SOCIAL NORMS AND THE ENFORCEMENT OF LAWS

Daron Acemoglu
Massachusetts Institute of Technology

Matthew O. Jackson
Stanford University,
and Santa Fe Institute

Abstract
We examine the interplay between social norms and the enforcement of laws. Agents choose a behavior (e.g., tax evasion, production of low-quality products, corruption, harassing behavior, substance abuse, etc.) and then are randomly matched with another agent. There are complementarities in behaviors so that an agent's payoff decreases with the mismatch between her behavior and her partner's, and with overall negative externalities created by the behavior of others. A law is an upper bound (cap) on behavior. A law-breaker, when detected, pays a fine and has her behavior forced down to the level of the law. Equilibrium law-breaking depends on social norms because detection relies, at least in part, on whistle-blowing. Law-abiding agents have an incentive to whistle-blow on a law-breaking partner because this reduces the mismatch with their partners' behaviors as well as the negative externalities. When laws are in conflict with norms and many agents are breaking the law, each agent anticipates little whistle-blowing and is more likely to also break the law. Tighter laws (banning more behaviors), greater fines, and better public enforcement, all have counteracting effects, reducing behavior among law-abiding individuals but increasing it among law-breakers. We show that laws that are in strong conflict with prevailing social norms may backfire, whereas gradual tightening of laws can be more effective in influencing social norms and behavior. (JEL: C72, C73, P16, Z1)

1. Introduction

Laws often go unenforced because they conflict with prevailing social norms. For example, many British laws went unenforced, or became badly distorted in the colonies, because they contradicted the local social norms and legal customs (e.g., see Barfield 2010 on Afghanistan and Parsons 2010 on India and Kenya). In other

The editor in charge of this paper was Nicola Gennaioli.

Acknowledgments: We thank Pascual Restrepo for excellent research assistance and comments. We also gratefully acknowledge financial support from the NSF grants SES-0961481 and SES-1155302, and ARO MURI Award No. W911NF-12-1-0509. We thank various seminar participants for comments, and Harry Di Pei, Qianjun Lyu, Bentley MacLeod, Stephen Nei, and Eric Rasmusen for valuable suggestions. Acemoglu and Jackson are Senior Fellows at CIFAR.

E-mail: daron@mit.edu (Acemoglu); jacksonm@stanford.edu (Jackson)

Journal of the European Economic Association April 2017 15(2):245–295 DOI: 10.1093/jeea/jvw006

instances, however, social norms appear to have changed significantly and rapidly following the introduction of certain laws. A telling example is the impact of civil rights legislation on norms and practices in the US South, where as recently as the late 1950s, racism, racial stereotypes, discrimination, and racial slurs were highly common (e.g., Woodward 1955; Wright 2013). But the enforcement of federal antidiscrimination and antiracist laws, even if not completely eliminating such behaviors, fundamentally changed the norms, with transformative effects on economic decisions, language, and social relations.

Similar issues arise in the context of economic decisions. Authorities often promulgate laws to discourage tax evasion or production of low-quality, unsafe products by businesses. But there are huge differences in the success of such laws across societies. For instance, the IMF (International Monetary Fund) estimates that, in 2011, 30% of taxes were evaded in Greece, whereas the same number was 7% in the United Kingdom (IMF 2013; HM Revenues and Customs 2013). This is not only because a large part of economic transactions in Greece take place in the shadow economy (29.5% of GDP in Greece between 1996 and 2006, compared to 12.9% in the United Kingdom, see Schneider, Buehn, and Montenegro 2010), but also because of differences in social norms.[1] Because authorities lack the resources to audit more than a trivial fraction of businesses to detect low-quality products and tax evasion, they rely on whistle-blowing by private citizens and other businesses. Suppose, for example, that after a choice related to the quality of products or tax evasion, each producer matches with another to form a business partnership. Business partners observe each other's behavior and can whistle-blow on behavior outside of the law. Low-quality, unreliable products, or tax evasion, create negative externalities on the rest of society, and in addition, a mismatch between two producers in terms of product quality or how much of their business is illicit creates a cost for each (each will have to adjust to the different behavior of their partner, which tends to be costly). These costs create an additional incentive for some businesses and individuals to whistle-blow to force law-breakers' behavior in line with the law, reducing the mismatch costs and the negative externality they suffer. But if potential whistle-blowers themselves are law-breakers, such whistle-blowing may be discouraged.

We introduce a model of the interplay between law enforcement and social norms, focusing on this role of cooperation from private citizens with law enforcement in the form of "whistle-blowing". We follow a common definition of social norms in sociology as "a rule or a standard that governs our conduct in the social situations in which we participate. It's a societal expectation." (Bierstedt 1963, p. 222). More specifically, a social norm (or simply a norm) in our model is the distribution of anticipated payoff-relevant behavior. We show that when laws conflict with prevailing norms—for instance, they attempt to restrict behavior excessively relative to the distribution in the society—then most people prefer to break the law. This then results in less whistle-blowing, which reduces the effectiveness of laws and encourages further

1. The New Yorker, for example, writes: "The reason tax reform will be such a tall order for Greece, in sum, is that it requires more than a policy shift; it requires a cultural shift." (2011).

law-breaking. Thus, as laws are broken by more people, whistle-blowing becomes even less likely, and law-breaking snowballs.

Though there are various facets of interactions between laws and norms, we believe that the aspects emphasized by our model are both important and novel. In the examples mentioned above, the effectiveness of laws is curtailed by lack of private cooperation; and, as shown by the statistics in Section 2.3, whistle-blowing plays a prevalent and broad role in law enforcement.[2]

We start our analysis with a static version of our model. Agents choose a behavior and then are matched uniformly at random with another agent. Utility depends negatively on the average behavior of other agents (higher levels of the behavior are "bad" for society) and on the mismatch between the behaviors of the two partners (there is a complementarity in behavior across people). A (simple) law bans behaviors above a certain threshold, denoted by L in the paper, and a law-breaker, when detected, pays a fine and has her behavior forced down to L. Law-abiding agents have an incentive to whistle-blow because this will reduce their partner's behavior, curtailing the negative externality and ameliorating the mismatch. When others break the law, each agent has further incentives to do so as well, because there will be less whistle-blowing. This introduces the first major interaction between social norms and laws, which can lead to multiple equilibria. Despite the potential multiplicity of equilibria, the model is tractable and enables a tight characterization of all equilibria in terms of a threshold for law-breaking, and a range of comparative statics of the lowest compliance equilibrium. For example, tighter laws (banning more behaviors) lead to greater law-breaking and lower behavior among law-abiding agents. In contrast, greater fines for law-breaking reduce law-breaking, but interestingly, increase behavior among law-breakers—because the composition of law-breakers changes toward higher types, and since each law-breaker chooses their behavior with the hope of matching other law-breakers, this changing composition induces each law-breaker to choose a higher behavior. Given the interdependence in incentives, it can also be that slight changes in laws or fines result in discontinuously large changes in behavior.

Though we start with a static model to illustrate the workings of the model, some of our most important results, and those that relate to our motivation above, come from our dynamic model, in which each individual matches with agents from both previous and future generations. Now prevailing norms are given by the distribution of behavior in the previous generation (as in Acemoglu and Jackson 2015). After establishing that steady-state equilibria of this dynamic model are identical to the equilibria in our static setup, we study the dynamic interaction between social norms and laws. Echoing some of our discussion above, we show that laws that are in strong conflict with existing norms backfire: abrupt tightening of laws causes significant lawlessness, whereas gradual imposition of laws that are more in accord with prevailing norms

2. Our model focuses on the interplay between laws and norms, and thus laws and fines are modeled in a simple form (rigid caps and one level of fine) and taken as exogenous. Enriching laws and their origins is an important area for further study as we discuss in Section 7.

can successfully change behavior and thus future norms.[3] The dynamic model also generates "social multipliers" in law-breaking (as emphasized in, inter alia, Glaeser, Sacerdote, and Scheinkman 1996): once there is high law-breaking, in the next period there will be less private cooperation with law enforcement, increasing law-breaking further.

In an extension, we also show how different types of laws interact when law-breakers in one type of behavior are discouraged from whistle-blowing on another type of behavior. This interaction leads to less law-breaking in one behavior when there is a loose law (as opposed to no law) in the other behavior because law-abiding gives an option to the agent to whistle-blow on her partner in the other behavior if her partner's action is very high. However, our analysis further shows that badly-designed—excessively tight—laws for one type of behavior (e.g., small-scale drug crime in inner cities) can make laws against other types of behaviors completely ineffective (e.g., laws against larceny or gangs).[4]

Our theory also suggests why laws relying on private enforcement (whistle-blowing) can be more effective in some activities than others depending on the degree of "assortative matching". When law-breakers are more likely to match with other law-breakers (which is typically the case in activities such as dueling, smuggling, or racketeering), the power of private enforcement is more limited than in activities where businesses breaking the law are more likely to match with law-abiding businesses such as tax evasion or investment in the safety and reliability of products.

