

On the Regulation of Social Norms*

Abstract

A model is developed to understand how norms can be influenced by "norm entrepreneurs", e.g., law-makers, government agencies, unions etc. Two instruments of influencing the dynamics of norm-following behavior are analyzed, namely transforming the (monetary) incentives and changing the meaning or the reputational value of following a norm. Both forms of norm regulation are incorporated into Akerlof's model of social custom (1980) and the comparative static properties of norm destruction and norm creation for different types of norms are derived. In particular, it is shown how norms should be regulated when almost everybody follows them and when they take the form of bandwagon and snob norms.

Keywords: social norms, regulation, conformity

Dorothea Kübler
Humboldt-University
Department of Economics
Institute for Economic Theory I
Spandauer Straße 1
D-10178 Berlin
kuebler@wiwi.hu-berlin.de

(...) in the absence of scientific history, short-lived man has little better means of ascertaining whether custom is quietly changing, than the fly, born today and dead tomorrow, has of watching the growth of the plant on which it rests. (p.640)

We change our customs more quickly than our forefathers did, and we are more conscious of our changes and more willing to convert our customs into legal enactments, and to make them uniform. (p.638)

(Alfred Marshall, *Principles of Economics*, 8th ed., London: Macmillan, 1961)

1 Introduction

Social norms have recently gained much attention from economists and legal scholars as an important driving force of individual behavior. For a long time, economists have treated norms as exogenous or, as Marshall puts it, so sticky that changes are barely observable.¹ But norms and customs undergo changes, and at an increasing pace – as Marshall noted at the end of the 19th century. More recently, the processes underlying norm formation have been formally investigated by economists, e.g., by analyzing the evolution of norms and standards in certain environments.² However, little attention has been paid by economists to modelling deliberate attempts by certain members or groups of society to change the existing set of norms. Concluding his paper on the evolution of social norms, Peyton Young (1998) writes:

(...) social change occurs for a complex variety of reasons. It is driven to some degree by the opinions and actions of influential people (role models) who precipitate change because they are widely noticed and imitated. Furthermore, changes in one sphere of interaction may have important spillover effects on other spheres and are related to broad underlying trends in society. (...) we have abstracted away from such complications.

In the following it is studied how a new code of behavior can be established or an old existing code can be destroyed by actions of "norm entrepreneurs"³ such as government agencies, unions, NGOs, and interest groups or single influential individuals.⁴ These questions have been addressed by legal scholars in a number of articles investigating how formal rules or government actions can change the existing set of social norms.⁵ In this paper a model is developed to formalize some issues of this informal discussion and to derive some new results. In particular, the model serves to illuminate the dynamics of social customs in the presence of a norm-regulating agency using two distinct regulation techniques, one directed at the individual monetary incentives and the other at the reputational value of following a certain norm.

Social norms that involve pecuniary disadvantages for all individuals adhering to them do not necessarily disappear. Akerlof (1980) introduced a model to show that norms can, but do not necessarily erode if they are disadvantageous for everybody in the group, because not following the norm may involve a loss of reputation. Two aspects of Akerlof's model, which will be used to analyze norm regulation, are worth emphasizing. First, a norm is not defined as a certain behavior followed by a number of people as in the literature on evolutionary game theory. Rather, it is a moral expectation shared by a group of people, entailing social stigmatization or at least moral indignation aimed at those who deviate. Second, in Akerlof's model the survival of a norm does not only depend on the material attractiveness or some other inherent property of following the norm, but also on the number of followers at the initial stage. For social norms, this seems to be a more adequate approach than looking only at the actual payoff received from following a certain code as it is by no means clear that norms always promote either self-interest, common interest, or genetic fitness.⁶ And even if the benefits of a norm are collective in nature, individual reinforcement does not necessarily work in the right direction.⁷

Norms are constituted by common understandings and expectations of a group of individuals. In order to change these taken-for-granted expectations and to establish a new code of behavior,

it is typically not sufficient that one individual changes his or her behavior. Rather, a group of individuals must embrace and follow the new rule to give it the status of a social norm. But to contribute to the public good of establishing a better norm (for example) is costly for individuals due to the old norm itself that stigmatizes any deviation from it. To put it more generally, a norm can solve a collective action problem, but for a norm to become effective (or to disappear), such a collective action problem has to be solved first. This is the task of norm regulation.

When inefficient norms persist, there is room for welfare improvement by changing the social meaning attached to certain actions.⁸ And regulating norms may be cheaper and more effective than other means of increasing social well-being.⁹ In the model below, two possible instruments of influencing the dynamics of norm formation and erosion are analyzed.¹⁰

First, incentives to follow or break a norm can be established by subsidizing, punishing, or prohibiting a certain behavior. For example, norm erosion can be achieved by granting monetary rewards to those who deviate from a prevailing norm.¹¹ Changing the incentives to break a norm, i.e., a change in relative prices of norm following vs. norm breaking, is referred to as "regulation of motives". For example, a norm of hiring women only if there is no (equally qualified) man applying for the job can be discouraged by rewarding those employers who hire women with a better chance to receive government contracts. Or ecological norms can be strengthened by refunds for returned glass bottles.¹² Another example are labor supply norms for men and women. The German tax system gives tax breaks to one-breadwinner households, but not to households with two incomes and thus favors an asymmetric division of labor. Although the tax system is formally gender neutral, it can be argued that it reinforces preexisting gender roles and discourages women from participating in the labor market.

The second technique to change norm-following behavior involves changing the content of the norm or convincing actors that following the norm is inappropriate. For example, the government can try to convince actors of the inefficiency or inappropriateness of a norm by performing symbolic

acts itself. If the government action has some moral weight of its own, the force of the norm, its *reputational value*, declines. This is modelled by assuming that the social pressure exercised by other norm followers becomes weaker when there is a norm entrepreneur who discredits the norm. In other words, deviation from the norm is vindicated by the norm entrepreneur's action, which is referred to as "regulation of meaning."

Returning to the above example of the norm requiring gender-based hiring decisions, state agencies can use non-discriminatory hiring practices or affirmative-action programs to express their criticism of the existing norm.¹³ Or they may attempt to change norms via education programs, for example by adjusting textbooks or instructing teachers to question certain codes of behavior. The behavioral code of "safe sex" was established as a response both to the spreading of AIDS and to a dysfunctional system of norms governing sexual relations, mainly by extensive advertising campaigns associating condom use with responsible behavior. In the domain of free speech, civil disobedience and free expression can be rewarded, for example, by creating prizes given by independent juries for the courage to stand up for one's beliefs and for resisting the norm "don't stick your neck out." Or, to press for the 5-day work week in Germany, unions launched an advertising campaign with a child saying: "On Saturdays, daddy is mine!" in order to associate the 5-day work week with family friendliness.¹⁴

The norm entrepreneur can also manage the quantity and kind of information people have about both the content of the norm and the proportion of people complying with it. People sometimes attach importance to official statements by a government agency because they believe that it has superior information on the issue at hand. In the case of AIDS, health officials were considered a credible source of advice, which helped to change norms governing sexual practices.¹⁵ Moreover, information about how many people follow a certain behavior might not be available to everybody. Revealing that only few people follow a code that is thought to be followed widely, can seriously undermine this code. Conversely, laws can strengthen norms by publicizing a societal consensus,

i.e., that the majority disapproves of a certain behavior. Or, depriving people of information about the actual (high) number of norm followers or of the information they need to enforce a certain norm (e.g., by protecting privacy) can weaken compliance with it.¹⁶

All of these norm-transforming acts imply that even if the number of people following the norm is rather high, belief in the norm deteriorates and, in the long run, less people are going to abide by it.

In the following, both techniques of norm regulation are incorporated into a model of social custom in order to analyze their short- and long-run effects on the proportion of norm followers.¹⁷ Some simple rules for decision makers concerning the optimal strategy of norm management are derived. The normative implications of the model for norm regulation are discussed by pursuing two examples in more detail: First, xenophobia is widespread in certain German communities, and tolerating unfriendly acts against foreigners, if not committing them oneself, is quite a powerful norm in certain groups. The question whether the punishment of crimes against foreigners should be more severe or whether education programs and symbolic acts such as politicians visiting the site of the crime are more effective, is open and widely discussed. Second, norms governing smoking will be analyzed as the history of smoking regulation in the U.S. is rich and well studied.¹⁸ The antismoking campaign beginning in the late 1960's relied on a mix of policies, among others by tying smoking to unhealthy behavior in advertisements and by a ban of smoking in public spaces.

Finally, two special cases of norms are analyzed in order to derive rules of optimal norm management:

A bandwagon norm is characterized by the property that once a critical number of norm followers is reached, the reputational value of a norm or the social pressure exercised by it increases sharply.¹⁹ Many norms seem to exhibit this property, for example norms governing smoking. As long as a large enough group of people smokes, smoking is considered entirely acceptable. But

when the group of smokers is small in a society, nonsmokers might feel more justified in imposing restrictions on smoking. Using the model it is shown that an inefficient bandwagon norm, adhered to by a large proportion of people, can best be weakened by introducing incentives to break it. On the other hand, to create a bandwagon norm a group of sufficient size must simultaneously adopt the new norm to get the bandwagon started. For example to establish recycling as a norm, a critical mass of people must start returning empty bottles in order to convince others to do the same.

