Comput Math Organ Theory (2010) 16: 1–28
DOI 10.1007/s10588-009-9061-7

Does ignorance promote norm compliance?

Patrick Groeber · Heiko Rauhut

Published online: 15 December 2009

Abstract A large extent of undetected norm violations may have positive effects for society. If many norm violations are hidden, society seems to be in good order so that actors are more willing to comply with the norms themselves. In this sense, ignorance promotes norm compliance. We challenge this view by arguing that in scenarios, in which norms are controlled and enforced by third parties who receive rewards for their success, the opposite is true: Ignorance promotes norm violations. The reason is that unsuspicious inspectors who believe that little hidden norm violations are committed will spend less effort for detection, some formerly detected norm violations will go undetected, norm targets will be less deterred from the lower detection probability and will commit more norm violations over time. This article develops a respective mathematical model and confirms the above described intuition.

Keywords Social norms · Dark figure · Self-fulfilling prophecy · Crime · Control · Punishment

1 Introduction

Imagine, everybody would be detected for every deviant act, for every peccadillo, for every foolish lapse. While most of us probably like to know about the slips and misconduct of our fellow men, we prefer to keep our own secrecies better for us so that a dark field of norm violations prevails. Individually, it may be tempting to be in

P. Groeber
Chair of Systems Design, ETH Zurich, Kreuzplatz 5, 8032 Zurich, Switzerland
e-mail: pgroeber@ethz.ch

H. Rauhut (✉)
Chair of Sociology, in Particular of Modeling and Simulation, ETH Zurich, Universitätsstrasse 41, 8092 Zurich, Switzerland
e-mail: rauhut@gess.ethz.ch

the position to know every indecency; however, it is open to research to which extent societies should try to uncover every norm violation.

This detection problem is crucial in the context of many norm violations like tax evasion, doping in sports or fare dodging. Further, its importance is increasing due to the increasing establishment of high tech control technologies in modern societies such as CCTV video systems, medical detection methods for drug use in professional sports, detection technologies for weapons at airports, electronic finger prints or data mining methods to trace criminals. Although there have been debates in social philosophy about such changes towards a "control society" in modernity (cf. Foucault 1977; Cohen 1985; Garland 2001; Hudson 2002), the theoretical understanding of the underlying mechanisms is still weak.

1.1 Why ignorance can promote norm compliance

While the missing knowledge of the actual extent of norm violations may appear as a nuisance at first sight, it can have positive effects for society. Popitz (1968) published a meanwhile classical essay on this argument by pointing out that ignorance may lead to norm compliance. If norm violations are hidden, society seems to be in good order and actors are willing to adhere to these norms as well. Loosely based on the saying "what the eye does not see, the heart does not grieve over", the dark field of norm violations has a preventive effect.

The early conjecture of Popitz (1968) has been analyzed in more detail in the literature on herding behavior in social groups. This literature analyzes how actors may draw wrong conclusions from public behavior to the unknown extent of this behavior in private and the respective beliefs behind. Such misperceptions may set off dynamics between public behaviors and private beliefs such that a large unknown extent of norm violations discourages actors to violate the social norm themselves. The herding literature in economics primarily analyzes *informational cascades* (Banerjee 1992; Bikhchandani et al. 1992). Here, widespread uncertainty leads people to follow others due to the false assumption that these others have more accurate information. Social psychologists primarily studied *pluralistic ignorance*, which explains herding behavior by the agents' false assumption that most others will approve of what the majority publicly complies with.[1] In simpler words, informational cascades describe the false assumption that everybody knows, pluralistic ignorance, that everybody believes.

Follow-up work analyzed the combination of herding with punishment. Punishment can even provoke dynamics which result in wide-spread normative compliance with largely unpopular norms (Bicchieri and Fukui 1999). A large potential of unknown, private norm violations can urge individuals to signal their sincerity by punishing others. This punishment may, however, pressure many individuals to comply and even to enforce such social norms which they privately disapprove. Centola et al. (2005) specified this idea in an agent based model and Willer et al. (2009) tested it experimentally. More specifically, the idea conceptualizes *punishment* as a *signal of sincerity* of "posers" who privately oppose but publicly comply with the norm.

[1] See Allport (1924), Miller and McFarland (1987) and O'Gorman (1986) for a discussion of the theoretical arguments and Kitts (2003) and Miller and Morrison (2009) for empirical studies.

Their punishment can signal their credibility and trustworthiness as a true believer; however, it pressures others to comply with the norm as well. As a consequence, the punishment mechanism can amplify the above described herding dynamics. While a large extent of unknown norm violations may reinforce norm compliance, the additional effects of punishment may even reinforce norm compliance of largely unpopular norms.[2]

1.2 Why ignorance can promote norm violations

While the described literature above confirms the notion that ignorance promotes norm compliance, we challenge this view. The underlying mechanism of the previous literature on herding implies that agents' misperception of the actual extent of norm violations spreads in their social group and affects others' norm compliance. Our research question, however, raises the issue as to whether these dynamics change for groups with competing interests. If two groups have opposing incentives, social influence among both groups may neutralize each other. We think of professionalized norm control personnel like police-men, conductors, guards, nichtwatchmen or doping testers. Such inspectors have incentives to perform successful controls while the norm targets have the competing incentive to commit successful norm violations. Our claim that the results will change for competing groups is based on game theoretical reasoning on zero-sum games. We know from theoretical (Tsebelis 1989, 1990) and empirical analyses (Rauhut 2009a, 2009b) of inspection games that punishment effects become inverted if there is competition between norm targets and inspectors.[3]

Our proposed model, therefore, analyzes situations in which norm targets and inspectors have competing interests and form beliefs about the unknown extent of norm violations. Both parties apply these beliefs to meet their decisions to commit a norm violation or to perform inspections. The dynamics consist of the following repulsive forces: On the one hand, if norm targets believe in normative compliance of other norm targets, this may reinforce their normative behavior. On the other hand, if inspectors believe likewise in norm compliance of norm targets, they will reduce their inspection activities. This, however, may reduce the detection probability of norm targets and, in turn, result in increasing norm violations over time. In the reversed case, if norm targets belief in a large extent of hidden norm violations, they may become infected to commit more norm violations themselves, which, however, incites inspectors to inspect more extensively, finally resulting in less norm violations over time. Thus, we expect predator-prey like dynamics, as Rauhut and Junker (2009) have demonstrated for simpler inspection scenarios between norm targets and inspectors. Our aim is to predict the resulting stable states of these dynamics in society: Does the belief in others' norm compliance reinforce normative behavior (self-fulfilling prophecy) or does it even result in less normative behavior (inverse self-fulfilling prophecy)?

[2]For a related simulation model of the emergence of unpopular norms see (Kitts 2006, 2008).

[3]For analyses of combined scenarios of centralized (formal) and decentralized (informal) social control of norm violations see Bendor and Mookherjee (1987); Kitts (2006).

1.3 Intuitive description of our model and main results

The remainder of the article and our main results are as follows. In a first step, we develop a mathematical framework for modeling the interactions between norm targets and inspectors. We assume that norm targets can decide to commit a norm violation or not. If they decide to violate the norm, they can spend more or less effort to conceal their norm violation from the inspector. The inspector, on the other hand, can spend more or less effort to reveal the norm violations of the norm targets. If a norm violation of a norm target is detected, the norm target receives a punishment while the inspector a reward.

Our first analysis describes the simplified situation that norm targets and inspectors know the actual extent of norm violations. The comparative statics of the model return results which are intuitive and therefore confirm the plausibility of our general model: Higher benefits for norm violations increase the rate of norm violations and its detection rate. In contrast, higher benefits for successful detections and higher punishments decreases the rate of norm violations.

Our second analysis extends the model to analyze the core research question of how ignorance in the sense of unknown norm violations affect inspections and norm violations. We differentiate between the detected and the overall rate of norm violations, while the overall rate consists of the detected and the undetected rate of norm violations. Our key parameter is λ, which is the constant suspiciousness of the agents, describing whether they believe that there is a small or large extent of undetected norm violations in the population. The smallest level of suspiciousness would be that agents believe that the detected rate of norm violations is similar to the overall rate of norm violations. The largest level of suspiciousness describes the agents' belief that all norm targets commit a norm violation although only some of them have been detected. Thus, our model specifies the belief λ of the agents in the range between the rate of detected and the total number of possible norm violations. We find that a higher suspiciousness of inspectors always decreases the overall rate of norm violations. Further, in most cases a higher suspiciousness of norm targets decreases the overall rate of norm violations as well. We conclude that the belief that many others commit hidden norm violations will *decrease* the actual rate of norm violations in the population. The reason is that suspicious inspectors will spend more effort to detect norm violations, formerly hidden norm violations will be detected and norm targets will be deterred to commit norm violations over time.

Our analysis suggests that it depends on the enforcement mechanism in the group whether ignorance promotes norm compliance. If the norms are enforced by group members themselves, ignorance seems to promote norm compliance; however, if the norms are enforced by third parties who receive rewards for their success, ignorance promotes norm violations.

2 The model

In our model, the population consists of n norm targets (agents) and one inspection institution, called the inspector.[4] We assume that at each time step t, a norm target's strategy $x_i(t) = (d_i(t), e_{a_i}(t))$ consists of two components: First there is a binary choice between violating the norm ($d_i = 1$), and adhering to the norm ($d_i = 0$). We refer to d_i as the *behavior* of norm target i. Second, $e_{a_i} \in [0, \infty)$ captures the norm target's expenditures on concealing her fraud. Here, a rational individual's choice of adhering to the norm implies no expenditures on concealing the norm violation, i.e. $e_{a_i} = 0$ if $d_i = 0$. The inspector has no option to refrain from inspection. Hence, her strategy is restricted to her expenditures on control effort $e_c(t) \in [0, \infty)$. An increase of control effort could reflect a larger sample of inspected norm targets or more sophisticated controls. In our model we assume that detection of norm violation is neither guaranteed nor occurs with a fixed probability. Instead, the detection probability depends on the interaction between the norm targets and the inspector: For a norm violator i ($d_i(t) = 1$), the probability of being detected by the inspector at time t is

$$p_i\big(e_{a_i}(t), e_c(t), q(t)\big) \tag{1}$$

where $q(t) = \frac{1}{n}\sum_i d_i(t)$ denotes the proportion of norm violators among the norm targets. Hence, the detection probability depends on the strategy of both parties: It varies with the norm target's concealment effort and the inspector's detection effort. In addition, the detection probability may also depend on the overall strength of the norm captured by how many norm targets adhere to it, for example in the case of *indirect detections*. By this we mean that in addition to a direct detection of norm target i's norm violation, it is also possible to discover indirectly the norm violation by the detection of a distinct norm target j. The reason is that i and j might have a common supplier of instruments for norm violation or concealment, e.g. a common provider of drugs in case of doping or respective accounts for illegal earnings at a common bank.[5]

For norm target i, the expected utility of a strategy $x_i = (d_i, e_{a_i})$ is defined by

$$u_a(x_i) = d_i\big(b_a - sp(e_{a_i}, e_c, q) - e_{a_i} - c_i\big) - rb_a q \tag{2}$$

for fixed control effort e_c and fixed proportion of norm violators q. Here, $b_a > 0$ denotes the benefit of norm violation if the norm target is not detected. Norm violation imposes a disadvantage on all norm targets numbered by $rb_a q$ whereas the factor $r \geq 1$ measures the extent of welfare loss for the population caused by norm violation and corresponds to the multiplier for public contributions in the standard public goods game. Thus, the only Nash equilibrium is the state where all norm targets violate the norm. As norm violation decreases welfare, this is a dilemma scenario and there is a

[4] Although we consider one singular inspection institution, it may consist of several inspectors. All inspections actions refer to the whole corporate actor, i.e. the inspection institution.