Our paper relates to a number of distinct literatures. First, there is a large law and economics literature focusing on the design of rules, punishments and monitoring structures in order to discourage certain types of behavior. This literature pioneered by Becker (1968) and Becker and Stigler (1974), and surveyed, for example, in Shavell (2004), typically does not investigate how private attitudes affect enforcement.

Our paper is more closely related to a small literature on multiple equilibria in law enforcement and law-abiding behavior.[5] One branch of this literature (e.g., Sah

3. Using the French Revolution as an example, Acemoglu et al. (2012) point to potential effectiveness of radical institutional change related to breaking the existing political equilibrium. The advantage of gradual changes in laws here abstracts from this political channel.

4. This extension also provides a perspective on the debate on the broken windows theory of Kelling and Wilson (1982), which claimed that the high incidence of serious crime in inner cities was a result of permissive attitudes toward small-scale crimes such as vandalism, graffiti, fare-dodging on the subway, or certain types of antisocial behavior. That theory calls for much stricter enforcement of laws against small-scale crimes, and was the inspiration of the tough policing strategies used in high-crime cities such as New York. Our theory suggests that pervasive law-breaking on such things as drugs or small-scale vandalism can indeed spill into law-breaking in other dimensions. Yet crucially, it sees the faultline not in the fact that there is such behavior in inner cities, but in the presence of laws that criminalize a large fraction of society in these areas. It would be much more effective, according to our theory, to decriminalize such behaviors, so that a large fraction of individuals do not automatically become law-breakers and can have greater willingness to cooperate with law enforcement in other dimensions of behavior of greater import to society.

5. There is also a vast literature on social norms and culture, with roots in the writings of Simmel (1903, 1908), Sorokin (1947), Morris (1956), Williams (1960), Bierstedt (1963), and Gibbs (1965) and a

1991; Acemoglu 1995; Glaesar, Sacerdote, and Scheinkman 1996; Rasmusen 1996; Calvo-Armengol and Zenou 2004; Ferer 2008; Aldashev et al. 2012) shows how law-breaking, corruption and rent-seeking type activities become profitable when others engage in such behavior, potentially leading to multiple equilibria.

Third, our work is also related to several recent works modeling the evolution of culture, social norms, and cooperation, such as Acemoglu and Jackson (2013), Bisin and Verdier (2001), Doepke and Zilibotti (2008), Galor (2011), Tabellini (2008), and Voth and Voigtlander (2012). In Tabellini, for example, moral values affect whether there is cooperation in a prisoner's dilemma type game, and parents' decisions concerning which types of values to inculcate in their children, along the lines of Bisin and Verdier's (2001) approach, is affected by the prevailing set of values in other agents in society. Also related in this context are papers emphasizing the emergence of equilibrium social norms without formal legal institutions (e.g., Ellickson 1991; Bernstein 1992; Pistor 1996).

A more recent fourth branch is even more closely related to our work, and investigates the interactions between laws and social norms. Benabou and Tirole (2011) develop a model in which norms encourage certain types of behavior because of agents' desire to signal their intrinsic types to others or to themselves, and laws in that context both interact with this signaling role of norms and may themselves signal the society's attitudes to individuals (see also Posner 1997; Cooter 1998; Posner 2002 for related perspectives, and McAdams and Rasmusen 2007 for a survey). What distinguishes our paper from all four of these literatures is our focus on how social norms are shaped by laws while at the same time critically constraining the effectiveness of laws. This two-way interaction is at the root of both our nonmonotonic comparative statics and the key results that excessively tight and abruptly introduced laws can backfire.

A number of other papers also discuss how social norms are important for the effectiveness of laws, but without identifying this key two-way interaction or developing any of our key results. For example, Hay, Shleifer, and Vishny (1996) and Hay and Shleifer (1998) mention—among several other factors—the importance of achieving congruence between prevailing social norms and laws, especially in the context of the transition from socialism to a market economy, arguing "A further reason that private parties in Russia refuse to use the legal system is that they operate to some extent extra-legally to begin with and, hence, do not want to expose themselves to the government" (Hay and Shleifer 1998, p. 399) and "whenever possible, laws must agree with the prevailing practice or custom. If public laws violate the practice, then private parties may refuse to enforce them either on their own or with ultimate referenced courts." (p. 402). Akerlof and Yellen (1994) anticipate the same point in their analysis of criminal behavior, and write "the major deterrent to crime is not an active police presence but rather presence of knowledgeable civilians, prepared to

substantial literature in social psychology on law-abiding behavior, for example, discussed in Tyler (1990). The different approaches to social norms within economics are discussed in Mailath and Samuelson (2006). But these literatures inside or outside economics do not focus on the interplay between norms and the enforcement of laws.

report crimes and cooperate in police investigations." Berkowitz, Pistor, and Richard (2003) also touch on these issues in their discussion of how transplantation of common law legal systems may not work in the context of certain prevailing customs. Parisi and von Wangenheim (2006) and Carbonara, Parisi, and von Wangenheim (2008) emphasize how social values constrain laws and how this might call for gradualism (as in our dynamic model). Finally, Dyck, Morse, and Zingales (2010), who discuss the importance of social norms in determining whistle-blowing behavior, is also related. Nevertheless, none of these papers, and no others that we are aware of, consider the two-way interactions between social norms and the enforcement of laws, which is our main focus here.

The rest of the paper is organized as follows. The next section introduces our baseline static model. Section 3 presents the analysis of the static model and contains our main results. We extend this baseline model to a dynamic setup in Section 4, and discuss the interplay between historically determined norms and laws, as well as how laws in conflict with norms may backfire. Section 5 contains several extensions, including the one concerning the interplay between laws regulating different types of behavior. Section 6 briefly discusses welfare properties of the simple laws we focus on in this paper, whereas Section 7 concludes. All of the proofs are contained in the Appendix.

2. The Static Model

We first present our baseline static model, which introduces the main economic forces.

2.1. Agents, Matching, Laws, and Payoffs

There is a finite population of agents, $N = \{1, \ldots, n\}$, with $n \geq 2$ taken to be even. In the baseline model, we consider a simple (uniformly random) pairwise matching of the agents represented by a matching function m: $N \to N$. Throughout, $m(i)$ denotes the match partner of agent i, with $m(i) \neq i$ (and with the usual convention that $m(m(i)) = i$).

In the baseline model, we focus on a single dimensional behavior. In particular, agent i chooses a base behavior $b_i \in [0, 1]$ before the matching stage (thus agent i does not know $m(i)$ when choosing b_i). We refer to b_i as the *base behavior* because the agent's *actual behavior* may be forced away from this level ex post to some B_i because of the enforcement of a law.[6]

Agent i has type $\theta_i \in [0, 1]$, distributed according to a cumulative distribution function F. Type draws are i.i.d. across agents, and for simplicity, we assume that F is strictly increasing and continuous on [0, 1], with $F(0) = 0$ and $F(1) = 1$. This is the

6. An agent might end up bearing some harm from the base behavior of her partner, $b_{m(i)}$, beyond his actual behavior, $B_{m(i)}$. Provided that some of the harm is from the actual behavior and can thus be reduced by enforcement, this would have no major effect on our results.

agent's preferred level of behavior, which will determine his or her law-abidingness in equilibrium.

A law, $L \in [0, 1]$, is set by the government and is an *upper bound* on the behaviors of agents, meaning that any behavior above L is prohibited.[7] However, the government only has limited ability to enforce laws without private cooperation (because it does not always observe individual behaviors). In particular, we assume that the government observes behavior in any pair with probability $\eta \in [0, 1)$, and in this case, the behavior of any law-breaking party in the relationship will be pushed down to L. In addition, when there is no law enforcement (with probability $1 - \eta$), an agent can whistle-blow to the government that her partner's behavior is above the law L. Agent i can only do so if her partner's behavior exceeds L; that is, if $b_{m(i)} > L$. We also assume that the agent herself must be law-abiding to be able to whistle-blow; that is, $b_i \leq L$.[8] We denote the whistle-blowing decision of agent i by $w_i \in \{0, 1\}$, with 1 indicating whistle-blowing.

If $b_i > L$ and $w_{m(i)} = 1$, meaning that agent i is breaking the law and her match whistle-blows (which, as just stated, presumes that $b_{m(i)} \leq L$), then she pays a fine $\varphi > 0$ and her behavior is adjusted down to the highest level consistent with the law, L. We should also note that fines can be interpreted as either deadweight losses or transfers. If fines are transferred back to the population, then there will be an additional term in the utility function. These would create an additional incentive to whistle-blow (similar to the externality term), but this would have no impact on our analysis until we turn to welfare in Section 6.