A snob norm is a norm that yields the most reputation to its followers when only a small number of people follow it. It is shown that increasing the incentives to break a snob norm raises its reputational value. For example, when cocaine gets more expensive, cocaine consumption might become even more prestigious for those who can still afford it. Thus, to substantially decrease the number of cocaine users, the price of cocaine must be very high to compensate for these reputation gains. Policies that try to convince people of the dangers of cocaine consumption can be cheaper and more effective. In contrast to bandwagon norms, the regulation of snob norms is more effective when such norms are discredited than when the incentive structure is changed.

The plan of the paper is the following: In Section 2, the model is introduced and the equilibria are characterized. For a simple linear reputational value function, the equilibrium dynamics are derived and the comparative static properties of the two policies of norm regulation are analyzed. In Section 3, the assumption of a specific function for the reputational value of norm following is relaxed, leading to a general comparative static result and to the analysis of bandwagon and snob norms. Section 4 concludes the paper.

2 The Model

In order to analyze norm regulation building on Akerlof's (1980) model, two new variables are introduced. First, regulation of motives (denoted by the variable m) will be identified as a policy affecting short-run utility maximization. Second, regulation of meaning (denoted by r) will be shown to influence the long-run dynamics of norm formation.

2.1 Short-run equilibrium

For individual behavior in the short run, consider a situation in which a norm is clearly defined and well known to everybody. Suppose for simplicity that it is only possible to either follow the norm or not. Take the case of an inefficient norm that prevents people from consuming a private good yielding utility $y > 0$.²⁰ The index variable A takes on the value of 0 if the individual follows the norm (and does not consume the good) and of 1 if he or she breaks it. Furthermore, μ represents the fraction of the population (of size normalized to 1) that believes in the norm. This fraction is fixed in the short run and is known by all members of the population. The taste parameter a_i follows a uniform distribution on the unit interval $I = [0, 1]$. It expresses how much the individual cares about his or her reputation, i.e., about what others believe to be the right behavior. Finally, the norm entrepreneur can introduce an incentive to disobey the norm denoted by $m(>-y)$, e.g., a monetary reward for breaking the norm. Individual utility is given by

$$U(a_i) := (y - a_i \mu + m)A. \quad (1)$$

Each individual is assumed to maximize this utility function, which constitutes the short-run equilibrium.²¹

The norm contradicts the individual's private interest as $y > 0$ (and $m = 0$ for the moment). Thus, absent reputational concerns ($a_i = 0$), a person would always break the code. The higher the proportion of believers in the code and the higher a_i , the greater the disutility from not

following the code. It is assumed that those with the highest taste for reputation, a_i , are also the ones who believe in the social norm. Thus, the strength of the societal belief in a norm, μ , is the aggregate of those individuals who believe in the norm.²²

The model also encompasses the case where nobody follows a norm because it is individually costly to do so and there is no social pressure to follow it. In this situation, a new norm can be established by a norm entrepreneur setting $m < 0$ such that individuals who do not follow the code are punished (with the maximum punishment smaller than the private benefit from breaking the code).

Individual i 's optimal decision is

$$A = \left\{ \begin{array}{ll} 0 & \text{if } a_i > \frac{y+m}{\mu} \\ 1 & \text{otherwise.} \end{array} \right\}$$

Suppose there exists a critical parameter a_c at which $y - a_c \mu + m = 0$. Then the proportion of norm followers at time t is given by

$$x_t = \left\{ \begin{array}{ll} 1 - a_c = \frac{\mu_t - (y+m)}{\mu_t} & \text{if } \mu_t > y + m \\ 0 & \text{otherwise.} \end{array} \right\} \quad (2)$$

The number of norm followers is a piecewise continuous decreasing function in the private benefit from breaking the norm, y , and the policy variable m . By introducing incentives to break the norm, m is increased from 0 to $m > 0$. In this case, it follows directly from (2) that the number of norm followers x_t decreases (if it is not zero already). As every individual gets some monetary reward for breaking the code, some types a_i change their behavior in response.²³ However, when the incentives to break a norm change, the loss of reputation due to violating an existing norm, $a_i \mu$, remains unchanged. Only in the long run does this reputation term adapt to changes in norm-following behavior.

2.2 Long-run equilibrium

Now consider the long run. First define the long-term reputational value of a norm, denoted by $v(\cdot) \in [0, 1]$, as the social meaning of a certain action.²⁴ It depends both on the number of people performing the specific action, x , as well as on policies directed at defining this meaning, r . This reputational value of a norm drives the dynamics of belief in the code, i.e., the short-term reputation effect of a certain action denoted by μ . Thereby, the reputational value of an action affects individual utility as defined in (1).²⁵ The long-run dynamics of the model are described by the following equation:²⁶

$$\frac{\partial \mu}{\partial t} = \delta(v(x_t, r) - \mu_t). \quad (3)$$

Thus, if the reputational value of a code differs from the current proportion of believers, this proportion μ will increase or decrease at rate $\delta \in (0, 1)$ until they are equal in equilibrium.²⁷ Note that δ is exogenous and parameterizes the inertia of the adjustment process, e.g., due to the time it takes until people find out how many others follow the norm.²⁸ And remember that $v(\cdot)$ adjusts to a change in μ because the number of people following the norm, x , depends on μ as specified in the utility function (1).

The relationship between short and long run is the following: In the beginning of each period t , the proportion of believers, μ_t , is fixed and individuals decide whether to follow the norm or not, given this proportion of believers, by maximizing their respective utility function (1). Between period t and the next instant in time, beliefs of the new period are formed according to Equation (3). Thus, in the case that $\partial v / \partial x > 0$, the more people follow a norm the more people believe in it and the higher the utility loss of norm breakers in the next period.²⁹

For the effect of the norm entrepreneur's action, r , on the long-term reputational value of a norm, $v(x_t, r)$, assume that $\partial v / \partial r \geq 0$.³⁰ For example, if the norm entrepreneur releases information that discredits the norm, the social meaning of an action can be altered. Smoking can

have various meanings attached to it such as masculinity or recklessness. This is influenced by advertisements of tobacco companies as well as antismoking campaigns. And violence directed at foreigners can be considered as heroic or as deserving contempt, depending on the social meaning of having a different passport. Again, regulating the social meaning directly (instead of influencing individual behavior which then affects the meaning of a norm) can alter the reputational value of xenophobic acts.

Note that regulation of meaning influences individual actions only in the long run by changing the set of long-run equilibria. This is due to the fact that if r is reduced for a given number of norm followers x_t , fewer people believe in the norm in the next period and even fewer in the period after the next period... until a long-run equilibrium is reached. And every time μ_t decreases, less people follow the norm.

To obtain closed form solutions, in this section of the paper the reputational value is assumed to take the functional form

$$v(x_t, r) = \left\{ \begin{array}{ll} 0 & \text{if } x_t + r < 0, \\ x_t + r & \text{if } 1 \geq x_t + r \geq 0, \\ 1 & \text{otherwise.} \end{array} \right\} \quad (4)$$

In this simple case, the norm entrepreneur's effort to change norm compliance has an identical effect on the dynamics of μ as an increase or decrease in the number of norm followers.³¹ For example, an advertisement warning against smoking has the same effect on belief in a smoking norm as a reduction in the number of smokers. In Section 3 this assumption is relaxed.

For stationarity of the process of norm change, i.e., $\partial\mu/\partial t = 0$, it is necessary and sufficient that $\mu_t = x_t + r$, by Equations (3) and (4). And remember that the number of norm followers x_t is directly related to the number of believers μ_t via individual utility maximization, which is summarized in Equation (2).

Proposition 1 1. If the monetary incentives to break the norm are weak, i.e., $y + m \leq r$, the only stable equilibrium is characterized by $x^* = 1 - (y + m)$ and $\mu^* = 1$.

2. If the monetary incentives to break the norm are of intermediate strength, i.e., $\frac{1}{4}(1 + r)^2 > (y + m) > r$, two stable long-run equilibria exist:

$$x^* = 0, \quad \mu^* = 0 \quad \text{and} \quad (5)$$

$$x^* = \frac{1}{2}(1 - r + [(1 + r)^2 - 4(y + m)]^{1/2}), \quad (6)$$

$$\mu^* = \frac{1}{2}(1 + r + [(1 + r)^2 - 4(y + m)]^{1/2}).$$

3. If the monetary incentives to break the norm are strong, i.e., $(y + m) \geq \frac{1}{4}(1 + r)^2$, the only stable equilibrium is characterized by $x^* = 0$ and $\mu^* = v(0, r)$.

Proof see Appendix. ■

The proposition characterizes one stable interior equilibrium point as well as two stable corner solutions. The equilibria can be illustrated in an (x, μ) -diagram shown in Figure 1.

