



**No. 10-2010**

**Emanuela Carbonara, Francesco Parisi, and  
Georg von Wangenheim**

**Inexpressive Law**

This paper can be downloaded from  
[http://www.uni-marburg.de/fb02/makro/forschung/magkspapers/index\\_html%28magks%29](http://www.uni-marburg.de/fb02/makro/forschung/magkspapers/index_html%28magks%29)

Coordination: Bernd Hayo • Philipps-University Marburg  
Faculty of Business Administration and Economics • Universitätsstraße 24, D-35032 Marburg  
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: [hayo@wiwi.uni-marburg.de](mailto:hayo@wiwi.uni-marburg.de)

Emanuela Carbonara<sup>\*</sup> - Francesco Parisi<sup>♦</sup> - Georg von Wangenheim<sup>♠</sup>

## Inexpressive Law<sup>\*</sup>

January 2010

### Abstract:

According to expressive law theories, expression of values is an important function played by the law. Expressive laws affect behavior, not by threatening sanctions or promising rewards, but by changing individual preferences and tastes and, in some cases, by affecting social norms and values. New laws, however, can run against sticky social norms, failing to achieve their expressive effects. By developing a dynamic model, in this paper we show that inexpressive laws (laws whose expressive function is undermined by sticky norms) can not only be ineffective but can push the values of society away from those expressed by the law. We study the effects of legal intervention on the values shared by members of society, considering the feedback effects between laws and social norms. Just like expressive laws can foster consensus in heterogeneous groups, inexpressive laws can create a social divide, even in previously homogeneous societies.

*JEL* Classifications: *K10, K42, D70, B52, Z13.*

Keywords: *Social Norms, Countervailing Effect, Expressive Law, Civil Disobedience*

*“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.”*

*Henry David Thoreau (1849)*

*“In a legal system structured as ours is, criminalization can work against the very norms on which it rests, meaning that popular norms may tend to move in the opposite direction from the law. Criminal law's relationship with popular norms may sometimes be perverse”*

*William Stuntz (2000)*

---

<sup>\*</sup> University of Bologna, Department of Economics.

<sup>♦</sup> University of Minnesota Law School and University of Bologna, Department of Economics.

<sup>♠</sup> University of Kassel, Faculty of Law.

\* This paper was previously circulated under the title “Unjust Laws and Illegal Norms”. We thank Robert Cooter, Tzili Dagan, Ben Depoorter, Giuseppe Dari-Mattiacci, Yuval Feldman, Nuno Garoupa, Mark Grady, Claire Hill, Robert MacCoun, Richard McAdams, Brett McDonell, Gideon Parchomovsky, Hans-Bernd Schäfer, Avraham Tabbach, Christian Traxler and the participants at the Hebrew University and Bar-Ilan University Conference on Law and Social Norms, the University of Minnesota Conference on Law and Empathy, the Conference of the Latin American and Caribbean Law and Economics Association (ALACDE) in Barcelona, the Conference of the European Association of Law and Economics (EALE) in Rome, and faculty workshops at the Universities of Amsterdam, Bologna, Bonn and Hamburg for insightful comments.

## 1. Introduction

Studies of individual behavior in response to legal commands generally support the argument that the law affects human choice by creating external incentives and promoting the individual internalization of the values expressed by the law. The economic model of law enforcement analyzes the incentives created by the relative prices of legal and illegal activities (Becker 1968, Polinsky and Shavell 2000a). By altering such relative prices, a more severe law will always induce individuals to substitute illegal activities with legal ones. The opposite, of course, will occur in case of new, more lenient laws. The growing expressive law literature (Cooter 1998 and 2000, McAdams 1997 and 2000), argues that the enactment of a new legal rule may lead individuals to internalize the values expressed by the law and trigger self-enforcement mechanisms that foster compliance, even in the absence of enforcement.

According to both streams of literature, the law produces the effects intended by the legislator even if it fails to express the values shared by members of the society. In many situations, however, laws that depart from the current values of society may run against “sticky norms,” failing to produce the internalization and self-enforcement hypothesized by the expressive law literature and possibly leading to backlash effects (Kahan 1998 and 2000). Recent work in experimental and behavioral law and economics provides several examples of this type of reaction to legal innovation.<sup>1</sup>

In this paper we consider the dynamic impact of the law in the presence of both internalization and backlash effects. We show that legal intervention which does not account for social reaction to legal innovation may not only be ineffective but may even have a negative impact. A law that introduces a new sanction, for example, may actually lead to an increase in the sanctioned behavior, as well as to a shift in social norms that goes in the opposite direction than the law. In the interest of linguistic symmetry, we shall refer to laws that fail in their expressive mission as “inexpressive laws.” Expressive law can lead to a convergence of views in society, leading to a consensus even among heterogeneous groups. Inexpressive law, to the contrary can trigger a social divide, such that a society that initially shares fairly homogeneous opinions may gradually split into two groups following opposite behavior and upholding different values. It is interesting to stress that inexpressive laws can exert their perverse effects even in the presence of enforcement.

To understand how this can happen, one should consider how society may react when legal systems generate laws that do not reflect widespread consensus and are perceived as unjust. A legal rule may be perceived as unfair in two alternative ways – defect or excess. A law fails in defect when it is too lenient or when it does not provide adequate punishment for behavior that the people in that society consider harmful and

---

<sup>1</sup> Gneezy and Rustichini (2003) provide experimental evidence of behavior opposite to that predicted by the deterrence model. Also, the results of their field study in day-care centers are widely known (Gneezy and Rustichini 2000): they found that imposing a fine on parents who were late picking up their children resulted in a significant increase in the number of tardy parents. Bohnet, Frey and Huck (2001) present an experiment where increasing the probability of detection in environments with weak institutions crowds out virtuous social norms and trustworthy behavior.

undesirable; it fails in excess when it is too strict or it punishes a conduct that is perceived by members of society as harmless or even desirable. When the society views the legal sanction as too lenient relative to the preexisting social norm, individuals may be inclined to “add” social sanctions to the behavior (e.g. disapprobation or further punishing those committing the action targeted by the law). When society perceives the legal sanction as too harsh, individuals may be inclined to “subtract” social sanctions (e.g. approbation or tolerance of infringers) to the behavior. Particularly, the enactment of an “unjust” law may trigger public opposition that reveals social disapproval of that legal rule. If social disapproval of a new legal restriction is sufficiently strong, it may outweigh the incentive and expressive effects of the law and produce countervailing effects, i.e., outcomes opposite to those intended by the lawmaker.<sup>2</sup>

There are several real-life examples that illustrate our point, including those of file sharing and music download, terrorism and abortion, all of which have been characterized by countervailing effects in the course of time. We will present them in greater detail in Section 2. There are several other examples, including gender discrimination, army recruitment, drug and alcohol consumption, and smoking.<sup>3</sup>

Our results shed new light on the effects of legal intervention when the law conflicts with incumbent social norms. Lawmakers can avoid problem areas and adopt appropriate legal instruments for enacting new laws if they open themselves to an understanding of the characteristics of situations where countervailing effects are more likely. We may have paradoxical cases in which reducing the severity of a law may lead to a decrease in crime.

### **1.1 Related Literature**

Our paper is related to the law and economics literature analyzing the interdependence between the law and social norms. First of all, our paper is related to the literature on expressive law (Cooter, 1998 and 2000). Expressive law theories revisit the traditional price-theory conceptions of law as an incentive mechanism, developing a richer theory of how legal rules can affect human behavior. According to expressive law theories, expression of values is an important function played by the law. Through expression the law can trigger the emergence of other incentives through the internalization of the values it embodies. Expressive laws affect behavior, not by threatening sanctions or promising rewards, but by changing individual preferences and tastes and, in some cases, by affecting social norms and values. This distinguishes expressive law theories from traditional

---

<sup>2</sup> Public opposition may take several forms, from verbal expression of dissent to violent acts and illegal behavior either against enforcers in the case of too strict laws or against those who commit the action in case of too lenient laws.