[5] An example are the occurrences in the run-up to the Tour de France in 2006 where many cyclists were suspended after a physician was accused by the police of conducting autologous blood transfusions.

demand to enforce the norm.[6] $s > 0$ is the punishment imposed on a norm violator when her fraud is detected.

Additional to the expenditures e_a for the concealment effort, a norm violator i receives a disutility $c_i > 0$ which can be interpreted as *moral costs*. This disutility is randomly assigned to each norm target according to a distribution function F and is therefore not explained by the model. c_i reflects norm target i's initial adherence to the norm and does not change over time. It captures the extent to which the benefit of norm violation has to exceed the expenditures on concealment and the expected sanction cost if norm violation is detected for norm target i. For example, some athletes know that doping would pay off but have internalized the anti-doping norm to an extent that makes them refrain from drug use. Similarly, there are taxpayers who do honestly report their income despite considerable economic incentives for norm violation (Andreoni et al. 1998, p. 822). Note that by these assumptions, an increase in opportunity costs for adhering to the norm (e.g. by reduced control effort leading to a decrease of detection probability for all norm targets) leads to an increase of norm violation in the population. Therefore, the norm targets' behavior is in accordance with the *low-cost hypothesis* (cf. North 1986; Diekmann and Preisendörfer 2003; Rauhut and Krumpal 2008) which postulates that the congruence of the acceptance of a social norm with the corresponding normative behavior decreases with increasing costs.

In our model, the inspector's strategy only consists of the control effort e_c. We assume that she gets a reward for each detected norm violator which constitutes an incentive for wide controls. Nevertheless, high expenditures on control effort will only pay off if the proportion of detected norm violators is sufficiently high so that the inspector's utility also depends on the actual proportion of norm violators. Hence, for fixed concealment efforts e_{a_i} and a proportion of norm violators q, the inspector's expected utility is defined by

$$u_c(e_c) = b_c\left(\frac{1}{n}\sum_{d_i=1} p(e_{a_i}, e_c, q)\right) - e_c. \tag{3}$$

Hence, the inspector's reward is proportional to the expected proportion of detected norm violators denoted by

$$\tilde{q} = \frac{1}{n}\sum_{d_i=1} p(e_{a_i}, e_c, q). \tag{4}$$

The reward for a detected norm violator is therefore $\frac{b_c}{n}$. We can interpret u_c as the variable part of the inspector's income whereas the fixed part does not influence the control effort. In case that all norm targets violate the norm and are detected, the inspector achieves the maximum reward $b_c > 0$. As $u_c(0) \geq 0$, i.e. zero control effort always implies non-negative utility for the inspector, we obtain $u_c(e_c^*) \geq 0$ for the utility maximizing effort e_c^*. Hence, the inspector's expenditures on control effort

[6] One might claim that norm violation can also lead to a welfare increase for the norm targets (e.g. induced by $r < 1$). With respect to doping, extraordinary achievements enhanced by drug use might lead to an increase of the aggregated income of all athletes by additional advertising revenue.

will not exceed the expected reward $\tilde{q}\, b_c$ if she maximizes her expected utility. This can be interpreted as an implicit budget constraint for the inspector.

With respect to the norm targets' and the inspector's knowledge, we make the following assumptions:

(K1) The utility functions u_a of a norm target and u_c of the inspector and the parameters b_a (benefit of norm violation), s (sanction cost) and b_c (benefit of detection) are common knowledge.
(K2) The moral costs c_i of norm violation are private information of norm target i. Further, neither any norm target nor the inspector know the distribution function F of these costs.
(K3) The proportion $\tilde{q}(t)$ of detected norm violators at time t is common knowledge at time $t+1$.
(K4) The proportion $q(t)$ of norm violators at time t *cannot* be observed by any norm target or the inspector at time $t+1$ (or any other timestep).
(K5) The are no false-positive detections of norm violators by the inspector.

We assume that norm targets and the inspector maximize their respective utility given the above knowledge constraints. All strategies, i.e. each norm target's behavior and concealment effort on one hand and inspector's control effort on the other, are chosen simultaneously at all time steps and cannot be observed afterwards. Due to our assumptions, the interaction between the norm targets and the inspector is partly strategic. For example, the norm targets know that the inspector's benefit increases with the proportion of detected norm violators and account for this knowledge in their decision-making. On the other hand, the inspector and other norm targets are not able to foresee whether norm target i violates the norm or not for given strategies of the other norm targets and the inspector as c_i is unknown. Further, a prediction of norm violation is not only impossible on the individual level—also the overall extent of fraud captured by the proportion q of norm violators in the population cannot be foreseen by the norm targets and the inspector as they do not know F. But as the inspector's benefit and possibly also the detection probability depend on q, we assume that both parties have to estimate the proportion of norm violators $q(t)$ at time t and use it within their decision-making process at the next time step. We denote these estimations by $\hat{q}_{a_i}(t)$ (norm targets) and $\hat{q}_c(t)$ (inspector).

If the distribution function F of the moral costs was common knowledge, we could formulate our model as a Bayesian game whereas F represents the norm targets' and the inspectors prior beliefs about each norm target's moral cost c_i. According to Harsanyi (1967–1968), we could reformulate this game of incomplete information as a game of imperfect information and apply standard equilibrium concepts (see Fudenberg and Tirole 1991, Chap. 8). However, as we cannot identify specific prior beliefs in tax evasion or doping, we argue that our behavioral approach is more suitable to model the norm targets' and the inspector's decision-making. Further, our results do not depend on the actual shape of F.

The estimations of the real proportion of norm violators, which are then plugged in the utility functions to determine the resulting optimal decisions, are based on the proportion of *detected* norm violators $\tilde{q}(t)$ at time t that is common knowledge by (K3). This degree of information resembles real world examples: Although nobody

knows the extent of tax evasion or doping in the population, actors will have vague or intuitive beliefs resulting from common knowledge about how many taxpayers or athletes have been detected as norm violators in a certain period.[7] Note that as we excluded false positive detections by (K5), a high proportion of detected norm violators $\tilde{q}$ induces more information about the real proportion of norm violators q compared to a low proportion $\tilde{q}$. The reason is that a low value of $\tilde{q}$ might on the one hand result from a low value of q. On the other hand, we may observe the same proportion of detected norm violators if there is a high proportion of norm violators paired with a low detection probability which can result from a high concealment effort or a low inspection effort. As $\tilde{q}$ is only a lower bound for q, all estimations $\hat{q}_{a_i}(t)$, $\hat{q}_c(t)$ have to fulfill

$$\tilde{q}(t) \leq \hat{q}_a(t), \hat{q}_c(t) \leq 1. \tag{5}$$

In general, an actual estimation by the norm targets or the inspector might depend on many factors. In Sect. 3.2 we specify a simple estimation procedure that combines objective information ($\tilde{q}$) with an individual's (external) subjective belief about the extent of norm violation in the population.

Overall, we developed a model to analyze how the dark figure of norm violation influences the interaction between the norm targets and the inspector assuming that norm violation is only detected with a certain probability depending on the two parties' respective effort. Both parties act partly strategically in the sense of knowing and considering the incentives of all involved individuals. However, they are only boundedly rational as certain parameters of their utility function are unobservable (the proportion of norm violators q) or private information (the moral cost c_i). In the following section, we specify further assumptions on the probability of detection ((P1)–(P7)) and the estimation procedure ((E1)–(E4)) and analyze the resulting dynamics. First, we investigate a scenario where the proportion of norm violators is observable at any time step. In Sect. 3.2 we analyze how the dynamics change if there is a dark field of norm violators, i.e. the proportion of norm violators is not observable.

3 Analysis

We model the interaction between the norm targets and the inspector as follows: At each time step, every norm target i chooses a strategy $x_i(t) = (d_i(t), e_{a_i}(t))$ where e_{a_i} denotes the concealment effort and d_i indicates whether norm violation pays off ($d_i = 1$) or not ($d_i = 0$). Simultaneously, the inspector chooses the control effort $e_c(t)$. Both parties maximize their respective expected utility according to (2) and (3)

[7] One may make the objection that detected deviance is not common knowledge (i.e. I know that you know that I know that you know the observed level of deviance in the previous discrete time step). While objective statistics may indeed exist for the level of doping or tax fraud in the real world, it may be practically hard to access this information and to remember it correctly. We argue that this simplified assumption is helpful, because the average belief of the rate of detected deviance may still resemble the actual rate with some random noise around.

Table 1 All parameters in the model. The first two sections describe the norm targets' payoff for violating and the inspector's payoff for inspecting the social norm. The third section describes the parameters referring to the detection problem. Some of these variables are introduced later in the text. The key variable is λ, which is a constant personality variable describing the belief of the extent of undetected norm violations. It describes whether the agents believe in a low extent of undetected norm violations (close to $\bar{q}$) or in a large extent of undetected norm violations (close to 1, meaning all norm targets violate the norm). The last section refers to Example 1 and summarizes the parameters of the exemplary functional form of the detection probability

Norm targets	
d	behavior (0 = norm adherence, 1 = norm violation)
b_a	benefit from norm violation
e_a	effort to conceal norm violation
s	punishment for norm violation
c	moral cost for norm violation
F	distribution of moral cost in population
r	welfare loss due to norm violations
u_a	expected utility of norm targets
Inspector	
b_c	benefit from successful detection
e_c	effort to detect norm violation
u_c	expected utility of inspector
Detection probability and unknown extent of norm violation	
p	detection probability of norm violation
α	baseline detection probability (zero efforts of norm targets and inspectors)
q	all norm violations
$\bar{q}$	detected norm violations
λ_a	norm target's suspiciousness (constant belief of extent of undetected norm violations)
λ_c	inspector's suspiciousness (constant belief of extent of undetected norm violations)
$\hat{q}_a$	norm target's updated estimate of undetected norm violations
$\hat{q}_c$	inspector's updated estimate of undetected norm violations
Parameters of functional form of detection probability p in Example 1	
β	measures how concealment and inspection effort affect detection probability
γ	measures how the extent of norm violations affects detection probability

on the basis of the observed consequences of their actions in the last period, namely the proportion of norm violators or detected norm violators. Within one time step, we assume strategic interaction between the norm targets and the inspector: Both parties have knowledge of each others utility function and can incorporate it in their respective optimization. Nevertheless, we restrict their strategic horizon to the actual time step by (K2): The norm targets and the inspector do not take possible actions in future periods into account as neither party has knowledge about the moral cost

distribution F (and therewith the actual proportion of norm violators)—a norm target i only knows her own value c_i.