Place Fig. 1 here

Reformulating Equation (2) yields the short-run equilibrium curve described by

$$\mu_t = \begin{cases} \frac{y+m}{1-x_t} & \text{if } 1-x_t > y+m \\ 1 & \text{otherwise,} \end{cases} \quad (7)$$

i.e., the number of believers is a piecewise continuous increasing function of the number of norm followers.³² This function is represented by the curved lines with arrows in Figure 1. If $r = 0$, long-run stability is represented by the 45°-line. Together with the assumption that high a -types are believers in the norm, this means that when $r = 0$, all followers of a norm are also believers and vice versa in the long run.

If there are two interior long-run equilibria, i.e., the short-run equilibrium schedule and the long-run schedule intersect twice, the upper equilibrium is asymptotically stable whereas the lower equilibrium is unstable (indicated by the small arrows in Figure 1). This can be seen by investigating points on the short-run equilibrium curve above and below the $\mu_t = x_t + r$ schedule. If $\mu_t > x_t + r$, μ_t falls according to Equation (3), which entails a fall of x_t . But below the schedule where $\mu_t < x_t + r$, both variables rise. Thus, if the number of norm followers is at or below the lower equilibrium, $\frac{1-x}{2} - \frac{1}{2}[(1+r)^2 - 4(y+m)]^{1/2} \geq x^*$, the norm erodes, while starting at any proportion of followers above it, the number of norm followers eventually reaches the stable long-run equilibrium.

If the convex schedule is just tangent to the locus of long-run equilibria, i.e., the expression in brackets is zero, only one interior equilibrium exists, which is unstable. But there is a stable equilibrium at $(x^* = 0, \mu^* = v(0, r))$. Similarly, if the short-run equilibrium schedule lies everywhere above the long-run schedule, no solution in real numbers exists and the only long-run equilibrium is $(x^* = 0, \mu^* = v(0, r))$. Finally, if the short-run equilibrium schedule lies everywhere below the long-run schedule, the only equilibrium is $(x^* = 1 - (y + m), \mu^* = 1)$. Everybody believes in the norm and if the direct utility from breaking the norm is greater than zero, only those people don't follow it who are rather insensitive to social pressure.

Destruction of an existing norm can be partial by pushing down the upper (stable) equilibrium and thus reducing the number of norm followers. Completely destroying an existing norm requires reaching the tipping equilibrium and thus necessitates a large enough change in policies. Similarly, establishing a new norm requires pushing down the lower (unstable) equilibrium to $x = 0$ such that just one person following the code can set off the process of norm-generation, characterized by a growing number of believers and followers over time. Two different types of policy can bring about these changes.

2.3 Regulation of motives (variation in m)

Figure 1 represents the effect of regulating motives, i.e., giving individuals an incentive m to break or follow the norm.³³ Subsidizing norm-breaking behavior moves the short-run equilibrium schedule upwards and to the left and thus shifts the equilibria. In particular, the upper (stable) equilibrium moves downwards and the lower (unstable) equilibrium upwards along the $\mu_t = x_t + r$ schedule. For example, when the number of norm followers in the upper equilibrium is decreased, this implies that in the long-run equilibrium the number of believers μ_t is reduced too. If the two schedules do not intersect any more, people do not follow the norm for any a_i and μ_t because the reward for breaking the norm is sufficiently large, and the norm vanishes. Therefore, taking a stable equilibrium with $x^* > 0$ as a starting point, the effect of creating incentives to break the code depends on the size of m . Either the proportion of followers and believers is simply moved down to the new stable equilibrium or the norm dies out completely.

As an example, consider the norm justifying or even encouraging crimes against foreigners prevailing in certain communities in Germany. This norm condones violence against people with a non-German passport or of non-German origin, based on various notions of their inferiority or the harm they supposedly inflict on Germans (such as taking away their jobs).³⁴ These crimes are often not prevented by passers-by and are hard to solve because many witnesses don't find them objectionable.³⁵ One possible instrument used to weaken this norm are more severe punishments.³⁶ Assume that the present state can be characterized as an equilibrium in which a large fraction of the relevant group (e.g., inhabitants of a town or members of a certain age group) act according to the racist norm that violence against foreigners should be tolerated or even deserves praise. Then, heavier sentences for those committing the crimes can weaken the norm by deterring some people from committing them and thus reducing the number of believers. However, to destroy the norm completely, very severe punishments would be needed.

Now consider a situation in which a norm is not followed by anybody, $x^* = 0$. Setting $m < 0$, i.e., introducing positive sanctions for norm followers, pushes down the tipping equilibrium. Once this unstable equilibrium reaches $x = 0$, a single norm follower is enough to induce a norm change leading to the only remaining stable equilibrium with a positive proportion of norm followers.

In the U.S. smokers have been rather successfully painted as "pariahs" and as "weak, reckless, or without self-control" (Lessig: 1029) for putting their own health at risk and for harming third parties with second-hand smoke. This has been achieved by, among other things, banning smoking in public places.³⁷ In Germany, smoking is not banned in many public places such as airports, train stations etc., but some recent attempts to change this are noticeable.³⁸ The model predicts that to establish a new antismoking norm by regulating only the motives of people, i.e., by increasing the "price" of smoking, the restrictions must be so severe that the utility from breaking the antismoking norm is zero.³⁹ In other words, when not smoking does not have any reputational value, the direct costs of smoking must be high enough to at least cancel out the benefits from smoking.

2.4 Regulation of meaning (variation in r)

Regulating the meaning of a norm only affects the long-run dynamics and equilibrium by reducing the reputational value of a norm for a given proportion of norm followers. A certain norm-guided behavior can be discredited by advertisement campaigns or symbolic acts of influential people, thereby reducing the social pressure to follow the norm exercised by norm followers.⁴⁰ Thus, in equilibrium a given number of followers is associated with a smaller proportion of believers than without the norm-transforming policy. Under the assumption that $v(x_t, r) = x_t + r$, meaning regulation amounts to a parallel shift of the $\mu_t = x_t + r$ schedule.

Place Fig. 2 here

Figure 2 depicts the effect of increasing or decreasing the reputational value of a norm for a given number of norm followers. Notice that the qualitative changes are the same as with regulating motives: After reducing $v(x_t, r)$ by setting $r < 0$, the upper equilibrium shifts downwards along the curved schedule while the lower unstable equilibrium shifts upwards. Again, depending on the size of r , norm-following behavior is either reduced or erodes completely.

Returning to the norm of xenophobia, there have been attempts by the German government and NGOs to regulate the social meaning of attacks on foreigners. For example, high-ranking politicians visit the sites of the crimes, giving interviews and speeches in which they condemn the motives of those responsible for the crime.⁴¹ Several NGOs are running education programs in schools to increase the knowledge and to influence the attitudes of pupils concerning foreigners in Germany. Again the model predicts that the norm can be weakened and even destroyed by these efforts. First, its reputational value is decreased, therefore less people believe in the norm, which reduces the social acceptability of violence against foreigners.

If a new norm is to be established, increasing r shifts the 45°-line upwards, thus decreasing the critical number of norm followers necessary to set off a process of norm formation. Simultaneously, the stable upper equilibrium moves further up. Considering the regulation of smoking, Lessig (1995: 1030) argues that the difference in smoking habits between the United States and Europe cannot be due to differences in knowledge, but rather to "different social meanings regulating smoking behavior." In the U.S. there have been various successful campaigns to discredit smoking.⁴² In light of the model, these campaigns increased the reputation gain from not smoking, for example by tying it to responsible behavior. In Germany, none of the antismoking campaigns has been very successful.⁴³ Thus, an antismoking norm never took off and even "price" increases

can only marginally change behavior (unless they are extreme), because the social meaning of smoking is ambiguous and there is no significant loss of reputation associated with it.

2.5 Equilibrium dynamics

In this section, the adjustment process leading to a new long-run equilibrium after a norm is regulated is analyzed in more detail. The equilibrium dynamics in terms of x_t are found by differentiating Equation (2) with respect to t and plugging in Equation (3):

$$\frac{\partial x_t}{\partial t} = \frac{\delta(x_t + r - \mu_t)(y + m)}{\mu_t^2} \quad \text{if } \mu_t > 0. \quad (8)$$

(If $\mu_t = 0$, then $x_t = 0$ is reached immediately by the definition in Equation (2) as $y + m \geq 0$.) The larger y , m , r , and x_t , the faster x_t changes over time; that is, for example, the higher the utility from the private good, the faster the norm erodes. Similarly, if the number of norm followers increases as $x_t + r > \mu_t$, this increase is faster the bigger y , m , r , and x_t .

For example, an antismoking norm will spread faster, the larger the individual utility from smoke-free air, the higher the "price" of smoking (including bans to smoke in public places etc.), and the more effective antismoking campaigns are.

Regulation of both motives and meaning can lead to a deterioration of the norm. However, the process leading to a long-run equilibrium is different under each policy. Modifying the motivational structure immediately changes the behavior of those individuals for whom breaking the norm becomes attractive although all norm breakers incur a disutility of $-a_i\mu$ from a loss of reputation. For example, when some members of a right wing youth group stop attacking foreigners out of fear to be convicted for it, they will lose social status within the group. Thus, when only the incentives are changed, a small δ means high reputational losses for the norm breakers as the norm disappears slowly.