Thus, the resulting action of agent i can be written as

$$B_i(w_{m(i)}, b_i) = \begin{cases} L & \text{if } b_i > L, \text{ and } w_{m(i)} = 1 \\ & \quad \text{or if there is public enforcement (probability } \eta), \\ b_i & \text{otherwise.} \end{cases}$$

With this notation, agent i's payoff is given by[9]

$$u_i(\theta_i, B_i) = -a\left(B_i - \theta_i\right)^2 - (1-a)(B_i - B_{m(i)})^2 \\ -\zeta_m B_{m(i)} - \zeta_o \sum_{j \neq i, m(i)} B_j - \left(\eta + (1-\eta)\, w_{m(i)}\right) \mathbf{I}_{\{b_i > L\}} \varphi. \quad (1)$$

The parameters $\zeta_m, \zeta_o \geq 0$ capture negative externalities from the behaviors of others and allow these negative externalities to be different from the behavior of a

7. That laws only prohibit high behaviors is for simplicity. We could also allow for laws that ban both high and low behaviors, such as speed limits on highways.

8. This is easy to motivate (in the presence of nontrivial fines). For example, when authorities attempt to enforce lower behavior on a law-breaker, they may also observe the behavior of her partner, thus leading to potential fines for the whistle-blower. We relax the assumption that law-breakers cannot whistle-blow and also discuss the issue of amnesty for whistle-blowers in Section 5.4.

9. Such coordination games are studied in a variety of contexts, for instance, see Alonso, Dessein, and Matouschek (2008) for an application to the organization of a firm.

direct partner than the rest of the population. For example, a person may suffer greater negative externalities from the racism or sexism of a partner or from their smoking behavior (represented by ζ_m) than from similar behavior by someone with whom they are not closely associated (represented by ζ_o).

The parameter $a \in (0, 1)$ is an (inverse) measure of "social sensitivity", and regulates the relative importance of own preference versus matching the prevailing "norm", which in the static model simply corresponds to "distribution of the behavior" of other agents in the economy that each player would partially like to match.

The $-(1 - a)(B_i - B_{m(i)})^2$ term captures the complementarities in actions. For instance, if one party to a transaction wants to underreport by half a transaction to authorities and the other party wants only to underreport by a quarter, then that makes the transaction more complicated. Similarly, if two police are partnered and one is willing to take small bribes and the other not, then that makes it more difficult for both involved than if they are both either willing to take small bribes, or both prefer not to.

The last term subtracts the fine imposed on law-breakers conditional on public enforcement (probability η) or when there is no public enforcement (probability $1 - \eta$) but there is private whistle-blowing.

Because high behaviors have negative externalities on the population, there is a reason for the government to discourage such behaviors, and laws acting as upper bounds on behavior are the government policy on which we focus. Equation (1) also clarifies the two reasons why an agent may whistle-blow on the high behavior of her partner: first, she reduces the negative externality she suffers (from the term $\zeta_m B_{m(i)}$); and second, she may be able to reduce the mismatch between her behavior and her partner's (from the term $(1 - a)(B_i - B_{m(i)})^2$).[10]

2.2. *Motivating Examples*

It is useful to have some running examples to fix ideas. One simple application illustrating some of the main trade-offs is that of a business partnership.

Suppose that each of the partners can choose a level of compliance with a tax, for instance a value-added tax. The nature of the business affects the ease of tax evasion (e.g., how much of the business is based on cash), which can be thought of as the agent's type. Evading taxes is easier if the companies involved in a transaction are evading at equal levels, since it is more difficult to disguise transactions to the extent that the reports of the partners regarding the transaction differ. Thus, there is an improvement in payoffs by matching a partner's level of tax evasion. Because tax authorities cannot police all balance sheets, whistle-blowing plays a critical role in the

10. If fines are transfers, there would also be a third reason, as whistle-blowing will generate revenues from fines, which as noted above, could be redistributed to the agents.

This discussion also clarifies why collusion is not a major issue in our setting: by not blowing the whistle on a law-breaker, an individual would be strictly worse off, and the law-breaker does not have a direct way of compensating her. If we allowed transfers, it might be difficult for a law-abiding agent to commit to not blow the whistle after receiving a transfer.

detection and prevention of tax evasion. Each agent's tax evasion would then lead to negative externalities on other agents through lost tax receipts (and a behavior even lower than the law can be interpreted as contributions to public goods, charities or other high social value activities). Moreover, following detection, a tax evader will be forced to pay taxes as in our model.

As a related example, suppose that each agent decides how much effort to exert to produce a high-quality, safe and reliable product, and then matches with another agent. The matched partners then combine their products to produce a consumer good, the quality of which will be a combination of the qualities of the two products. Low-quality, unreliable products create a negative impact on the utility of the population, but are potentially profitable for the producer because producing them is less costly and some fraction of customers do not observe the quality of the consumer good they purchase.

In this example as well, there are natural reasons for why each agent would like their behavior (quality of product) to match that of their partner. In particular, those with high-quality products would not like to see the quality of the consumer good they produce being brought down by the low-quality product of their partner. Conversely an agent with a low-quality product will suffer additional costs because of the incompatibility between her product and the high-quality product of their partner. Relatedly, it may be easier for two firms that are cooperating to produce to similar tastes and quality preferences than to have one producing a higher quality. (It could even be that the low quality producer would prefer that her partner not try to hold the product to higher standards, which could produce significant frictions in the relationship.)

Though the government would like to regulate product quality, it has an imperfect ability to do so because it does not always directly observe qualities. It can legislate a minimum quality standard on each product, and then enforce it through random inspections and whistle-blowing behavior of business partners. If an agent is detected to have a low-quality or unsafe product, she will be subject to a fine and will also have to take costly action to bring her product into compliance with the law. The business context also motivates our assumption that whistle-blowers who are law-breakers will themselves be detected. For example, if a law-breaking firm whistle-blows on its partner's low-quality product or tax evasion, the subsequent investigation will likely reveal its own transgressions. In this example too, as in our model, following detection a firm with a low-quality product will be forced to bring its product in line with regulations.

Both examples further clarify the meaning of "norms" in our model. As discussed in the Introduction, norms correspond to "external norms", summarizing the distribution of expected (or in the dynamic model, past as well as future) behavior in the population in a setting where such behavior has important payoff consequences. In addition to external norms, our model could also be used to study the interplay between laws and "internal norms" (i.e., norms that individuals adhere to for moral reasons, e.g., Hoffman 1977; Tyler, 1990). In this case, individuals would whistle-blow because the behavior of their partners or of others that they observe are in dissonance with their moral values and expectations. For example, we could assume that internal norms adapt

to an individual's own behavior and some function of average behavior in society, and if an individual sees a behavior very far from their internal norms, they feel compelled to try to prevent it.

Yet another example that fits our model arises in the context of safety in the workplace, as enforced, for instance by OSHA (Occupational Safety And Health Administration) in the United States. Safety in the workplace is a concern for workers because they do not wish to be injured, and for firms because injuries can lead to lost production, increased costs, or potential damage to customers, and so forth. We can think of the behavior of the firm as corresponding to the riskiness of the workplace (do they educate their workers, take out proper insurance, inspect their worksites, provide safe and up-to-date equipment, etc.), and the behavior of the worker as how many risks they take on the job (e.g., does a roofing worker wear a safety harness). The θs then correspond to the workers' and firms' tolerances for risk. When a worker is matched with a firm that provides an excessively risky workplace, her main recourse is to whistle-blow in order improve the workplace conditions, and once again, the ex post behavior of a law-breaking firm would be brought in line with regulations if detected.[11]

It is straightforward to see that many other types of behaviors also fit the model, for example, dueling, tax compliance, consumption of illegal substances, disruptive behavior, harassment, corruption, and so on. In all of these cases, behaviors are not typically observable directly by the government but are more commonly detected by some other agents (roommates, business partners, employees, etc.), and agents care about the behavior of those with whom they are closely associated (matched).

2.3. The Importance of Whistle-Blowing

Our focus on whistle-blowing is motivated by its central role in the detection of a range of crimes and thus in the enforcement of corresponding laws. For example, according to the Association of Certified Fraud Examiners' data from 2014, 42.2% of initial fraud detection was by informant/whistle-blower tip, whereas the next highest category of detection was via a management review, at 16.0%; only 2.2% of detection came directly from detection by law enforcement agencies.[12] Several high-profile cases of whistle-blowers, from "Deepthroat's" role in Watergate, to Snowden's release of classified security agency documents, illustrate the importance of whistle-blowing in many diverse settings. Several organizations also rely heavily on whistle-blowing from their employees for detection of such transgressions as discrimination, favoritism,

11. It is straightforward to adapt our model further to this example by allowing matches to be between two separate groups of agents and allowing for some differences across sides. For instance, we could allow workers to whistle-blow, and employers to fire workers. Such a model is a straightforward extension and so we stick with the simpler setting.

12. See the "Report to the nations on occupational fraud and abuse: 2014 Global Fraud Study", retrieved on January 10, 2015 from http://www.acfe.com/rttn/docs/2014-report-to-nations.pdf.

harassment, work-safety issues, and abuse. In the United States, OSHA houses an "Office of the Whistleblower Protection Program".