We now specify further assumptions with respect to the shape of the detection probability function p. First, we require that for every norm target, the marginal probability $\frac{\partial p}{\partial e_a}$ of being detected with respect to her effort e_a does not depend on the inspector's expenditures e_c and vice versa. This leads to

(P1) $p(e_a, e_c, q) = (1-\alpha) f_c(e_c, q) - \alpha f_a(e_a, q) + \alpha$

with $f_a, f_c : \mathbb{R}_+ \times [0,1] \to [0,1]$ denoting the effect of the respective effort on the probability of catching a norm violator. The parameter $\alpha = p(0,0,q)$ measures the probability of being detected when both parties' effort is zero. Hence, a low value of α leads to an initial advantage for the norm violators as the revelation probability is low in this situation. However, the marginal effect of every norm target's effort is decreased by a low value of α so that her initial advantage disappears with increasing efforts. For the extreme example $\alpha = 0$ ($\alpha = 1$) detection is impossible (guaranteed) for zero efforts of both parties, while any effort of the norm targets (the inspector) has no effect. Considering the example of fare dodging, the probability of detecting a passenger without a ticket is mainly affected by the inspector's effort which in this case represents the frequency of ticket examination. In contrast, the passenger's ability to conceal his fraud in case of inspection is very limited which overall is reflected in $\alpha \approx 0$ in this situation.

Further, for $i = a, c$, f_i has to fulfill the following requirements:

(P2) f_i is continuous and twice continuously differentiable on its interior.
(P3) $f_i(\cdot, q)$ is strictly concave for all q.
(P4) $f_i(0, q) = 0$ for all q.
(P5) $\lim_{e_i \to 0} \frac{\partial f_i}{\partial e_i}(e_i, q) = \infty$ for all q.

While (P2) is a purely technical assumption, (P3) guarantees that the detection probability is strictly decreasing (strictly increasing) in the concealment (control) effort and that there is a decreasing marginal effect of both kinds of investments on the revelation probability of norm violation.[8] (P4) reflects that without any investments, neither the inspector nor the norm targets can influence the probability p of catching a norm violator. Finally, (P5) assures that the marginal effect of investments is infinitely high in case of zero expenditures. Hence, the inspector and a norm violator will always exhibit non-zero investments. The assumptions (P2)–(P5) basically guarantee that there is always a unique optimal concealment and control effort for any proportion of norm violators.

Additionally, we confine the dependence of the detection probability on the proportion of norm violators to effects caused by indirect detection:

(P6) $\frac{\partial f_a}{\partial q} \leq 0$, $\frac{\partial f_c}{\partial q} \geq 0$.
(P7) $\frac{\partial^2 f_i}{\partial e_i \partial q} \leq 0$, $\frac{\partial^2 f_c}{\partial e_c \partial q} \geq 0$.

[8] Strict concavity, domain $\mathbb{R}_+$ and a non-negative codomain imply that f_i is strictly increasing.

By this assumptions, the effect and the marginal effect of expenditures on concealing a norm violation are non-increasing with the proportion of norm violators q while the effect and the marginal effect of inspection investments do not decrease with q.

Example 1 An example for a function p satisfying all the above assumptions is

$$p(e_a, e_c, q) = (1-\alpha)\left[1-\exp\left(-e_c^{\beta}\right)\right] - \alpha\left[1-\exp\left(-e_a^{\beta}\right)\right](1-\gamma q^2) + \alpha$$

with parameters $\beta, \gamma \in (0, 1)$.

Here, the effect of the effort expenditures e_a and e_c on the probability of detecting a norm violator increases with β while γ determines the extent to which the effect of the concealment effort depends on the proportion of norm violators q. In Fig. 1 we depict some properties of p for $\alpha = \beta = \gamma = 0.5$. Note that we use this example only to illustrate certain properties of our model—all following results hold for any function p fulfilling the above assumptions and do not depend on the particular choice of p.

Based on these assumptions, we now derive the dynamics for the key variable, namely the proportion of norm targets in the population who do not adhere to the norm.

3.1 Observable proportion of norm violators

First we analyze the model assuming that the proportion q of norm violators is observable by the norm targets and the inspector. The inspector's optimal effort $e_c^*(t)$ requires

$$q(t)\frac{\partial f_c}{\partial e_c}\left(e_c^*(t+1), q(t)\right) = \frac{1}{(1-\alpha)b_c} \tag{6}$$

whereas (P3)–(P5) guarantee its existence and uniqueness for all parameters. In the following, $e_c^*(q(t))$ denotes the inspector's optimal effort at time $t+1$ for fixed α and b_c. Note that it also depends on $q(t)$ if the detection probability is independent of the proportion of norm violators.

Norm target i's optimal effort $e_{a_i}^*(t+1)$ has to fulfill

$$\frac{\partial f_a}{\partial e_a}\left(e_{a_i}^*(t+1), q(t)\right) = \frac{1}{\alpha s} \tag{7}$$

whereas for all parameters, the optimum always exists and is unique as (P3)–(P5) hold. As each norm target observes the same proportion of norm violators $q(t)$, the optimal concealment effort is identical for all norm targets. In the following, $e_a^*(q(t))$ denotes a norm target's optimal effort at time $t+1$. Note that this optimal effort is independent of $q(t)$ in case the detection probability is independent of the proportion of norm violators in the population. Additionally, the inspector's decision influences whether the norm target actually violates the norm or not. Norm target i will violate the norm at time $t+1$ if

$$u_{a_i}\left((0,0)\right) < u_{a_i}\left(\left(1, e_a^*(q(t))\right)\right)$$

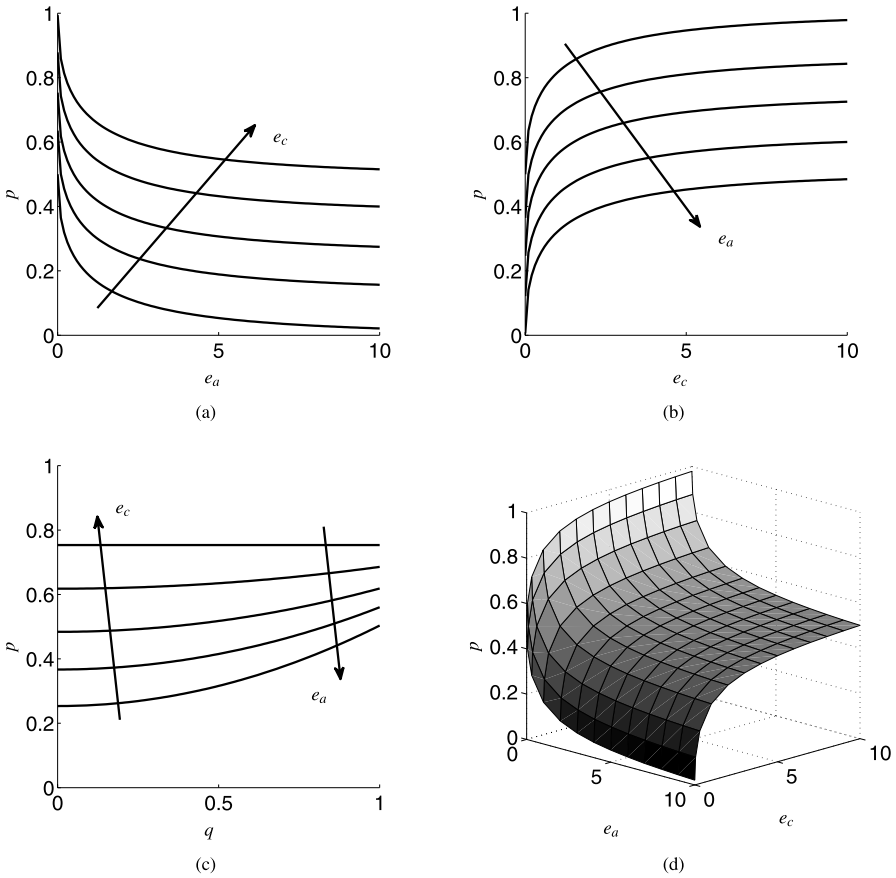


Fig. 1 The probability $p(e_a, e_c, q)$ of being detected dependent on the concealment effort e_a, the control effort e_c and the proportion of norm violators q for Example 1 with parameters $\alpha = \beta = \gamma = 0.5$. In (**a**) and (**b**), we depict p as a function of e_a (or e_c respectively). With q fixed, we choose different values of e_c (e_a) whereas the *arrow* indicates the direction of increase. In (**c**), we fix e_a and e_c and depict p as a function of q where $p(e_a, e_c, \cdot)$ decreases with e_a and increases with e_c. (**d**) shows p as a function of e_a and e_c for q fixed

or equivalently

$$c_i < b_a - sp\big(e_a^*(q(t)), e_c^*(q(t)), q(t)\big) - e_a^*(q(t)). \tag{8}$$

Here, the norm targets are able to determine the inspector's optimal effort as $q(t)$ and the inspector's utility are common knowledge. Note further that a norm target's decision whether to violate the norm or not does not depend on the factor r and is thus not affected by the extent of welfare loss for the population caused by norm violation.

One can easily verify how the optimal efforts respond to a change in the proportion q of norm violators (cf. Fig. 2):[9]

[9] All proofs are provided in the Appendix.

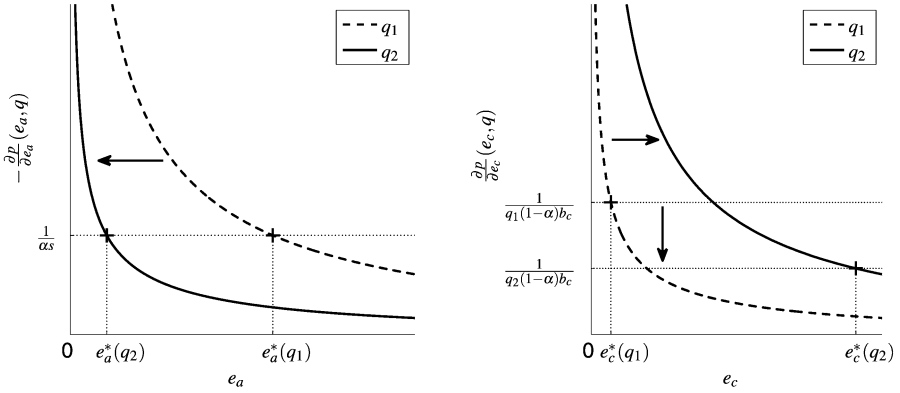


Fig. 2 Illustration of Proposition 1. An increase in the proportion of norm violators from q_1 to q_2 leads to a decrease of the optimal concealment effort e_a^* and an increase of the optimal control effort e_c^*

Proposition 1 *For all* $b_a, b_c, s \in \mathbb{R}_+$, $\alpha \in [0, 1]$,

(i) $e_a^*(q)$ *is non-increasing in* q,
(ii) $e_c^*(q)$ *is strictly increasing in* q

for any proportion of norm violators $q \in [0, 1]$.