On the other hand, transforming the meaning or social value of a norm only affects indi-

vidual behavior indirectly by reducing the reputational gain from following the norm and thus the proportion of believers.⁴⁴ The speed at which belief in the norm adjusts to actual behavior is parameterized by δ . If δ is small, regulation of meaning leads only to a slow change of norm-following behavior. But after a regulation of meaning individuals do not experience any disutility from social sanctions for breaking a still powerful norm as in the case of regulation of motives.

The optimal choice of the two instruments thus also depends on whether its effects should be instantaneous or whether long-term effects are sufficient. Consider the management of norms directed against foreigners in Germany. While more severe punishments for crimes with a xenophobic motivation change the behavior of people immediately, which may be desirable, education programs and symbolic acts by politicians have an effect only in the long run. On the other hand, in order to reduce smoking, long-run effects may be sufficient.⁴⁵

Place Fig. 3 here

Figure 3 illustrates the adjustment process after increasing m and decreasing r simultaneously, starting in the upper stable equilibrium. By introducing incentives to break the norm, the number of norm followers jumps to the left immediately, represented by the long horizontal arrow. In the long run, this leads to a deterioration of beliefs in and obedience to the norm, seen by the move along the short-run equilibrium curve according to the small arrows. The process ends at the stable point on the new long-run equilibrium line, which lies below and to the left of the stable point that would have been reached without a decrease in r .

2.6 Comparative statics (variation in m and r)

Investigating the marginal effects of both regulation techniques, complements the graphical analysis performed in the subsection above. Consider a situation where a norm is followed by the same proportion of people in every period, i.e., (a) a stable long-run equilibrium with $x^* > 0$ or, (b) a stable equilibrium with no norm-following behavior, $x^* = 0$.

Proposition 2 (a) *Creating incentives to break an existing norm, i.e., raising m marginally, decreases the stable long-run equilibrium number of norm followers and believers x^* and μ^* by the same amount,*

$$\frac{\partial \mu^*}{\partial m} = \frac{\partial x^*}{\partial m} = -[(1+r)^2 - 4(y+m)]^{-1/2} < 0. \quad (9)$$

(b) *Creating incentives to follow a new norm starting from $x^* = \mu^* = 0$ by lowering m leads to a new equilibrium at $(x^* = 1 - (y+m)\mu^* = 1)$ iff*

$$r \geq y + m. \quad (10)$$

Proof see Appendix. ■

As an increase in m results in a new stable equilibrium on the $\mu_t = x_t + r$ line, raising m has the same effect on the equilibrium number of followers and believers, x^* and μ^* . In the stable equilibrium in which nobody follows the norm, a marginal change of m may have no effect as the tipping point has to be reached to start the process of norm formation. And in particular, if $r = 0$, m has to be lowered to $y = -m$ such that all types of individuals are just indifferent between following and not following the norm at $\mu_t = 0$. Similarly, destroying the norm completely necessitates reaching the tipping equilibrium, i.e., a change in policies such that $\frac{1}{4}(1+r)^2 = (y+m)$.

Now consider the effect of transforming the meaning of a norm (a variation of r):

Proposition 3 (a) *Discrediting a norm, i.e., reducing r , decreases the stable long-run equilibrium number of believers μ^* more than the number of norm followers x^* as:*

$$\frac{\partial x^*}{\partial r} = -\frac{1}{2} + \frac{1}{2}[(1+r)^2 - 4(y+m)]^{-1/2}(1+r) > 0 \quad (11)$$

$$\frac{\partial \mu^*}{\partial r} = \frac{1}{2} + \frac{1}{2}[(1+r)^2 - 4(y+m)]^{-1/2}(1+r) > 0. \quad (12)$$

(b) *To establish a new norm, the raise in r must be sufficient to satisfy Inequality (10).*

Proof see Appendix. ■

Regulating the social meaning leads to a movement along the short-run equilibrium schedule and thus changes the proportion of believers to followers. The same tipping points as with the regulation of motives have to be reached in order to destroy an existing norm or create a new one.

Summarizing these results, both measures can be used to influence norm guided behavior. Regulation of meaning leads to a change in the equilibrium proportion of believers to followers, by influencing the beliefs rather than actual behavior. Regulation of motives leaves the equilibrium proportion unchanged. Going back to the example of smoking, banning smoking in public places increases the equilibrium number of non-smokers and of believers in the non-smoking norm by the same amount. Advertisement campaigns against smoking, on the other hand, lead to a stronger increase in the number of people believing in antismoking norms than to a decrease in smoking. If beliefs are important to the norm entrepreneur, for example, because she is worried about what people teach their children, regulation of meaning is relatively more effective.

Now consider the effect of a mix of both policies.

Proposition 4 *Regulation of motives and meaning are complementary measures to change norm-following behavior.*

Proof see Appendix. ■

The marginal effect of m on μ^* and x^* is larger, the bigger r is. If for example m is increased and r reduced to erode the power of the norm, the effect of motive regulation is enhanced by the regulation of meaning (and vice versa). Thus, the two policies are complements, which is illustrated in Figure 3. The optimal policy of a norm entrepreneur will include both measures, with the exact amount depending on their respective marginal costs.⁴⁶ For example, xenophobia in Germany is best addressed by both, heavier sentences for the culprits and symbolic acts of politicians expressing condemnation for the crimes. And according to the model, when politicians visit the site of the crime, this reinforces the effect of stricter sanctions (and vice versa). Similarly, smoking is best reduced by increasing its price (levying taxes on tobacco or banning smoking in public places) *together with* clever advertisement campaigns.

The question remains as to how general these comparative static properties are. This is investigated in the following section.

3 Generalized reputational value function

Some of the above results rely on the specific function for the reputational value of following a norm, $v(x_t, r) = x_t + r$. But many other specifications seem justifiable when thinking about examples in the real world. A simple modification is the case where the effect of regulating the meaning of a norm is proportional to the number of norm followers, e.g., $v(x_t, r) = x_t r$ with $r \geq 0$. Here, regulation of meaning has a greater effect, the more people follow the norm. But the reputational value can also be non-linear in x_t , for example, if there is a critical number of norm followers necessary for others to feel forced to follow the norm as well. Which of the comparative static properties described above rely on the particular linear additive specification? The general results are summarized in

Proposition 5 (a) *Creating incentives to break an existing norm by raising m as well as dis-*

crediting a norm by reducing r decreases the number of norm followers x^* in an interior stable equilibrium in the following way:

$$\frac{\partial x^*(m, r)}{\partial m} = \frac{-\frac{1}{1-x^*}}{\frac{y+m}{(1-x^*)^2} \cdot \frac{\partial v}{\partial x^*}} < 0 \quad (13)$$

$$\frac{\partial x^*(m, r)}{\partial r} = \frac{\frac{\partial v}{\partial r}}{\frac{y+m}{(1-x^*)^2} \cdot \frac{\partial v}{\partial x^*}} > 0. \quad (14)$$

(b) In an equilibrium with a high number of norm followers, i.e., where x^* is sufficiently close to 1, regulation of motives is more effective at the margin than regulation of meaning.

Proof see Appendix. ■

First, the signs of the marginal effects of regulation of motives and meaning derived for the special case in the last section hold in the general case. Second, regulating motives is more effective than regulating meaning in the limit when the proportion of norm followers approaches one. Of course, as the costs of both policies are not specified, this comparison cannot be made in absolute terms. However, the intuition for this result is straightforward: When almost everybody follows the norm, the critical individual (who is just indifferent) is rather insensitive to social pressure (i.e., a_c is small). Thus, changing the reputational value of the norm has little effect on the marginal individual's decision, and it is better to influence this individual via monetary incentives.

Consider a town where hatred against foreigners is widespread. Almost everybody believes in the norm that racist crimes should be tolerated or even encouraged. In such a case, the model predicts that more severe punishments for those who aid and abet verbal attacks or violence directed at foreigners will be quite effective in changing the behavior of critical individuals who are almost indifferent between following the norm or not following it. Their change in behavior then influences the beliefs of other individuals, thus weakening the norm. Politicians or other public figures expressing condemnation for the behavior prescribed by the norm will not have

an equally strong effect on norm-following behavior, because the critical people in a town where almost everybody follows the norm are those who do not care much about social pressure and reputation. (However, when the two measures are complements, they should both be used to weaken support for the norm and to eventually destroy it.)

Similarly, when almost everybody smokes and there is a norm that only smokers are cool, relaxed people, making smoking "costlier" (e.g., increasing taxes, banning smoking in restaurants) is more effective at the margin than antismoking campaigns. Higher costs immediately discourage those people from smoking who don't smoke because many other people smoke, but mainly because they get some utility from smoking itself. Those people don't care much about what others do, which is why it is not too unpleasant for them to stop smoking although almost everybody else continues to do so.

Two types of a reputational value function seem of special interest. I will analyze them in more detail below.⁴⁷

3.1 Bandwagon norms

Bandwagon norms are characterized by the fact that once a critical proportion of the population follows the norm, it becomes very costly in terms of reputation not to follow it, which is illustrated in Figure 4. For a large number of norm followers, there is a diminishing marginal effect on reputation. Similarly, as long as only a few people exhibit a certain behavior, there is no effect on others in terms of reputation losses for not doing the same.

