *Virginia Law Review* 86: 1577.

Depoorter, Ben and Sven Vanneste (2005). "Norms and Enforcement: The Case Against Copyright Litigation." *Oregon Law Review* 84: 1127-1180.

Gneezy, Uri and Aldo Rustichini (2000). "A Fine is a Price." *Journal of Legal Studies* 29(1) 1-17.

Gneezy, Uri and Aldo Rustichini (2003). "Incentives, Punishment and Behavior," in *Behavioral Economics* (edited by Colin Camerer, George Loewenstein and Matthew Rabin), Princeton University Press, Princeton, New Jersey, USA.

Harel, Alon and Alon Klement (2007). "The Economics of Stigma: Why More Detection of Crime May Result in Less Stigmatization." *Journal of Legal Studies* 36: 355-377.
Kahane, L.H. (2000). "Anti-Abortion Activities and the Market for Abortion Services. Protest as a Disincentive." *American Journal of Economics and Sociology* 59: 463-485.
Koh, Harold H. (2003). "Rights to Remember. How Has September 11th Changed America's Approach to Human Rights?" (excerpts from the 2003 John Galway Foster lecture delivered in London on October 21st, 2003), The Economist, 30th October.
Kuran, Timor (1989). "Sparks and Prairie Fires: A Theory of Unanticipated Political Revolution." *Public Choice* 61: 41-74.
Leeds, Jeff (2007). "Labels Win Suit Against Song Sharer." *The New York Times*, October 5, http://www.nytimes.com/2007/10/05/business/media/05music.html.
McAdams, Richard H. (2000). "A Focal Point Theory of Expressive Law." *Virginia Law Review* 86: 1649.
Oksanen, Ville and Mikko Välimäki (2007). "Theory of Deterrence and Individual Behavior. Can Lawsuits Control File Sharing on the Internet?" *Review of Law and Economics* 3(3):Article 4. http://www.bepress.com/rle/vol3/iss3/art4
Parisi, Francesco and Georg von Wangenheim (2006). "Legislation and Countervailing Effects from Social Norms" in *Evolution and Design of Institutions* (edited by C. Schubert and G. von Wangenheim), Routledge, UK: 25-55.
Posner, Eric (2000). "Law and Social Norms: The Case of Tax Compliance." *Virginia Law Review* 86: 1781.
Robinson, Paul H. (2000). "Why Does the Criminal Law Care What the Layperson Thinks Is Just - Coercive Versus Normative Crime Control." Virginia Law Review 86: 1839.
Stuntz, William J. (2000). "Self-Defeating Crimes." Virginia Law Review 86: 1871.
Sugden, Robert (1998). "Normative Expectations: The Simultaneous Evolution of Institutions and Norms" in *Economics, Value, and Organization* (edited by Ben-Ner, A. and Putterman, L.), Cambridge: Cambridge University Press.
Sunshine, Jason and Tom R. Tyler (2003). "The Role of Procedural Justice and Legitimacy in Shaping Public Support for Policing." *Law and Society Review* 37: 555-589.
Thoreau, Henry David (1894), Civil Disobedience, in Walden and Other Writings by Henry David Thoreau (Joseph Wood Krutch ed., 1981).
Tyler, Tom R. (1990). *Why People Obey the Law*. New Haven: Yale.
Tyler, Tom R. (1994). "Psychological Models of The Justice Motive." *Journal of Personality and Social Psychology* 67: 850-863.
Tyler, Tom R. and Y.J. Huo (2002). *Trust in the Law*. N.Y.: Russell-Sage. 31
Weber, Max (1968). *Economy and Society*. (G. Roth and C.Wittich, eds.). Berkeley: University of California Press.
Weidlich, W. and G. Haag (1983), *Concepts and Models of Quantitative Sociology*. Springer: Berlin/Heidelberg/New York.