As both parties know $q(t)$, they can immediately determine their respective optimal effort level. To check whether norm violation pays off, each norm target additionally can use her knowledge about the inspector's utility to obtain $e_c^*(t)$ and thus insert it in (8). Hence, the new proportion of norm violators

$$q(t+1) = F\big(b_a - sp\big(e_a^*(q(t)), e_c^*(q(t)), q(t)\big) - e_a^*(q(t))\big) =: g_q(q(t)) \qquad (9)$$

is completely determined by the respective proportion at the previous time step (see Fig. 3). Any equilibrium proportion of norm violators q^* thus is a fixed point of g_q, i.e. $q^* = g_q(q^*)$. Further, we denote the proportion of detected norm violators in equilibrium by

$$\tilde{q}(q^*) = q^* p\big(e_a^*(q^*), e_c^*(q^*), q^*\big).$$

We know how the iteration function g_q responds to a change in the proportion of norm violators:

Proposition 2 *If* F *is differentiable in all except finitely many points, the proportion of norm violators* $q(t+1)$ *at time* $t+1$ *decreases with the proportion of norm violators* $q(t)$ *at time* t.

Note that by allowing a finite number of points where F is not differentiable we do not exclude for instance a uniform distribution on an interval. Further, the proof shows that our assumptions (P6) and (P7) are necessary for the above proposition as the effect of an increase of $q(t)$ on the detection probability might be ambiguous if one of these assumptions did not hold.

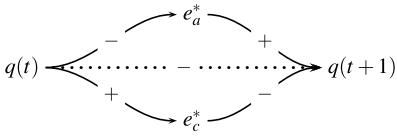


Fig. 3 Dynamics for an observable proportion of norm violators. A plus (minus) sign indicates that an increase of the respective input variable causes an increase (decrease) in the dependent variable whereas the monotonicity is not strict. The *dotted line* pictures the multi-level relation between the proportion of norm violators at time t and the proportion of norm violators at time $t+1$

We now show that this equilibrium is unique for arbitrary parameters and derive its comparative statics.

Theorem 1 *If F is differentiable in all except finitely many points, g_g has a unique fixed point $q^*(b_a, s, b_c, F)$ which is*

(i) *increasing in b_a,*

(ii) *decreasing in s and b_c*

for all parameters $b_a, s, b_c > 0$. Further, the associated proportion of detected norm violators $\tilde{q}(q^*(b_a, s, b_c, F))$ in equilibrium is increasing in b_a.

Hence, the assumption of an observable proportion of norm violators leads to a unique equilibrium q^* which increases with the benefit of norm violation b_a and decreases with the sanction cost s for a detected norm violator and the inspector's maximum reward b_c for detected norm violation. Note that this holds for all parameters and in particular for any distribution of moral costs. We can also predict that the proportion of detected norm violators $\tilde{q}^*$ in equilibrium increases with b_a as this implies more norm violators and a higher detection probability. In contrast, the effect of an increase of s or b_c on $\tilde{q}^*$ is ambiguous: The proportion of norm violators is reduced in both cases, but the increased quality of inspections counteracts this effect and prohibits any general prediction.

For a finite number of norm targets, the assumption of a continuous moral cost distribution is obviously violated. Here, the equilibrium may not be reached as the fraction of drug users is always rational in this situation. Nevertheless, one can interpret the equilibrium q^* as the expected proportion of norm violators when the moral cost value c_i of each norm target i is randomly drawn from a continuous distribution function F fulfilling the assumptions of Theorem 1.

The equilibrium is not necessarily stable, i.e. for an initial proportion of norm violators which is different from q^* convergence is not guaranteed. For Example 1, dependent on the parameters one can also observe alternating proportions of norm violators. In Fig. 4, we compare two scenarios whereas the parameters only differ in the punishment s for detected norm violators. For a low value of s, we observe convergence towards the equilibrium proportion of norm violators q^*. An increase of s leads to destabilization as q alternates between zero and a positive value. Here, there is initially no incentive for controls as all norm targets adhere to the norm. The norm targets anticipate this behavior, and norm violation pays off for a non-zero proportion

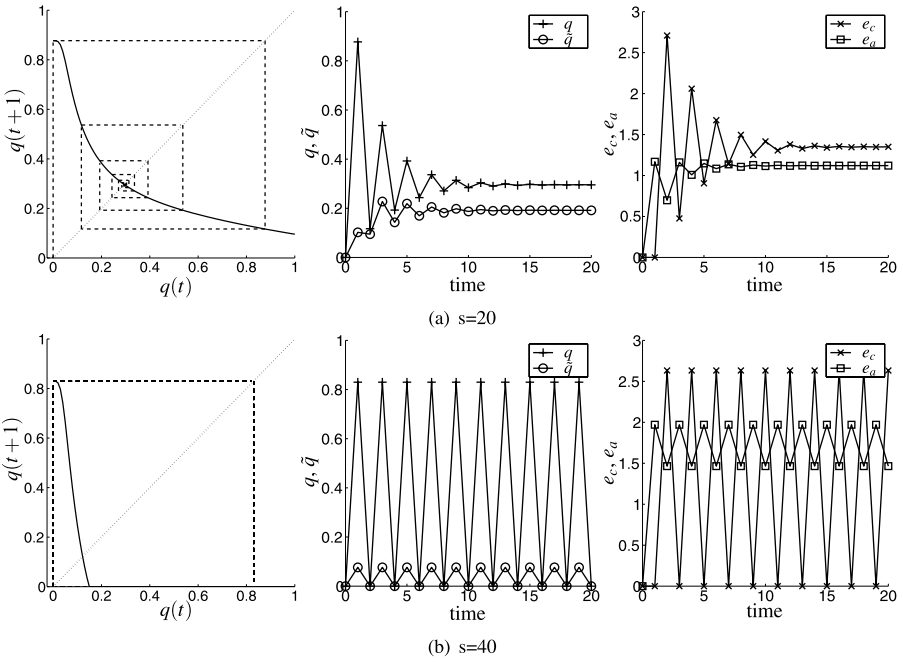


Fig. 4 Dynamics of the proportion of norm violators q, the proportion of detected norm violators $\tilde{q}$ and the concealment/control effort e_a/e_c for Example 1 (including cobweb diagram for q) with $b_a = b_c = 20$, c_i uniformly distributed on $[0, b_a]$ and $\alpha = 0.2$, $\beta = 0.8$, $\gamma = 0.5$ whereas initially all norm targets are norm adherent. In (**a**), the equilibrium is stable, while an increase in the punishment s for detected norm violators leads to a destabilization of the fixed point in (**b**). Note that the optimal concealment effort can be positive although all norm targets adhere to the norm

of the population. At the next time step, this can be observed by the inspector who therefore increases the control effort. Due to the high punishment of norm violation, this is sufficient to make all norm targets refrain from norm violation and we will hence observe oscillating behavior instead of convergence towards the equilibrium.

As we did in general not specify the function of probability of detection p in detail, we cannot determine specific ranges within the parameter space that yield a stable equilibrium for arbitrary p. Nevertheless, the proof of Proposition 2 shows that $|g'_q(q)|$ increases with s and decreases with b_c.[10] Thus, increasing punishment for detected norm violators destabilizes the system whereas an increase of the inspectors benefits leads to the opposite effect.[11]

3.2 The dark figure of norm violation

So far we assumed that at any time step, the norm targets and the inspector know the actual proportion of norm violators q and therefore are able to determine their optimal

[10] Cf. (A.1).

[11] For any fixed point q^* of g_q, $|g'_q(q^*)| < 1$ implies stability.

decisions based on that value. In many situations this information is not available for either party, e.g. in the context of tax fraud or doping in sports. In order to determine their optimal effort at time $t+1$ according to (7) and (6) respectively, they need to estimate $q(t)$ based on the public information of the proportion of detected norm violators $\tilde{q}(t)$ in the previous period. Any estimation has to be consistent with that information: If the proportion of detected norm violators at time t is $\tilde{q}(t)$, a norm targets' estimation $\hat{q}_{a_i}(t)$ and the inspector's estimation $\hat{q}_c(t)$ of the real proportion of norm violators $q(t)$ at that time naturally have to fulfill (5) as this always holds for the estimated value q in absence of false-positive detections.

The norm targets and the inspector now have to choose a feasible estimation of q according to (5). We assume that this choice is determined by the moral confidence, i.e. the belief of the extent of norm compliance in the population. In our model, a norm target's moral confidence is measured by the *suspiciousness* $\lambda_{a_i} \in [0,1]$ which denotes the antonym (i.e. the belief of the extent of norm violation). The inspector's suspiciousness towards the extent of norm violation is measured by the parameter $\lambda_c \in [0,1]$. Note that these parameters are exogenous and not explained by our model. Additional requirements to the estimation procedure are the following:

(E1) A norm target's (or the inspector's) estimation $\hat{q}_{a_i}(t)$ ($\hat{q}_c(t)$) of the real proportion of norm violators is increasing in the proportion of detected norm violators $\tilde{q}(t)$ for any given suspiciousness λ_{a_i} (λ_c) at any timestep t.
(E2) A norm target's (or the inspector's) estimation $\hat{q}_{a_i}(t)$ ($\hat{q}_c(t)$) of the real proportion of norm violators is increasing in the suspiciousness λ_{a_i} (λ_c) for any given proportion of detected norm violators $\tilde{q}(t)$ at any timestep t.
(E3) The suspiciousness of a norm target (of the inspector) is private information and does not change over time.

For the sake of simplicity, we assume that the estimation of the real proportion of norm violators is an affine-linear function of the suspiciousness and the proportion of detected norm violators respectively:[12]

$$\hat{q}_j(t) = \tilde{q}(t) + \lambda_j\,(1-\tilde{q}(t)), \quad j = a_i, c. \tag{10}$$

For $\lambda_{a_i} = 0$ ($\lambda_{a_i} = 1$), norm target i assumes the minimum (maximum) control level, i.e. $\hat{q}_{a_i} = 1$ ($\hat{q}_{a_i} = \tilde{q}$). For $\lambda_{a_i} = 0.5$, the norm target's estimation is the arithmetic mean of the extreme values (analogously for λ_c). λ_{a_i} (λ_c) is also the estimated proportion of norm violators of a norm target (the inspector) if no norm violators are detected. Hence, the norm targets and the inspector combine private beliefs (which are exogenous and constant) and public information (which is explained by the model) to estimate the proportion of norm violators. By (E3), we bound the norm targets' and the inspector's rationality as any individual (norm target or inspector) is only aware of her own suspiciousness and therefore cannot foresee other individuals' estimations of the proportion of norm violators q. Instead we assume that she plugs in her own estimation in case the decision-making requires another individual's estimation of q

[12] Without changing the results one could also assume any function $z : [0,1] \times [0,1] \to [\tilde{q}, 1]$ that is increasing in both arguments to locate the estimation $\hat{q} = z(\tilde{q}, \lambda)$ of q in a more general way.