Place Fig. 4 here

Remember that the shape of the long-run equilibrium schedule is determined by those (x_t, μ_t) -pairs for which $v(x_t, r) = \mu_t$. Thus, the reputational value and therefore the equilibrium number of believers displays a steep increase in a critical range of norm followers, x_t . It may be that this property describes norm-following behavior better than the linear case in most instances. It approximates the extreme description that a norm is either perceived as valid or not by (almost) everybody in the group. Or it can describe a situation in which people want to do what the majority does. Consider the example of antismoking norms which can exhibit critical mass effects. If the majority of people in a room or in a population smoke, nonsmokers will typically not try to stop them because almost nobody believes in the antismoking norm. However, when the proportion of nonsmokers exceeds a certain critical point, belief in the norm against smoking suddenly becomes very strong.

What are the marginal effects of norm regulation on bandwagon norms? First, consider an existing norm that is followed by almost everybody. Introducing incentives to break it has a strong effect on the number of norm followers as the schedule of the reputational value is almost flat in this upper portion. Thus, belief in the norm and its reputational value are barely affected by m , in contrast to the linear case. On the other hand, lowering the reputational value of the norm using r decreases x_t and μ_t along the short-run decision schedule, leading only to a small change in x_t . Thus, regulation of motives is more effective at the margin than regulation of meaning to change behavior.

Starting at the stable equilibrium with $x^* = 0$, regulation of motives to establish norm-following behavior by pushing the short-run decision schedule downwards must again be of sufficient size. But note that in contrast to the linear case, setting $v(0, r) \geq y + m$ may not be sufficient to get the bandwagon started. In particular, if $\frac{\partial \mu_t}{\partial x_t} > \frac{\partial v}{\partial x_t}$ for small x_t (i.e., the marginal increase in reputational value of one person following the new code is smaller than the private benefit from not following the norm), then it is necessary and sufficient for establishing a new norm that

$v(x_t, r) \geq \frac{y+m}{1-x_t}$ for all x_t , implying that more people have to follow the new code at once to start the dynamics of norm formation. This is important for decision makers who want to establish a new norm: If critical-mass effects are expected to play a role, it is necessary to change motives and meaning to a sufficient degree, otherwise the policies will have no effect at all.

In the case of new codes of ecological behavior, the monetary incentive to recycle empty bottles together with the increase in the reputational value of doing so (brought about by raising public awareness etc.) must be strong enough right away to set the process of norm formation in motion. For example, a German company recycling plastic, tin cans, etc. launched a big advertisement campaign in order to convince people to use special containers for these materials⁴⁸ and it lowered the cost of recycling by providing households with their own containers which are emptied every week by the company. Together with campaigns by Greenpeace and other NGOs as well as media coverage of the issue, a recycling norm was established.

3.2 Snob norms

A snob norm confers the highest reputation to its followers when the number of those following the code is small. Thus, the reputation gain from following the norm is negatively correlated with the number of followers once a certain number of followers x^s is reached. The reputational value of a snob norm peaks at this intermediate number of followers x^s as shown in Figure 5.

Place Fig. 5 here

An example for such a norm are group norms where certain actions qualify people as members of the in-group. If the group becomes too large, the social gains from adhering to it diminish. Different classes of society often develop such norms to differentiate themselves from others. For

example huge money donations lead to reputation gains because (among other reasons) only a very limited number of people engages in this behavior. The fashion world abounds with examples for snob norms as there is a constant creation of new trends that increase the reputation of its followers (i.e., wearers) only as long as there are not too many of them. Now consider the effect of the regulation of motives and meaning under this assumption.

Suppose that a snob norm exists and that it is followed by a stable proportion of people that is significantly larger than the critical number x^s , $x^* > x^s$. Introducing incentives for breaking it has the following effects: Although the number of followers is decreased by increasing m , the reputational value increases. This happens whenever the stable equilibrium is on the downward-sloping portion of the $v(x_t, r)$ schedule. A decrease in the number of followers benefits those adhering to the code by giving them an even higher reputation, which is the reverse of the linear case. Note also that whenever $x^* = x^s$, a slight change of incentives can destroy the norm completely because the number of norm followers becomes too small to exert any social pressure. A possible example is juvenile drug use. If using drugs confers social status or group membership, increasing the price of drugs or increasing the punishment decreases the number of drug users only by little, but increases the reputation gain of those still taking drugs. Only when the number of drug users can be reduced significantly such that it is below x^s , does the norm vanish.⁴⁹

What are the effects of transforming the social meaning of the snob norm in order to destroy it? A snob norm can be discredited such that the reputation gain from following it is lower for every proportion of norm followers (e.g., if $v(\cdot)$ is linear in r). This amounts to shifting the curve downwards and leads to a reduction in the long-run equilibrium proportion of norm followers. But just as with regulation of motives, this decline of norm-following behavior is small compared to the case of bandwagon norms *ceteris paribus*. This can be taken from Proposition 4 as the denominator of the marginal effects is larger for snob norms than for bandwagon norms, because $\partial v / \partial x_t < 0$. Thus, destroying a snob norm is difficult, but should in most cases not rely on

monetary incentives alone. The model suggests for example that peer pressure to take drugs among young people can often better be counteracted by information campaigns about the health hazards of taking drugs than by prohibiting them and/or raising their price.

How can a snob norm be established that is non-existent, for example a norm requiring rich people to make huge money donations to public schools? Again, introducing incentives to follow this norm (tax exemptions for example), i.e., decreasing m to push down the short-run equilibrium schedule, can be successful if it is done to a sufficient degree. Alternatively or as a complement to this measure, the norm can be established for example by Bill Gates announcing that he will give a big sum of money to public schools. This can increase the reputational value of contributing by raising $v(x_t, r)$ sufficiently. Then, even if almost nobody gave money to public schools before (x_t close to 0), some people start following his example and a norm develops.

4 Conclusions

The process of norm formation and erosion is a product of both collective action and individual attempts to challenge, support, discredit or introduce a certain code of behavior. This paper presents a simple model aimed at a better understanding of the interaction between these two elements of the norm dynamics. In particular, it is an attempt to endogenize norms by investigating how individual actors or groups of individuals (norm entrepreneurs) can influence the set of relevant norms in a society. Two different ways of norm regulation are introduced, namely changing the incentive structure and changing the social meaning and reputational value of a norm. A number of principles to guide decision makers concerned with norm management can be derived from the analysis.

A first simple rule states that the two instruments, the regulation of motives and of meaning, reinforce each other in many cases and should then be used simultaneously. Only if the comple-

mentarity is very weak and there are high fixed costs involved, can it be optimal to use a single instrument.

A further result of the paper concerns the question how the optimal mix of policies depends on how many people follow a certain norm. Consider the example of smoking norms. The model suggests that when the majority of people smoke, a change of incentives (like the prohibition to smoke in certain public areas or in restaurants) is most effective to discourage the critical individuals, i.e., those who are rather insensitive to social pressure and are most likely to stop smoking even if all others continue.

The main result of the paper is that the optimal policy mix depends on the type of the norm under consideration. If a norm belongs to the class of bandwagon norms, erosion of the norm can be initiated most powerfully by changing the incentive structure. But destroying snob norms by changing the relative prices can be very costly compared to regulating their meaning by advertisement campaigns, etc. Regarding the creation of new norms, the more important a critical mass as in bandwagon norms, the more people must be simultaneously convinced to adopt the new code. Thus, the measures for norm regulation must be of sufficient size. Creating snob norms on the other hand requires only a few individuals to adopt an action considered as praiseworthy. This can make it attractive for others to take the same action because a small number of norm followers confers high reputational gains.

A number of questions remain unaddressed. Among them is a full welfare analysis, incorporating the costs and benefits of regulating norms. A related question is how long a certain policy must be continued to establish or destroy a norm. Furthermore, it is conceivable that the assumption of full information concerning beliefs of others is often violated. Sometimes norms can persist because people think that almost everybody else believes in it and follows it, but when polls are published presenting the true figures of norm following behavior, the norm may unravel rapidly.

Appendix

Proof of Proposition 1 If $r \geq y + m$, the short-run equilibrium schedule is below the long-run equilibrium schedule at $x_t = 0$ and to the right of the long-run schedule at $\mu_t = 1$. As the short-run schedule is convex, it is everywhere below the long-run schedule and $\mu_t < x_t + r$ for all x_t . Thus, μ_t increases up to $\mu^* = 1$ with $x^* = 1 - (y + m)$. This proves the first part of the proposition.

Rearranging Equation (2) together with $\mu_t = x_t + r$ for x_t yields

$$x_t^2 + (r-1)x_t + y + m - r = 0.$$

Solving the quadratic equation leads to

$$x_{1,2}^*(m, r) = -\frac{r-1}{2} \pm \left[\left(\frac{r-1}{2} \right)^2 - (y + m - r) \right]^{1/2}.$$

Rearranging this yields

$$x_{1,2}^*(m, r) = \frac{1-r}{2} \pm \frac{1}{2} [(1+r)^2 - 4(y+m)]^{1/2},$$

which characterizes the two interior equilibrium points. Only the upper equilibrium is stable, thus

$$x^*(m, r) = \frac{1-r}{2} + \frac{1}{2} [(1+r)^2 - 4(y+m)]^{1/2}. \quad (15)$$

At $(1+r)^2 = 4(y+m)$ the two schedules are tangent and only one equilibrium exists. When $y+m = r$, the two curves intersect only at the corners. Thus, two interior intersections exist whenever $\frac{1}{4}(1+r)^2 > y+m > r$. This proves the second part.