<sup>3</sup> When searching for a common characteristic of all these examples, we notice a potential for social conflict and intergroup controversy. For instance, hardly anybody in a society thinks that armed robbery and violent crimes in general may be good or desirable for society. There may be a debate on how harsh punishment should be for a given crime but such debate on appropriate sentencing would not send any mixed signal with respect to the social disapproval for the action itself. It is however interesting to notice that some actions have been raised to the status of “uncontroversial crimes” only in recent times (e.g., honor crimes, indecent assault and harassment).

theories focused as they are on the role of law as an instrument for creating external incentives, such as taxes, sanctions and rewards. According to expressive law theories, internalized rules may trigger private enforcement mechanisms and change observed patterns of behavior even in the absence of other external incentives. Private enforcement mechanisms include first-party, second-party and third-party enforcement. These three interrelated mechanisms are important ingredients of the expressive effects of law. Expressive law theories, by shedding light on the role of law in shaping social values and norms, point to the relevance of legal intervention and the important responsibility of lawmakers as prospective norms-entrepreneurs.<sup>4</sup> Opposing the expressive law view, Kahan (1998 and 2000), Lessig (1998) and McAdams (1995) argue that behavior signals one's values and beliefs to others and can produce approval and acceptance in social groups. Outlawing behaviors with high signaling power will not have an impact on the prevalence of that particular behavior unless outlawing the behavior undermines intra-group approval. Kahan (2000) deals with the "sticky-norms problem": when legislators enact laws aimed at changing existing social norms, often the likely result is the irrelevance of the law for behavior. In fact, enforcers may be reluctant to apply such legal changes, and their reluctance increases with the severity of the law. This leads to the policy prescription of using softer laws (what he calls "gentle nudges") to revert social norms, i.e., moderate penalties. In Kahan's (2000) analysis, however, "gentle nudges" may be more effective than hard laws (what he calls "hard shoves"), but in his setup, both forms of legal intervention are (weakly) better than no intervention (Kahan, 200, p. 619).<sup>5</sup> We show instead that laws triggering opposition may lead to countervailing effects.

Our paper is related also to the law and economics literature analyzing social norms as an alternative to law in regulating behavior (Bernstein 1992, Ellickson 1991, Posner 2000, among others). Unlike this literature, however, we argue that social norms, when countervailing, can lead to violations of legal rules and to social unrest.

Recent law and economics literature has brought to light cases in which the enactment of more severe laws produces effects opposite to those auspicated by the lawmaker. Particularly, higher sanctions may produce an increase in the sanctioned activity. This literature, however, does not consider the possible feedback between social and legal norms.<sup>6</sup>

---

<sup>4</sup> McAdams and Nadler (2005) provide experimental evidence that the law has the power to attract attention to particular actions, by publicly endorsing them. Thus, the law induces compliance by helping coordination around a focal equilibrium.

<sup>5</sup> Kahan's model for example cannot be used and is "not meant to challenge the conventional economic premise that the level of crime varies inversely with the expected penalty for it." (Kahan, 2000, p. 619)

<sup>6</sup> For instance, Nussim and Tabbach (2009) show that larger punishment may increase crime if potential criminals invest in avoidance activities. Borck (2004) shows how stricter enforcement may increase tax evasion. With stricter enforcement, low income voters may prefer higher tax rates, since the tax system becomes more progressive. This may induce richer voters to evade more. Jost (2001) proves that increased investigation may increase the prevalence of illegal behavior, if the individual propensity to engage in illegal activities depends on the behavior of others and the police has limited budget.

Our paper is linked to the literature on information cascades and opinion formation (Bikhchandani, Hirschleifer and Welch 1992 and 1998). It is also related to Kuran's (1989) paper and (1995) book on sudden revolutions. Departing from Kuran's contributions, whose results are based solely on the possible discrepancy between true and revealed preferences, we consider the process of endogenous formation of individual preferences and their interaction with action, the main target of our analysis. Moreover, in our model there is no incentive to misrepresent one's preferences, as in Kuran's.<sup>7</sup> The literature viewing social and legal norms as complements and substitutes is also related to our paper (Posner 1998, Zasu 2007). Differently from this literature, in our model social norms evolve as a result of legal innovation.

Finally, Polinsky and Shavell (2000b) include a notion of the fairness of sanctions into the standard model of public enforcement. In their model, however, such notion of fairness is exogenous and is not influenced in any way by new legal rules.

We organize the paper as follows: Section 2 presents some examples and case studies of countervailing norms. Section 3 models the effects of legal change on human behavior. In Section 4, we describe how, in a dynamic setting, sticky norms eventually move, but sometimes, like a springboard, may trigger countervailing effects. We identify the conditions under which legal intervention may be counterproductive and where legal enforcement could generate negative net effects. Section 5 develops some extensions, discussing two effects that may entail sudden large reactions to legal change. We call them the announcement effect and the outcry effect. Consideration of these effects allows us to present in Section 6 policy strategies that may reduce the occurrence of countervailing norms. Section 7 concludes the paper. Appendix A provides proofs for the existence and stability of equilibria. Appendix B generalizes the model, whereas Appendix C contains some technical material.

## **2. Fairness of laws and legal compliance**

Laws can more easily achieve their goals when their content is in alignment with the existing social values. Studies of people's behavior in response to laws generally support the argument that the alignment of legal precepts and decisions of authorities with current social values has a positive influence on compliance (Tyler, 1990, with respect to law; and Tyler & Huo, 2002, with respect to the decisions of authorities). The belief of individuals in the legitimacy of the legal system is of critical importance for compliance. Legitimacy is undermined when the content of the law departs from social norms, be they based on moral, ethical, or merely cultural values.<sup>8</sup>

---

<sup>7</sup> The insurgence and outcomes of insurrections has been extensively analyzed also by Grossman (1995). His models, however, are static, general equilibrium models where well defined social groups (aristocracy, incumbent rulers, peasants, worker families) may be involved in a revolution when the level of predation by the incumbent rulers becomes intolerable. Grossman's models never deal with social norms nor with mechanisms of opinions formation and his main concern is to study the redistribution of productive resources after insurrections.

<sup>8</sup> According to Tyler (1990, p. 25), legitimacy represents an "acceptance by people of the need to bring their behavior into line with the dictates of an external authority".

As discussed in Section 1, a legal rule may be perceived as unfair for prohibiting a conduct that is perceived as harmless or desirable or when being too lenient and failing to prohibit behavior that is considered harmful and undesirable. Zwiebach (1975) provides examples of these two forms of civil disobedience. An example of the first form of disobedience is disobedience of laws perceived to be too restrictive and infringing upon constitutionally protected liberties. Examples range from the case of an African American person that refuses to sit in a segregated section of a courtroom, or a student wearing a black armband as a symbol of political protest in violation of school regulations, or a doctor prescribing birth control in violation of a state law. In each of these instances, violations of the law take place predominantly to assert dissatisfaction with the law, rather than to fulfill the violator's selfish interest in breaching the law. The second form of reaction to law takes place when the law is too lenient and fails to protect individuals. In this case, a protester asserts that a right which current law does not recognize ought to be recognized, or that, if a law already exists, legal protection must be effectively implemented and enforced by authorities. Examples of this type exist in the various forms of protest in the area of human rights. Historically, this form of protest has been very valuable to society, allowing the gradual recognition and protection of new rights that would not easily arise through traditional political or lawmaking processes.

Stuntz (2000) provides other interesting illustrations of the law's impact on social norms. Stuntz considers how criminal law can defeat itself due to the lack of alignment between criminal laws and laypeople's values and norms, generating disobedience rather than obedience. As the author describes it, the relationship between laws and social norms "is not always the relationship of car to driver or driver to car, but rather the relationship of one side of a seesaw to the other. To put the point simply, some crimes may be self-defeating. Sometimes, the best way for the legal system to advance or reinforce norms may be to ignore them (Stuntz, 2000: 1872-73)." Stuntz considers the examples of vice crimes (e.g. drinking during Prohibition, gambling) and highly divisive "morals" crimes (e.g. slavery), suggesting that the opposition to these laws was further exacerbated by the perception that these policies were driven by racial or class bias rather than moral justice, corroding the authority of the law for a larger portion of the public.

Besides the examples already considered in the literature, we provide further illustrations of how the law can positively or negatively influence social norms, considering other situations where legal change triggered countervailing effects and reaction in behavior opposite to that normally expected. We try to distinguish cases where countervailing effects are produced by positive social sanctions from cases where negative social sanctions triggered the undesired consequences.