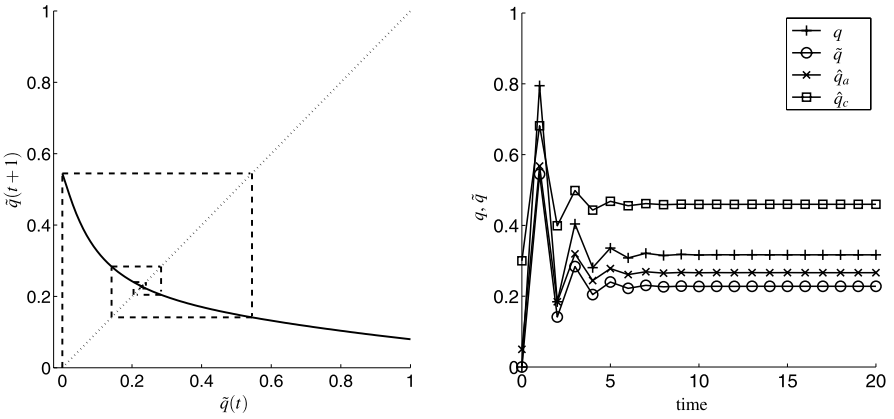


Fig. 5 Dynamics of the proportion of detected norm violators $\tilde{q}$, the proportion of norm violators q and the estimated proportion of norm violators $\hat{q}_a$ and $\hat{q}_c$ if the proportion of norm violators is unobservable for Example 1 (including cobweb diagram for $\tilde{q}$) with $\lambda_a = 0.05$, $\lambda_c = 0.3$, $b_a = b_c = s = 20$, c_i uniformly distributed on $[0, b_a]$ and $\alpha = 0.2$, $\beta = 0.8$, $\gamma = 0.5$ whereas initially all norm targets are norm adherent

(e.g. a norm targets has to consider the inspector's incentives which depends on the proportion of norm violators). Additionally and similar to assumption, we require that the suspiciousness is homogeneous among the norm targets:

(E4) The suspiciousness of all norm targets is identical, i.e. $\lambda_{a_i} = \lambda_a$.

Hence, all estimations of the real proportion of norm violators by the norm targets are identical, i.e. $\hat{q}_{a_i} = \hat{q}_a$.

With these assumptions, the dynamics of our model in case of an unobservable proportion of norm violators can now be rewritten in terms of the proportion of detected norm violators $\tilde{q}$. With $\tilde{q}(t)$ denoting this proportion at time t, we obtain

$$\tilde{q}(t+1) = q(t+1)\, p\big(e_a^*(\hat{q}_a(t)), e_c^*(\hat{q}_c(t)), q(t+1)\big) =: g_{\tilde{q}}(\tilde{q}(t); \lambda_a, \lambda_c) \qquad (11)$$

where $\hat{q}_a(t) = \hat{q}_a(\tilde{q}(t))$ and $\hat{q}_c(t) = \hat{q}_c(\tilde{q}(t))$ depend on $\tilde{q}(t)$, λ_a, λ_c via (5) and (10) and $q(t+1) = g_q(\hat{q}_a(t))$. Note that a fixed point $\tilde{q}^*$ of $g_{\tilde{q}}$ induces an equilibrium proportion of norm violators that we denote by $q^*(\tilde{q}^*)$. Figure 5 depicts the dynamics according to (11) for Example 1.

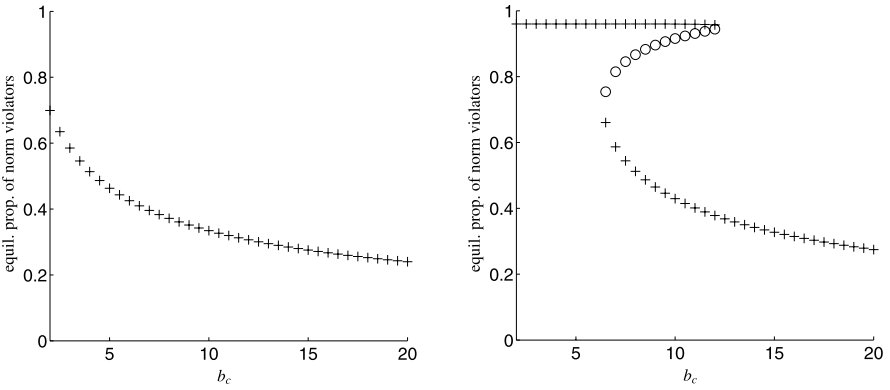
Further, we can specify how the iteration function $g_{\tilde{q}}$ depends on the parties' respective suspiciousness:

Proposition 3 *For all* $\tilde{q}, \lambda_a, \lambda_c \in [0, 1]$,

(i) $g_{\tilde{q}}(\tilde{q}, \lambda_a, \lambda_c)$ *is increasing in* λ_c,
(ii) $q(\tilde{q})$ *is decreasing in* $\tilde{q}$

where $q(\tilde{q})$ *denotes the proportion of norm violators resulting from a proportion* $\tilde{q}$ *of detected norm violators.*

After having introduced uncertainty by the unobservability of the proportion of norm violators, the first question is whether Theorem 1 still holds for the new dy-



(a) equilibrium proportion of norm violators for observable (left) and unobservable (right) $q(t)$

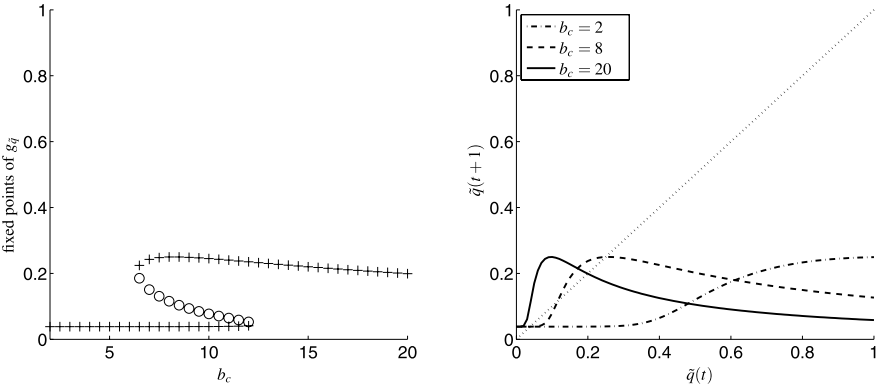


Fig. 6 Example 1 with $b_a = s = 10$, $\lambda_a = \lambda_c = 0.01$, c_i uniformly distributed on $[0, b_a]$ and $\alpha = 0.04$, $\beta = 0.9$, $\gamma = 0.1$. (**a**) pictures how the equilibrium proportion of norm violators change with the detection reward b_c for observable (*left*) and unobservable proportion of norm violators $q(t)$. A plus indicates stable, a circle unstable equilibria. In (**b**), we show the equilibrium proportions of detected norm violators (the fixed points of $g_{\tilde{q}}$) for unobservable $q(t)$ and the iteration function $g_{\tilde{q}}$ for $b_c = 2$ (one fixed point), $b_c = 8$ (three fixed points: two stable, one unstable) and $b_c = 20$ (one fixed point)

namics, i.e. if there is always a unique equilibrium proportion of norm violators.[13] The answer is no: In Fig. 6 we show that there are parameter settings where multiple fixed points of $g_{\tilde{q}}$ can arise in Example 1 when there is a dark figure of norm violation. In this example, the number of fixed points depends on the inspection reward b_c: for $0 < b_c < b_c^1 \approx 6.4$, there is a unique stable fixed point of $g_{\tilde{q}}$ and therefore also a unique stable equilibrium proportion of norm violators. For $b_c^1 < b_c < b_c^2 \approx 12.2$, we observe two stable equilibria: one where approximately 96.

The reason is that $g_{\tilde{q}}$ is not necessarily decreasing in $\tilde{q}$ (see Fig. 7): An increase of $\tilde{q}$ will also increase the parties' estimation $\hat{q}_a$ and $\hat{q}_c$ of the real proportion of norm violators q (cf. (5), (10)). According to Proposition 1, this leads to an increase

[13] The Brouwer fixed point theorem guarantees the existence of an equilibrium.

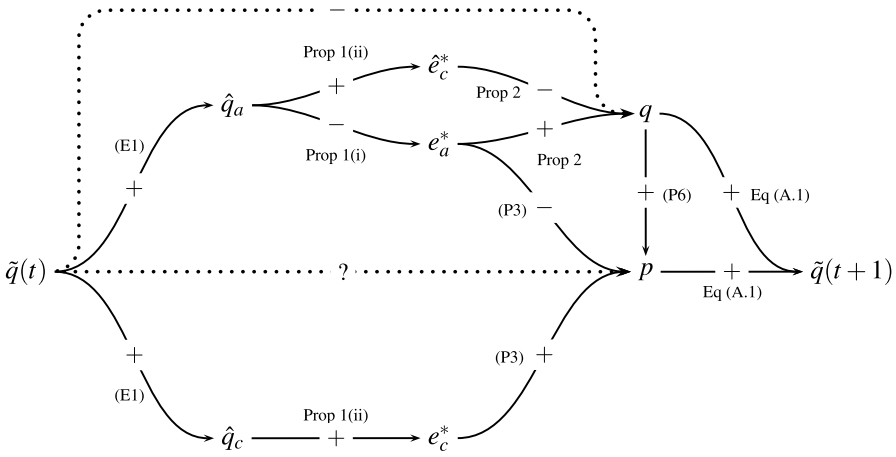


Fig. 7 Dynamics for an unobservable proportion of norm violators. A plus (minus) sign indicates that an increase of the respective input variable causes an increase (decrease) in the dependent variable whereas the monotonicity is not strict. A question mark denotes that the effect of a change in the input variable cannot be predicted in general. Multi-level relations as the response of the proportion of norm violators to a change in the proportion of detected norm violators are pictures by a *dotted line*. The *arrow* label indicates the origin of the respective relation

of e_c^*, a decrease of e_a^* and in addition to a decrease of the resulting proportion of norm violators q as g_q is decreasing (with argument $\hat{q}_a$). However, the probability of being detected decreases in this situation. Hence, we obtain less norm violators but also a lower probability of being detected. Consequently we cannot make general predictions about the new proportion of detected violators.

Our basic research question is how estimation of the dark figure of norm deviance influences the actual strength of a norm. Regarding our model, we can reformulate this question: How does a change in the norm targets' and the inspector's estimation procedure of the real proportion of norm violators, i.e. a change in λ_a or λ_c, influence the proportion of norm violators in equilibrium? As we only made very basic assumptions about the dynamics by not specifying the dependence of the probability of detecting norm violation on the parties' respective effort and the proportion of norm violators in detail, an all-encompassing answer is not possible. Nevertheless we are able to make very general predictions in Theorem 2 about how a "typical" stable equilibrium $\tilde{q}^*$ (i.e. $\tilde{q}^*$ is asymptotically stable and hyperbolic[14]) of detected norm violators will respond to small changes in the suspiciousness parameters λ_a and λ_c of the norm targets and the inspector. There, we make use of the following lemmas referring to that type of fixed point.

Lemma 1 *Let $x_{t+1} = f(x_t)$ denote a difference equation with $f : [0,1] \to [0,1]$ continuous. If x^* is an asymptotically stable fixed point of f, there is $\delta > 0$ with*

(i) $f(x) < x$ *if* $x \in (x^*, x^* + \delta)$,

[14] See e.g. Elaydi (1996) for a definition of an asymptotically stable and hyperbolic equilibrium.

(ii) $f(x) > x$ *if* $x \in (x^* - \delta, x^*)$.