If $\frac{1}{4}(1+r)^2 \geq (y+m)$, there is at most one point of tangency between both schedules. Thus, the short-run schedule lies everywhere above the long-run schedule, i.e., $\mu_t > x_t + r$, leading to decrease of μ_t and x_t until $(x^* = 0, \mu^* = v(0, r))$ is reached. This proves the third part of the proposition. ■

Proof of Proposition 2 (a) Partially differentiating Equation (15) with respect to m leads

to

$$\frac{\partial x^*}{\partial m} = \frac{1}{4}[(1+r)^2 - 4(y+m)]^{-\frac{1}{2}}(-4).$$

μ^* is derived by solving Equation (7) together with $x_t = \mu_t - r$. Analogous to the proof of Proposition 1, first rearrange the expression to

$$\mu_t^2 - (1+r)\mu_t + y + m = 0.$$

Solving this leads to

$$\mu_{1,2}^* = \frac{1+r}{2} \pm [(\frac{1+r}{2})^2 - y - m]^{\frac{1}{2}}, \quad (16)$$

where in the stable upper equilibrium, it must hold that

$$\frac{\partial \mu^*}{\partial m} = \frac{1}{4}[(1+r)^2 - 4(y+m)]^{-\frac{1}{2}}(-4).$$

As the term in brackets is positive in an interior equilibrium, the derivative is negative.

(b) See the proof of part 1 of Proposition 1. ■

Proof of Proposition 3 (a) Partially differentiating Equation (15) with respect to r leads to

$$\frac{\partial x^*}{\partial r} = -\frac{1}{2} + \frac{1}{4}[(1+r)^2 - 4(y+m)]^{-\frac{1}{2}}(2(1+r)),$$

which yields Equation (11).

Partially differentiating Equation (16) for the upper equilibrium with respect to m leads to

$$\frac{\partial x^*}{\partial r} = \frac{1}{2} + \frac{1}{4}[(1+r)^2 - 4(y+m)]^{-\frac{1}{2}}(2(1+r)),$$

which yields Equation (12).

(b) See the proof of part 1 of Proposition 1. ■

Proof of Proposition 4 The cross derivatives of Equation (15) and Equation (16) are

$$\frac{\partial^2 \mu^*}{\partial m \partial r} = \frac{\partial^2 x^*}{\partial m \partial r} = [(1+r)^2 - 4(y+m)]^{-3/2}(1+r) \quad (17)$$

for the interior stable equilibrium. As the term in brackets is positive when $x^* > 0$, the cross effect is positive. ■

Proof of Proposition 5 Equation (7) together with $v(x_t, r) = \mu_t$, which must hold in a long-run equilibrium, yields $v(x^*(m, r), r) \equiv \frac{y+m}{1-x^*(m, r)}$. Implicit differentiation of this identity with respect to m and r yields

$$\begin{aligned} \frac{\partial v}{\partial x^*} \frac{\partial x^*}{\partial m} &= \frac{1-x^* + (y+m) \frac{\partial x^*}{\partial m}}{(1-x^*)^2} \\ \frac{\partial v}{\partial x^*} \frac{\partial x^*}{\partial r} + \frac{\partial v}{\partial r} &= \frac{(y+m) \frac{\partial x^*}{\partial r}}{(1-x^*)^2}. \end{aligned}$$

Rearranging these equations leads to the marginal effects of m and r in a stable equilibrium with $x^* > 0$, Equations (13) and (14). Notice that the two terms in the denominator are just equal to the slopes of the short-run and the long-run equilibrium schedules. If the slope of the short-run equilibrium schedule is bigger than the slope of the long-run schedule, the denominator is positive, which is always the case as the short-run equilibrium schedule intersects the long-run schedule from below in any stable interior equilibrium. Thus, $\frac{\partial x^*}{\partial m} < 0$ for $x^* < 1$ and $\frac{\partial x^*}{\partial r} > 0$ as $\frac{\partial v}{\partial r} > 0$. This proves part (a) of the proposition.

Dividing Equation (13) by Equation (14) yields

$$\frac{\frac{\partial x^*}{\partial m}}{\frac{\partial x^*}{\partial r}} = \frac{1}{-\frac{\partial v}{\partial r}(1-x^*)}.$$

As the equilibrium proportion of norm followers x^* approaches 1, this expression approaches infinity, i.e., the effect of m on x^* gets infinitely bigger than the effect of r . This proves part (b) of the proposition.

Notes

* Thanks to Iris Bohmet, Wieland Müller, Jörgen Weibull, Georg Weizsäcker, Bengt-Arne Wickström, two anonymous referees, the editor, and participants of the workshop "Economics

and Social Norms” at Stockholm University. Financial support from Deutscher Akademischer Austauschdienst (DAAD) and Deutsche Forschungsgemeinschaft (SFB 373) is gratefully acknowledged. Part of the paper was written while I was visiting U.C. Berkeley in 1997/98 and KSG, Harvard University in 1999/2000. My thanks also go to the University of Bergen for its hospitality and to NFR-Ruhrgas for funding.

1 See for example the papers published in the *University of Pennsylvania Law Review*, 1996, vol. 144, and in the *Journal of Legal Studies*, 1998, vol. 27.

2 For an example of the evolutionary approach applied to norms see Young (1998) who shows that evolutionary forces may lead to norms implying efficient and more or less egalitarian outcomes. Huck (1998) studies the impact of legal rules on the evolution of preferences.

3 See Sunstein (1996b: 909) who coined this term.

4 There is a difference between government interventions to change the set of norms and actions of private parties. For the government, there is a “liberal constraint on social meaning making” (Lessig, 1995: 1016) as people have strong feelings against manipulations by the state, reminding them of totalitarian propaganda so pervasive in the past century. Therefore, the government is often more successful in conveying certain messages when it does so indirectly, e.g. by letting school teachers speak out against racial discrimination or doctors in favor of using condoms to prevent the spreading of AIDS. But a certain asymmetry in the evaluation of government actions can be observed: If the status quo or reigning orthodoxy is supported, there is not much resistance, but any attempt to challenge or even change the dominant view is seen as problematic. Also, corporate propaganda is ubiquitous (see Corneo and Jeanne (1997) who model a monopolist creating a “consumption norm” for his product). Although there is and should be a difference between the appropriateness of government and corporate strategies to influence social norms, it does not seem obvious why the government should have no role in the process at all. Sunstein (1996b) who defines and defends the role of law in “norm management” argues that individual

rights are constraints on norm management by the government.

5 For a discussion and critique of this issue see Basu, Jones, and Schlicht (1987). Schlicht (1998) provides an extensive discussion of changes of custom resulting from fuzziness of customs, self-interested behavior, and changes in the incentive structure.

6 Several mechanisms can bring inefficient norms into existence: Technological, economic or social change might render previously useful norms inefficient and efficient norms do not develop immediately, but often with considerable time lags involved (see Posner (1996) who discusses Demsetz' famous article about fur-trading native tribes in Canada and argues that the common pool problem due to trading furs arose much earlier than property rights developed, which led to a depletion of fur areas during that period). Also, there is no mechanism securing that an efficient norm develops whenever it is beneficial for all participants. Also, norms may be generated on the basis of certain principles such as salience, prominence or analogy which do not guarantee, but are even unrelated to their optimality (see Sugden, 1989). Moreover, if third parties who are affected by the norm do not participate in the norm-generating process, the resulting norm may be inefficient. For example, cartel rules or codes of business ethics usually do not reflect the interests of consumers (see Schäfer and Ott, 1993). Finally, McAdams (1997) argues that people seek the esteem of others and this drive for esteem produces both inefficient norms and needless conformity.

7 For an extended discussion along these lines see Elster (1989: 114) who argues for the "autonomy of norms".

8 For a defense of "norm management" by the government see Sunstein (1996b).

9 But norm regulation also takes place when interest groups try to change a code of behavior (e.g., by lobbying government agencies) that is disadvantageous to its members, even if the code is beneficial from a social welfare perspective. For these reasons, Posner (1998: 795) argues that the state should be endogenized in an analysis of norm management.