A first example that has dominated the news several times in recent years is the impact that restrictive copyright laws have had on the amount of file sharing via peer-to-peer technologies over the internet. There are several empirical papers showing that legal enforcement has had an unintended effect. For instance, Depoorter and Vanneste (2005) and Oksanen and Välimäki (2007) provide evidence of an increase in file sharing after the

music companies started suing users of peer-to-peer technology for illegal downloading. In this example, we have all the elements required by our model: the increase in sanctions against downloaders of illegal files sparked protest both inside the community of file sharers and outside of it.<sup>9</sup> This revealed that the prevailing social norm admitted file sharing and encouraged more people to engage in the activity and to increase the level at which they performed it. In fact, in the summer 2003, the RIAA started suing an increasing number of people for downloading and sharing music files. There was strong negative reaction and protest among both music downloaders and non-downloaders. An OECD report issued in 2004 shows that the number of users of peer-to-peer technologies has continued to grow between 2002 and 2004 notwithstanding attempts to stop this phenomenon with increasingly severe sanctions.<sup>10</sup> Moreover, Depoorter, Parisi and Vanneste (2005) provide experimental findings indicating that legal sanctioning of file sharing produces a backlash effect on norms and, consequently, on behavior. “If sanctions have a countervailing effect”, they conclude “exclusive reliance on punishment might be unfortunate.”<sup>11</sup>

A second episode, where protest itself was criminalized and, as a result, protest increased (albeit temporarily) can be found in West Germany, in the 1970s, when a high level of terrorist activity was observed. In an attempt to reduce terrorism, criminal law and criminal procedural law were made substantially stricter in all fields related to terrorism. Not only was the founding of and membership in a terrorist organization penalized, but the “support of, or recruitment for” such organizations was also penalized. All suspects in pre-trial custody and inmates found guilty of violating these provisions were (or could be) subjected to severe restrictions of their rights as prisoners. In the first years after enactment, the new law was used in a relatively large number of cases, in particular to sanction lesser forms of support of terrorism that could not be subsumed to any other criminal offense. The effect, though, was a clear increase in the openly expressed opposition against these laws. At the same time, there was an increase in the number of individuals actively criticizing custody conditions of terrorists and in the number of individuals supporting terrorists in more severe ways. All this was encouraged by an increasing number of persons opposing to the strict stance the state took on such activities.<sup>12</sup>

Finally, we have an example of a countervailing effect that produced a reinforcement of the law. The promulgation of laws supporting a woman’s right to terminate a pregnancy has given rise to political and social battles over abortion rights in many countries. In the US, this happened in 1973, with the Supreme Court decision *Roe v. Wade*. This decision has sparked and still provokes much protest by prolife groups. Protest ranges from open expression of dissent to attempts to directly discourage or prevent women from having

---

<sup>9</sup> See, for instance, postings on the New York Times after a woman from Minnesota was found liable for copyright infringement for sharing music online and condemned to pay a fine of \$222,000 in damages (see Leeds 2007: [http://news.blogs.nytimes.com/2007/10/05/sued-for-sharing-music/.](http://news.blogs.nytimes.com/2007/10/05/sued-for-sharing-music/))

<sup>10</sup> See Depoorter and Svanneste (2005) and Oksanen and Välimäki (2007).

<sup>11</sup> Depoorter, Parisi and Vanneste (2005), pag. 367.

<sup>12</sup> See Oppenheimer (1978).



access to abortion services. Such attempts are sometimes classifiable as true harassment and include chemical attacks, death threats and even homicide. There is empirical evidence that such antiabortion activities have had a significant impact both on the supply and on the demand of abortion services in the US, which would have been higher in the absence of protest of antiabortion activists (Kahane, 2000).<sup>13</sup>

### 3. The Model

Imagine a simple framework, where there is an action  $a$ , which the law can either allow or forbid. For instance, we might think of actions like smoking in public spaces, illegally copying recorded music, carrying out abortions, or supporting terrorists by advocating their freedom of expression and public communication. We assume that in the initial period action  $a$  is legal and that in a subsequent period a new law is enacted, which forbids  $a$  and levies a sanction  $S$ . 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. A value  $S = 0$  means a law without a sanction, whereas  $S$  becomes larger as the severity of the sanction or the probability of its enforcement increase.<sup>14</sup> The primary effect of a sanction is to deter individuals from engaging in the illegal behavior. Sanctions increase the relative price of the illegal conduct and possibly lead to a substitution towards other legal activities. The decision to carry out the action, however, depends not only on the legal sanction, but also on the value individuals assign to the action and on current social norms.

In line with Bicchieri (2006), we define social norms as informal rules followed by a sufficiently large subset of the population and supported by *normative expectations* (i.e., individuals believe that a sufficiently large subset of the population expects them to conform and may sanction deviations). In our model, both current behavior and expressed opinions constitute a “social norm indicator”. This, we believe, is a more realistic assumption than letting individuals learn social norms either from behavior (as in Kahan 2000) or from expressed opinions only (Kuran 1989 and 1995). People may in fact be confused by the mere consideration of observed behavior, which may be motivated by self interest (e.g., crime).<sup>15</sup>

We now turn to defining individual behavior and expressed opinions and to characterize the interdependence of individual choices, social norms and the law.

---

<sup>13</sup> The evidence provided by Kahane (2000) is not without counterexamples in the literature. For instance, Oltmans Ananat, Gruber and Levine (2004) provide evidence that the legalization of abortion has produced a decline in fertility rates of women who had their peak childbearing years during the early 1970s. The interpretation of this evidence is, however, problematic. The reduction in fertility could be attributable to better access to contraceptive techniques and to the information that was made available to women after the legalization of abortion.

<sup>14</sup> 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.

<sup>15</sup> If expressed opinions and behavior diverge (e.g., the majority of the people voices the opinion that corruption is bad but still corruption is pervasive), the indication about the appropriate behavior may become blurry and decisions are more and more based on individual beliefs and material payoffs.

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 individual deviation from this common value.<sup>16</sup>  $B-v_a$  is therefore the benefit that an individual derives from committing the action. We assume that  $v_a$  varies among individuals according to the distributions function  $G(v_a)$ , with the corresponding density  $g(v_a)$ . Analytically, this implies that all those with

$$B - v_a > S + \beta n - \alpha x \quad (1)$$

choose to carry out the action, where  $\alpha$  and  $\beta$  are non-negative constants and measure the weight that the individual attaches to social norms in the utility function. Particularly,  $\alpha$  is the weight attached to expressed aggregate social norms (defined by  $n$ , a variable that we shall explain in detail shortly), whereas  $\beta$  is the weight attached to observed behavior (defined by  $x$ , the proportion of individuals carrying out the action). Then the proportion of people who are willing to violate the law and engage in the sanctioned behavior in a given period is  $G(B-S-\beta n+\alpha x)$ . The choice to engage in the illegal behavior adapts over time according to changes in the current state. Thus, given  $x$ , the change in the proportion of individuals carrying out the illegal action in each period is

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

where the dot denotes the derivative with respect to time.<sup>17</sup>

We further assume that each person also assigns a normative valuation  $v_n$  to the activity, where  $v_n$  can be interpreted as the size of the legal sanction that the person deems appropriate for the action. The lower the normative valuation of the activity, the less an individual would like to see the action being performed by others and hence the larger her most preferred sanction  $v_n$ . Let the distribution of normative valuations in the society be given by  $F(v_n)$ , with density  $f(v_n)$ .<sup>18</sup> 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  $a$ . Opinions and social reaction can either express approval or disapproval and contribute to the establishment of social norms.

Let  $q$  be the proportion of individuals who disapprove action  $a$  to the point of actually manifesting discontent and protest against the law at time  $t$  for being too lenient, considering the sanction for action  $a$  to be too small. Similarly, let  $p$  be the proportion of individuals who approve action  $a$  to a degree that they would manifest discontent and protest against the law at time  $t$ , or otherwise openly support violations of the law by others,

---

<sup>16</sup> 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.

<sup>17</sup> We omitted the time index for notational simplicity. 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.

<sup>18</sup> 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 carry out  $a$ . Positive correlation allows individuals to infer normative valuations of others not only from their expressed opinions but also from their actions. However, to keep our model as general as possible, we do not assume any form of correlation and consider  $v_a$  and  $v_n$  as independent.

because they deem the sanction for action  $a$  to be inappropriate or excessively high. We define the aggregate expressed social norm at time  $t$  as  $n = q - p$ . That is, the expressed social norm with respect to action  $a$  is given by the difference between the rates of social disapproval and approval for  $a$ .

Expressed opinions and the extent to which individuals are willing to reveal their values by supporting or opposing a legal rule changes over time to respond to the current state of the society. To understand their dynamic evolution, one should consider that the expression of opinions and protest are costly. Some of these costs are financial opportunity costs, like taking time off work or leisure to express one's opinions and protesting against an unjust law. Some others are psychological costs, like the cost of expressing opinions contrary to prevailing social norms inferred from the opinions expressed by 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. Thus, an individual with a preference for a stricter law will be willing to express her opinion only when the utility she derives from such stricter law is high enough to offset the costs of expression, i.e. when

$$v_n - S > c - \lambda n_t + \gamma x \quad (3)$$

Similarly, an individual preferring a more lenient law will express her opinion only when

$$S - v_n > c + \lambda n_t - \gamma x_t \quad (4)$$

The right-hand side of the last two inequalities represents the total cost of expression and protest and consists of a constant part,  $c$ , and a variable part depending on current social norms. The parameters  $\lambda \geq 0$  and  $\gamma \geq 0$  respectively measure the impact of expressed social norms and others' behavior on the individual's cost of protest.