The role of whistle-blowing is even greater for economic crimes. Tax agencies strongly encourage tips regarding tax evasion. The US Internal Revenue Service is an exemplar with its "Whistleblower—Information Award" of up to 30% of recovered taxes and penalties,[13] which has resulted in numerous high-profile cases, such as a former banker at UBS getting over 100 million dollars as a reward for whistle-blowing on a tax evasion scheme.[14]

Importantly, whether a behavior is viewed as appropriate—for example, whether employees see safety regulations as appropriate or excessive—critically depends on the social norms of an organization or society, and not simply on the laws. This makes certain laws ineffective because whistle-blowing or more generally cooperation with the law will not be forthcoming.[15]

2.4. Equilibrium

The game we have set up consists of two stages: in the first, each agent chooses a behavior conditional on their type, and then in the second stage, they decide whether to whistle-blow on their partner.

It is straightforward to observe that, in our baseline model, it is a strictly dominant strategy for an agent i with $b_i \leq L$ and $b_{m(i)} > L$ to whistle-blow (because this reduces both the externality and the mismatch with the partner's behavior). As a consequence, in this multistage complete information game, any sequential equilibrium will have whistle-blowing as part of the equilibrium whenever the opportunity arises.

Given this observation, we define an equilibrium for the first stage only, using the standard notion of a pure-strategy symmetric Bayesian equilibrium. Such an equilibrium described by a strategy β: [0, 1] $\rightarrow$ [0, 1], with $\beta(\theta_i)$ indicating the action taken by type θ_i.[16]

In what follows, whenever we refer to *equilibrium*, it is to such a pure-strategy symmetric Bayesian equilibrium of the induced one-stage game.

13. http://www.irs.gov/uac/Whistleblower-Informant-Award, accessed January 12, 2015.

14. New York Times, September 11, 2012, "Whistle-Blower Awarded $104 million by IRS".

15. We do not model potential repercussions on a whistle-blower. Many governments have whistle-blower protection laws, but those are not always effective, nor do they protect a whistle-blower from other forms of social retaliation or ostracism. This can lead to further ineffectiveness of a law (e.g., Benoît and Dubra 2004). It may be interesting to extend our model to allow for potential endogenous costs of whistle-blowing, which might be governed by their own norms.

16. We take strategies to be measurable functions of type. Since F is continuous (and thus does not have any atoms) and agents are only indifferent at one type, there will be no nontrivial mixing, so the focus on pure-strategy equilibria is without loss of much generality. The possibility for asymmetric equilibria (and hence the qualifier "symmetric") is a consequence of having a finite number of agents. With a continuum, all equilibria are in symmetric strategies (agents have essentially unique best responses and face the same behavior of others, given that they are each of measure 0).

3. Analysis of the Static Model

In this section, we characterize the equilibria of the static model presented in the previous section, and present our main comparative static results.

3.1. Existence of Equilibrium

Strategies are said to be *monotone* if $\beta(\theta_i) \geq \beta(\theta_i')$ whenever $\theta_i > \theta_i'$. Our first result shows that all equilibria are in monotone strategies. An implication of monotone strategies is that, in any equilibrium, there exists some threshold type θ^* such that all types below θ^* are law-abiding, and all types above θ^* are law-breaking (and this threshold can be one or zero, corresponding, respectively, to all agents being law-abiding and law-breaking).

PROPOSITION 1. *An equilibrium exists, and each equilibrium is in monotone strategies and is characterized by a threshold θ^* above which all types break the law and below which they obey the law.*

Like all of our other results, this proposition's proof is given in the Appendix. There, Lemma A.1 establishes the monotonicity of strategies, and existence follows straightforwardly given the threshold characterization in Proposition 4.

It is important to note that ours is not a game of strategic complementarities because a higher behavior of an agent's (potential) partner can encourage her to obey the law in order to be able to whistle blow.

3.2. Equilibrium without Laws

Next, we characterize the equilibria without laws, or equivalently the case where $L = 1$ (so that laws are not binding). In this benchmark case, there is a unique equilibrium described in the next proposition.

PROPOSITION 2. *Without any law ($L = 1$), there is a unique equilibrium and it is linear in own type and described by*

$$\beta(\theta_i) = a\theta_i + (1 - a)\mathbb{E}[\theta].$$

Equilibrium strategies are depicted in Figure 1 as a function of own type, θ_i. Linearity here is a simple consequence of the fact that, without laws, payoffs are quadratic.

A feature that is paralleled by equilibria with laws is that the parameter a, regulating the "social sensitivity" of payoffs, determines the slope of equilibrium strategies. The distribution of types F only affects the form of the equilibrium via the threshold value of θ between law-abiding and law-breaking (which is not relevant here), and via $\mathbb{E}[\theta]$ (and thus the intercept in Figure 1). This implies that agents always choose a convex combination, with weights a and $1 - a$, between their preferred action given by their

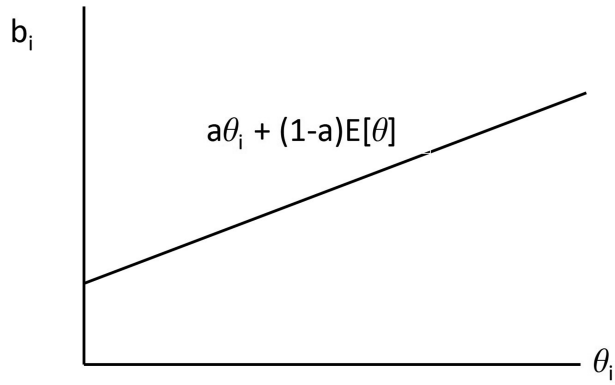

FIGURE 1. Equilibrium without laws.

type, θ_i, and the "social norm", that is, the distribution of expected behavior in society, which in this case is summarized by $\mathbb{E}[\theta]$.

In light of this benchmark, it becomes clear that "social norms" here refer to equilibrium behaviors, incorporating a spectrum of play, which is influenced by the expectations of others' behaviors—potentially very strongly, for example, for low a.

3.3. Equilibrium with Laws

With this benchmark in place, for the remainder of the paper we focus on the case in which laws are binding, that is, $L < 1$. The next two propositions characterize the equilibria in this case, showing that, compared to the case without laws, there is a crucial choice between law-abiding and law-breaking behavior, and the impact of laws on the behavior of others will also change the expected behavior that enters into the calculations of law-abiding and law-breaking agents.

PROPOSITION 3. *For any $L \in (0, 1)$, there exists $\bar{\varphi} \geq 0$, such that*

- *if $\varphi > \bar{\varphi}$, then there is a unique equilibrium, which involves full compliance (all types obey the law);*
- *if $\varphi < \bar{\varphi}$, then there are multiple equilibria: one with full compliance and (at least two) other equilibria ordered by the threshold above which all types break the law.*

When the fine against law-breakers that are detected is sufficiently large, that is, $\varphi > \bar{\varphi}$, the unique equilibrium involves law-abiding behavior (full compliance) by all agents. The intuition for this case is instructive. For any $L > 0$, there will be some agents who always obey the law, because their type, θ, is below L, and they can always whistle-blow and bring down the behavior of their law-breaking partner to L (and this implies that their behavior will be lower than the economy without any law characterized in Proposition 2). This means that any law-breaking agent faces a strictly

positive probability of paying the fine φ (even if the direct government detection of law breaking is 0), which discourages law-breaking for very large fines. Even when $\varphi < \bar{\varphi}$, the full compliance equilibrium continues to exist. In this case, even though the fine is not large enough to discourage law-breaking with only a small fraction of agents obeying the law and whistle-blowing, when all agents are law-abiding a law-breaker will be reprimanded with probability 1 (either due to public enforcement or whistle-blowing by her partner), and so will also choose a behavior capped by L. Nevertheless, when $\varphi < \bar{\varphi}$, there are also equilibria in which a nontrivial fraction of the population breaks the law.[17]

The multiplicity of equilibria highlighted in this proposition has a different economic intuition than the multiplicities already emphasized in the literature (e.g., congestion effects in policing or the legal system, or peer effects), and highlights the critical role of social norms. A social norm of expecting others to break the law encourages each individual to do the same, whereas instead a social norm of law abiding and whistle-blowing encourages each individual to follow that norm.

Although the equilibrium threshold for law-breaking and the fraction of law-breakers may not be unique, the next proposition shows that the form of the equilibrium is the same in all equilibria.

PROPOSITION 4. *Each equilibrium is of the following form. There exists* $\theta^* \in [L, 1]$ *such that*

$$\beta(\theta_i) = \beta_{abiding}(\theta_i) \equiv \min[a\theta_i + (1-a)x, L] \text{ if } \theta_i < \theta^* \tag{2}$$

and

$$\beta(\theta_i) = \beta_{breaking}(\theta_i) \equiv a\theta_i + (1-a)\mathbb{E}[\theta|\theta > \theta^*] \text{ if } \theta_i > \theta^*, \tag{3}$$

where x *is the unique solution to* $x = \mathbb{E}[\min(\beta(\theta), L)]$.