10 For a non-formal discussion of different techniques of norm regulation see Lessig (1995), Posner (1996), Sunstein (1996b), Posner (1998) and Cooter (1998). In addition to the two mechanisms discussed below, which Posner (1996) calls norm-violation and norm-transformation, he also describes norm-circumvention as bargaining around inefficient norms by private parties. The government can facilitate norm-circumvention by lowering bargaining costs, e.g., refining property rights. Lessig distinguishes between behavioral and semiotic techniques of norm change, which are then differentiated into the semiotic techniques of tying and ambiguation as well as the behavioral techniques of inhibiting and inducing a certain behavior. Sunstein discusses five tools the government can use to change norms: education, persuasion, taxes and subsidies, time, place, and manner restrictions as well as straightforward coercion. In the framework of his theory of social norms arising from signalling games, Posner (1998) mentions four different effects of legal interventions, namely changes in the costs of signalling and in the senders' and receiver's gains from cooperation as well as changes in beliefs about the distribution of types and the introduction of new signals. Cooter focuses on expressive law as a source of focal points when norms create multiple equilibria. If law succeeds in coordinating people, a move to a Pareto superior equilibrium is possible.

11 Hirschman (1985) argues that raising the cost of an action does not stigmatize it in the same way as making it unlawful. A law prohibiting an action without monetary sanctions would fall under the category of regulation of meaning (discussed in the next paragraph), whereas monetary sanctions change the incentive structure.

12 However, simple monetary incentives can also be ineffective. Consider the case where norms against genetically manipulated food exist, e.g., it is considered risky to give this food to children, although using the technology is legal. Suppose that prices for genetically manipulated food are lower, either due to subsidies or because of a more efficient production. Nevertheless, the existence of the norm and the slow process of norm change may cause losses to firms using the technology.

Finally, the norm may erode or the technology is driven out of the market.

13 See Sunstein (1996a) for a general discussion of the expressive role of law. In the context of norm regulation, Cooter (1998) argues that law can help to reach a tipping point by expressing certain values.

14 In German the slogan was "Samstags gehört Papi mir!"

15 Releasing information can also be very effective, because norms and statements of fact are closely related: When almost everybody smokes, it is hard to believe that smoking is very harmful. See Sunstein (1996b: 931) on this issue.

16 For a thorough discussion of these issues see McAdams (1997: 425ff.)

17 Certain sanctions for norm breakers can be interpreted as regulation of motives *and* of meaning as they affect both, incentives and reputation. Kahan (1996: 650) proposes to combine fines or community service with shaming to change the expressive role of these sanctions, i.e., when fines are tied to a punishment carrying the meaning of condemnation, both the price of an action and its reputational value are affected.

18 For an overview see Lessig (1995: 1025-34).

19 Sunstein (1996b) refers to "norm bandwagons" when a norm disappears very fast such that finally its followers are stigmatized for adhering to it.

20 This is the "intrinsic value of an option" (Sunstein 1996b: 916).

21 The main new feature of this particular functional form, introduced by Akerlof (1980) and later used by Romer (1984), Naylor (1989), and Corneo (1995) among others, lies in the direct effect of the proportion of believers μ on individual utility. For reasons of tractability, the formulation of the utility function assumes linearity in μ . The remaining arguments, y and m , are normalized such that the effects of changes in their values are comparable in size to changes in μ .

22 Note that making the utility dependent on the proportion of believers in the norm instead of on the proportion of norm followers allows for inertia in the model. When the behavior of

some people changes, norm following behavior of others does not change immediately, but only after the societal belief in the norm has changed as well. Moreover, distinguishing between the number of followers and believers allows for a richer description of the reputational value of norm following, which will be introduced below. However, for simplicity it is assumed that believers and non-believers have the same utility function. The model can be extended in a straightforward manner to allow for believers suffering more from breaking the code than non-believers, due to cognitive dissonance for example. See Naylor (1989).

23 Analogously, individual i decides to follow a new norm although it is costly to him or her if

$$y - a_i \mu + m < 0.$$

Thus, by decreasing m sufficiently, the norm entrepreneur can induce more types to follow the new code.

24 Social meanings are understood as "the semiotic content attached to various actions, or inactions, or statuses, within a particular context." See Lessig (1995: 951).

25 See Corneo and Jeanne (1997) for a discussion of the reputational value of a norm.

26 See Akerlof (1980) and Corneo and Jeanne (1996) among others for a similar specification of the dynamics.

27 This specification of the dynamics implies that the adjustment process slows down, the more believers there are as $\partial(\partial\mu/\partial t)/\partial\mu < 0$. Thus, some people are harder to convince that a norm is valid even if its reputational value is high.

28 The assumption of exogeneity is made for the sake of simplicity. However, δ could be endogenized to account for norm management that influences how fast information about norm-following behavior spreads. This could be addressed by future attempts to model norm regulation.

29 See McAdams (1997: 368) who argues that this "feedback effect" can lead to rapid changes in norm compliance. However, this effect does not necessarily apply to snob norms: When more

people follow a snob norm, this does not necessarily imply a higher reputation gain from following it as it is possible that $\partial v/\partial x < 0$ (see Section 3.2). Nevertheless, the model is consistent with the main features of McAdams' "esteem theory of norms" (1997: 355) according to which a norm persists when the loss of esteem is greater than the private benefit of acting against the consensus in the population.

30 Of course, the effort to transform a norm may also be resisted, e.g., because people believe that a certain issue should not be regulated by the government or, as Sunstein (1996b: 919) puts it, because "the source of the effort is disqualifying". Also, the norm entrepreneur may not have convincing arguments in his campaign. In this case, the policy has either no effect or is even counterproductive, $0 \geq \frac{\partial v}{\partial r}$. For a different and more detailed description of the unpredictability of the law's effect on behavior and beliefs see Posner's (1998) signalling model of norms.

31 As x_t necessarily lies between 0 and 1, the variable r is normalized for this property to hold.

32 Moreover, for $1-x_t > y+m$ the short-run equilibrium curve is convex as

$$\frac{\partial \mu_t}{\partial x_t} = \frac{y+m}{(1-x_t)^2} > 0 \quad \text{and} \quad \frac{\partial^2 \mu_t}{\partial x_t^2} = \frac{2(y+m)}{(1-x_t)^3} > 0 \quad (18)$$

for $x_t \in [0, 1)$.

33 For a similar treatment of the impact of incentives on social customs see Schlicht (1998: 46-49).

34 Posner (1998: 785) argues that when black migration threatens the position of whites in the labor market, discrimination against blacks signals loyalty to each other.

35 For example, when a house for asylum seekers was put to fire in Rostock-Lichtenhagen by young right wing radicals in August 1992, the crowds watched the fire cheering and shouting "Deutschland den Deutschen" (Germany to the Germans). For several days the police did not stop the aggressions and the riots and violence went on.

36 The German courts have only recently started to investigate more carefully whether certain

crimes have a political background, leading to heavier sentences of right wing radicals. It is also discussed whether the Criminal Code should be changed to allow for longer terms of imprisonment in severe cases.

37 Sunstein (1996b) calls this tool to change norms "time, place, and manner restrictions".

38 For example in Berlin, smoking was banned in subway stations in 2000.

39 Notice that this condition for a norm to evolve differs from the condition for an existing norm to be destroyed. I am assuming that in Germany neither a smoking norm nor an antismoking norm exists.

40 For a more general discussion of moral suasion and education as instruments of economic policy see Frey and Kirchgässner (1994: 403).

41 On the other hand, some politicians have argued and still argue in public that Germany should reduce the number of asylum seekers it admits into the country claiming that there are already so many that the country cannot cope with them any more. Political slogans like "The boat is full." ("Das Boot ist voll.") can be interpreted as a (partial) vindication of xenophobic acts. This is an example of norm entrepreneurs rationalizing the use of discrimination, see Posner (1998: 786).

42 Sunstein (1996b) argues that the striking difference between smoking habits among white teenagers and African-American teenagers in the United States may be partly due to a private advertisement campaign. For example in subways in Harlem the slogan was posted "They used to make us pick it. Now they want us to smoke it." (1996b: 906)

43 In the 1980's, posters saying "Who likes to kiss nicotine?" ("Wer küsst schon gerne Nikotin?"), printed by health insurance companies, were ridiculed for their tacky design and because health insurances were not considered a good source of advice in questions of friendship and love.

44 For a related analysis of the direct and indirect effects of legal institutions see Huck (1998).

In an evolutionary framework he shows that punishments not only deter a certain action in the short run, but can also have long-run effects. For certain punishment levels, types who feel remorse are materially better off than others. Thus, the evolutionary process will lead to an increase in the proportion of types feeling remorse.

45 Of course, the model does not specify how long one instant t is and how fast adjustment takes place, measured by δ . Nevertheless, it follows from the model that regulation of motives affects behavior faster than regulation of meaning.

46 Only if the fixed cost of one instrument is high, it might be optimal to employ only the other. For example, regulation of meaning can require high up-front investments (devising a new advertising campaign or writing new schoolbooks), such that it is better to regulate only the incentives.

47 I owe the general ideas of this section to the analysis of externalities in consumption by Leibenstein (1950) and Corneo and Jeanne (1997).

48 The slogan on posters showing a bottle opener or a toy made out of tin was: "I used to be a can." ("Ich war eine Dose.")

49 Another example is dueling, which is governed by a complicated set of norms and described as a practice that served to distinguish members of the upper class from others. Lessig (1995) argues that attempts by non-members of the elite to ban dueling increased solidarity and support for dueling in the American South.

