International Review of Law and Economics 60 (2019) 105848

Contents lists available at ScienceDirect

International Review of Law and Economics

On esteem-based incentives☆

Ali Mazyaki[a,b], Joël van der Weele[c,*]

[a] Department of Economics, Allameh Tabataba'i University, Shahid Beheshti, Tehran, Iran
[b] Department of Economics, Institute for Management and Planning Studies, Niyavaran, Tehran, Iran
[c] University of Amsterdam and Tinbergen Institute, The Netherlands

ARTICLE INFO

Article history:
Received 23 August 2018
Received in revised form 20 May 2019
Accepted 25 June 2019
Available online 8 July 2019

JEL classification:
D02
H41
K42

Keywords:
Legal sanctions
Signaling
Incentives
Reputation

ABSTRACT

The rise of the internet, increased connectivity and higher availability of personal data increases the relevance of incentives based on reputation and the allocation of esteem. However, their use is controversial: critics argue that shaming can lead to a loss of control over the size of the sanction and to mob justice. We use the signaling model of social behavior by Bénabou and Tirole (2011) to explore the effect of esteem-based incentives and their interaction with traditional fines. We show that the use of esteem and stigma can indeed lead to a loss of control by generating multiple equilibria, some of which feature high levels of compliance and high levels of stigma. Moreover, the deterrent effect of monetary and esteem incentives is interdependent. If both types of incentives are costly to implement, esteem incentives should optimally be used relatively more for rare behaviors and in societies that have more heterogeneous values.

1. Introduction

The desire for a good reputation is an important driver of human behavior. Policy makers attempt to harness this motive through regulations that affect the allocation of esteem and stigma. For instance, law-makers determine the visibility of transgressions by the decision to either expunge criminal records or to make them available to employers and credit agencies. A more extreme example is the public registration of sex offenders, for instance via Megan's law in the United States.[1] With the advent of the internet, the availability of ever more personal information, social networks and modern surveillance techniques increases the opportunities to affect esteem. Several social credit score systems are currently being tested in China, which give each citizen a public score based on "good behavior" and apply (so far modest) penalties to low scores.[2]

Legal scholars have vigorously debated the desirability of incentives based on reputation. Proponents like Brennan and Pettit (2004) argue for increased use of esteem and stigma as it is a cheap and powerful tool of deterrence. Similarly, Kahan and Posner (1999: 366) argue that "Shaming [...] may offer a cost-effective and politically acceptable alternative to the short terms of imprisonment that such offenders now typically receive." By contrast, critics argue that shaming is a blunt and unpredictable instrument. By delegating punishment to the public, the effect of such sanctions is hard to control.[3] In an influential critique, Whitman (1998) writes:

☆ The authors thank Zachary Grossman for useful comments, Paolo Crosetto for software advice and Ivar Kolvoort for research assistance. Joel van der Weele gratefully acknowledges financial support through a personal VIDI grant from the Dutch Science Foundation.

* Corresponding author at: University of Amsterdam and Tinbergen Institute, The Netherlands.
E-mail addresses: mazyaki@atu.ac.ir (A. Mazyaki), vdweele@uva.nl (J. van der Weele).

[1] Other examples abound. Kahan (1996: 635) document a rise in shaming sanctions in the U.S. for a variety of offenses, taking forms such as visible community service, rituals for disgracing the offender, or forced wearing of symbols publicizing their crime. Other examples include judges' decisions to make offenders place yard signs or ads in local newspapers announcing their crimes, or special license plates for drunk drivers (see "Crime and Punishment: Shame Gains Popularity", by Jan Hoffman, NYT January 16, 1997). More recently, Daughety and Reinganum (2010) provide a list of policy examples, such as the UC Berkeley Law Department rule to limit information on class rankings, the shaming of those who waste water during draughts in Georgia, and the shaming of speeders by public display of license plates.

[2] See http://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion (accessed April 3rd, 2018).

[3] In addition, Nussbaum (2009) argues that shaming sanctions are cruel and illiberal. Kahan (2006) similarly retracts his earlier endorsement of shaming sanctions,

https://doi.org/10.1016/j.irle.2019.06.001
0144-8188/

> "There is no way to predict or control the way in which the public will deal with [an offender], no rhyme or limit to the terms the public may impose. Shame sanctions, in this regard, are very different from prisons or fines." (p. 1090–91).

Again, these concerns become more pertinent with the availability of electronic surveillance and the possibility to share information on social platforms. Hess and Waller (2014) and Ronson (2015) discuss the unpredictable nature of online trolling and mobbing, leading to outsized punishment for relatively small infractions.

In this paper, we analyze the use of esteem and the potential loss of control in a signaling model, based on Bénabou and Tirole (2011). This is a simple and tractable framework to capture reputation and compliance behavior that is increasingly popular in theoretical work on ethical decision making and social norms. In the model, an authority interacts with a continuum of agents, who decide whether to engage in an activity that is personally costly but generates positive externalities. Agents are assumed to have three motivations. First, they care about the personal (monetary) payoffs from the action. Second, agents have different "values" or "intrinsic motivation" for complying with the law, and this preference type is private information. Third, people care about their reputation, modeled as the expectation other people have about their values. This expectation is conditional on the observed action, giving rise to an incentive effect of reputation.

In this framework, we investigate the deterrent effect of two policy instruments: a traditional monetary incentive and a reputation-based "esteem incentive". The latter consists of an increase in the visibility of the agents actions, thereby increasing the amount of esteem that agents can reap by complying with formal rules, or conversely, the amount stigma associated with breaking them. We analyze the deterrent effect of both incentives in the context of a perfect Bayesian, semi-separating Nash equilibrium, and study how the two interact.

Our results provide several new insights to the use of esteem and stigma as a deterrence tool in economic and legal settings. First, we show the conditions under which both incentives are mutually reinforcing. This is the case for relatively high levels of compliance, where incentives raise the informativeness of behavior about the underlying character of the agent. Second, we show that unlike monetary incentives, increasing the visibility of actions can indeed lead to a "loss of control" for the authority. A high level of such incentives can induce multiple equilibria with different compliance levels, especially in situations where compliance is not too costly, and the distribution of values in society is relatively homogeneous.

Third, we show that if both incentives are costly to implement, reputation-based incentives are used relatively more for extreme behaviors, as these are most informative about the underlying character of the agent and hence yield a stronger deterrent effect. This explains why shaming is used mostly for rare and extreme behaviors like sex offenses. We also show that reputation-based incentives are used relatively more in a society with more heterogeneous or polarized values, as in such societies actions reflect larger differences moral values, and hence generate more stigma.

With these results, we contribute to the literature on stigma and shame in legal policy, which we review below. Contrary to previous literature, our model abstracts from the exact use of reputation, for instance in the labor market. In return, it allows a clear characterization of how penalties based on esteem interact with more traditional incentives, the relative use of different policy instruments, and the conditions under which a "loss of control" can occur. While these insights are only one step in understanding the complex interplay of different policies, they are important in a world with ever increasing connectivity.

2. Reputation incentives in the economics literature

Economists have traditionally focused on monetary incentives as the main tool to influence behavior and reach policy goals. The analysis of esteem has been picked up only recently, both empirically and theoretically. On the empirical side, there has been a lot of study into the effects of esteem from peers, showing that it promotes pro-social behavior both in the lab (e.g. Rege and Telle, 2004; Andreoni and Petrie, 2004; Andreoni and Bernheim, 2009; Ariely et al., 2009) and in the field (e.g. Harbaugh, 1998; Lacetera and Macis, 2010; Karlan and McConnell, 2014).

On the theoretical side, our paper is based on the model by Bénabou and Tirole (2011), who investigate how the presence of reputational concerns influences optimal monetary incentives. They give a central to the case where the authority has superior information about the distribution of values in society, making sanctions a signal of the distribution (see also Sliwka, 2007 and Van der Weele, 2012a). This fits in a wider literature that analyses the impact of visibility on the effectiveness of monetary incentives (e.g. Bénabou and Tirole, 2006; Ariely et al., 2009; Bowles and Polania-Reyes, 2012). Bénabou and Tirole (2011) focus on the optimal level of monetary incentives, and assume that esteem concerns are "not too high" to guarantee uniqueness of equilibrium. By contrast, we explicitly explore the multiplicity of equilibria resulting from incentives based on stigma and esteem.

A few papers analyze the use of esteem incentives or variations in privacy in a signaling context. Bénabou and Tirole (2006) show that higher visibility of legal (or prosocial) actions increases compliance by inducing low types to behave better. The policy is partially self-defeating however, since the additional compliance generated by reputational concerns weakens the signal of altruism sent by complying. Daughety and Reinganum (2010) explore the tradeoffs between incentives provided by visibility and the conformism this induces on agent behavior, which leads to possible over-investment in the public good. Jann and Schottmüller (2016) show that reduced privacy leads to impaired information aggregation. However, none of these studies address the potential loss of control from esteem-based sanctions, or the optimal joint level of the two incentives.