The general form of equilibria is depicted in Figure 2. The form of the strategies $\beta_{abiding}(\theta_i)$ and $\beta_{breaking}(\theta_i)$ is intuitive. Recall that without laws (cfr. Proposition 2), an agent chose a convex combination of her preferred behavior given by her type, θ_i, and expected behavior in society. For law-abiding agents, the calculus is still similar, except that as shown by the expression, $a\theta_i + (1-a)x$, the expected behavior is replaced by x.[18] This is the expectation of the actual behavior of other agents (rather than their base behavior, b_i) and thus takes into account that the agent herself can whistle-blow (or there is public enforcement) on any partner who chooses the behavior above L and force them down to L. By definition of these agents being law-abiding,

17. Note, however, that $\bar{\varphi}$ can be equal to zero, in which case such equilibria will not exist. A trivial example is when L is arbitrarily close to 1.

18. Note that x is implicitly defined as a fixed point of an increasing mapping, since $\beta(\theta)$ is itself an increasing function of x. This implies that x may not be unique. However, we show in the Appendix that it is defined uniquely.

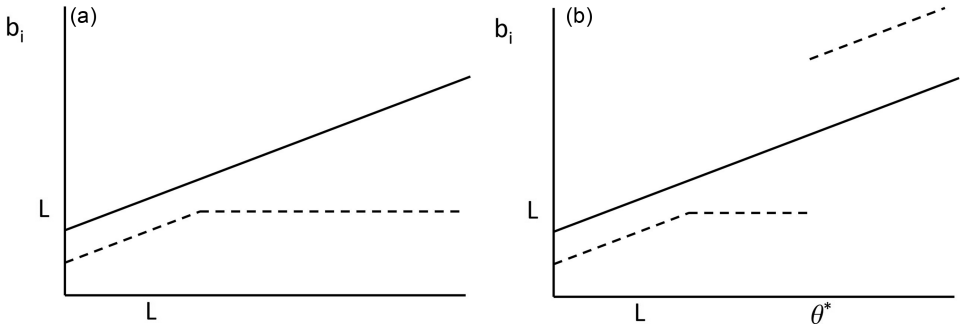

FIGURE 2. Multiple equilibria: The black line is the equilibrium without any law. The dashed lines indicate equilibrium behavior. These are the two extreme equilibria. In (a) everyone obeys, and in (b) it is those below θ^* who obey, where θ^* is the lowest compliance equilibrium threshold. Equilibria can be ordered in terms of θ^*.

their behavior is equal to $a\theta_i + (1 - a)x$ only if this expression is below L (and otherwise their behavior is capped at L).

The calculus of law-breaking agents is different. These agents know that their partner, if she is law-abiding, will whistle-blow on them, setting their behavior down to L. Thus, when choosing their behavior, these agents only reason *conditionally*—conditional on matching with another law-breaking agent and not being subject to public enforcement. As a result, their behavior is a convex combination of their own type, θ_i, and the expected behavior of their partners conditional on their partner being law-breaking, which is $\mathbb{E}[\theta|\theta > \theta^*]$. Note that the parameter η does not explicitly enter equations (2) and (3). This is because, as just explained, law-breakers reason conditionally, whereas law-abiders know that they can always whistle-blow on a law-breaking partner. The law-breaking threshold θ^* naturally depends on the probability of public enforcement, η. This discussion also highlights that, in contrast to the case without laws, the relevant social norm is no longer summarized by $\mathbb{E}[\theta]$, but depends on the entire distribution of behavior in society.

Equilibria in Proposition 4 are conditional on the law-breaking threshold θ^*. It is also straightforward to derive an expression for this threshold, which balances out the costs and benefits of law-breaking for the threshold agent at θ^*, and this is given by equation (A.10) in the Appendix.

Although Proposition 3 highlights that there are multiple equilibria (in the case where $\varphi < \bar{\varphi}$), Proposition 4 shows that all equilibria are characterized by a law-breaking threshold θ^* (which may be equal to one, in which case all obey the law). Equations (2) and (3) also indicate that different equilibria are ranked in terms of their threshold for law-breaking, θ^*. If we consider two equilibria with different levels of θ^* (for given levels of L and φ), then in the one with lower θ^*, the behavior of all law-abiding citizens will be the same (because x does not change); there will be more law-breaking (more agents above θ^*); but law-breakers will choose lower behaviors (because $\mathbb{E}[\theta|\theta > \theta^*]$ is lower). In what follows, we often focus on the

lowest compliance equilibrium, meaning the equilibrium with the lowest θ^* and thus the highest amount of law-breaking.[19]

3.4. Comparative Statics

We now provide some simple comparative statics for the lowest compliance equilibrium.

COROLLARY 1. *Consider a setting in which there is a lowest compliance equilibrium with partial compliance ($\theta^* \in (0,1)$). Then, for that equilibrium:*

(1) *A small increase in φ (higher fine), ζ_m (greater within-match externality), and/or η (higher likelihood of public enforcement):*[20]
- *increases θ^* and so lowers the fraction of agents breaking the law;*
- *leaves behavior by each agent who was obeying the law unchanged;*
- *but leads to higher behavior among those still breaking the law.*

(2) *A large increase in φ, ζ_m, and/or η eliminates the partial compliance equilibrium and results in full compliance and reduces overall average behavior.*

(3) *There exists $\bar{\zeta}_m > 0$ such that if $\zeta_m < \bar{\zeta}_m$, a small decrease in L (a stricter law):*
- *decreases θ^*, increasing the fraction of agents breaking the law;*
- *leads to lower behavior by each agent who was already breaking the law; and*
- *leads to lower average behavior by those obeying the law.*

(4) *If $\zeta_m > \bar{\zeta}_m$, then the comparative statics of a stricter law are reversed.*

The corollary focuses on the lowest compliance equilibrium, but the same comparative statics apply to equilibria that are stable under best-response dynamics as discussed in footnote 19.

Several aspects of this proposition are worth stressing. First, higher fines, greater public enforcement, and tighter (more restrictive) laws all have unambiguous effects on the fraction of law-breakers, as depicted in Figure 3, but more nuanced impacts on levels of behavior.

19. We might also study the stability of equilibria under best-response dynamics to investigate which equilibria are more likely to be robust to small perturbations. Under some regularity conditions, both the full compliance equilibrium and the lowest compliance equilibrium are stable under this notion of stability.

20. A small increase is defined to be small enough so that, after this change, the lowest compliance equilibrium has a law-breaking threshold in a small neighborhood of θ^* (the law-breaking threshold before the change). In particular, suppose that there exists a neighborhood $\mathcal{N}$ of the parameter vector $(L, a, \varphi, \eta, \zeta_m)$ and a continuous function $\Theta^* : \mathcal{N} \to [0,1]$ such that $\theta^* = \Theta^*(L, a, \varphi, \eta, \zeta_m)$ and there is an equilibrium with law-breaking threshold $\Theta^*(L', a', \varphi', \eta', \zeta_m')$ for all $(L', a', \varphi', \eta', \zeta_m') \in \mathcal{N}$. Then there always exists small changes in φ, η, and ζ_m (within $\mathcal{N}$) that still lead to an equilibrium nearby, and the statements here refer to comparisons to this nearby equilibrium. This requirement holds generically and simply rules out situations in which the lowest compliance equilibrium is defined by a point of tangency rather than the intersection of the two curves defined in the proof of Proposition 4. When this requirement is not satisfied, any change in parameters is a "large" change in terms of this proposition.

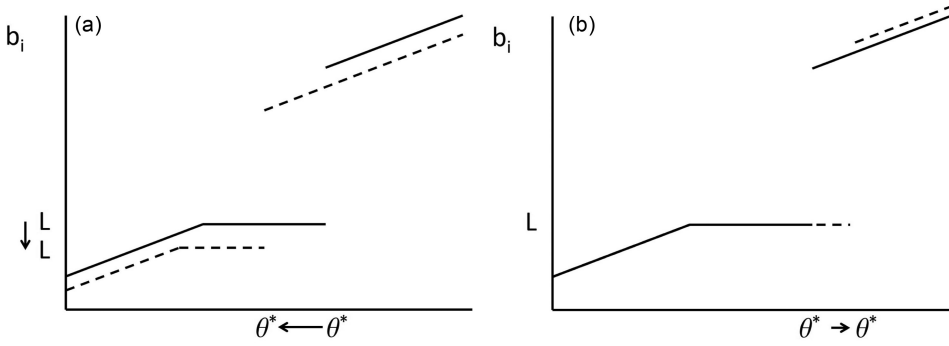

FIGURE 3. Comparative statics. (a) A decrease in the (tightening of the) law L leads the equilibrium to change from the solid to the dashed lines. (b) An increase in φ; externality, ζ_m or public enforcement, η, change the equilibrium from the solid to the dashed lines.