Lemma 2 *Let* $f : [0,1] \times [0,1] \to [0,1]$ *be continuous and let* $x_{t+1} = f_\lambda(x_t)$ *denote a difference equation with* $f_\lambda := f(\cdot, \lambda)$ *continuously differentiable for all* λ. *If* $x^*(\lambda)$ *is an asymptotically stable hyperbolic fixed point of* f_λ, *then for all* $\delta_x > 0$ *there is* $\delta_\lambda > 0$ *so that for all* $\lambda' \in [0,1]$ *with* $|\lambda' - \lambda| < \delta_\lambda$ *there is a unique fixed point* $\tilde{x}$ *of* $f_{\lambda'}$ *in* $[x^* - \delta_x, x^* + \delta_x]$, *i.e.*

$$|\lambda' - \lambda| < \delta_\lambda \;\Rightarrow\; \exists \tilde{x} \in [x^* - \delta_x, x^* + \delta_x] \text{ with } f_{\lambda'}(\tilde{x}) = \tilde{x} \text{ and } |f'_{\lambda'}(\tilde{x})| < 1.$$

Moreover, $\tilde{x}$ *is asymptotically stable and hyperbolic.*

If we observe convergence towards an equilibrium $\tilde{q}^*$, then this fixed point is typically asymptotically stable and hyperbolic. Nevertheless we cannot exclude possible exceptions such as semi-stable or asymptotically stable but non-hyperbolic equilibria (or starting in an unstable equilibrium). But for any given functional form of $p(e_a, e_c, q)$ it can easily be verified whether an equilibrium $\tilde{q}^*$ of $g_{\tilde{q}}$ has the designated property.[15]

Theorem 2 *Let* $\tilde{q}^*(\lambda_a, \lambda_c)$ *denote an asymptotically stable hyperbolic fixed point of* $g_{\tilde{q}}$ *for suspiciousness parameters* $\lambda_a, \lambda_c \in [0,1]$ *and let* $g_{\tilde{q}}$ *be continuously differentiable at* $\tilde{q}^*(\lambda_a, \lambda_c)$. *Then there is* $\delta > 0$ *with*

$$\tilde{q}^*(\lambda_a, \lambda'_c) \begin{cases} \leq \tilde{q}^*(\lambda_a, \lambda_c) & \text{if } \lambda'_c \in [\lambda_c - \delta, \lambda_c], \\ \geq \tilde{q}^*(\lambda_a, \lambda_c) & \text{if } \lambda'_c \in [\lambda_c, \lambda_c + \delta], \end{cases}$$

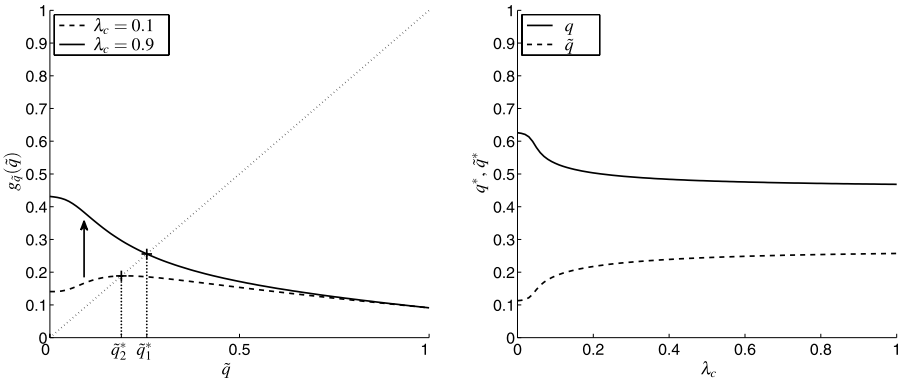
whereas $\tilde{q}^*(\lambda_a, \lambda'_c)$ *denotes the unique fixed point of* $g_{\tilde{q}}$ *on* $[x^* - \delta, x^* + \delta]$ *for suspiciousness parameters* λ_a, λ'_c.

This means that a small increase in the inspector's suspiciousness λ_c leads to an increase of the equilibrium proportion of detected norm violators $\tilde{q}^*$. The reason is that the iteration function $g_{\tilde{q}}$ increases with λ_c as for a given proportion of detected norm violators, the inspector's estimation of the proportion of norm violators increases with λ_c (cf. Fig. 8(a)). By this shift of $g_{\tilde{q}}$, the equilibrium proportion of detected norm violators must increase.[16] As the equilibrium proportion of norm violators q^* decreases with the norm targets' estimation $\hat{q}_a^*$ which is positively affected by the increase of $\tilde{q}^*$ (cf. Fig. 7), this will finally lead to a decrease of the proportion of norm violators q^* in the new equilibrium.

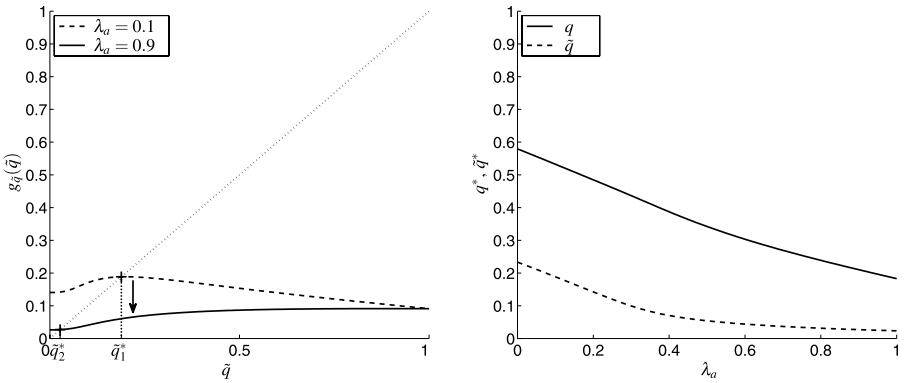
Corollary 1 *Let* $\tilde{q}^*(\lambda_a, \lambda_c)$ *denote an asymptotically stable hyperbolic fixed point of* $g_{\tilde{q}}$ *for suspiciousness parameters* $\lambda_a, \lambda_c \in [0,1]$ *and let* $g_{\tilde{q}}$ *be continuously differentiable at* $\tilde{q}^*(\lambda_a, \lambda_c)$. *Let further* $q^*(\lambda_a, \lambda_c) = q(\tilde{q}^*(\lambda_a, \lambda_c))$ *denote the proportion of*

[15] $|g'_{\tilde{q}}(\tilde{q}^*)| < 1)$ is a sufficient condition.

[16] This conclusion is possible if the old equilibrium is asymptotically stable and hyperbolic and the change in λ_c is sufficiently small.



(a) $\lambda_a = 0.1$



(b) $\lambda_c = 0.1$

Fig. 8 Change in the norm targets' and the inspector's suspiciousness for Example 1 with $b_a = b_c = s = 10$, c_i uniformly distributed on $[0, b_a]$ and $\alpha = \beta = 0.9$, $\gamma = 0.5$. The *left column* pictures the iteration function $g_{\tilde{q}}$, the *right column* the equilibrium proportion of norm violators q^* and detected norm violators $\tilde{q}^*$ depending on the respective suspiciousness. In (**a**) we increase the inspector's suspiciousness λ_c from 0.1 to 0.9 while in (**b**) the norm targets' suspiciousness λ_a is increased

norm violators in that equilibrium. Then there is $\delta > 0$ *with*

$$q^*(\lambda_a, \lambda_c') \begin{cases} \geq q^*(\lambda_a, \lambda_c) & \text{if } \lambda_c' \in [\lambda_c - \delta, \lambda_c], \\ \leq q^*(\lambda_a, \lambda_c) & \text{if } \lambda_c' \in [\lambda_c, \lambda_c + \delta], \end{cases}$$

whereas $\tilde{q}^*(\lambda_a, \lambda_c')$ *denotes the unique fixed point of* $g_{\tilde{q}}$ *on* $[x^* - \delta, x^* + \delta]$ *for suspiciousness parameters* λ_a, λ_c'.

Note that in general, we can only guarantee the existence of the threshold δ for a change in the inspector's suspiciousness where a prediction regarding the change in the equilibrium proportion of norm violators is possible. Depending on the functional form of $p(e_a, e_c, q)$, the threshold could be arbitrarily small, whereas in other scenarios this prediction is possible for any change in the inspectors suspiciousness.

Further, Theorem 2 might not be valid in case of a more sophisticated estimation procedure of the proportion of norm violators by the norm targets and the inspector. For example, both types of individuals could use their knowledge of their past actions or their counterpart's estimation procedure and thereby shorten the interval of feasible estimations of the proportion of norm violators. One can show that this more sophisticated procedure does not fulfill assumption (E1) for arbitrary detection probability functions p.

Without further requirements on p it is not possible to deduce a similar result with respect to a change in the norm targets' suspiciousness λ_a: It is not clear whether an increase of λ_a increases or decreases the equilibrium proportion of detected norm violators $\tilde{q}^*$ (cf. Fig. 8). The reason is that the iteration function $g_{\tilde{q}}$ is not decreasing in λ_a. However, regarding the dependence of the equilibrium proportion of norm violators q^* on λ_a, we obtain

$$\begin{aligned}\frac{\partial q^*}{\partial \lambda_a}(\lambda_a, \lambda_c) &= \frac{\partial}{\partial \lambda_a} g\big(\hat{q}_a^*(\lambda_a, \lambda_c)\big) \\ &= \underbrace{g_q'\big(\hat{q}_a^*(\lambda_a, \lambda_c)\big)}_{\leq 0}\Bigg[(1-\lambda_a)\frac{\partial \tilde{q}^*}{\partial \lambda_a}(\lambda_a, \lambda_c) + \underbrace{1-\tilde{q}^*(\lambda_a, \lambda_c)}_{\geq 0}\Bigg].\end{aligned}$$

Here, $1-\tilde{q}^*(\lambda_a, \lambda_c)$ is non-negative and reflects the marginal effect of a change in λ_a on the norm targets' estimation $\hat{q}_a^*$ of the proportion of norm violators (with $\tilde{q}^*$ fixed) while the sign of $\frac{\partial \tilde{q}^*}{\partial \lambda_a}(\lambda_a, \lambda_c)$ cannot be determined in general. Hence, $q^*(\lambda_a)$ increases with λ_a only if the change in $\tilde{q}^*$ is negative and overcompensates the increase of $\hat{q}^*$ by λ_c. Thus, we expect that an increase in the norm targets' suspiciousness causes a decrease of the proportion of norm violators in most cases.

Overall, we can predict that an increase in the norm targets' or the inspector's suspiciousness usually leads to a decrease in the equilibrium proportion of norm violators in the population. We provide a precise definition of "usually" whereas the definition with respect to a change in the inspector's suspiciousness differs from that with respect to a change in the norm targets' suspiciousness.

4 Conclusion

In this paper, we investigated how the belief in others' norm compliance affects own normative behavior. Popitz (1968) argued that ignorance of others' norm compliance promotes normative behavior, because actors are more willing to adhere to social norms if they believe others do so as well (self-fulfilling prophecy). We challenge this view by considering strategic interaction between norm targets and inspectors. If inspectors believe in a low extent of undetected norm violations, the detection probability for norm violations decreases and norm violations will increase (inverse self-fulfilling prophecy).