Like our paper, Ali and Bénabou (2016) also use a signaling model to study the "loss of control". In their model, agents have individual knowledge about the "quality" of different public goods, which may change over time. The effect of increasing visibility is to increase contributions to currently valued public causes. However, this increase in compliance obscures dynamic shifts in the quality of different public goods, which in turn introduces uncertainty about the optimal level of sanctions and the strength of disapproval generated by a given level of visibility. By contrast, in our setup preferences are fixed, and the unpredictability of the effect of esteem-based incentives arises because they can lead to multiple equilibria.

In the field of law and economics, there is a mostly theoretical literature investigating the interaction between legal rules and informal norms cemented by mechanisms of esteem and reputation (for overviews, see e.g. Kahan, 1997; Ellickson, 1998; Posner, 2000, and Van der Weele, 2012a). Like the present study, Harel and Klement (2007) study the relationship between the use and intensity of stigma. They show that liberal use of stigma undermines its value as a deterrent as employers will start hiring stigmatized people. Dur and van der Weele (2013) show that penalties raising the cost of particular criminal activities will change the signal associ-

on the ground that community based punishments have a populist underpinning that may undermine affirmation of individualistic values.

ated with those activities, causing subtle spillover effects on other criminal activities. Mungan (2016a) shows that criminalization of acts that have low correlation with having a bad character, or, as an extreme case of this, the conviction of innocent people (Mungan, 2017), reduces the stigma associated with having a criminal record, and hence it's deterrent effect. Cooter and Porat (2001) discuss how reputation loss of convicts should affect awarded damages.

Several papers look at the impact of deterrence on stigma in a labor market context using models of asymmetric information (e.g. Rasmusen, 1996; Funk, 2004). Iacobucci (2014) also uses the logic of signaling to point out that an increase in legal sanctions may affect the esteem associated with breaking the law, an idea that also underlies the current paper. Mungan (2016b) generalizes the models of Iacobucci (2014) and Rasmusen (1996), demonstrating that there is an ambiguous effect of formal sanctions on reputational sanctions. Compared to these papers, we do not focus of the effect of labor market institutions on optimal deterrence. Instead, we study the deterrent effect of both monetary incentives and esteem incentives and outline the interaction between the two in more detail.

3. Model

The model in this section, as well as the notation, is almost identical to that by Bénabou and Tirole (2011, henceforth BT). There is a large population of agents of measure 1, each of which simultaneously decides whether to comply with the law ($a_v = 1$) or not ($a_v = 0$). We consider two policy instruments that an authority can use to influence this decision and deter non-compliance. One instrument is a *monetary penalty* of size $y \geq 0$, which applies to lawbreakers only. Note that we do not explicitly model any monitoring effort or uncertainty in getting the penalty, so y can be interpreted as an expected penalty. The other incentive is an *esteem incentive* of size $s \geq 0$, which influences the visibility of compliance and lawbreaking in the community. Thus, a higher s may reflect the public naming or shaming of offenders.

Each agent has a preference type v, defined by the following utility function, which depends on her own action a_v and the actions of the other types a_{-v}:

$$U_v(a_v, a_{-v}) = \underbrace{[c-y](1-a_v) + e\bar{a}}_{\text{Material}} + \underbrace{va_v}_{\text{Intrinsic}} + \underbrace{s\mu E(v|a_v)}_{\text{Esteem}}. \quad (1)$$

The utility function has four components. First, compliance leads to a monetary benefit of y, since no penalty is incurred, minus some personal cost c associated with compliance (i.e. the benefit of crime). Second, we assume complying with the law has a positive externality, so each agent benefits proportionally to the fraction of compliers the population $\bar{a}$. Here, the parameter $e \geq 0$ measures the importance of the externality. Third, the 'type' of the agent v reflects the 'intrinsic utility' that an agent gets when she complies with the law, and can be interpreted as the degree of 'altruism' or 'civic-mindedness' of the agent. Agents are distributed over the type space according to the continuous cdf $F(v)$ with full support on $v \in [\underline{v}, \bar{v}]$.

The final component of utility is a concern for "reputation" or "esteem", which is the inferred value of her type by the external observer(s). We assume that each agent's type v is private information and cannot be directly observed by others. However, observers can condition their inference on the agent's action. These inferences are determined endogenously in equilibrium by Bayes' rule, based on the equilibrium profile of actions by each type of agent (see below). The parameter $\mu > 0$ measures the importance of such reputation or esteem to the agent. Reputation is also multiplied by the policy parameter s that determines visibility. We model μ and s separately, to make the point that the authority does not have perfect control over the importance of esteem.

So far we have used the words "esteem", "stigma" and "reputation", so it is worth clarifying these terms. Our use of esteem as a policy instrument is closely related to that of a "reputational sanction", as defined in Iacobucci (2014) and Mungan (2016b). Iacobucci (2014) takes such a sanction to arise from "pure self-interest", because reputation damage will reduce attractiveness as a trading partner. In our paper, the term $s\mu E(v|a_v)$ can also be interpreted as a reputational sanction, that is, as a reduced form expression for the continuation value of reputation in future (unmodeled) interactions. By making actions more widely visible, the government can influence the importance of reputation in such interactions. We take "esteem" to be the currency of reputation, where higher esteem means a better reputation and vice versa. By contrast, we will sometimes use "stigma" to refer to the converse of esteem, i.e. the negative reputation associated with law violations. In this, we are consistent with Mungan (2016b), who argues that "stigma" also falls in the category of reputational sanctions. Note that our definition of esteem can be broadened beyond that of external reputation, as $s\mu E(v|a_v)$ may also describe psychological losses from feelings of shame.[4]

The next part of the paper will consider the effect of exogenous changes in the incentives s and y on agents' decisions. In Section 7 we look at an authority that sets sanctions before agents take their decision in order to maximize welfare. In Section 8 we discuss an alternative interpretation of the model in a workplace context.

4. Equilibrium and the esteem premium

We apply the solution concept of (perfect) Bayesian Nash equilibria. Each agent's (or type's) action maximizes her expected utility given her beliefs about the strategies of the other agents and the corresponding inferences of the observers, that are formed by Bayes' rule. That is, every agent of type v chooses $a_v^* \in \text{argmax}_{\{0,1\}} U_v(a_v, a_{-v}^*)$. We focus on equilibria characterized by threshold type $v^* \in [\underline{v}, \bar{v}]$ such that types $v \geq v^*$ prefer comply with the law, and types $v < v^*$ do not. In an equilibrium with threshold type v^*, the corresponding compliance level is $a^* = 1 - F(v^*)$. The behavior in this equilibrium is indeed optimal for each type if and only if the following equilibrium condition (EC) is satisfied

$$\begin{aligned} v^* + y - c + s\mu E[v \mid v > v^*] &= s\mu E[v \mid v < v^*] \\ v^* &= c - y - s\mu\Delta(v^*). \end{aligned} \quad (EC)$$

Here, $\Delta(v^*) := E[v \mid v > v^*] - E[v \mid v < v^*]$ is the difference between the expected type of those who break the law and those who do not, and is a measure of the informativeness of behavior. In the remainder, we will refer to $\Delta(v^*)$ as the *esteem premium*, as it reflects the gain in esteem that is associated with behaving legal compliance, or conversely, the stigma associated with breaking it.

The threshold type v^* is the type who is just indifferent between incurring the net cost of compliance and obtaining esteem $\Delta(v^*)$. From the EC it follows that the threshold v^* will be in the interior if $\underline{v} \leq c - y - s\Delta(\underline{v})$ and $c - y - s\Delta(\bar{v}) \leq \bar{v}$. If either of these two conditions is violated, this will result in a pooling equilibrium with either no compliance or full compliance, respectively. Existence of the threshold is guaranteed by the continuity of $\Delta(v^*)$, and the fact that $\Delta(v^*)$ is positive and bounded.[5]

[4] The government may influence the size of such feelings either through the visibility of actions, or through campaigns emphasizing the importance of good character (Kaplow and Shavell, 2007).