In particular, starting from $\varphi < \bar{\varphi}$, a small increase in φ (which keeps us in the same region) reduces the fraction of law-breakers because the penalty is now higher. The impact of this increase in fines on behavior is more subtle, highlighting the rich interactions between laws and behavior in our model. A higher φ has no impact on β_{abiding} (x is independent of both φ and θ^* as noted above), but increases θ^* and thus $\mathbb{E}[\theta|\theta > \theta^*]$. This implies that the function β_{breaking} shifts up, highlighting that even though all strategies are monotone, the impact on the behavior of law-breakers is nonmonotone. Intuitively, some law-breakers switch to law-abiding behavior in response to the higher fine, and the remaining law-breakers now expect to match with a relatively higher θ (with consequently higher behavior) and so they adjust their behavior upward. So, fewer agents break the law, but do so by a greater amount (i.e., law-breakers increase their behavior). This leads to an ambiguous average impact that depends on the specifics of the distribution (as illustrated in Figure 4 for the case of changes in L).

A large increase in φ, on the other hand, takes us above the threshold $\bar{\varphi}$ thus destroying all equilibria except the full compliance equilibrium (and thus avoiding the somewhat paradoxical effect of encouraging high behavior among law-breakers).

Greater public enforcement (higher η) has the same effects as a greater fine for law-breaking, increasing the law-breaking threshold and also increasing behavior among law-breakers. A greater externality in behavior (higher ζ_m) also acts similarly to a greater fine, although for different reasons: it induces people to obey the law so that they can whistle-blow on their match. We note that ζ_o does not affect the behavior of the agents, since their choice of action and whistle-blowing cannot affect individuals other than those whom they observe (their match).

A stricter law acts quite differently from increasing fines or public enforcement. Let us first focus on the case where the externality is small, that is, $\zeta_m < \bar{\zeta}_m$. In this case, a stricter law increases the fraction of law-breakers, because obeying the law is now more onerous (requires lower behavior). However, now in addition to increasing

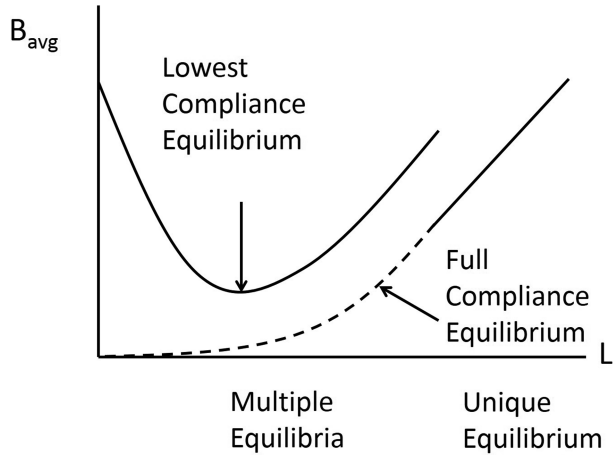

FIGURE 4. Average behavior as a function of the law. Average behavior decreases in the full compliance equilibrium as a law is tightened, but may increase or decrease in the lowest compliance equilibrium. There can also be discontinuities in behavior: near the point of discontinuity, a slightly *looser* law, can result in a dramatic lowering of average behavior and full compliance.

law-breaking, a tighter law reduces the behavior of each law-abiding agent (because L is lower), and switches the highest types—who were choosing the highest behavior among the law-abiding agents—to law-breaking. This reduces average behavior among law-abiding agents through both channels.

Nevertheless, because those switching from law-abiding to law-breaking increase their behavior, the impact on overall average behavior is again ambiguous. The effect of a tighter law on average behavior is shown in Figure 4. In particular, this figure illustrates that, compared to no law, a very permissive law—which is associated with a unique full compliance equilibrium—necessarily reduces average behavior. But a further tightening of the law may eventually lead to a discontinuous jump in average behavior at the point where full compliance ceases to be the unique equilibrium and there emerges a low compliance equilibrium with a positive fraction of agents breaking the law. Further tightening of the law reduces average behavior in the lowest compliance equilibrium, but this effect is again reversed eventually, when the increase in law-breaking starts to overwhelm the reduction in behavior among law-abiding and law-breaking agents. These results highlight the possibility that excessively tight laws may achieve the opposite of their objectives: increasing, rather than reducing, average behavior in society. We further discuss the potential paradoxical effects of strict laws in the context of our analysis of dynamic settings and settings with multiple types of behavior in Section 5.1.

The multiplicity of equilibria involves a discontinuity, so that there are situations where a tiny *loosening* of a law could eliminate the low compliance equilibrium, and lead to a substantial lowering of average behavior and transition from significant noncompliance to full compliance. This results from the feedback in behaviors: people

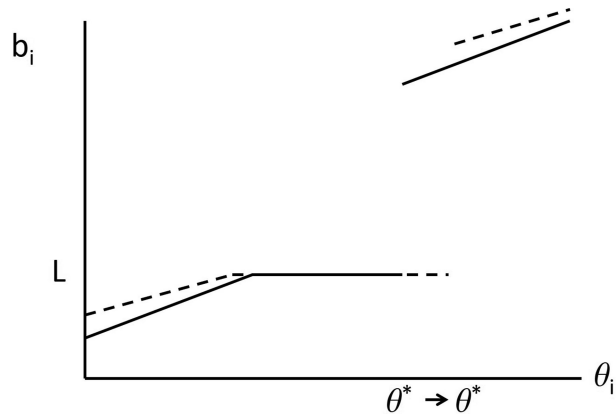

FIGURE 5. Increasing the sensitivity to the partner's action (a decrease in a) changes from solid to dashed lines.

are only willing to break the law if a substantial fraction of others also do, and slight changes in the law can lower that fraction past a tipping point, completely eliminating partial compliance equilibria.

Also worth noting is that the impact of a change in law reverses for high enough externalities ζ_m (i.e., when $\zeta_m > \bar{\zeta}_m$). This reversal is a consequence of the fact that when the externality is large, the desire to whistle-blow on other agents and reduce the externality they create overwhelms the other effects, and so agents prefer to obey the law in order to be able to whistle-blow and reduce the behavior of others.

We finally note that the impact of a change in the parameter a, which regulates the social sensitivity of behavior, is more subtle because this parameter affects both the slope of the equilibrium and the cutoff threshold θ^*, as pictured in Figure 5.

4. Dynamics

The static model transparently exposits how laws and expected behavior of other agents influence behavior. Though the expected behavior of other agents acts as a "norm," anchoring the behavior of each individual in society, the static setup does not permit an analysis of how past "norms" (patterns of behavior) influence the evolution of future behavior. In this section, we consider a dynamic generalization of the static model, which enables us to demonstrate how laws that are in conflict with prevailing norms can backfire, whereas more moderate laws can be successful in changing behavior and norms in society.

4.1. Dynamic Model

Our dynamic model is a straightforward generalization of our static setup. We consider an overlapping generations model, with each generation consisting of n agents. An

agent i of generation t, denoted (i, t), when born, has a type $\theta_{i,t}$ drawn according to some atomless F_t (satisfying the same assumptions as above, and drawn independently across time and agents within a generation). When the generation is clear we write θ_i to simplify notation. Agent (i, t) after observing $\theta_{i,t}$ chooses a base behavior $b_{(i,t)}$, and then matches with an agent of the previous generation, $t-1$, and an agent of the next generation, $t+1$.[21] We denote the matching partner of agent (i, t) at time t (from generation $t-1$) by $m_t(i, t)$ and at time $t+1$ (from generation $t+1$) by $m_{t+1}(i, t)$.

We assume that the base behavior of each agent is "sticky", in particular meaning that it is chosen only once by each agent.[22] For instance, using one of our motivating examples, each agent decides how much investment to make in a high-quality product at the beginning of their business life.

The rest of the setup closely follows our static model (and in fact is chosen to maximize this parallel), except that to simplify the notation in the dynamic model and with little loss given our focus here, we set $\zeta_m = 0$ and $\zeta_o = \zeta$.

Specifically, there is a law L_t that governs the play in period t. To begin with, we assume that the sequence of laws is known to all agents, but we also consider an unanticipated change in the sequence of laws below. Each agent can whistle-blow against either/both her partner from the previous and the next generation, and we assume that an agent who is caught breaking the law (either because of whistle-blowing or public enforcement) is only forced to change her behavior in the specific interaction in which her behavior is detected.[23] In other words, even though the base behavior, $b_{(t,i)}$, is constant, the actual behaviors of agent (i, t) against her partner from the previous and the next generations can be different, and are denoted, respectively, by $B^t_{(i,t)}$ and $B^{t+1}_{(i,t)}$ (where we are using the natural convention that the interaction between generations $t-1$ and t takes place at time t). We also assume that, when laws are time-varying, individuals are bound by the laws that were in operation in their youth.[24] So an individual born at time t will have broken the law in both of her interactions at time t and $t+1$ if her base behavior $b_{(i,t)}$ is above L_t. And if she

21. Matching with own generation, as well as previous and next generations, complicates the analysis further, without modifying any of our main results, so we simplify matters by focusing on an environment in which matching is only with previous and next generations.