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

$$\dot{q} = \sigma(1 - F(S + c - \lambda n + \gamma x) - q) \quad (5)$$

and

$$\dot{p} = \sigma(F(S - c - \lambda n + \gamma x) - p) \quad (6)$$

where  $\sigma$  is the number of individuals expressing their opinion in every time period.<sup>19</sup>

Combining equations (5) and (6), expressed opinions  $n = q - p$  evolve over time according to the following dynamics:

$$\dot{n} = \sigma(1 - F(S - c - \lambda n + \gamma x) - F(S + c - \lambda n + \gamma x) - n). \quad (7)$$

---

<sup>19</sup> Recall that, in equation (2), we assumed that only one individual per period decided whether to take action  $a$ . As individuals may decide on the expression of their opinions faster or more slowly than on their actions, we introduce the parameter  $\sigma$ . If  $\sigma > 1$ , opinions change faster than actions. The opposite occurs if  $\sigma < 1$ .

To simplify the graphical presentation of this problem, we note that the dynamic system consisting of equations (2) and (7) perfectly reflects the dynamic system consisting of equations (2) through (6), both with respect to the location and the stability of the equilibria.<sup>20</sup>

By definition, the equilibria of the system are reached when both the proportion of individuals carrying out action  $a$  and the proportion of individuals expressing their opinion become constant over time, 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 simplify the presentation of our results, we introduce some assumptions.<sup>21</sup> Particularly, we assume that  $\alpha=0$  in equation (2), meaning that the choice of carrying out action  $a$  does not depend on  $x$ , and  $\lambda=0$  in equation (5), meaning that the change in opinions is not influenced by the current social norm.

Equations (2) and (7) therefore reduce to

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

and

$$\dot{n} = \sigma(1 - F(S - c + \gamma x) - F(S + c + \gamma x) - n). \quad (7')$$

By posing  $\dot{x} = 0$  in equation (2') we define the equilibrium proportion of individuals carrying out action  $a$  as a function of the current social norm:

$$x^*(n) = G(B - S - \beta n) \quad (8)$$

and by posing  $\dot{n} = 0$  in equation (7') we define the equilibrium social norm as a function of the current proportion of individuals who engage in activity  $a$ :

$$n^*(x) = 1 - F(S - c + \gamma x) - F(S + c + \gamma x) \quad (9)$$

We further assume that both  $f(\cdot)$  and  $g(\cdot)$  are uni-modal and rather concentrated around the mode. Finally, we assume that the cost parameter  $c$  is small. The latter two assumptions guarantee that both functions  $x^*(n)$  and  $n^*(x)$  have the shape depicted in Figure 1 (reverse S-shape). Since equilibria are defined as the values of  $x$  and  $n$  that solve  $\dot{x} = 0$  and  $\dot{n} = 0$  simultaneously, they are located where the functions  $x^*(n)$  and  $n^*(x)$  intersect. Figure 1 provides an example with three equilibria, where stable equilibria are represented by black circles and unstable ones by empty circles, as one may infer from the arrows denoting the directions of the dynamics described by equations (2') and (7').<sup>22</sup>

We should note that the equilibria need not be so nicely distributed across the entire range of  $x$  and  $n$ , but may instead be very concentrated, for example in the upper left quarter of Figure 1, which would be the case if about one half of the population would never carry

---

<sup>20</sup> Proof in Appendix A.

<sup>21</sup> Our results do not hinge upon the simplifying assumptions used in this Section. These assumptions will be relaxed in Appendix B, where we will generalize our model showing that the nature of our results does not change.

<sup>22</sup> The shape of the functions  $x^*(n)$  and  $n^*(x)$  and the stability of equilibria are analyzed in Appendix A.

out the action and never protest against the law for being too strict. Clearly, a uni-modal distribution implies that the majority of individuals share the same behavior and beliefs. It is therefore interesting to find that, starting from such an initial distribution, the system can end up in equilibria where society is split in two distinct groups: some individuals whose prior values were closer to the new law become law-abiders, while other individuals whose prior values were further away from the new law react, moving away from the law and converging to a law-violation behavior. This would happen, for instance, if all three equilibria were close to the  $x$  – axis. In such a case, legal intervention triggers a social divide, so that a society that initially shares fairly homogeneous opinions splits into two groups following opposite behavior and upholding different opinions.<sup>23</sup>

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

In this Section we study the impact that the incentives created by sanctions and social norms have on deterrence when the adoption of a new law generates “social feedback.” When legal intervention triggers social feedback, both the reinforcement of a legal sanction and the enactment of more lenient laws can have countervailing effects on behavior, increasing criminal activities or reducing prevalence of an activity that the legislator would like to encourage.

We shall proceed by considering the feedback between expressed opinions  $n$  and action  $x$  when a lawmaker implements an increase in the severity of the sanction  $S$ . From equation (8), the change in the prevalence of activity  $a$  is

$$\frac{dx^*(n)}{dS} = -g(\cdot) < 0 \quad (10)$$

whereas the change in protest for any level of  $x$  is, from equation (9)

$$\frac{dn^*(x)}{dS} = -(f(S - c + gx) + f(S + c + gx)) < 0 \quad (11)$$

Thus, an increase in  $S$  produces a decrease in  $x$  for any given level of  $n$  and also a decrease in  $n$ , due to the simultaneous increase in the protest for law strictness  $p$ , and decrease in the protest for a too lenient law,  $q$ . In fact, all other things being equal, a higher  $S$  reduces the expected payoff from committing the action and, for any given social norm  $n$ , reduces the proportion of individuals carrying out  $a$ . Similarly, an increase in  $S$  would make it less likely that an individual with valuation  $v_n$  expresses an opinion in favor of a stricter law (thus decreasing  $q^*$ ) and would instead increase individual incentives to protest for the excessive strictness of the newly enacted law, thus increasing  $p^*$ .

To understand how countervailing effects may arise, let us suppose that initially the social system is in the equilibrium marked as  $e_1$  in Figure 2. In  $e_1$  the value of  $n$  is relatively high (meaning that the proportion  $q$  of people protesting for a stricter law is relatively

---

<sup>23</sup> Such result would be all the more likely if we started from a bimodal distribution, as we argue in Appendix A. By assuming a uni-modal distribution, we therefore put ourselves in the most difficult initial conditions.

higher than the proportion  $p$  of people expressing the opposite opinion) and  $x$  is low. Assume now that the legislator passes a new law increasing the sanction  $S$  to  $S'$ . According to equations (10) and (11), the function  $x^*(n)$  shifts leftwards to  $\tilde{x}^*(n)$ , whereas function  $n^*(x)$  moves downwards. Following a (possibly small) increase in the sanction  $S$ , if the function  $n^*(x)$  shifts much more than the function  $x^*(n)$  in response to the legal change (like in the situation illustrated in Figure 2, where  $n^*(x)$  moves to  $\tilde{n}^*(x)$ ), then the only equilibrium would be  $\tilde{e}_3$ , where  $x$  is high and  $n$  is negative and relatively large (implying  $p$  much greater than  $q$ ). This implies that the direct feedback in norm formation is large and clearly larger than the feedback in action decisions.<sup>24</sup> Hence, the proportion of individuals carrying out the action would increase substantially and the social norm would turn sharply against the law.

A similar effect (though opposite in sign) could be observed in case of a reduction in the sanction  $S$ . In that case, the function  $x^*(n)$  would move rightward, whereas  $n^*(x)$  would shift up. If the shift in  $n^*(x)$  were much larger than the move in  $x^*(n)$ , we could observe a transition towards a state in which the number of people actively protesting against the law for being too lenient increases substantially and the proportion of individuals who choose to carry out the action decreases notwithstanding the more permissive law.

In both cases presented above, the countervailing effects of legal innovation occur as the social reaction is so strong to produce a positive net effect of legal sanctions on deterrence. If the shift in the function  $n^*(x)$  were smaller (so that  $n^*(x)$  moved to  $\tilde{n}^*(x)$ ), starting from equilibrium  $e_1$ , society would reach  $\tilde{e}_1$  and not  $\tilde{e}_3$ , and no countervailing effect would occur. In equilibrium  $\tilde{e}_1$  in Figure 2, although protest against the strictness of the law increases (and  $n$  decreased compared to the initial equilibrium  $e_1$ ) such an increase is not enough to compensate for the incentive effect of a higher  $S$ , so that the change in the law reduces  $x$ .

Copyright infringement by participants to peer-to-peer networks is one case in which the reaction to stricter sanctions and enforcement produced countervailing effects. The empirical evidence suggests that individuals who engage in file sharing do not generally consider their activities as serious copyright violations that would justify a criminal sanction and often regard them as means to damage corporations, seen as entities with monopoly power which they use to take advantage of consumers (see Depoorter, Parisi and Svanneste 2005 and Oksanen and Välimäki 2007). In terms of our model this implies that

---

<sup>24</sup> In terms of equations (2) and (7) this is equivalent to requiring that the parameter  $\gamma$  is much smaller than the parameter  $\beta$ , i.e., that the influence of observed action on social norms is much weaker than the reverse influence, and that the parameter  $\lambda$  is large, so that there is larger direct feedback in norm formation than in action-decisions.

in the face of strict copyright enforcement,  $n$  is very low, possibly negative.<sup>25</sup> Moreover, the parameter  $\beta$  in equation (8) is likely to be high, meaning that behavior is very sensitive to expressed social norms.