[5] Pooling equilibria may exist alongside the threshold equilibrium, where the former are supported by low off-equilibrium beliefs for either compliance or non-compliance. Pooling equilibria on non-compliance are not likely to survive standard

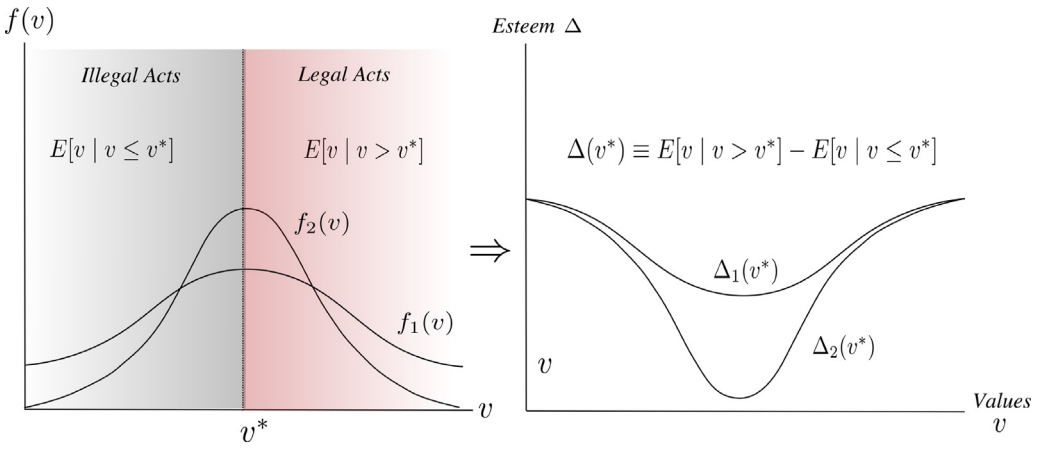


Fig. 1. The esteem function $\Delta(v^*)$ of the truncated normal distribution. The function gets flatter if the variance σ_v^2 increases.

The shape of esteem premium plays an important role in the analysis and is closely related to the distribution of values $F(v)$. Here, we will make the assumption that values are distributed according to a normal distribution, i.e. we assume that v is distributed $v \sim \mathcal{N}(0, \sigma_v^2)$ with support on $[\underline{v}, \bar{v}]$, with $\underline{v} = -\bar{v}$.[6] The focus on the normal distribution seems reasonable, as in most communities members will share moral values and their intrinsic valuation of different actions will not vary widely. That is, the further the intrinsic motivation is from the average, the less frequent it is. However, the parameters of the distribution may differ between communities. For instance, highly diverse communities, perhaps made up of different ethnicities or nationalities may have a wider dispersion of values than a homogeneous community with little outside influence. These differences can be captured by the distribution variance σ_v^2, which will play an important role in our analysis. Note that our qualitative findings do not depend on the assumption of a normal distribution, as any single peaked symmetric functions with an interior maximum will yield similar results.[7]

Under these assumptions, we can derive the following formula for the esteem function:

Lemma 1. *For the truncated normal distribution, the esteem function is given by*

$$\Delta(v^*) = \frac{\sigma_v^2 h(v^*)}{F(v^*)}, \tag{2}$$

where $h(v^*) := \frac{f(v^*)}{1-F(v^*)}$. This function is graphed in Fig. 1. The higher the variance of the truncated normal, the flatter and higher the esteem function becomes, because for any intermediate levels of v^*, it is more likely that the type of the observed agent is somewhere in two tails. Note that in the extreme case, where the distribution is uniform, the esteem function is constant.

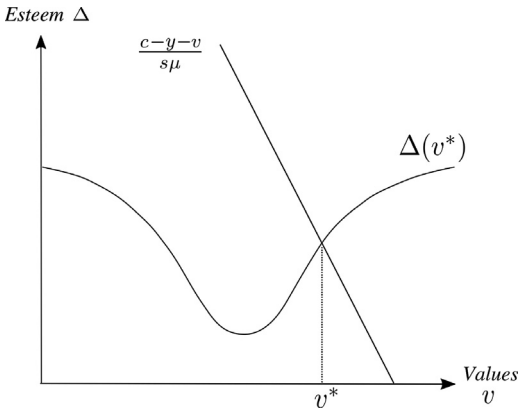


Fig. 2. Equilibrium is found on the intersection of the (straight) EC-line and the esteem premium $\Delta(v^*)$.

The shape of the esteem function reflects the change in esteem when the compliance level in society changes. When average compliance is low (v^* is high), compliance is very informative as it signals an exceptionally high type. Similarly, illegal actions generate strong negative esteem when v^* is low, as they signal a very low type. By contrast, when v^* is in an intermediate range, illegal actions convey relatively little information and the esteem premium is low. Fig. 1 also shows that the information conveyed by criminal acts is muted when the tails of the distribution get thicker, as the esteem premium gets flatter if the variance σ_v^2 increases.

The equilibrium condition (EC) is depicted graphically in Fig. 2. The straight, downward sloping line in Fig. 2 is given by $\frac{c-v-y}{s\mu}$, that we will refer to as the EC-line. It represents all pairs $(v^*, \Delta(v^*))$ such that the threshold type is exactly indifferent between being complying or not, given the material costs c of contributing as well as the levels of y and s. Thus, equilibrium of the game is found on the intersection of this line with the $\Delta(v^*)$ curve. As we show below, policies y and s will determine compliance levels by shifting the EC line.

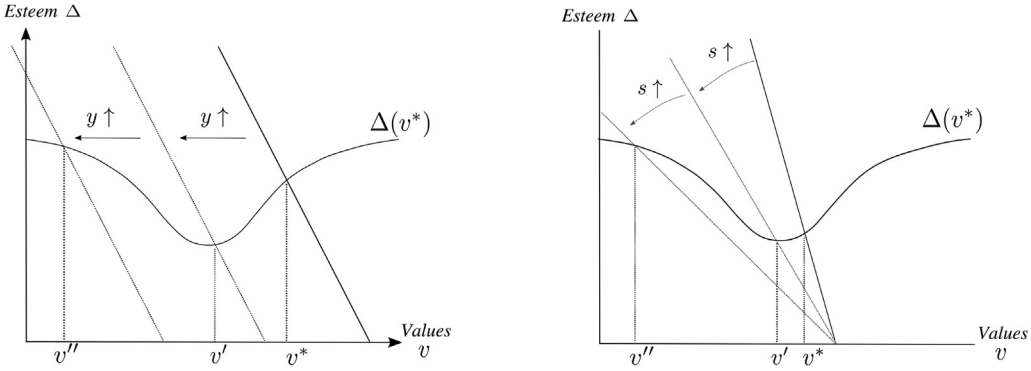
5. Incentives and compliance in equilibrium

We will derive our results in two steps. In the first step we investigate the effect of incentives on the equilibrium compliance level. This provides insight into how different incentives affect behavior, and whether they reinforce or dampen the deterrent effect of

equilibrium refinements like D1, as it is mostly in the interests of high types to comply. Pooling equilibria on full compliance may exist, but as Adriani and Sonderegger (2015) point out, forward induction arguments will favor the semi-separating equilibrium we study here.

[6] Technically, our normal distribution is truncated, but we will assume that $F(\underline{v})$ and $1 - F(\bar{v})$ are small, and abstract from it in our analysis. To have a density function on $[\underline{v}, \bar{v}]$, we have to normalize the density by dividing by $1 - 2F(\underline{v})$. Abstracting from these normalizations does not affect our qualitative results. Moreover, to have $\Delta(v^*)$ defined and continuous at the boundaries we define $\Delta(\bar{v}) := \lim_{v \to \bar{v}} \Delta(v)$ and $\Delta(\underline{v}) := \lim_{v \to \underline{v}} \Delta(v)$. By setting $E_v = 0$, the aggregate value of reputation is 0. Thus, increasing visibility does not affect the total amount of esteem, and we avoid the question whether esteem is a good or bad thing in itself.

[7] As proved in Jewitt (2004) and BT, the esteem function has an interior minimum whenever $f(v)$ has an interior maximum. Adriani and Sonderegger (2015) provide an extensive discussion of the relation between the shape of the type distribution and the esteem function.



(a) The monetary incentive shifts the EC curve to the left.

(b) The esteem incentive rotates the EC curve around its intersection with the x-axis.

Fig. 3. Graphical illustration of the effect of the monetary incentive (left panel) and the esteem incentive (right panel) on the compliance level.

the other incentive. In the next section we take the second step and derive the optimal government incentive, given the reaction of the agents. We focus on interior equilibria in which the first order conditions hold.