22. Although full stickiness makes the analysis transparent, a partial friction still produces similar effects, though with additional complications (see Acemoglu and Jackson 2015 for a more detailed discussion).

23. This implies that the government cannot automatically force an individual to change her behavior in future interactions just because she was whistle-blown in her "youth". This could be, for example, because the government cannot direct its enforcement toward monitoring certain agents more closely, and we can imagine that an agent can in fact change her behavior in the future by paying some fixed costs, so that the government cannot be sure that this past transgressor has indeed broken the law again without monitoring her.

24. All of our major insights also hold with the alternative assumption of no grandfathering, which makes each agent responsible to match the current law (in fact, all of the results with constant laws are trivially the same). More importantly, a version of one of our key results in this section, Proposition 5, also holds in this case, and in fact becomes somewhat more obvious, since a tightening of the law automatically makes many people in the previous generation law-breakers and thus unable to whistle-blow on the next generation. This mechanical effect is eliminated with our grandfathering assumption.

is caught in either period, her behavior is forced down to L_t (and she pays the fine φ_t). This assumption also implies that when there is a change in laws, a generation is automatically grandfathered provided that they obeyed the law in place at the time of their birth.

Behavior at time 0 is taken as given and described by some strategy $\beta_0(\cdot)$. Let $w^k_{m_k(i,t)} = 1$ designate that agent (i, t)'s partner at time k whistle-blows on her in time period k, and $w^k_{m_k(i,t)} = 0$ indicate that her partner does not whistle-blow.

The payoffs to agent (i, t) are a natural generalization of those in the static model and are given by

$$\begin{aligned} u_{(i,t)} = -(1-\lambda)\Bigg[& a\left(B^t_{(i,t)} - \theta_{(i,t)}\right)^2 + (1-a)\left(B^t_{(i,t)} - B^t_{m_t(i,t)}\right)^2 \\ & - \zeta \bar{B}^t_{(i,t)} + \left(\eta + (1-\eta)\, w^t_{m_t(i,t)}\right) \boldsymbol{I}_{\{b_{(i,t)} > L_t\}} \varphi_t \Bigg] \\ & + \lambda \Bigg[a\left(B^{t+1}_{(i,t)} - \theta_{(i,t)} +\right)^2 + (1-a)\left(B^{t+1}_{(i,t)} - B^{t+1}_{m_{t+1}(i,t)}\right)^2 \\ & + \zeta \bar{B}^{t+1}_{(i,t)} + \left(\eta + (1-\eta)\, w^{t+1}_{m_{t+1}(i,t)}\right) \boldsymbol{I}_{\{b_{(i,t)} > L_t\}} \varphi_t \Bigg], \end{aligned} \tag{4}$$

where

$$\bar{B}^t_{(i,t)} \equiv \sum_{j \neq m_t(i,t)} B^t_{(j,t-1)} + \sum_{j \neq i} B^t_{(j,t)} \text{ and}$$

$$\bar{B}^{t+1}_{(i,t)} \equiv \sum_{j \neq m_{t+1}(i,t+1)} B^{t+1}_{(j,t+1)} + \sum_{j \neq i} B^{t+1}_{(j,t)},$$

which are the analogs of the negative externality term in the static model and ignore the effects of behavior of (i, t)'s matches, as well as own behavior, to simplify the analysis. The relationship between base and actual behaviors is given by

$$B^k_{(i,t)}\left(w^k_{m_k(i,t)}, b_{(i,t)}\right) = \begin{cases} L_t & \text{if } b_{(i,t)} > L_t, \text{ and } w^k_{m_k(i,t)} = 1 \\ & \quad \text{or there is public enforcement (prob. } \eta), \\ b_{(i,t)} & \text{otherwise,} \end{cases}$$

where $k = t$ or $t + 1$.

In addition, $1 - \lambda$ and λ are the payoff weights on, respectively, past and future interactions, and are inclusive of time-discounting. Laws and fines are allowed to be time-varying as noted above (and to interpret the timing of the terms at the end of each line, recall the convention that the interaction between generations $t - 1$ and t takes

In the case in which a law is relaxed, the results are essentially identical under both grandfathering assumptions.

place at time t, and that an agent born at time t is subject to the laws and fines imposed at time t).

In the text, we focus on the case in which $\lambda = 0$, so that behavior is purely backward looking and history drives behavior (but we specify the full utility function to ensure well-defined behavior; namely, even though initial behavior is backward looking, agents need to choose optimal whistle-blowing behavior in the second period of their lives in equilibrium). The more general case is discussed in the Appendix, where we show that steady-state equilibria are always identical to the equilibrium of the static model, and also characterize dynamic equilibria.

Whistle-blowing is no longer automatic in cases in which laws differ across periods. For example, an agent obeying a less tight law may wish to allow an agent of the previous generation to break a more stringent law. Thus, strategies now must include not only the base behavior as a function of an agent's type, but also the whistle-blowing choices of a law-abiding agent as a function of the behavior of her matches from the previous and the next generations.

We also assume that agents only observe their own types, and not the history of behaviors of the agents preceding them (and, for instance, do not know the realizations of the strategies of the previous generation). Allowing for some historical observation would complicate the analysis, without adding substantial insight, since strategies would depend on each possible history of realized actions, rather than best responding to a distribution of recent actions. In our analysis, agents are already correctly predicting the strategy functions used in all generations, just not the actual realizations of types, which makes the analysis tractable, and allows us to compare it to the static model.

Thus, a pure-strategy of a player can be summarized by a triplet

$$(\beta_t(\theta_i), W_{t,t-1}(\theta_i), W_{t,t+1}(\theta_i)) \in [0, 1]^3,$$

where $\beta_t(\theta_i)$ is the base behavior of the agent, $w_{t,t-1}(\theta_i)$ is the cutoff behavior from the agent's first-period match above which they whistle-blow (where generation t is matched with generation $t-1$), and $w_{t,t+1}(\theta_i)$ is the cutoff behavior from the agent's second-period match above which they whistle-blow.

A *dynamic equilibrium* is then defined as a pure-strategy perfect Bayesian equilibrium that is symmetric with respect to the agents of any given generation,[25] and is summarized by a collection of functions $\{\beta_t(\theta), W_{t,t-1}(\theta), W_{t,t+1}(\theta)\}_{t=0}^{\infty}$ describing base behavior choices, as well as whistle-blowing decisions.[26]

25. Because, as mentioned above, whistle-blowing behavior now occurs after observing other agents' behaviors, our equilibrium notion incorporates sequential rationality. With just the requirement of Bayesian equilibrium, there exist trivial equilibria in which all agents obey laws and whistle-blow on any deviations (even though this may not be in their interest ex post, since whistle-blowing then ends up off the equilibrium path). Because aspects related to Bayesian updating of beliefs are not relevant in our context, we essentially just use Bayesian equilibrium strengthened with subgame perfection (thus avoiding technical details related to updating beliefs on sets of measure zero).

26. β_0 is an initial condition, and need not be a best response to the future plays, but can be (as in the case of a steady-state equilibrium). We focus on equilibria in which play is symmetric within a generation,

4.2. Social Norms and the Effectiveness of Laws

We now establish one of our main results, showing that introducing laws that are in too strong a conflict with the prevailing norms may backfire and significantly increase law-breaking, whereas more moderate laws that are not in discord with prevailing norms may reduce behavior without causing as much lawlessness—because they change social norms in the process.

Though the general insight concerning the interplay between social norms and laws holds for all parameter values, the analysis with changing laws and behaviors is quite complex, and this motivates our focus in the next proposition on a subset of the parameters for which the sharp contrast between abrupt and gradual tightenings of laws can be transparently demonstrated. Again, since our emphasis is on how the best response of the current generation is shaped by the prevailing social norms determined by the behavior of past generations, we focus on the case in which $\lambda = 0$, which also implies that the equilibrium is unique.[27] We also focus on a stationary setting in which $F_t = F$ and $\varphi_t = \varphi$ for all t and isolate the effects of a change in laws, from a steady-state law of $L_t = L$ for all t, to some other law L'.