We consider the detection problem of norm violations in the context of public goods dilemmas. As in the context of tax evasion or doping in sports, the extent of norm violations is unknown to both the norm targets (e.g. taxpayers or athletes) and

the inspectors. We model the interaction between norm targets and inspectors of a social norm. Our model links the belief of the extent of undetected norm violations with the actual rate of norm violations and with inspection behavior. The norm targets and the inspectors combine their respective (exogenous) belief of the undetected norm violations on the one hand and public information about the proportion of detected norm violators on the other. The agents use their beliefs and knowledge to estimate the overall extent of detected and undetected norm violations in the population.

Our analysis suggests that Popitz' (1968) intuitive reasoning that ignorance promotes norm compliance does not hold in general. We can derive an "inverse" self-fulfilling prophecy effect: An increase in the inspector's belief in the norm targets' compliance always *decreases* the actual overall proportion of norm compliance in equilibrium. In addition, we typically observe that an increase in the norm targets' belief of others' compliance *decreases* the proportion of norm compliance in equilibrium. Our results can be understood with considering the competitive incentive structure of norm targets and inspectors. The inspectors lose incentives for control efforts when there are (or they believe that there are) little undetected norm violations. Thus, the effect of moral stabilization on the side of the norm targets is overcompensated by the increased incentive for norm violations due to the low detection probability.

As a second result, we demonstrate that the dark field of norm violations can induce multiple equilibria of committed norm violations. This result can be linked to the recent debate that punishment operates dissimilar in different societies (Herrmann et al. 2008), which may be due to distinct punishment cultures (Gintis 2008).

The interpretation of our results relates to recent developments in doping in sports. While in the decades before approximately 1990, very few athletes were detected for drug use, there is strong evidence that doping was performed area-wide in a number of countries, in particular in the German Democratic Republic (Franke and Berendonk 1997). While the number of detected drug using athletes recently increased (e.g. Brissonneau and Depiesse 2006, p. 164), the actual absolute and relative extent of doping has probably decreased (e.g. Leonard 2001). Although an increased number of detections could be interpreted as a hint for more violations of the anti-doping norm, the opposite may be true due to an increased effectiveness of controls. In the context of our model, this development may be explained by an increase in the belief of the number of undetected drug using athletes, possibly due to intensified media reports on doping.

Prospective research could validate our theoretical results with empirical data. The most promising method were laboratory experiments because of the measurement problems of the dark field of undetected norm violations. In the field, data is usually only available for detected tax evasion, detected doping or detected crime in general. Survey data on self-reports of norm violations, on the other hand, are typically unreliable because these behaviors are understated in surveys.[17] In contrast, undetected norm violations could be measured in the laboratory more precisely.[18] As a baseline scenario, a public goods game could be used to measure detected and

[17] See Elffers et al. (1987) for an example with respect to a survey on tax evasion.

[18] For a related experiment on detection of norm violations see Rauhut (2009a, 2009b).

undetected norm violations. One type of subjects (the norm targets) could either contribute a fixed amount of money to a public good or violate the cooperation norm by contributing nothing. If a norm violation is detected by inspectors (the other type of subjects), the respective norm target would be punished. Norm violators could invest in concealment to reduce their detection probability and inspectors could invest in the quality of controls in order to receive rewards in case of success.

Acknowledgements The authors thank Volker Müller-Benedict and María Sáez-Martí for helpful suggestions and comments and Kerstin Press for fruitful discussions. Further, we thank Thomas Voss for his encouragement to pursue our work at an early stage of the paper. Finally, we thank two anonymous reviewers and the area editor for valuable comments and suggestions, which improved the paper significantly.

Appendix

Proof of Proposition 1 From (7) and (6) we can conclude that

$$\frac{\partial e_a^*}{\partial q} = -\frac{\frac{\partial^2 f_a}{\partial e_a \partial q}}{\frac{\partial^2 f_a}{\partial e_a^2}} \leq 0,$$

$$\frac{\partial e_c^*}{\partial q} = -\frac{q\frac{\partial^2 f_c}{\partial e_c \partial q} + \frac{\partial f_c}{\partial e_c}}{\frac{\partial^2 f_c}{\partial e_c^2}} > 0$$

as f_a, f_c are strictly concave and we required (P6) and (P7). □

Proof of Proposition 2 We have

$$g_q'(q) = \underbrace{F'(b_a - s\, p(e_a^*(q), e_c^*(q), q) - e_a^*(q))}_{\geq 0}\, c(q)$$

with

$$\begin{aligned} c(q) &= -s(1-\alpha)\left(\frac{\partial f_c}{\partial e_c}(e_c^*(q), q)\frac{\partial e_c^*}{\partial q}(q) + \frac{\partial f_c}{\partial q}(e_c^*(q), q)\right) \\ &\quad + s\alpha\left(\frac{\partial f_a}{\partial e_a}(e_a^*(q), q)\frac{\partial e_a^*}{\partial q}(q) + \frac{\partial f_a}{\partial q}(e_a^*(q), q)\right) - \frac{\partial e_a^*}{\partial q}(q) \\ &= -s\Bigg[\underbrace{\frac{1}{q\, b_c}\frac{\partial e_c^*}{\partial q}(q) + (1-\alpha)\frac{\partial f_c}{\partial q}(e_c^*(q), q)}_{>0} - \alpha\underbrace{\frac{\partial f_a}{\partial q}(e_a^*(q), q)}_{<0}\Bigg] \\ &\leq 0 \end{aligned} \tag{A.1}$$

and thus $g_q'(q) \leq 0$ if g_q is differentiable at q by using Proposition 1, assumptions (P6) and (P7), (7) and (6). Hence, g_q is decreasing as it is continuous in the finitely many points in which it is not differentiable. □

Proof of Theorem 1 We have to show that g_q has a unique fixed point. The Brouwer fixed point theorem immediately guarantees the existence of a fixed point q^*. The uniqueness of q^* is ensured by the fact that g_q is decreasing according to Proposition 2.

As g_q is decreasing for all b_a, s, b_c, monotonicity of g_q in one of this parameters implies a corresponding monotonicity of $q^*(b_a, s, b_l)$ in that parameter. One can verify that $\frac{\partial g_q}{\partial b_a} \geq 0$, $\frac{\partial g_q}{\partial s}$, $\frac{\partial g_q}{\partial b_c} \leq 0$ (where F is differentiable) which proves the monotonicity properties of q^*. For $\tilde{q}^* = q^*\, p(e_a^*(q^*), e_c^*(q^*), q^*)$ we obtain

$$\frac{\partial \tilde{q}^*}{\partial b_a} = \frac{\partial q^*}{\partial b_a}\left[p\big(e_a^*(q^*), e_c^*(q^*), q^*\big) + q^*\left(\frac{\partial p}{\partial e_a}\frac{\partial e_a^*(q^*)}{\partial q^*} + \frac{\partial p}{\partial e_c}\frac{\partial e_c^*(q^*)}{\partial q^*} + 1\right)\right] \geq 0$$

where F is differentiable. □

Proof of Proposition 3 First, the norm targets' estimation $\hat{q}_a$ of the proportion of norm violators does not depend on λ_c. Hence, we obtain

$$\frac{\partial g_{\tilde{q}}}{\partial \lambda_c}(\tilde{q}, \lambda_a, \lambda_c) = q(\hat{q}_a)\,\frac{\partial p}{\partial e_c}\big(e_a^*(\hat{q}_a), e_c^*(\hat{q}_c), q(\hat{q}_a)\big)\frac{\partial e_c^*}{\partial \hat{q}_c}(\hat{q}_c)\frac{\partial \hat{q}_c}{\partial \lambda_c}(\tilde{q}, \lambda_c) \geq 0$$

as p increases with e_c^*, e_c^* increases with $\hat{q}_c$ (cf. Proposition 1(ii)) and

$$\frac{\partial \hat{q}_c}{\partial \lambda_c}(\tilde{q}, \lambda_c) = (1 - \tilde{q}) \geq 0.$$

Second, $q(\tilde{q}) = g_q(\hat{q}_a(\tilde{q}))$ leads to

$$\frac{dg_q}{d\tilde{q}}(\hat{q}_a(\tilde{q})) = \frac{dg_q}{dq}(\hat{q}_a)\,(1 - \tilde{q}) \leq 0$$

as $\frac{dg_q}{d\hat{q}_a} \leq 0$ according to Proposition 2. □

Proof of Lemma 1 If for all $\delta > 0$ there is $x \in (x^*, x^* + \delta)$ with $f(x) = x$, x^* cannot be attractive and thus not asymptotically stable. Hence let us assume that we always find x with $f(x) > x$ in the same interval. As f is continuous, this implies that there is $\varepsilon > 0$ with

$$f(x) > x \text{ for } x \in [x^*, x^* + \varepsilon]. \tag{A.2}$$

Then x^* cannot be stable: for all $\delta' > 0$ (with $\delta' < \varepsilon$) we can choose $x_0 \in (x^*, x^* + \delta')$. If we assume $|x_t - x^*| = x_t - x^* < \varepsilon$ for all t, $(x_t)_t$ must converge with

$$x^* < \lim_{t\to\infty} x_t \leq \varepsilon.$$

But then the limit is a fixed point of f:

$$f(\lim_{t\to\infty} x_t) = \lim_{t\to\infty} f(x_t) = \lim_{t\to\infty} x_{t+1} = \lim_{t\to\infty} x_t$$

as f is continuous. This is a contradiction to (A.2), hence we proved (i). The proof for (ii) is analogous. □

Proof of Lemma 2 According to Lemma 1 and the fact that x^* is an asymptotically stable hyperbolic fixed point, we can find $\delta > 0$ ($\delta \leq \delta_x$) with

$$f_\lambda(x) \begin{cases} < x & \text{if } x \in [x^*, x^* + \delta], \\ > x & \text{if } x \in [x^* - \delta, x^*] \end{cases} \tag{A.3}$$

and

$$|f'_\lambda(x)| < 1 \quad \text{for } x \in I$$

whereas $I := [x^* - \delta, x^* + \delta]$. As f_λ is continuous on the compact set I, we can define

$$\varepsilon_1 := 1 - \max_{x \in I} |f'_\lambda(x)| > 0.$$

Further we can choose $x_-, x_+ \in I$ with $x_- < x^*$ and $x_+ > x^*$ and define

$$\varepsilon_2 := \min(f_\lambda(x_-) - x_-, x_+ - f_\lambda(x_+)) > 0$$

according to (A.3). As $\frac{\partial f}{\partial x}$ is uniformly continuous on the compact set $I \times [0, 1]$, there is $\delta_1 > 0$ so that for $(x_1, \lambda_1), (x_2, \lambda_2) \in I \times [0, 1]$

$$\|(x_1, \lambda_1) - (x_2, \lambda_2)\| < \delta_1 \;\Rightarrow\; |f'_{\lambda_1}(x_1) - f'_{\lambda_2}(x_2)| < \varepsilon_1$$

holds whereas $\|\cdot\|$ denotes the Euclidean norm. In particular, we thereby obtain

$$|\lambda' - \lambda| < \delta_1 \;\Rightarrow\; |f'_{\lambda'}(x) - f'_\lambda(x)| < \varepsilon_1$$

for all $x \in I$ and $\lambda' \in [0, 1]$ as $\|(x, \lambda') - (x, \lambda)\| = |\lambda' - \lambda|$.