5.1. Monetary incentives and compliance

We first derive the effectiveness of monetary incentives on compliance. In any interior equilibrium, we can compute the first order derivative of the equilibrium compliance level a^* with respect to the monetary incentive y.

$$\frac{\partial a^*}{\partial y} = -f(v^*)\frac{\partial v^*}{\partial y} = \frac{f(v^*)}{1 + s\mu\Delta'(v^*)}. \tag{3}$$

This expression replicates Eq. (6) in Bénabou and Tirole (2011: 7). It shows that the effect of the monetary incentive depends on the slope of the esteem premium at the threshold, $\Delta'(v^*)$. If $\Delta'(v^*) > 0$, which is the case for low compliance rates (or high v^*), the effect of esteem counteracts the effect of the incentive. The reason is that the shift in behavior induced by monetary incentives "dilutes" the expected type of the small number of compliant people more quickly than it dilutes the expected type of the non-compliant majority. Conversely, for high levels of compliance, the esteem premium increases with the amount of compliance, as the expected type of the remaining criminals drops quickly with the threshold. The strength of these multiplier effects depend on the importance of esteem or stigma, which is determined by $s\mu$.

The left panel of Fig. 3 demonstrates these results graphically. An increase in y shifts the EC line leftward. The shift in the equilibrium threshold depends on the slope of the esteem premium. A negative multiplier reduces the effects of a shift in the EC line that occurs on the upward-sloping part of the esteem function. The consequence is a modest increase in compliance from v^* to v'. An equivalent raise in y that occurs on the downward sloping part of the esteem function benefits from a positive multiplier, and hence produces a much larger shift in compliance from v' to v''.

5.2. Esteem incentives and compliance

We next investigate the corresponding effect of reputation incentives on compliance. In any interior equilibrium,

$$\frac{\partial a^*}{\partial s} = -f(v^*)\frac{\partial v^*}{\partial s} = \frac{f(v^*)\mu\Delta(v^*)}{1 + s\mu\Delta'(v^*)}. \tag{4}$$

For the esteem incentive, the multiplier effect in the denominator is the same as for the monetary incentive. However, there is an additional (numerator) effect which depends on the *level* of esteem $\Delta(v^*)$. The higher the esteem premium, the higher the effect of raising the visibility of actions.

Both effects can be verified in the right panel of Fig. 3, which graphically demonstrates the effect of the esteem incentive. An increase in s rotates the EC line inwards around the intersection with the v-axis. Increased visibility means that the esteem premium necessary to convince the agent to comply is now lower for any level of intrinsic values v. Again, the shift in the equilibrium threshold depends on the slope of the esteem premium. A rotation in the EC line that occurs on the low and upward-sloping part of the esteem function produces only a modest increase in compliance from v^* to v'. An equivalent raise in y that occurs on the high and downward sloping part of the esteem function produces a much larger shift in compliance from v' to v''.

Thus, depending on the levels of μ and $\Delta(v^*)$, the esteem-based incentive can either be more or less effective than the monetary incentive. Generally, it will be most effective when esteem is high, which is the case for either very low or very high compliance levels.

5.3. Mutual reinforcement between esteem and monetary incentives

Our previous results already showed that the two incentives are interdependent. We now investigate this interdependence in more detail, by looking at the crossderivative $\frac{\partial^2 v^*}{\partial y \partial s}$. If this expression is negative, then an increase in one of the incentives makes the other incentive more effective in increasing compliance direction and we call the incentives "mutual reinforcers".

Proposition 1. *The two incentives are mutual reinforcers if and only if*

$$\Delta'(v^*) < -s\frac{\partial v^*}{\partial s}\Delta''(v^*). \tag{5}$$

The condition in Proposition 1 implies that reinforcement depends on both the first and second derivative of the esteem premium. These derivatives matter for different reasons. The reason the first derivative matters is that the esteem incentive depends on the informativeness of actions, i.e. the height of the esteem premium. If actions are uninformative, i.e. $\Delta(v^*)$ is low, increasing visibility does not do much to increase deterrence. Thus, a (monetary) incentive that increases the esteem premium by decreasing v^* (i.e. $\Delta'(v^*) < 0$) will make the esteem incentive s more efficient. The reason the second derivative appears in (5) because the effectiveness of both types of incentives depends on the *slope* of the esteem premium, as shown above. Thus, if increasing compliance makes the esteem premium more negatively sloped (increase faster

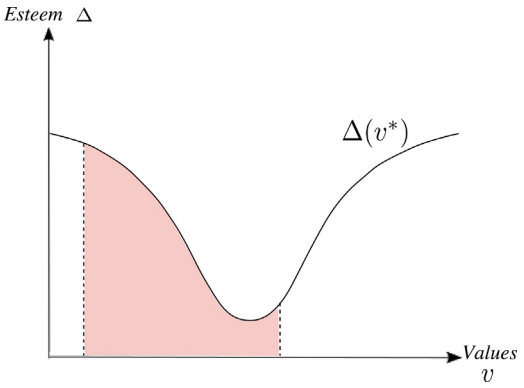


Fig. 4. The two incentives reinforce each other only in the shaded region.

with compliance), this will increase the social multiplier of both incentives.

Fig. 4 demonstrates graphically the shaded area where the two incentives are mutual reinforcers. Either incentive increases the effect of the other incentive only when compliance is relatively high, i.e. for actions where v^* is low. Roughly speaking, an increase in compliance in this area results in a more informative signal, raising the level and the slope of $\Delta(v^*)$. By contrast, if compliance is low, the effect of increasing compliance is a decrease in the esteem-premium, which makes both incentives less effective. Similarly, if compliance is very high, raising incentives further flattens the esteem premium, lowering the effectiveness of both incentives.

Summary 1. The effect of both monetary and esteem incentives depends on the slope of the esteem premium, which depends on the amount of compliance. In addition, the effect of esteem incentives depends on the size of the esteem premium, which is higher for extreme actions. The two incentives are reinforcers if compliance is relatively high, as the esteem premium increases in the level of compliance this domain.

6. Multiple equilibria and loss of control

We now turn to the potential for a "loss of control" that various legal scholars have associated with the use of stigma. We operationalize the idea of "loss of control" by looking at the existence of multiple interior equilibrium points v^*, which make the effect of incentives fundamentally unpredictable.[8] Before we develop our mathematical results, we illustrate the main ideas graphically in Fig. 5. The left panel shows how an increase in s can lead to multiple equilibria. In the shaded zone, the EC line cuts the $\Delta(v^*)$ three times, twice from above and once from below. Of these three equilibrium points, only the most extreme ones, marked A and C, are stable.[9]

Fig. 5 elucidates the conditions for multiple equilibria to occur. First, the esteem premium must decline steeply for relatively high compliance levels, i.e. increase fast with compliance. Moreover, the EC line needs to be relatively flat, that is, esteem must be important enough. If these conditions are satisfied, there is a possibility of co-existence of equilibria with either low levels of compliance (high v^*) and a low esteem premium $\Delta(v^*)$, and equilibria with high compliance and a high esteem premium.

Multiplicity of equilibria is an obvious problem for policy-making. Unless there are grounds to predict that agents can coordinate on a given equilibrium, welfare maximization becomes impossible. The right panel of Fig. 5 illustrates this problem, by showing the compliance levels associated with the different equilibria points A and C depicted in the left panel as a function of s: for $s > \underline{s}$, the authority cannot predict the level of compliance and thus suffers a loss of control.

Fig. 5 allows a few more observations. First, the occurrence of multiple equilibria necessitates relatively high levels of the esteem incentive, such that the EC line is relatively flat. When s is low and the EC line is near vertical, multiple equilibria cannot occur for any level of the monetary incentive y. It is thus clear that a loss of control is indeed associated with high stigma, in line with the intuition of legal scholars.

Second, the additional equilibria that appear when s increases have relatively high compliance levels and are associated with a high esteem premium. Thus, they can be considered a form of mobbing or "crowd justice", where increased visibility leads to high levels of shaming or stigma for a small minority, even for relatively minor offenses (for examples see Hess and Waller, 2014, and Ronson, 2015). While such strong applications of esteem generate high compliance, this is not necessarily efficient, as compliance may overshoot the optimal level, as we discuss in more detail below. Thus, these multiple equilibria reflect the concerns of critics like Whitman (1998, cited in the introduction) that outsourcing justice to the crowd may lead to unpredictable and heavy punishment.

Finally, note that an increase in σ_v^2 flattens the esteem premium, as discussed in Fig. 1. Thus, multiple equilibria are more likely to occur in populations with more homogeneous values. The intuition here is that when types are very concentrated, changes in compliance will have a large impact on the expected types for each given action, and hence on the esteem premium. Thus, tightly-knit communities of like-minded people might be more prone to instance of mobbing for deviant behavior.

6.1. Uniqueness

We now investigate the conditions for multiple equilibria to occur more precisely. Lemma 2 characterizes a sufficient condition for uniqueness of an interior fixed point v^*, for any level of the policy variables s and y.