The results of the above analysis thus suggest that sudden and extreme changes in social norms can lead to sudden large reactions in behavior. In the framework we have presented so far, such extreme changes are persistent and protest does not subside over time. In the next Section, we shall present alternative explanations for the countervailing effects of legal sanctions on actual behavior, explanations that do not necessarily require a persistent increase in protest and might provide a “microfoundation” for the different speed in the reaction of social norms and actions.

## 5. The Countervailing Effect of Legal Announcements and Public Outcry

In this Section, we discuss the conditions that may entail large reactions to minor legal changes. All conditions we describe are based on the countervailing effects of social protest and are observed when the initial enforcement of a new legal sanction has a greater or faster impact on protest than on the level of compliance (action). According to their theoretical basis, we call these effects the “announcement effect” and the “outcry effect.”

### *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 in the case of statutory law. When a new law is announced but has not yet come into force, it is generally unable to produce any direct deterrent effect on current behavior.<sup>26</sup> The announcement effect occurs when the announced law, although not yet effective, triggers an immediate reaction of social norms. When a new law comes into force with some lag, social norms may react to the legal change before the new 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, 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. Consider a social system similar to that described by the equilibrium labeled  $e_1$  in Figure 2, where the prevalence of “crime” before the legal change is limited and  $n$  is positive and relatively large, revealing a strong approval

---

<sup>25</sup> This conjecture is validated by evidence that support for file sharers caught and fined in court is strong and opposing voices very few. For instance, Oksanen and Välimäki (2007) report the case of an American teenager who created a search engine for music files. He was discovered, sued by RIAA (the Recording Industry Association of America), and forced to settle his case for \$12,000, an amount which equaled his entire savings for college. This case provoked the angry reaction of the file-sharing community, which organized a fund-raising campaign that raised enough money to pay the settlement.

<sup>26</sup> This is especially true when the new legislation is not retroactive and is thus unable to have effects on behavior which takes place before its entry into force.

for strict laws. Now suppose that an increase in the sanction is announced, with a delayed implementation. Prior to the entry into force, behavior is not yet deterred by the new sanctions (the schedule  $x^*(n)$  still describes the levels of  $x$  towards which the system is drawn for given values of  $n$ ), but is already affected by the observed social reaction, characterized by a quick increase in the protest against the stricter law ( $p$  increases, reducing  $n$ , so that  $n^*(x)$  shifts downwards, to  $\tilde{n}^*(x)$ ). The schedules  $x^*(n)$  and  $\tilde{n}^*(x)$  intersect only once at  $\tilde{e}_3$  and, in the transitional period, the dynamics of behavior and protest is driven towards this unique equilibrium. Expressed opinions will become increasingly less supportive of the high sanction and, as a consequence, social norms will evolve against the law, driving behavior in the opposite direction. Once the higher legal sanction becomes effective,  $x^*(n)$  shifts downward to  $\tilde{x}^*(n)$  and the social system would stay in the region of attraction of equilibrium  $\tilde{e}_3$ . As a result of the increase in the sanction, the frequency of the illegal activity in society would ironically increase and social norms would develop against strict laws and possibly in support of the illegal activity.

It is worth noting that legal rules produced through legislation (and civil law systems in general) are more vulnerable to announcement effects than judge-made law. The announcement effect is in fact typical of legislative processes that have to undergo complex procedures and have formal publication requirements, which is typical of codified law and parliamentary systems. Judge-made rules and common law rules in general become instead known to society through the court decision, without prior notice. This clearly eliminates the time lag and announcement effect that may exacerbate the rise of countervailing social norms. However, this by no means implies that common law courts can ignore the countervailing effects of their decisions.

***b. Getting Used to Unjust Law: The Outcry Effect***

Laws perceived as excessively strict or even unjust when enacted could later become tolerated and eventually accepted. The initial outcry triggered by more severe sanctions may eventually give way to acceptance and perhaps even internalization. In other words, initially sharp reactions and changes in the expressed social norm may fade away over time. The initial strong reaction, however, although temporary, may change the values of society and move it towards a different equilibrium, exactly as happens with the announcement effect.

Consider a situation where a new law is passed, increasing the prescribed sanction from a given level  $S_0$  to a level  $S > S_0$ . Assume that both  $q$  and  $p$  depend not only on the absolute size of the sanction  $S$  (as they did before), but also on how much it changed relative to the original value  $S_0$ . Formally, this assumption is introduced in our model by adding a term  $\kappa_i \Delta S$  to the size of the (changed) sanction  $S$  in the inequalities describing the decisions to express one's opinion. Then equation (7') becomes

$$\dot{n}_i = \sigma \left( 1 - F(S + \kappa_i \Delta S - c + \gamma x_i) - F(S + \kappa_i \Delta S + c + \gamma x_i) - n_i \right), \quad (7'')$$



where  $\Delta S = S - S_0$  is the size of the change in the legal sanction and  $S_0$  is the level of the sanction prior to legal change. The parameter  $\kappa_t > 0$  measures the way in which the change in  $S$  impacts protest at time  $t$ .

The outcry effect captures the fact that society reacts not only when a new law introduces a sanction that public opinion perceives to be unjust, but also when the new law departs substantially from the previous state of the law. For instance, if a given criminal action such as substance abuse was previously punished with six months in jail and then a new law is passed increasing the sanction to five years, public opinion will react much more than in the case where the sanction is raised from four to five years. *Ceteris paribus*, an unjust law will trigger a larger social reaction if it represents also a substantial departure from the prior state of the law.

The assumption that social norms react not only to the objective severity of the law but also to the extent of the change from prior law implies that people adjust to the law over time, even to those laws that were initially perceived as unjust. The social acceptance (or internalization) of the law could be a relatively slow process, and implies that  $\dot{\kappa}_t < 0$ . Restating this in words, internalization implies that society progressively reacts less to the difference between the new law and the original standard  $S_0$ .

In terms of Figure 2, then, public outcry causes a temporary shift of the function  $n^*(x)$  to  $\tilde{n}^*(x)$ . This effect partially withers away as time elapses (and the social-norm equilibrium schedule moves back to  $\hat{n}^*(x)$ , while  $x^*(n)$  only shifts to  $\tilde{x}^*(n)$ ). Then, as was the case for the announcement effect, the system tends towards the equilibrium described by the intersection of  $\tilde{n}^*(x)$  and  $\tilde{x}^*(n)$ , labeled  $\tilde{e}_3$  in the Figure 2. If the stronger initial reaction persists for a sufficiently long time, then the system would find itself in the region of attraction of  $\tilde{e}_3$  even when the unjust law is eventually accepted and the dynamics of the system is described by  $\hat{n}^*(x)$  and  $\tilde{x}^*(n)$ . At that point, a low-activity equilibrium with social norms closely aligned with the content of the law may emerge, as described in the figure by  $\tilde{e}_1$ , but the system remains close to  $\tilde{e}_3$  and, absent further disturbances, will reach this equilibrium.

The idea behind the outcry effect is straightforward. Recent experience of an excessively strict law induces a wave of protest that fades away as individuals get used to the higher legal sanction. Similar to what we observed with the announcement effect, such wave of protest may shift the social system to an equilibrium characterized by higher levels of violations and protest compared to the levels that preceded legal intervention.

This dynamic could explain the reaction that took place during the late seventies in West Germany, where many people engaged in public protest for the extreme increase in the sanctions for the ideological support of terrorism and the corresponding decline in procedural safeguards for individuals charged with these alleged crimes. The strong reaction triggered by the enforcement of these laws led to the expression of empathy for

jailed suspects, and eventually led to a widespread opposition of those laws by a greater number of individuals. While the discussion became less intense after some time (i.e., public outcry  $\kappa_i \Delta S$  vanished over time) a large number of people still opposed the new laws and many continued to express solidarity with the individuals accused or convicted for ideological support of terrorism until the eighties.

Similar effects may have played in the upsurge of violent reprisals against women seeking abortions and doctors practicing these procedures, discussed above. To describe the dynamics in this latter case, one should consider that the law allowing for abortion (where no pregnancy termination was legally admitted before) is equivalent to setting more lenient legal standards. In this case, we can assume that  $n$  is initially positive or near zero.<sup>27</sup> The decrease in  $S$  moves the functions  $x^*(n)$  upward and  $n^*(x)$  rightward in Figure 3. However, in this case society is likely to perceive  $\Delta S$  to be very large, since the law represented a sudden change from prohibition to liberalization (although with some restrictions) of abortions. This may have the effect of raising a considerable outcry in the prolife group, temporarily moving  $n^*(x)$  to such extent that the only possible equilibrium could be found in the most leftward intersection of  $\tilde{x}^*(n)$  and  $\tilde{n}^*(x)$ , labeled  $\tilde{e}_1$  in Figure 1.