References

- Akerlof, George. 1980. "A Theory of Social Custom, of Which Unemployment May Be One Consequence," 94 Quarterly Journal of Economics 49–75.
- Basu, Kaushik, Jones, Eric, and Schlicht, Ekkehart. 1987. "The Growth and Decay of Custom: The Role of the New Institutional Economics in Economic History," 24 Explorations in Economic History 1–21.

- Cooter, Robert D. 1996. "Decentralized Law for a Complex Economy: The Structural Approach to Adjudicating the New Law Merchant," 144 University of Pennsylvania Law Review 1643–96.
- Cooter, Robert D. 1998. "Expressive Law and Economics," 27 Journal of Legal Studies 585–608.
- Cooter, Robert D., and Thomas Ulen. 1997. Law and Economics (2nd ed.), Reading, MA: Addison–Wesley.
- Corneo, Giacomo. 1995. "Social Custom, Management Opposition, and Trade Union Membership," 39 European Economic Review 275–92.
- Corneo, Giacomo, and Olivier Jeanne. 1997. "Snobs, Bandwagons, and the Origin of Social Customs in Consumer Behavior," 32 Journal of Economic Behavior and Organization 333–47.
- Elster, Jon. 1989. "Social Norms and Economic Theory," 3 Journal of Economic Perspectives 99–117.
- Frey, Bruno, and Gebhard Kirchgässner. 1994. Demokratische Wirtschaftspolitik. München: Vahlen
- Hirschman, Albert. 1985. "Against Parsimony," 1 Economics and Philosophy 7–21.
- Huck, Steffen. 1998. "Trust, Treason, and Trials: An Example of How the Evolution of Preferences Can Be Driven by Legal Institutions," 14 Journal of Law, Economics, and Organization 44–60.
- Kahan, Dan M. 1996. "What Do Alternative Sanctions Mean?" 63 The University of Chicago Law Review 591–653.
- Kamiat, Walter. 1996. "Labor and Lemons: Efficient Norms in the Internal Labor Market and the Possible Failures of Individual Contracting", 144 University of Pennsylvania Law Review 1953–70.
- Leibenstein, Harvey. 1950. "Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand," 64 Quarterly Journal of Economics 183–207.
- Lessig, Lawrence. 1995. "The Regulation of Social Meaning," 62 The University of Chicago Law Review 943–1045.
- McAdams, Richard H. 1997. "The Origin, Development, and Regulation of Norms," 96 Michigan Law Review 338–433.
- Naylor, Robin. 1989. "Strikes, Free Riders, and Social Customs," 103 Quarterly Journal of Economics 771–85.
- de Neubourg, Chris, and Maarten Vendrik. 1994. "An Extended Rationality Model of Social Norms in Labour Supply," 15 Journal of Economic Psychology 93–126.
- Posner, Eric A. 1996. "Law, Economics and Inefficient Norms," 144 University of Pennsylvania Law Review 1697–1744.

- Posner, Eric A. 1998. "Symbols, Signals, and Social Norms in Politics and the Law," 27 Journal of Legal Studies 765–98.
- Romer, David. 1984. "The Theory of Social Custom: A Modification and Some Extensions," 99 Quarterly Journal of Economics 717–27.
- Rock, Edward B., and Michael L. Wachter. 1996. "The Enforceability of Norms and the Employment Relationship," 144 University of Pennsylvania Law Review 1913-52.
- Schlicht, Ekkehart. 1998. On Custom in the Economy. Oxford: Clarendon Press.
- Schäfer, Hans-Bernd, and Claus Ott. 1993. "Emergence and Construction of Efficient Rules in the Legal System of German Civil Law," 13 International Review of Law and Economics 285–302.
- Sugden, Robert. 1989. "Spontaneous Order," 3 Journal of Economic Perspectives 85–97.
- Sunstein, Cass R. 1996a. "On the Expressive Function of Law," 144 University of Pennsylvania Law Review 2021–53.
- Sunstein, Cass R. 1996b. "Social Norms and Social Roles," 96 Columbia Law Review 903–68.
- Young, H.Peyton. 1998. "Social Norms and Economic Welfare," 42 European Economic Review 821–30.

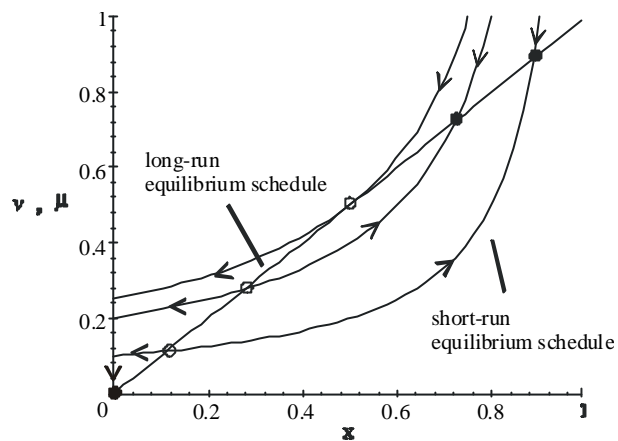


Figure 1: Equilibria and regulation of motives. Full circles represent stable equilibria, empty circles unstable equilibria.

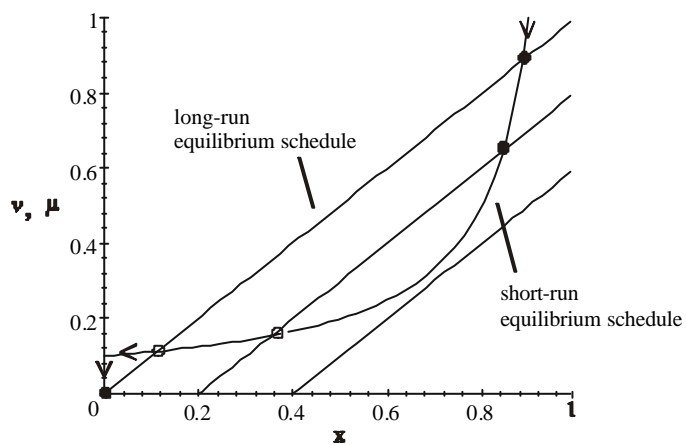


Figure 2: Equilibria and regulation of meaning. Full circles represent stable equilibria, empty circles unstable equilibria.

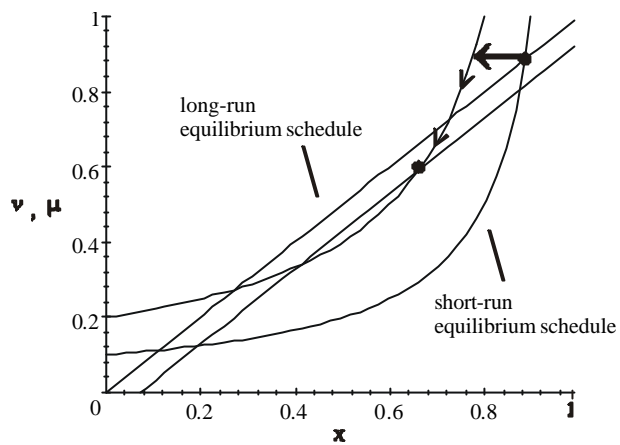


Figure 3: Adjustment process, cross effects. Full circles represent stable equilibria.

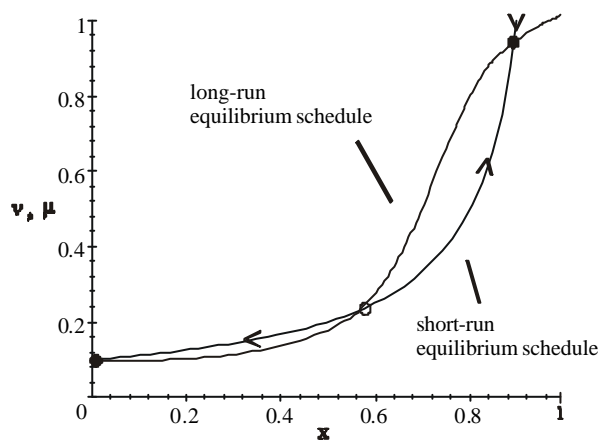


Figure 4: Bandwagon norms. Full circles represent stable equilibria, empty circles unstable equilibria.

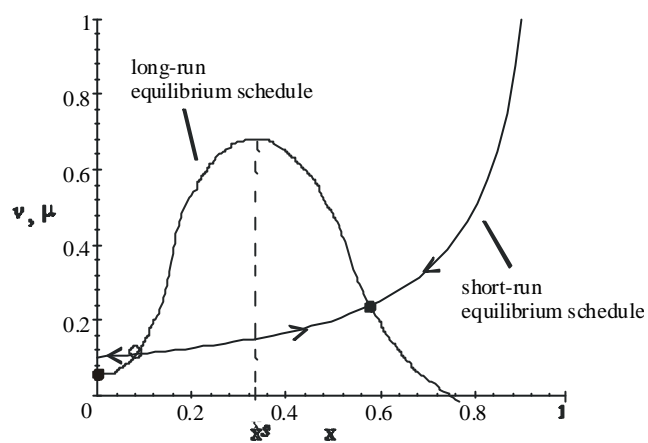


Figure 5: Snob norms. Full circles represent stable equilibria, empty circles unstable equilibria.