Lemma 2. *Let* $\hat{v} := \text{argmin}_{v \in [\underline{v}, \bar{v}]} \Delta'(v)$. *If*

$$\Delta'(\hat{v}) > \frac{\Delta(\hat{v})}{\hat{v} - c}, \tag{6}$$

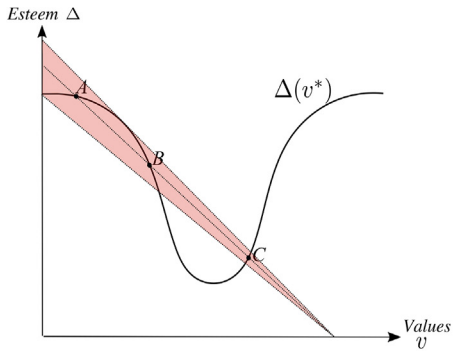
there is at most one internal equilibrium threshold $v^* \in (\underline{v}, \bar{v})$ *satisfying (EC), for all* $s, y \geq 0$.

Two parameters affect if (6) is likely to hold. First, the derivative $\Delta'(\hat{v})$ becomes more negative in a society with homogeneous values, i.e. a small σ_v^2. Thus, uniqueness is more likely to hold when types are dispersed, since the esteem premium does not vary much with the compliance level. This prevents the simultaneous existence of equilibria with low compliance (high v^*) and low esteem $\Delta(v^*)$, or with high compliance and a high esteem.

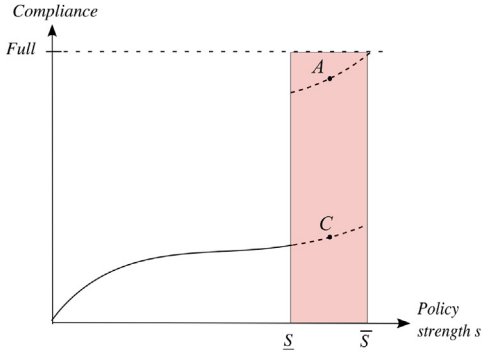
Second, (6) is more likely to hold if c is low, since the right hand side of (6) is now small or positive. The intuition is that multiple

[8] In BT (2011, p. 6), multiple equilibria are ruled out by imposing the condition that μ (the importance of esteem to the agent) is "not too large". This assumption makes sense in the context of their investigation of monetary incentives, but not in the present paper, which investigates the effect of (potentially high) stigma.

[9] Intersection point B is also an equilibrium, but it is unstable. To see why, note that when compliance declines a bit from B (e.g. v^* shifts upward, perhaps because the marginal types make a mistake), the esteem premium is now lower than the EC line. This means that compliance becomes less attractive, and additional, inframarginal types will shift to non-compliance. This logic leads compliance to unravel until a new stable point is reached at C. Similarly, if compliance shifts upward slightly from B, the esteem premium is now higher than the EC line, so compliance becomes more attractive and will increase until point A is reached. These arguments are reversed for the stable equilibrium point A or C. For instance, if compliance increases slightly from A, the esteem premium falls below the EC line, making compliance less attractive and pushing behavior back to the equilibrium point.



(a) When raising visibility, multiple equilibria occur in the shaded region. Only the extreme intersection points (like A and C) represent stable equilibria.



(b) Compliance levels as a function of the level of the esteem incentive. Only compliance levels of stable equilibria are depicted. Point A and C correspond to equilibria depicted in left panel.

Fig. 5. (a) When raising visibility, multiple equilibria occur in the shaded region. Only the extreme intersection points (like A and C) represent stable equilibria. (b) Compliance levels as a function of the level of the esteem incentive. Only compliance levels of stable equilibria are depicted. Point A and C correspond to equilibria depicted in left panel.

equilibria display either high and intermediate levels of compliance, since esteem rises fastest with compliance in this region. If costs are low, compliance is already relatively high, so no additional equilibria exists for intermediate compliance levels.

6.2. Loss of control

When Lemma 2 is not satisfied, high levels of the esteem incentive may lead to multiple equilibria and a loss control for the authority, as illustrated graphically in Fig. 5. The following formal result makes this intuition more precise, where again we use the definition $\hat{v} := \text{argmin}_{v \in [\underline{v}, \bar{v}]} \Delta'(v)$.

Proposition 2. *If*

$$\hat{v} - \frac{\Delta(\hat{v})}{\Delta'(\hat{v})} < c - y < \underline{v} - \frac{\Delta(\underline{v})}{\Delta'(\underline{v})}, \tag{7}$$

there exist $\underline{s} > 0$ *and* $\bar{s} > \underline{s}$*, such that there exist multiple stable equilibrium thresholds if* $s \in [\underline{s}, \bar{s}]$.

Proposition 2 shows that for multiple equilibria to occur, the monetary incentive for compliance $c - y$ can neither be too low nor too high. Both a graphical and mathematical proof are provided in the Appendix. The first inequality of (7) is the converse of (6) when $y = 0$, showing that (6) is indeed necessary to guarantee uniqueness for all y, $s > 0$. The second inequality in Proposition 2 implies that multiple equilibria cannot occur if monetary costs of compliance $c - y$ are extremely high. In this case, one needs very high levels of visibility s to generate compliance in the region around $\hat{v}$, where potential multiple equilibria occur. Depending on the slope of the esteem curve at $\underline{v}$, there may be no stable interior equilibrium to the left of $\hat{v}$. In this case, there may be an additional stable equilibrium featuring pooling by all types on compliance.[10]

Summary 2. Esteem sanctions may lead to a loss of control for the authority for intermediate monetary incentives for compliance $c - y$, and a society with relatively homogeneous values, i.e. a small σ_v^2. The additional equilibria are associated with high compliance and high levels of stigma. Thus, the model bears out the criticism that increasing visibility and/or shame may result in unpredictable levels of sanctions.

7. Optimal incentives and welfare

In this section we consider the decisions of a government or an authority. We assume that government maximizes aggregate social welfare, given by the function

$$W(y, s) = \bar{U} + (1 - \bar{a})y - (1 - \bar{a})(c_y(y) + c_s(s)) \tag{8}$$

where $\bar{U}$ is the average utility of the population. The term $(1 - \bar{a})y$ reflects the income from monetary sanctions applied to non-compliant agents, while the last term reflects the cost of operating the different incentive schemes. Costs are multiplied by $(1 - \bar{a})$, because incentives are only applied to agents who break the law.[11] We do not make explicit assumptions on the cost functions $c_y(y)$ and $c_s(s)$, but we will argue below that an interior equilibrium is likely to exist only if both $c_y(y)$ and $c_s(s)$ are sufficiently convex. Convex costs may arise as increased incentives may lead to higher efforts to conceal bad behavior and hence higher costs of implementing the sanction. Moreover, the political costs of reducing privacy are likely to be increasing in the amount of visibility.

We assume the government first sets incentive y and s, and then each agent chooses her action $a_v \in \{0, 1\}$, after which payoffs are realized. When multiple equilibria occur, the maximization of the welfare function may not be possible without imposing further assumptions. In the following, we therefore assume that the uniqueness condition (6) holds. Thus, we only consider the case where the esteem premium Δ is relatively flat, and assume away any problems related to the loss of control discussed above.

To study the nature of this equilibrium, we first show the properties of the optimal incentives in the cases $c_y = c_s = 0$. In this case, it does not matter which incentive is used by the authority. Thus the analysis is identical to finding optimal monetary incentive, as done in BT (Proposition 1). We can take the first order condition of the welfare condition with respect to y^* and set it to zero. This yields $v^* + e = c$, showing that the marginal cost of an additional contribution is equal to the marginal benefit, consisting of the externality plus the moral benefit of the threshold type. Substituting this optimality condition into the EC yields

$$y^* + s^* \mu \Delta(c - e) = e. \tag{9}$$

Eq. (9) shows that y^* and s^* can be traded off at rate $\mu\Delta$. Thus, if $\Delta(v^*)$ is high, i.e. only very few people comply or very few people do not comply, a small increase in s^* leads to a relatively large drop

[10] While throughout this analysis we have focused on interior equilibria, one may of course characterize such a situation as one with multiple equilibria. In that sense, the second inequality in Proposition 2 can be considered a less stringent condition.

[11] BT assume an almost identical social welfare function with linear costs.

in y^*. By contrast, if $\Delta(v^*)$ is low, i.e. about half of the population complies, a small increase in s^* is compensated by only a small drop in y^*. The reason is that the social multiplier on s is much higher in the former case.

For the second-best case where implementing both incentives is costly, we are not able to derive explicit expressions for the individual level of the two incentives. The existence of an interior equilibrium requires that Δ is relatively flat, implying a rather stable marginal effect of s and y on compliance as shown by (3) and (4). Moreover, the cost functions $c_y(y)$ and $c_s(s)$ need to be sufficiently convex so that each policy is used in equilibrium.[12]

If these conditions are satisfied, the first-order conditions are sufficient and we can derive the following result.