## 6. Overcoming the Reverse Effects of Legal Intervention

In the previous Sections, we have shown how changes in the law can trigger changes in social norms, producing countervailing effects and backlash. In all the cases of countervailing effects described above, the unexpected impact of legal innovation is due to the triggering of protest leading to changes in social norms that dominate relatively milder responses in behavior through incentives or deterrence. Countervailing effects are less likely to occur when lawmakers avoid drastic departures from current social norms and sharp changes from current law. This may imply applying lower initial sanctions when restricting a previously unregulated behavior, or carrying out only a partial liberalization of a previously restricted behavior.<sup>28</sup>

Particularly, piecemeal legislation and sunset clauses are especially effective strategies in case of potential announcement and outcry effects, since they might avoid or reduce the surge in protest. In the case of the announcement effect, the increase of protest between the time in which the new law is announced and the time in which it comes into effect is limited if the new law remains relatively close to the status quo. Due to the weak initial protest, the system would remain in the basin of attraction of the desired final equilibrium, regardless of the duration of the transitional phase. By introducing legal

<sup>27</sup> Medoff (2002) finds that voter preferences in abortion issues are bimodal. This implies that there are two active groups lobbying for opposite legal provisions. In terms of our model, this can be represented by  $q$  and  $p$  being rather similar, with society almost evenly split between prochoice and prolife groups and  $n$  near zero.

<sup>28</sup> We illustrate the dynamic interaction between social and legal norms without venturing into welfare analysis. Whether incumbent social norms or innovative legislation are more likely to result in efficient rules is a question that is orthogonal to our analysis. On this specific subject, see Posner (1996).

change in small consecutive steps, piecemeal legislation can carry out substantive changes in the law over time. Subsequent increases in the legal sanction would trigger an increase in protest, but this reaction would take place in an environment characterized by the higher levels of compliance obtained in the previous phase. The same legislative strategies would also work in the case of a possible outcry in response to a change in the law. Consider, for instance, the case in which the lawmaker wants to implement a substantial increase in sanction (from  $S_0$  to  $S_I$ , with  $S_0 \ll S_I$ ). In order to avoid an initial surge in protest with possible countervailing effects, the government may initially change the sanction from  $S_0$  to  $S_2$ , with  $S_0 < S_2 < S_I$ . Assume that the initial equilibrium is  $e_I$  in Figure 2. Crime and protest would then reach a short-term equilibrium in  $\tilde{e}_1$ . Individuals would then gradually adjust to the new legal rule and the final equilibrium will be a point along  $\tilde{x}^*(n)$  and above  $\tilde{e}_1$ , with a lower crime rate, as  $\kappa_t \Delta S$  becomes smaller and smaller over time and  $\tilde{n}^*(x)$  shifts up. The government could then move from  $S_2$  to  $S_I$ . Given that also this second legal change has a small  $\Delta S$ , there will be a limited increase in protest. This would allow the lawmaker to reach a long-run equilibrium with even less crime and only a slight decrease in  $n$ . In this way, the lawmaker can avoid countervailing effects by introducing stepwise legislation.<sup>29</sup>

The second possible strategy the lawmaker could adopt to avoid countervailing social norms is the enactment of temporary legislation and sunset clauses. When legislation contains a sunset provision, the effects of the law terminate after a specific date, unless the effectiveness of the law is extended by legislative action.<sup>30</sup> The adoption of sunset clauses potentially reduces protest and prevents the realization of announcement and outcry effects. People in fact rely on the fact that, when a new law is temporary, its effects will automatically end, unless the law is reenacted or expressly extended. This may have the effect of decreasing protest in the transition period before the law is actually in place, reducing the impact of the announcement effect. Similarly, sunset clauses, by limiting the period of time during which an “unjust law” is implemented, have the likely effect of decreasing  $\kappa_t$ , thus lowering the impact of the change  $\Delta S$  on protest.<sup>31</sup>

Two additional ways to tame the reaction triggered by the announcement of a new law would entail (i) shortening the lag between the enactment and the entry into force of the

---

<sup>29</sup> Things would change in a world of forward-looking individuals. With rational expectations, individuals would anticipate that the final objective of the legislator is  $S_I$  and would tailor their protest to such expected final level and not to the initial level. With perfectly rational individuals who are able to foresee the entire legislative scheme, other legislative strategies should be considered to control countervailing effects.

<sup>30</sup> Temporary legislation consists of statutes containing clauses limiting the duration of their validity. Such clauses are usually termed “sunset clauses”. Sunset clauses operate as a sort of termination date beyond which the law will expire and the sanction will go back to its original level. For analysis of the effects of other timing rules, see Gersen and Posner (2007) and Luppi and Parisi (2009).

<sup>31</sup> This was the strategy 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. 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 (see Koh in *The Economist*, 30<sup>th</sup> October 2003).

law, and (ii) recur to executive legal intervention. Both these strategies reduce the transition period. Executive legal intervention allows the government to promulgate a law that is immediately enforceable, even if it has later to be confirmed by the parliament. The incentive effects of legal intervention would then not be diluted or reversed by the announcement effect.<sup>32</sup>

Finally, when the legislator wants to avoid countervailing effects, (pigouvian) taxes or, possibly, rewards may be used in lieu of sanctions. Taxes, as well known, have a different framing power: unlike sanctions, taxes are not perceived as ways to punish a given behavior. Agents themselves choose to pay them, as a price for undertaking an action that is not condemned by the tax as illegal. Similarly, a lawmaker may use rewards to prevent countervailing effects when he wants to implement a stricter law or to liberalize a given action in the face of a strong social opposition. While it is plausible that positive and negative incentives (carrots versus sticks) might have comparable expressive and incentive effects, it is also likely that they will differ to the extent they trigger social reaction.

## 7. Conclusions

In this paper, we have analyzed the interaction between law and social norms. The law affects human choice by creating incentives and altering the cost of alternative behavioral choices. Legal rules can also affect behavior by promoting the individual internalization of the values expressed by the law. In this paper, we have shown how the effectiveness of legal intervention further depends on the “social response” triggered by the enactment and enforcement of a new rule. In general, it is likely that rules that depart from current social values may trigger opposition. The dynamic effects of such protest could lead to changes in behavior opposite to those intended by the lawmaker.

By engaging in protest and disobedience, individuals reveal their values to others. This may reinforce others’ dislike of the law and lead to an increased level of legal disobedience. When a lawmaker enacts laws that clash with existing social values, those social values may be crowded in. We have shown the interesting relationship between the expressive effects of the law and the feedback provided by the reaction of society to the new law. When society is very reactive to new laws, an increase in the strictness of the law may lead to countervailing effects featuring high protest and a higher rate of violations, whereas laws that are more lenient could actually help reduce the incidence of the activity. The public knowledge of a forthcoming legal change before its coming into force may exacerbate the countervailing effects. Given the complementarities between the social opposition and the actual violations of the law, a surge in protest may trigger an increase in the number of violations before the new law is able to exert its incentive effects. To avoid this, lawmakers should minimize the time lag between enactment and enforcement of new statutes, or else introduce sunset clauses or proceed with piecemeal legislation to minimize

---

<sup>32</sup> We could of course think of many 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. However, this would violate modern constitutional or democratic principles.

the countervailing effects triggered by widespread social protest. Often, social reaction depends on not only the absolute size of the sanction but also on the extent to which it changes prior law. Statutes intending to induce substantial shifts from current norms may thus have to proceed in a gradual fashion to avoid countervailing effects.

Finally, our paper points out an interesting and somewhat paradoxical result: lawmakers can exploit the countervailing effects of protest and social reaction to increase deterrence by means of softer legal enforcement. The idea is that the adoption of more lenient laws may prevent the rise of protest, thus avoiding the emergence of a perceived social norm that condones or even supports violations. This effect of an increased perceived social disapproval counteracts the reduced incentives of the lower sanction. When the impact of social norms is sufficiently high, a reduction in the sanction may actually reduce violations.

Future extensions should consider alternative instruments to optimize the net deterrence of intervention in a comparable dynamic setting. For example, when gradual adjustments are not possible, lawmakers could adopt instruments such as taxes to avoid triggering countervailing effects by taking advantage of the different framing power of taxes as opposed to fines or other legal sanctions. More generally, future extensions of our analysis should explore the different effect of positive and negative incentives. In incentive terms, we generally regard the rewarding of law-abiders as equivalent to the sanctioning of law-violators. This equivalence is lost when we take into account the expressive and countervailing effects of legal intervention. Our analysis of the countervailing effects of legal intervention may also help lawmakers choose the optimal combination of probability and magnitude of a penalty when the countervailing effects of legal intervention are present. This analysis might quite substantially change the conventional wisdom based on the standard deterrence model.