Further, the continuity of $f(x, \cdot)$ for all x guarantees the existence of $\delta_2 > 0$ with

$$|\lambda' - \lambda| < \delta_2 \;\Rightarrow\; |f_{\lambda'}(x_-) - f_\lambda(x_-)|, |f_{\lambda'}(x_+) - f_\lambda(x_+)| < \varepsilon_2$$

for $\lambda' \in [0, 1]$. Overall, this implies

$$\begin{aligned} &|\lambda' - \lambda| < \delta_\lambda \;\Rightarrow\; |f'_{\lambda'}(x)| < 1 \quad \text{for all } x \in I, \\ &f_{\lambda'}(x_-) - x_- > 0, \qquad f_{\lambda'}(x_+) - x_+ < 0 \end{aligned}$$

with $\delta_\lambda := \min(\delta_1, \delta_2)$. Hence, according to the intermediate value theorem, we obtain

$$|\lambda' - \lambda| < \delta_\lambda \;\Rightarrow\; \exists \tilde{x} \in (x_-, x_+) \text{ with } f_{\lambda'}(\tilde{x}) = \tilde{x} \text{ and } |f'_{\lambda'}(\tilde{x})| < 1,$$

i.e. for small variations of λ there is always an asymptotically stable hyperbolic fixed point near the old equilibrium x^*. Moreover, as $|f'_{\lambda'}(x)| < 1$ for $|\lambda' - \lambda| < \delta_\lambda$ and $x \in I$, we know that $\frac{\partial}{\partial x} f_{\lambda'}(x) - x < 0$, i.e. $\tilde{x}$ is the only fixed point of $f_{\lambda'}$ on I for $|\lambda' - \lambda| < \delta_\lambda$. □

Proof of Theorem 2 We define $h_{\lambda_c'}(\tilde{q}) := g_{\tilde{q}}(\tilde{q}, \lambda_a, \lambda_c')$ for $\lambda_c' \in [0, 1]$. Lemma 1 guarantees the existence of $\delta_{\tilde{q}} > 0$ with

$$h_{\lambda_c}(\tilde{q}) - \tilde{q} \begin{cases} > 0 & \text{if } \tilde{q} \in [\tilde{q}^*(\lambda_a, \lambda_c) - \delta_{\tilde{q}}, \tilde{q}^*(\lambda_a, \lambda_c)), \\ < 0 & \text{if } \tilde{q} \in (\tilde{q}^*(\lambda_a, \lambda_c), \tilde{q}^*(\lambda_a, \lambda_c) + \delta_{\tilde{q}}]. \end{cases}$$

According to Lemma 2, there is further $\delta_{\lambda_c} > 0$ so that for $|\lambda_c' - \lambda_c| < \delta_{\lambda_c}$ there is a unique fixed point of $h_{\lambda'}$ in $[\tilde{q}^*(\lambda_a, \lambda_c) - \delta_{\tilde{q}}, \tilde{q}^*(\lambda_a, \lambda_c) + \delta_{\tilde{q}}]$ which we refer to as $\tilde{q}^*(\lambda_a, \lambda_c')$.

If $\lambda_c' \in [\lambda_c - \delta_{\lambda_c}, \lambda_c]$, Proposition 3 implies that $h_{\lambda_c'}(q) \leq h_{\lambda_c}(\tilde{q})$ for $\tilde{q} \in [0, 1]$, which leads to

$$h_{\lambda_c'}(\tilde{q}) - \tilde{q} \leq h_{\lambda_c}(\tilde{q}) - \tilde{q} < 0 \quad \text{for } \tilde{q} \in (q^*(\lambda_a, \lambda_c), q^*(\lambda_a, \lambda_c) + \delta_{\tilde{q}}].$$

Hence, there is no fixed point of $h_{\lambda_c'}$ in that interval and we thereby obtain

$$\tilde{q}^*(\lambda_a, \lambda_c') \in \left[\tilde{q}^*(\lambda_a, \lambda_c) - \delta_{\tilde{q}}, \tilde{q}^*(\lambda_a, \lambda_c)\right].$$

Analogously one can show that

$$\tilde{q}^*(\lambda_a, \lambda_c') \in \left[\tilde{q}^*(\lambda_a, \lambda_c), \tilde{q}^*(\lambda_a, \lambda_c) + \delta_{\tilde{q}}\right]$$

for $\lambda_c' \in [\lambda_c, \lambda_c + \delta_{\lambda_c}]$. □

Proof of Corollary 1 The statement is an immediate consequence of Theorem 2 and the fact that

$$\frac{\partial q^*}{\partial \lambda_c}(\lambda_a, \lambda_c) = \frac{\partial}{\partial \lambda_c} g\left(\hat{q}_a^*(\lambda_a, \lambda_c)\right) = g'\left(\hat{q}_a^*(\lambda_a, \lambda_c)\right)(1 - \lambda_a)\frac{\partial \tilde{q}^*}{\partial \lambda_c}(\lambda_a, \lambda_c) \leq 0$$

by using Proposition 2. □

References

Allport FH (1924) Social psychology. Houghton-Mifflin, Boston

Andreoni J, Erard B, Feinstein J (1998) Tax compliance. J Econ Lit 36(2):818–860

Banerjee AV (1992) A simple model of herd behavior. Q J Econ 107(3):797–817

Bendor J, Mookherjee D (1987) Institutional structure and the logic of ongoing collective action. Am Polit Sci Rev 81(1):129–154

Bicchieri C, Fukui Y (1999) The great illusion: ignorance, informational cascades, and the persistence of unpopular norms. Bus Ethics Q 9(1):127–155

Bikhchandani S, Hirshleifer D, Welch I (1992) A theory of fads, fashion, custom, and cultural-change as informational cascades. J Polit Econ 100(5):992–1026

Brissonneau C, Depiesse F (2006) Doping and doping control in French sport. In: Spitzer G (ed), Doping and doping control in Europe. Meyer & Meyer Sport, Aachen, pp 145–167

Centola D, Willer R, Macy M (2005) The emperor's dilemma: a computational model of self-enforcing norms. Am J Soc 110(4):1009–1040

Cohen S (1985) Visions of social control: crime, punishment, and classification. Polity Press, Blackwell

Diekmann A, Preisendörfer P (2003) The behavioral effects of environmental attitudes in low-cost and high-cost situations. Ration Soc 15(4):441–472

Elaydi SN (1996) An introduction to difference equations. Springer, New York

Elffers H, Weigel RH, Hessing DJ (1987) The consequences of different strategies for measuring tax evasion behavior. J Econ Psychol 8(3):311–337
Foucault M (1977) Discipline and punish: the birth of the prison. Pantheon Books, New York
Franke W, Berendonk B (1997) Hormonal doping and androgenization of athletes: a secret program of the German Democratic Republic government. Clin Chem 43(7):1262–1279
Fudenberg D, Tirole J (1991) Game theory. MIT Press, Cambridge
Garland D (2001) The culture of control: crime and social order in late modernity. Clarendon, Oxford
Gintis H (2008) Punishment and cooperation. Science 319(5868):1345–1346
Harsanyi JC (1967–1968) Games with incomplete information played by Bayesian players, parts I–III. Manag Sci 14:159–182 320–334, 486–502
Herrmann B, Thoni C, Gächter S (2008) Antisocial punishment across societies. Science 319(5868):1362–1367
Hudson B (2002) Punishment and control. In: Maguire M, Morgan R, Reiner R (eds), The Oxford handbook of Criminology. Oxford University Press, Oxford, pp 233–263
Kitts JA (2003) Egocentric bias or information management? Selective disclosure and the social roots of norm misperception. Soc Psychol Q 66(3):222–237
Kitts JA (2006) Collective action, rival incentives, and the emergence of antisocial norms. Am Soc Rev 71(2):235–259
Kitts JA (2008) Dynamics and stability of collective action norms. J Math Soc 32:142–163
Leonard J (2001) Doping in elite swimming: a case study of the modern era from 1970 forward. In: Wilson W, Derse E (eds), Doping in elite sport: the politics of drugs in the Olympic movement. Human Kinetics Publishers, Champaign, pp 225–239
Miller D, McFarland C (1987) Pluralistic ignorance: when similarity is interpreted as dissimilarity. J Personality Soc Psychol 53(2):298–305
Miller DT, Morrison KR (2009) Expressing deviant opinions: believing you are in the majority helps. J Exp Soc Psych 45(4):740–747
North DC (1986) The new institutional economics. J Inst Theor Econ 142:230–237
O'Gorman HJ (1986) The discovery of pluralistic ignorance: an ironic lesson. J Hist Behav Sci 22(4):333–347
Popitz H (1968) Über die Präventivwirkung des Nichtwissens: Dunkelziffer, Norm und Strafe. Mohr, Tübingen
Rauhut H (2009a) Higher punishment, less control? Experimental evidence on the inspection game. Ration Soc 21(3):359–392
Rauhut H (2009b) Stronger incentives for control reduce crime. A lab experiment on paradoxical effects of incentives and a game theoretical explanation. Mimeo ETH, Zürich
Rauhut H, Junker M (2009) Punishment deters crime because humans are bounded in their strategic decision-making. J Artif Soc Syst Soc 12(3)
Rauhut H, Krumpal I (2008) Die Durchsetzung sozialer Normen in Low-Cost und High-Cost Situationen. Z Soziol 5:380–402
Tsebelis G (1989) The abuse of probability in political analysis: the Robinson Crusoe fallacy. Am Polit Sci Rev 1:77–91
Tsebelis G (1990) Penalty has no impact on crime. A game theoretic analysis. Ration Soc 2:255–286
Willer R, Kuwabara K, Macy MW (2009) The false enforcement of unpopular norms'. Am J Sociol 115(2):451–490

Patrick Groeber is currently a Ph.D. student at the ETH Zurich, Chair of Systems Design. In his thesis, he investigates the emergence and enforcement of social norms and conventions by means of mathematical modeling, computer simulations and laboratory experiments. He earned a master degree in mathematics from the University of Bremen.

Heiko Rauhut is a Post-Doc researcher at the ETH Zurich, Chair of Sociology, in particular of Modeling and Simulation. His general interest concerns how social norms shape our lives. He is employing quantitative methodologies, such as laboratory experiments, survey data and agent based models for understanding of how social norms promote coordination and cooperation on the one hand, but conflict and collapse of social order on the other. Moreover, his interests expand to methodological questions such as quantitative social research methods, rational choice theory and empirical tests of game theoretical models. His current peer-reviewed publications deal with the enforcement of social norms (Zeitschrift für Soziologie, 2008), the strategic interaction between norm targets and norm guardians (Rationality and Society, 2009), and with learning dynamics in normative behavior (Journal of Artificial Social Systems and Societies, 2009). He is currently working on a book entitled 'Crime and Punishment from a Game Theoretical Perspective'.