Proposition 3. *In any interior equilibrium where the first order conditions are sufficient for a maximum, the optimal y and s of incentives satisfy*

$$\frac{c_s'(s^*)}{c_y'(y^*)} = \mu\,\Delta(v^*). \tag{10}$$

The expression in Proposition 3 equates the marginal benefits of each incentive to its marginal cost, where we know from Propositions (3) and (4) that the increase in compliance from a unit increase in s^* is $\mu\Delta(v^*)$ times the increase in compliance due to a unit increase in y^*.

Under the assumption that $c_y(y)$ and $c_s(s)$ are convex, Proposition 3 has several important implications. First, the use of the esteem incentive s is associated with 'extreme' levels of v^*, i.e. behaviors that either very few or very many people do.[13] The reason is that in this case the esteem premium is very high, and a change in visibility therefore has a large incentive effect. This rationalizes the use of shaming sanctions for very deviant acts.

Second, an exogenous increase in the importance of esteem μ implies a higher relative use of esteem-based incentives. While not surprising as a result, this implies that shaming sanctions are most effective in close-knit communities where news spreads fast and interactions are repeated. Furthermore, it gives support for an increased use of shaming sanctions for white-collar criminals (Skeel, 2001). White-collar workers such as businessmen have much more to lose from a ruined reputation, which is essential to secure business contacts.

Finally, an interesting corollary of Proposition 3 is that when values v become more heterogeneous the relative use of esteem-based sanctions increases. To make this more precise, we formally define heterogeneity.

Definition 1. Society 1 is more heterogeneous than Society 2 if $F_1(v)$ second-order stochastically dominates $F_2(v)$.

For the case of the normal distribution, this definition implies that an increase in σ^2 (while keeping the mean fixed), results in an increase in heterogeneity of a society. We can now make use of a result derived in Adriani and Sonderegger (2015, Lemma 3): if distribution $F_1(v)$ second-order stochastically dominates distribution $F_2(v)$ and the two distributions have an identical mean, then $\Delta_1(v) < \Delta_2(v)$ for all $v \in [\underline{v}, \bar{v}]$.

Corollary 1. *If Society 1 is more heterogeneous than Society 2, $s_1^* > s_2^*$ in any interior equilibrium.*

The intuition behind the result is simple: a more heterogeneous society will have higher levels of esteem $\Delta(v^*)$, as for any given v^*, the conditional expectations of types of compliant and non-compliant agents are further apart. Thus, an increase in the polarization of values, perhaps due to more ethnic diversity or cultural disagreements, implies an increased use of reputation and shame relative to other kinds of punishments.[14]

Summary 3. If the first order conditions are sufficient and both policies are costly to implement, both esteem and monetary incentives are used in equilibrium. Esteem incentives are associated with more extreme behaviors and are used more when the distribution of values is more heterogeneous.

8. Application to prosocial and organizational behavior

Throughout the paper, we have emphasized our results in terms of crime and legal sanctions. However, the model can also be applied to prosocial behavior in organizational contexts. In such an interpretation, the authority consists of (a team of) managers and the "citizens" can be employees. Our model fits these environments when $a_v = 1$ is interpreted a prosocial act, and $a_v = 0$ as an antisocial, but not necessarily illegal act. With regard to the incentives, y is a subsidy for pro-social behavior, while s reflects the presence of leaderboards or employee of the month displays, as well as ceremonies and symbolic rewards for virtuous behavior.

Our results above go through unaltered in such an interpretation. As before, esteem-incentives are associated with extreme behavior, so the model helps us understand why Medals of honor and awards are only given out for exceptionally virtuous behavior, not for simply doing one's job.

These results contribute to a recent strand of literature that investigates the use of symbolic rewards as motivators in such environments. Neckermann and Frey (2013) show that providing the prospect of an award has significant effect on stated willingness to contribute to a public good, especially if accompanied by a public ceremony. Kosfeld and Neckermann (2011) show that awards have a considerable impact on work effort in a laboratory environment (see also Kosfeld and Neckermann, 2011; Bradler et al., 2016). Markham et al. (2002) provides evidence that public recognition boosts attendance in a large manufacturing firm. In the management literature there is ample support for the idea that public recognition is a key motivator of employee performance (e.g. Holton et al., 2009). Ashraf and Bandiera (2018) discuss recent empirical literature showing the importance of the interaction of monetary and social incentives.

9. Discussion and conclusion

In this paper we have considered a framework to analyze the deterrent effect of both monetary and reputation-based incentives. This framework rationalizes some existing intuitions and provides a number of new insights. First, we show that the effect of esteem incentives and monetary incentives both depend on the compliance level. They reinforce each other when levels of compliance behavior are relatively high, as in this case they make illegal acts a more informative signal.

Second, the use of esteem and stigma can indeed lead to a loss of control for the authority, as critics of such incentives have pointed out. When values are homogeneous and the monetary incentives for compliance are not too high, ramping up levels of visibility

[12] One can show that some optimal policy exists by using the extreme value theorem. However, the exact conditions for the existence of an interior equilibrium are implicit as they depend on the shape of Δ and the cost functions $c_y(y)$ and $c_s(s)$.

[13] This result depends crucially on the shape of the esteem function, and hence the distribution of types. The same qualitative results from any single-peaked function with declining density on either side of the mode, see Jewitt (2004) and Adriani and Sonderegger (2015).

[14] Note that we are assuming that while intrinsic values or motivations for a given action are further spread out when polarization increases, agents still agree on which actions are worthy of esteem.

can lead to coexistence of equilibria with high levels of compliance and high levels of stigma for deviant actions, and equilibria with lower stigma and lower compliance. This captures incidents of "mob justice" or the "digital pillory", i.e. the unpredictable effect of shaming.

Third, the model provides some support for the conjecture by Kosfeld and Neckermann (2011: 97), who write that "it is likely that social status and monetary aspects reinforce each other and that optimal incentives are based on the combination of social as well as monetary elements." While we show that the two instruments are mutual reinforcers only for relatively high levels of compliance, we show that under some regularity assumptions, both incentives should indeed be used by the authority.

Fourth, we find that esteem incentives are associated in equilibrium with rare actions, for which esteem concerns are high. In this case, because extreme behaviors are committed by extreme types, actions are informative about character and esteem incentives are more effective. This is in accordance with observed real world practices. For instance, names and addresses of offenders are published only for extremely undesirable behavior like sex offenses against children, as several American States do under the so-called 'Megan's law', and the U.K. government (with some limitations) does under 'Sarah's law'.

Finally, an interesting implication of the model is that the heterogeneity of moral values in society matters for the optimal policy mix. Specifically, relatively homogeneous values imply a lower esteem premium, reducing the level of esteem incentives. Moreover, homogeneous values also make the esteem premium steeper on some parts of the domain and are more likely to lead to a loss of control.

The interplay between esteem and financial incentives offers much scope for further research. In this paper we have analyzed the deterrent effects of both incentive schemes. However, as Kahan (1996)) stresses, the expressive value of both types of sanctions may be equally important. To this end, one can analyze how the optimal policy mix is affected by information asymmetries between the authority and the agents about the distribution of values in society, as considered for example in Sliwka (2007), Bénabou and Tirole (2011), and Van der Weele (2012b). Moreover, the interplay between the two kinds of sanctions may be more intricate than we have assumed, as the size of the financial sanction may itself influence the visibility of an action and the amount of esteem associated with it. Another issue that would be interesting to incorporate in the model is recidivism. While reputational punishments may deter crime or anti-social behavior, it may also encourage recidivism by lowering the outside options of ex-offenders (Funk, 2004).

Appendix A. Appendix with proofs

Proof *(Proof of Lemma 1).* Suppose v has a truncated normal distribution $\mathcal{N}(0,\sigma_v^2)$ in $[\underline{v},\bar{v}]$ such that $\underline{v}=-\bar{v}$. The distribution is given by $f(v)=\frac{\alpha}{\sigma\sqrt{2\pi}}e^{-\frac{v^2}{2\sigma_v^2}}$ in which α is a multiplier that ensures the density to sum up to one.

It's first derivative is given by $f'(v)=-\frac{v}{\sigma_v^2}f(v)$, so we can write $\int_{\underline{v}}^{v^*} vf(v)dv=\int_{\underline{v}}^{v^*}-\sigma_v^2 f'(v)dv=\sigma_v(f(\underline{v})-f(v^*))$.