## Appendix A

### A1. Uniqueness of equilibria for $p$ and $q$

Lemma 1 proves that, in order to find the long-run equilibrium for  $p$  and  $q$  we just need to look for the long-run equilibria of  $n$ . Hence, in our analysis of long-run equilibria, we can study the equilibria of the variables  $n$  and  $x$ : the equilibria of the other variables follows.

LEMMA 1: *For every  $n$  satisfying  $\dot{n}=0$ , there are unique levels of  $p$  and  $q$  satisfying  $\dot{q}=\dot{p}=0$ .*

PROOF: Assume  $n^*$  is a value of  $n$  satisfying  $\dot{n}=0$ . Solving equations (5) and (6), respectively for  $q$  and  $p$  respectively, after substituting  $n=n^*$ , yields

$$q = 1 - F(S + c - \lambda n^* + \gamma x) \quad (12)$$

$$p = F(S - c - \lambda n^* + \gamma x) \quad (13)$$

It is immediate to see that there is a unique value of  $q$  and  $p$  that satisfies equations (12) and (13), since  $F(\cdot)$  is a constant for given  $n$  and  $x$ . Hence, for any equilibrium value  $n^*$  it is possible to determine a unique equilibrium value for  $q$  and  $p$ . ■

### A2. Shape of the functions $x^*(n)$ and $n^*(x)$

By differentiating equation (8), one gets

$$\frac{dx^*(n)}{dn} = -g(\cdot)\beta \quad (14)$$

which is always negative, since  $\beta > 0$  and  $g(\cdot)$  takes only positive values, being a probability density function. The second derivative is

$$\frac{d^2x^*(n)}{dn^2} = g'(\cdot)\beta^2 \quad (15)$$

Since  $g(\cdot)$  is single peaked, it will be negative for high values of the argument  $B - S - \beta n$  (i.e., for positive but low and negative values of  $n$ ) and positive for low values of the argument (i.e., for high values of  $n$ ), so that  $x^*(n)$  is convex for low  $n$  and concave for high  $n$ . Similarly,

$$\frac{dn^*(x)}{dx} = -\gamma [f(S - c + \gamma x) + f(S + c + \gamma x)] \quad (16)$$

which is always negative, since  $\gamma > 0$  and  $f(\cdot)$  takes only positive values. The second derivative is

$$\frac{d^2n^*(x)}{dx^2} = -\gamma^2 [f'(S - c + \gamma x) + f'(S + c + \gamma x)] \quad (17)$$

Then,  $n^*(x)$  is decreasing when either  $f'(S - c + \gamma x)$  and  $f'(S + c + \gamma x)$  are both positive or  $f'(S - c + \gamma x)$  is positive and greater than  $|f'(S + c + \gamma x)|$ . The function  $n^*(x)$  is

increasing when the opposite occurs (one should recall that  $f(\cdot)$  is also single peaked). Now,  $f'(S-c+\gamma x)$  and  $f'(S+c+\gamma x)$  are both positive for low values of  $x$  and both negative for high values of  $x$ , provided that  $c$  is not too large. Thus, the assumptions made in Section 3 ( $\alpha = \lambda = 0$ , single peakedness of  $g(\cdot)$  and  $f(\cdot)$  and small  $c$ ) all guarantee that  $x^*(n)$  and  $n^*(x)$  are as depicted in Figure 1.

### A3. Stability of Equilibria

To determine the dynamic stability of equilibria we take the first derivatives of equations (2') and (7')

$$\frac{d\dot{x}}{dn} = -\beta g(\cdot) \quad (18)$$

$$\frac{d\dot{n}}{dx} = -\sigma\gamma(f(S-c+\gamma x) + f(S+c+\gamma x)) \quad (19)$$

The derivative in equation (18) proves that  $\dot{x} < 0$  when  $x > x^*(n)$  and the proportion of individuals carrying out action  $a$  decreases over time. Similarly, from (19), when  $n > n^*(x)$ ,  $\dot{n} < 0$  and the difference between  $q$  and  $p$  decreases over time. The direction of change is illustrated by the solid arrows in Figure 1. Looking at the figure, it is immediate to verify that the top-left and the bottom-right equilibria are stable, whereas the middle equilibrium is unstable.

## Appendix B: Generalizing the model

In Sections 3 through 6 we have made a number of simplifying assumptions in order to get simple forms for equations (2') and (7') and easily analyze them. In this Appendix, we will discuss the effects of giving up such 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

In the model we set  $\alpha = \lambda = 0$  in equations (2') and (7'), so that the dynamics of  $x$  did not depend on the current value of  $x$  and the dynamics of  $n$  did not depend on the current value of  $n$ . 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  $x$  first. Equating the time derivative in equation (2) to zero, may then yield up to three solutions in  $x$  for any given  $n$ . Taking the inverse of the partial equilibrium function  $x^*(n)$  so obtained, we get

$$n(x^*) = (B - S + \alpha x^* - G^{-1}(x^*)) / \beta, \quad (8')$$

where  $G^{-1}(\cdot)$  stands for the inverse of  $G(\cdot)$ . It is possible to check that  $n(x^*)$  is a function (there is exactly one  $n(x^*)$  for every  $x^*$ ) and that this function is strictly increasing whenever  $\alpha g(B - S - \beta n(x^*) + \alpha x^*) > 1$  and decreasing elsewhere, since the slope of  $n(x^*)$  is given by

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

as the interested reader may easily verify by taking the total derivative of equation (2) after setting  $\dot{x} = 0$ . Saying that  $\alpha g(B - S - \beta n(x^*) + \alpha x^*) > 1$  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 1 through 0) 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 carry out the action. If either of these conditions is satisfied,  $n(x^*)$  has an upward sloping part and thus  $x^*(n)$  is not only  $S$ -shaped, but has a backward bending part in the middle. In this case, multiple equilibria become more likely, which makes our arguments stronger and renders all effects we discuss more likely.

Similarly, if  $\lambda(f(S - c - \lambda n + \gamma x) + f(S + c - \lambda n + \gamma x)) > 1$ , 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 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 1 through 0) 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 whereas all equilibria given by the intersections of the backward bending parts of  $x^*(n)$  with  $n^*(x)$  and of  $n^*(x)$  with  $x^*(n)$  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 simplifying restriction was the restriction to a “sufficiently small” level of the part of the costs of expressing one’s opinion that is independent of other individuals’ behavior and opinion expression, i.e. the parameter  $c$ . 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 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 reader will already anticipate 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.

## **References**

- Becker, Gary S. (1968). "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76: 169-217.
- 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.
- Bikhchandani, Sushil, David Hirschleifer and Ivo Welch (1998). "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades", *The Journal of Economic Perspectives* 12: 151-170.
- Bernstein, Lisa (1992). "Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry.", *Journal of Legal Studies* 21: 115-157.
- 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.
- Borck, Rainald (2004). "Stricter Enforcement May Increase Tax Evasion", *European Journal of Political Economy* 20: 725-737.
- 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, Francesco Parisi and Sven Svanneste (2005). "Problems with the Enforcement of Copyright Law: Is There a Social Norm Backlash?", *International Journal of the Economics of Business* 12: 361-369.
- Depoorter, Ben and Sven Vanneste (2005). "Norms and Enforcement: The Case Against Copyright Litigation" *Oregon Law Review* 84: 1127-1180.
- Ellickson, Robert C. (1991). *Order Without Law: How Neighbors Settle Disputes*, Cambridge, Mass: Harvard University Press.
- Gersen, Jacob E. and Eric A. Posner (2007). "Timing Rules and Legal Institutions." *Harvard Law Review* 121: 543.
- 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.
- Jost, Peter-J. (2001). "Crime, Coordination, and Punishment. An Economic Analysis", *International Review of Law and Economics* 21: 23-46.
- Kahan, Dan (1998). "Social Meaning and the Economic Analysis of Crime", *Journal of Legal Studies* 27: 609-622.
- Kahan, D (2000). "Gentle Nudges v. Hard Shoves: Solving the Sticky Norms Problem", *University of Chicago Law Review* 67: 607-646.
- 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 21<sup>st</sup>, 2003), *The Economist*, 30<sup>th</sup> October.
- Kuran, Timor (1989). "Sparks and Prairie Fires: A Theory of Unanticipated Political Revolution." *Public Choice* 61: 41-74.
- Kuran, Timor (1995). *Private Truths, Public Lies. The Social Consequences of Preference Falsification*. Cambridge, Mass. Harvard University Press.
- 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>.
- Lessig, Lawrence (1998). "The New Chicago School", *Journal of Legal Studies* 27: 661-691.
- Luppi, Barbara and Francesco Parisi (2009). "Optimal Timing of Legal Intervention: The Role of Timing Rules" *Harvard Law Review Forum* 122: 1-14.
- McAdams, Richard H. (1995). "Cooperation and Conflict: The Economics of Group Status Production and Race Discrimination". *Harvard Law Review* 108: 1003-1084.
- McAdams, Richard H. (1997). "The Origin, Development, and Regulation of Norms," *Michigan Law Review* 96: 338-433.