This implies that $E(v|v<v^*)=\frac{\int_{\underline{v}}^{v^*} vf(v)dv}{F(v^*)}=\frac{\sigma_v^2(f(\underline{v})-f(v^*))}{F(v^*)}$. Similarly, we can write $E(v|v>v^*)=\frac{\sigma_v^2(f(v^*)-f(\bar{v}))}{1-F(v^*)}$. Using symmetry $(f(\underline{v})=f(\bar{v}))$ and the definition of $\Delta(v^*):=E(v|v>v^*)-E(v|v<v^*)$, we find:

$$\Delta(v^*)=\frac{\sigma_v^2(f(v^*)-f(\underline{v}))}{(1-F(v^*))F(v^*)}.$$

We will approximate $f(\underline{v})$ to be zero, which will be true if $\underline{v}$ is small enough. Then, the first derivate can be written as

$$\Delta'(v^*)=-\frac{\Delta(v^*)}{\sigma_v^2}v^*+\frac{(\Delta(v^*))^2}{\sigma_v^2}(2F(v^*)-1).$$

□

Proof *(Proof of Proposition 1).* The two incentives are reinforcers if and only if $\frac{\partial^2 v^*(y,s)}{\partial y\partial s}<0$.

Since $\frac{\partial^2 v^*(y,s)}{\partial y\partial s}=\mu\frac{\Delta'(v^*)+s\mu(\Delta'(v^*))^2-s\mu\Delta(v^*)\Delta''(v^*)}{(s\mu\Delta'(v^*)+1)^2}$ the two incentives are reinforcers if and only if

$$\begin{aligned}\Delta'(v^*)+s\mu(\Delta'(v^*))^2-s\mu\Delta(v^*)\Delta''(v^*) &<0\\ \Delta'(v^*)\left(-\frac{1}{\frac{\partial v^*}{\partial s}}\right)-s\Delta''(v^*) &<0\\ \Delta'(v^*) &<-s\frac{\partial v^*}{\partial s}\Delta''(v^*).\end{aligned}$$

□

Proof *(Proof of Lemma 2).* We start with a few remarks. First, we define v^T as the threshold type associated with the internal equilibrium with the highest compliance level. Note that this equilibrium is always stable since $\Delta(\bar{v})>0$ and the EC line given by $\frac{-v-y+c}{s\mu}$ slopes downward. Thus, the EC line crosses $\Delta(v)$ from above, which is a sufficient and necessary condition for a stable equilibrium. Second, it is easy to verify that $\Delta'(\hat{v})$ is decreasing on $[\underline{v},\hat{v}]$ and $[-\hat{v},\bar{v}]$ and increasing on $[\hat{v},-\hat{v}]$. Third, our proof assumes $y=0$. This is in fact a sufficient condition, as it follows from the EC condition that we can replace c by $c-y$ whenever $y>0$, which will not cause a violation of (6).

We now show that (6) is sufficient for uniqueness. To rule out additional equilibria with compliance level below v^T, it is sufficient that the slope of $\Delta(v)$ is higher (less negative) on $[\underline{v},v^T]$ than the slope of the EC, where the latter is given by $\frac{\Delta(v^T)}{v^T-c}$. We now confirm that this is the case, where we restrict our analysis to $v^T<\hat{v}-\frac{\Delta(\hat{v})}{\Delta'(\hat{v})}$, since $v^T>\hat{v}-\frac{\Delta(\hat{v})}{\Delta'(\hat{v})}$ implies $c>\hat{v}-\frac{\Delta(\hat{v})}{\Delta'(\hat{v})}$, which violates (6). We distinguish two cases:

1 Suppose $v^T\in[\underline{v},\hat{v}]$. Since $\Delta'(\hat{v})$ is decreasing on this interval, $\Delta'(\hat{v})$ is lower than the slope of the EC on $[\bar{v},v^T]$. This rules out the existence of a second equilibrium to the left of v^T, and hence v^T is unique.

2 Suppose $v^T\in\left[\hat{v},\hat{v}-\frac{\Delta(\hat{v})}{\Delta(\hat{v})}\right]$. In this case, a sufficient condition for uniqueness is

$$\begin{aligned}\Delta'(\hat{v}) &>\frac{\Delta(v^T)}{v^T-c},\\ c &<v^T-\frac{\Delta(v^T)}{\Delta'(\hat{v})}.\end{aligned}\tag{A.11}$$

For this case, (6) implies $c < \hat{v} - \frac{\Delta(\hat{v})}{\Delta'(\hat{v})}$. Thus (6) implies (A.11) if

$$\begin{aligned} \hat{v} - \frac{\Delta(\hat{v})}{\Delta'(\hat{v})} &< v^T - \frac{\Delta(v^T)}{\Delta'(\hat{v})}, \\ \Delta'(\hat{v}) &< \frac{\Delta(\hat{v}) - \Delta(v^T)}{\hat{v} - v^T}. \end{aligned} \tag{A.12}$$

By the definition of $\hat{v}$, $\Delta'(\hat{v})$ is lower than $\Delta'(v)$ for all v, so (A.12) always holds. This insures that the slope of the EC is smaller than $\Delta'(\hat{v})$, and rules out the existence of a second equilibrium to the left of v^T. Hence, v^T is unique.

□

Proof *(Proof of Proposition 2).* A graphical illustration of this proof is provided below. To establish multiple equilibria, it is sufficient to find an $\underline{s}$ and $v_1 \in (\underline{v}, \hat{v}]$ such that the EC line is exactly tangent to $\Delta(v)$ at $(v_1, \Delta(v_1))$. Since $\Delta(v)$ is concave on $[\underline{v}, \hat{v}]$ as we noted in Lemma 2, there exists an $\epsilon > 0$ that pivots the EC line downward such that $\underline{s} + \epsilon$ will result in intersections between the EC and $\Delta(v)$ on either side of v_1. The intersection with the threshold $v < v_1$ is associated with a stable equilibrium, while the intersection with the threshold $v > v_1$ is associated with an unstable equilibrium as the EC crosses $\Delta(v^*)$ from below. However, since $\Delta(\bar{v}) > 0$ and the EC slopes downward, there exists another stable equilibrium where the EC crosses $\Delta(v^*)$ from above, associated with a higher equilibrium threshold than that of the unstable equilibrium. This establishes multiplicity.

We now investigate the conditions for the existence of $\underline{s}$. First, consider an equilibrium $v^* = \underline{v}$. The EC line crosses $\Delta(v)$ from above at $(\underline{v}, \Delta(\underline{v}))$ if $\Delta'(\underline{v}) > -\frac{\Delta(\underline{v})}{c-y-\underline{v}}$ or

$$c - y < \underline{v} - \frac{\Delta(\underline{v})}{\Delta'(\underline{v})}. \tag{A.13}$$

Second, consider an equilibrium $v^* = \hat{v}$. The EC crosses $\Delta(\hat{v})$ from below at $(\hat{v}, \Delta(\hat{v}))$ if

$$c - y > \hat{v} - \frac{\Delta(\hat{v})}{\Delta'(\hat{v})}, \tag{A.14}$$

(see also the proof of Lemma 2).

The concavity of $\Delta(v)$ on $[\underline{v}, \hat{v}]$ implies that

$$\begin{aligned} \frac{\Delta(\hat{v}) - \Delta(\underline{v})}{\hat{v} - \underline{v}} &< \Delta'(\underline{v}) \\ \frac{\Delta(\hat{v}) - \Delta(\underline{v})}{\Delta'(\underline{v})} &> \hat{v} - \underline{v} \\ \frac{\Delta(\hat{v})}{\Delta'(\hat{v})} - \frac{\Delta(\underline{v})}{\Delta'(\underline{v})} &> \hat{v} - \underline{v} \\ \underline{v} - \frac{\Delta(\underline{v})}{\Delta'(\underline{v})} &> \hat{v} - \frac{\Delta(\hat{v})}{\Delta'(\hat{v})}. \end{aligned} \tag{A.15}$$

Thus, there exists an nonempty interval V_0, such that if $c - y \in V_0$, both (6) and (A.14) are satisfied.

Consider $c - y \in V_0$. Since the slope of the EC equals $-\frac{1}{\mu s}$ and is continuous on $(-\infty, 0)$, and since $\Delta'(\hat{v}) < \Delta'(\underline{v})$, there exists an $\underline{s}$ such that the EC is exactly tangent to $\Delta(v)$.

Finally, the strict concavity of $\Delta(v)$ implies that the width of V_0 is positive and bounded away from zero. Therefore we can find an $\bar{s} > \underline{s}$ and bounded away from $\underline{s}$ such that multiple equilibria exist for any $s \in [\underline{s}, \bar{s}]$.

The proof is illustrated in Fig. A1. If $c - y$ lies in the shaded area, multiple equilibria will occur for some values of s.□

Proof *(Proof of Proposition 3).* We denote the equilibrium by $v^* = v^*(y, s)$. Welfare maybe rewritten by

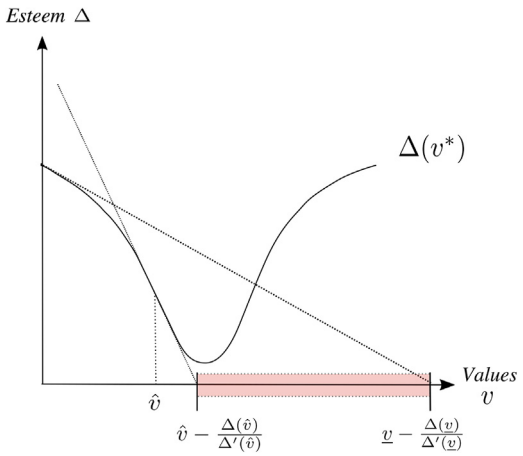


Fig. A1. When $c - y$ lies in the shaded area, multiple equilibria exists for some values of s.

$$W(y, s, v^*) = \int_{v^*}^{\bar{v}} (v - c + e) f(v) dv - \int_{\underline{v}}^{v^*} (c_y(y) + c_s(s)) f(v) dv. \tag{A.16}$$

We derive first order conditions of the welfare function w.r.t. s and y:

$$\frac{\partial W}{\partial y} = -\frac{\partial v^*}{\partial y} f(v^*)[e + v^* - c - c_y(y) - c_s(s)] - c_y'(y) F(v^*) = 0$$

$$\frac{\partial W}{\partial s} = -\frac{\partial v^*}{\partial s} f(v^*)[e + v^* - c - c_y(y) - s_y(s)] - c_s'(s) F(v^*) = 0$$

Combining (A.17) and (A.18), we obtain

$$\frac{\frac{\partial v^*}{\partial s}}{\frac{\partial v^*}{\partial y}} = \frac{c_y'(y^*)}{c_s'(s^*)}$$

Substituting the equations from Propositions 1 and 2, we obtain the desired result.

□

References

Adriani, F., Sonderegger, S., 2015. A theory of esteem-based peer pressure. Mimeo, University of Leicester.

Ali, N.S., Bénabou, R., 2016. Image Versus Information: Changing Societal Norms and Optimal Privacy (No. w22203). National Bureau of Economic Research.

Andreoni, J., Bernheim, D.B., 2009. Social esteem and the 50–50 norm: a theoretical and experimental analysis of audience effects. Econometrica 77 (5), 1607–1636.

Andreoni, J., Petrie, R., 2004. Public goods experiments without confidentiality: a glimpse into fund-raising. J. Public Econ. 88, 1605–1623.

Ariely, D., Bracha, A., Meier, S., 2009. Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. Am. Econ. Rev. 99 (1), 544–555.

Ashraf, N., Bandiera, O., 2018. Social incentives in organizations. Ann. Rev. Econ. 10, 439–463.

Bénabou, R., Tirole, J., 2006. Incentives and prosocial behavior. Am. Econ. Rev. 96 (5), 1652–1678.

Bénabou, R., Tirole, J., 2011. Law and Norms. NBER working paper 17579.

Bowles, S., Polania-Reyes, S., 2012. Economic incentives and social preferences: substitutes or complements? J. Econ. Liter. 50 (2), 368–425.

Bradler, C., Dur, R., Neckermann, S., Non, A., 2016. Employee recognition and performance: a field experiment. Manage. Sci. 62 (11), 3085–3099.

Brennan, G., Pettit, P., 2004. The Economy of Esteem: An Essay on Civil and Political Society. Oxford University Press.

Cooter, R., Porat, A., 2001. Should Courts Deduct Nonlegal Sanctions from Damages? J. Legal Stud. 30, 401–422.

Daughety, A.F., Reinganum, J.F., 2010. Public Goods, Social Pressure, and the Choice Between Privacy and Publicity. Am. Econ. J.: Microecon. 2, 191–221.

Jewitt, I., 2004. Notes on the Shape of Distributions, unpublished manuscript.

Karlan, D., McConnell, M.A., 2014. 2014 Hey look at me: the effect of giving circles on giving. J. Econ. Behav. Organ. 106, 402–412.
Dur, R., van der Weele, J.J., 2013. Status-seeking in criminal subcultures and the double dividend of zero-tolerance. J. Publ. Econ. Theor. 15 (1), 77–93.
Ellickson, R., 1998. Law and economics discovers social norms. J. Legal Stud. 27 (2), 537–552.
Funk, P., 2004. On the effective use of stigma as a crime deterrent. Eur. Econ. Rev. 48 (4), 715–728.
Harbaugh, W.T., 1998. What do donations buy? A model of philanthropy based on prestige and warm glow. J. Publ. Econ. 67, 269–284.
Harel, A., Klement, A., 2007. The economics of stigma: why more detection of crime may result in less stigmatization. J. Legal Stud. 36 (2), 355–377.
Hess, K., Waller, L., 2014. The digital pillory: media shaming of 'ordinary' people for minor crimes. Continuum: J. Media Cult. Stud. 28 (1), 101–110.
Holton, V., Dent, F., Rabbetts, J., 2009. Motivation and Employee Engagement in the 21st Century: A Survey of Management Views. Ashridge.
Iacobucci, E.M., 2014. On the interaction between legal and reputational sanctions. J. Legal Stud. 43 (1), 189–207.
Jann, O., Schottmüller, C., 2016. An informational theory of privacy. TILEC Discussion Paper No. 2016–030.
Kahan, D., 1996. What do alternative sanctions mean? Chicago Law Rev. 63, 591–653.
Kahan, D., 1997. Social influence, social meaning and deterrence. Virginia Law Rev. 83 (2), 349–395.
Kahan, D., 2006. What's really wrong with shaming sanctions. Texas Law Rev. 84, 2075.
Kahan, D.M., Posner, E.A., 1999. Shaming white-collar criminals: a proposal for reform of the federal sentencing guidelines. J. Law Econ. 42, 365–391.
Kaplow, L., Shavell, S., 2007. Moral rules, the moral sentiments and behavior: toward a theory of an optimal moral system. J. Polit. Econ. 115 (3), 494–514.
Kosfeld, M., Neckermann, S.S., 2011. Getting more work for nothing? Symbolic awards and worker performance. Am. Econ. J.: Microecon. 3, 86–99.
Lacetera, N., Macis, M., 2010. Social image concerns and prosocial behavior: field evidence from a nonlinear incentive scheme. J. Econ. Behav. Organ. 76, 225–237.
Markham, S.E., Dow Scott, K., McKee, G.H., 2002. Recognizing good attendance: a longitudinal, quasi-experimental field study. Person. Psychol. 55 (3), 639–660.
Mungan, M.C., 2016a. Stigma dilution and over-criminalization. Am. Law Econ. Rev. 18, 88–121.
Mungan, M.C., 2016b. A generalized model of reputational sanctions and the (ir)relevance of the interactions between legal and reputational sanctions. Int. Rev. Law Econ. 46, 86–92.
Mungan, M.C., 2017. Wrongful convictions, deterrence, and stigma dilution. Supreme Court Econ. Rev. 25, 199–2016.
Neckermann, S., Frey, B., 2013. And the winner is..? The motivating power of employee awards. J. Socio-Econ. 46, 66–77.
Nussbaum, M.C., 2009. Hiding from Humanity: Disgust, Shame, and the Law. Princeton University Press.
Posner, E.A., 2000. Law and Social Norms. Harvard University Press, Cambridge.
Rasmusen, E., 1996. Stigma and self-fulfilling expectations of criminality. J. Law Econ. 39 (2), 519–544.
Rege, M., Telle, K., 2004. The impact of social approval and framing on cooperation in public good situations. J. Publ. Econ. 88, 1625–1644.
Ronson, J., 2015. So You've Been Publicly Shamed. Picador, Londen.
Skeel, D.A., 2001. Shaming in Corporate Law. Univ. Pensylvania Law Rev. 149, 1811–1869.
Sliwka, D., 2007. Trust as a signal of a social norm and the hidden costs of incentive schemes. Am. Econ. Rev. 97 (3), 999–1012.
Van der Weele, J.J., 2012a. Beyond the state of nature: introducing social interactions in the economic model of crime. Rev. Law Econ. 8 (1), 401–432.
Van der Weele, J.J., 2012b. The signaling power of sanctions in social dilemmas. J. Law Econ. Organ. 28 (1), 103–125.
Whitman, J.Q., 1998. What is wrong with inflicting shame sanctions. Yale Law J. 107 (4), 1055–1092.