- McAdams, Richard H. (2000). "A Focal Point Theory of Expressive Law." *Virginia Law Review* 86: 1649.
- McAdams, Richard H. and Janice Nadler (2005), "Testing the focal point theory of legal compliance: The effect of third-party expression in an experimental hawk/dove game". *Journal of Empirical Legal Studies* 2, 87-123.
- Medoff, Marshall H. (2002). "The Determinants and Impact of State Abortion Restrictions", *American Journal of Economics and Sociology* 61: 481-493.
- Nussim, Jacob and Avraham D. Tabbach (2009). "Deterrence and Avoidance", *International Review of Law and Economics*, forthcoming.
- 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>.
- Oltmans Ananat, Elizabeth, Jonathan Gruber and Phillip B. Levine (2004). "Abortion Legalization and Lifecycle Fertility", *NBER Working Paper W10705*.
- Oppenheimer, Martin (1978). "The Criminalization of Political Dissent in the Federal Republic of Germany", *Contemporary Crises*, 2: 97-103.
- Polinsky, A. Mitchell and Steven Shavell (2000a). "The Economic Theory of Public Enforcement of Law", *Journal of Economic Literature* 38: 45-76.
- Polinsky, A. Mitchell and Steven Shavell (2000b). "The Fairness of Sanctions: Some Implications for Optimal Enforcement Policy", *American Law and Economics Review*, 2: 223-237.
- Posner, Eric (1996), "Law, Economics, and Inefficient Norms." *University of Pennsylvania Law Review* 144: 1697.
- Posner, Eric (1998). "Symbols, Signals, and the Social Norms in Politics and the Law", *Journal of Legal Studies* 27: 765-798.
- Posner, Eric (2000). *Law and Social Norms*. Cambridge, Mass. Harvard University Press.
- Stuntz, William J. (2000). "Self-Defeating Crimes." *Virginia Law Review* 86: 1871.
- Thoreau, Henry David (1849). "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. and Y.J. Huo (2002). *Trust in the Law*. N.Y.: Russell-Sage. 31
- Zasu, Yoshinobu (2007). "Sanctions by Social Norms and the Law: Substitutes or Complements?", *Journal of Legal Studies* 36: 379-396.
- Zwiebach, B. (1975), *Civility and Disobedience*, Cambridge University Press.

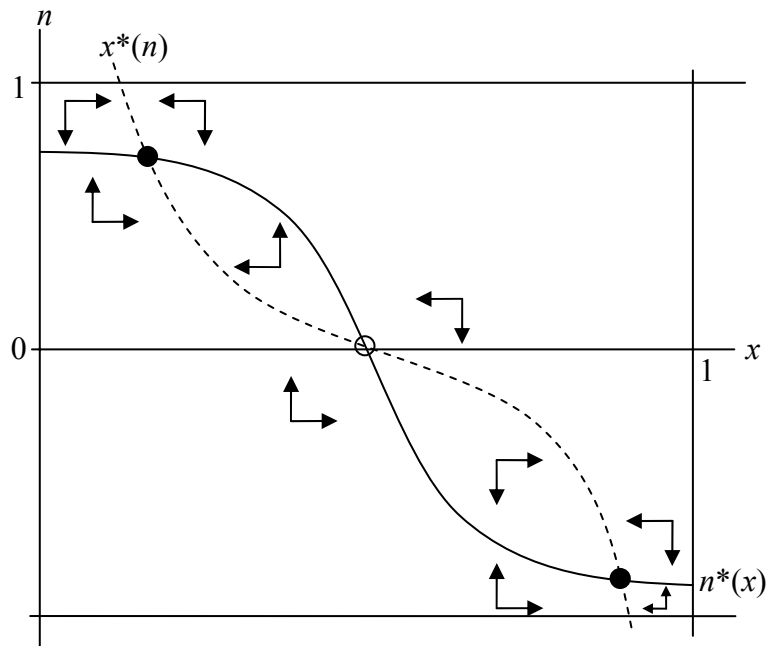


Figure 1: Equilibria: Solid arrows denote direction of dynamics, filled and empty circles denote stable and unstable equilibria, respectively.

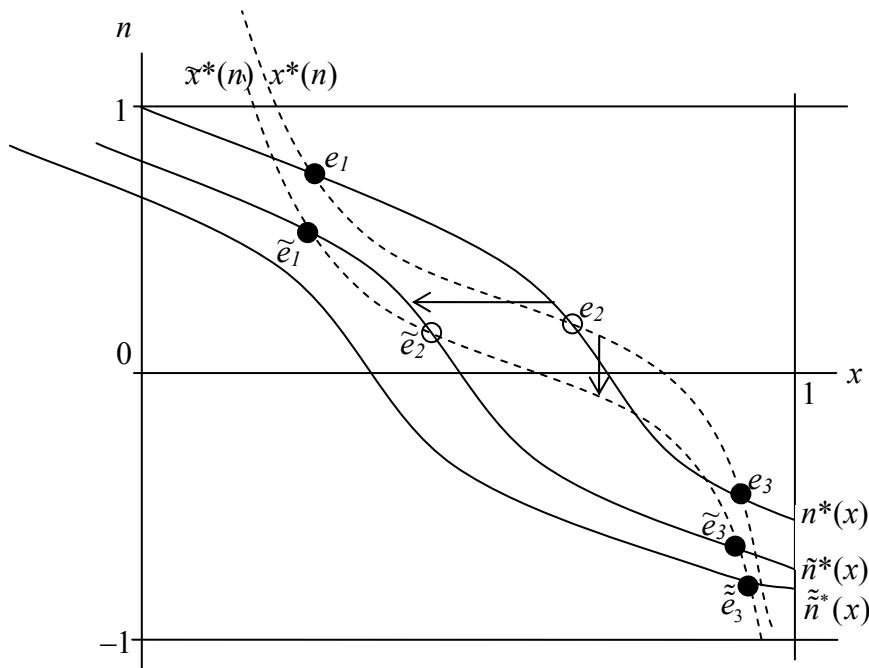


Figure 2: Shifting the equilibrium functions by an increase in  $S$ : Arrows indicate the shift of the equilibrium functions when  $S$  increases.

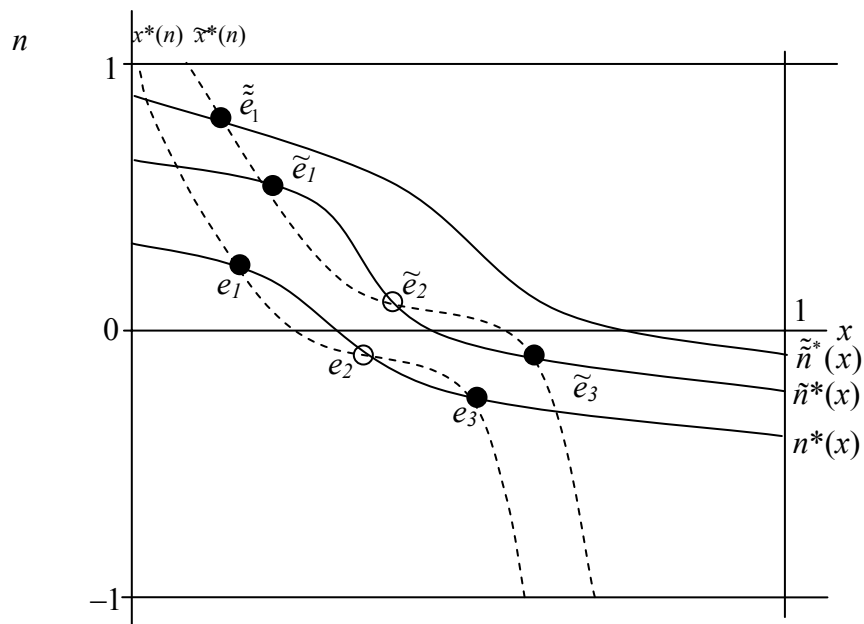


Figure 3: Outcry effect with a decrease in  $S$ . The equilibrium goes from  $e_3$  (where  $n$  is relatively close to 0 and prevalence relatively high) to  $\tilde{e}_1$ , where  $n$  is much larger than 0 and  $x$  has decreased.