PROPOSITION 5. *Fix F, $\varphi > 0$, and $\zeta \geq 0$, and let $\lambda = 0$. There exists $\bar{\eta} > 0$, and for each $\eta \in (0, \bar{\eta})$ there exists $\bar{a} > 0$ such that for $a \in (0, \bar{a})$ the following holds. Suppose that there is an initial law, $L \geq (1-a)\mathbb{E}[\theta] + a$ (so nonbinding), and a society starts at time $t = 0$ in the unique steady-state equilibrium corresponding to L. For any new law $L' < (1-a)\mathbb{E}[\theta] - a$ for which there is a less than full compliance steady-state equilibrium under L':*

- (Abrupt tightening of law) *If there is an unanticipated and permanent change to L' in period 1, then all agents break the law in period 1, and some agents continue to break the law in subsequent period in any equilibrium, and behavior converges (weak*) to the lowest compliance equilibrium associated with L' over time.*
- (Gradual tightening of law) *However, for any such L' there exists a (finite) decreasing sequence of laws $\{L_t\}_{t=1}^{T}$ with $L_{T+k} = L'$ for all $k \geq 0$, such that following a switch to this sequence of laws, all agents comply with the law at their birth and play converges (weak*) to the full compliance steady-state equilibrium associated with L'.*

A significant tightening of the law creates a conflict between the prevailing norms and the law, leading to an immediate and significant increase in law-breaking. This then

as equilibrium strategies are only subscripted by time and not players' identities. There is no similar requirement of symmetry across generations.

27. Even though when $\lambda = 0$, in making their first-period decisions, individuals do not take into account their utility from the second period, once they reach the second period they still have a well-defined objective function given by (4), and by the sequential rationality of equilibrium, choose an optimal whistle-blowing decision. Thus, they have the same whistle-blowing behavior (as a function of $b_{(t,i)}$) as in the case where $\lambda > 0$.

makes it impossible for society to achieve the full compliance steady-state equilibrium. For example, starting with the full compliance steady-state equilibrium for law L, a much tighter law L' will not be enforced. This is because agents in the previous generation, even though they are grandfathered, will force their partner down to L' if they whistle-blow, but L' is much lower than their current behavior, and thus forcing their partner down to L' would cause significant miscoordination. This implies that though the initial (grandfathered) generation can whistle-blow, it will choose not to do so against the behavior that would have been chosen by the next generation absent of the law. But then knowing that there will be no whistle-blowing forthcoming from this initial generation, the next generation ignores the law, leading to pervasive law-breaking. This extent of law-breaking cannot last, because the very low types in the next generation will want to obey the law and whistle-blow on law-breakers. Nevertheless, significant law-breaking persists and society converges (from below, meaning from greater law-breaking behavior) to the lowest compliance steady-state equilibrium, with positive law-breaking. This type of "reproduction" of high levels of law-breaking is related to the "social multiplier" effects mentioned in the Introduction and discussed in the Appendix.

In contrast, a series of gradual laws converging to L' can be much more effective in containing law-breaking and can achieve full compliance. This is because each gradual tightening of the law will have a small impact on behavior, and the next generation will be willing to abide by the law as this enables both coordination with the older generation and avoidance of the costs imposed by public law enforcement. This gradual sequence of tighter laws slowly changes the prevailing norms, and as norms change, these tighter laws become more and more powerful in restricting behavior.[28] This ensures the convergence of the dynamic equilibrium without ever deviating from full compliance. Both parts of the proposition are illustrated in Figure 6.

We stated the result for $L' < (1-a)\mathbb{E}[\theta] - a$ because in this case none of the agents in the older generation would want to whistle-blow. The result extends to higher values of L' ("looser" laws), but becomes more complicated as now some of the lowest types would prefer to abide by the new stricter law, and the determination of the sequence of cutoffs is no longer in closed-form.

5. Extensions

In this section, we discuss several extensions. To facilitate the exposition, these are presented in the context of the baseline static model, and we also set $\zeta_m = \zeta_o = \zeta$ to simplify notation.

28. There is no conflict between this result and Corollary 1, since the result here concerns the limit behavior following a sequence of tighter laws (starting from full compliance), not a comparative static result. Crucially, each gradual tightening of the law changes the social norm for the next generation, and it is this property that enables full compliance to be achieved ultimately.

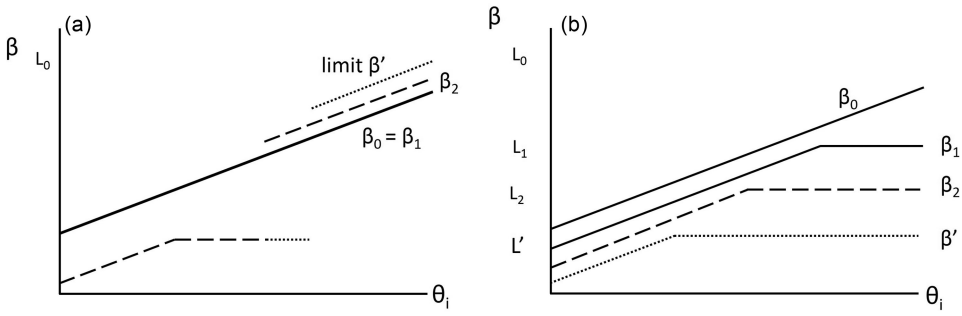

FIGURE 6. Dynamics of the effects of a law. (a) An abrupt tightening of the law leading to a switch from full compliance to the lowest compliance equilibrium. (b) A gradual tightening of the law while maintaining the full compliance equilibrium.

The first extension focuses on multiple types of behavior regulated by multiple laws, showing that badly designed laws in one dimension make other dimensions of laws ineffective. Other extensions show how assortative matching between individuals depending on their behavior changes the results, and highlight the implications of relaxing some other assumptions in the model.

5.1. Multiple Laws

We first consider an extension of the model to multiple behaviors.

Each individual has a two-dimensional vector of types (θ_i^1, θ_i^2), and chooses two behaviors $\left(b_i^1, b_i^2\right)$. Individual types are drawn from a joint type distribution G. To simplify the discussion in this extension and focus on the interaction in laws, we assume that G is such that $\theta_i^1 = \theta_i^2$ for all i (meaning that G puts probability 1 on the "diagonal" where each agent has the same type on both dimensions).

Each individual now matches (uniformly at random) with two others, corresponding to the two dimensions of behavior. We denote the partners of agent i in these two matches by $(m^1(i), m^2(i))$. There are two different laws applying to each dimension of behavior, denoted by (L^1, L^2). We take payoffs to be a direct generalization of (1),

$$u_i = \sum_{k=1,2} \Bigg[-a\left(B_i^k - \theta_i^k\right)^2 - (1-a)\left(B_i^k - B_{m^k(i)}^k\right)^2 - \zeta^k \bar{B}_i^k - \left(\eta + (1-\eta)\, w_{m^k(i)}^k\right) \boldsymbol{I}_{\{b_i^k > L^k\}} \varphi^k \Bigg], \tag{5}$$

where

$$\bar{B}_i^k \equiv \frac{\sum_{j \neq i} B_j^k}{n-1}.$$

As equation (5) shows, there is no direct payoff linkage between the two dimensions of behavior. The key linkage results from the fact that an agent cannot whistle-blow on her partner on either dimension if she has broken any laws (e.g., a firm that whistle-blows on their business partner for their environmental transgressions will also have her tax records scrutinized). We also assume that an agent only observes the behavior of her partner in the specific dimension for which they are matched with her.

Equilibria can again be defined similarly and are in monotone strategies. However, the linkage between the two dimensions of behavior makes equilibrium more subtle. In particular, compared to a world in which there is only one type of behavior, an individual may be more willing to obey the law because by doing so, she has the option to whistle-blow on the other dimension. But also an individual will be less willing to obey the law because others who are breaking the law on the second dimension will not be able to whistle-blow on her on the first dimension. Though it is not possible in general to determine whether the first or the second effect dominates (and thus whether the linkage between the two dimensions of behavior makes law-breaking more or less likely), the next proposition characterizes some cases in which the law on one dimension encourages law-abiding behavior on the other and also some cases in which it encourages law-breaking on the other dimension.

PROPOSITION 6. *Consider the model with multiple laws described above. Suppose that there is a nontrivial law on the first dimension ($L^1 < 1$) but no law on the second dimension ($L^2 = 1$), and an associated equilibrium, ($\beta^1(\theta)$, $\beta^2(\theta)$) with a law-breaking threshold $\theta^{1*} < 1$ on the first dimension. Then*

- *There exist $\bar{\delta}$ and $\underline{\delta}$ such that if a law $\tilde{L}^2 \in (L^1 - \underline{\delta}, L^1 + \bar{\delta})$ is imposed on the second dimension, then there is a new equilibrium $(\tilde{\beta}^1(\theta), \tilde{\beta}^2(\theta))$ that involves a law-breaking threshold $\tilde{\theta}^{1*} > \theta^{1*}$ (i.e., there is less law-breaking on the first dimension).*
- *There exists $\underline{b}L > 0$ such that if a law $\tilde{L}^2 < \underline{b}L$ is imposed on the second dimension, then the new equilibrium $(\tilde{\beta}^1(\theta), \tilde{\beta}^2(\theta))$ involves a law-breaking threshold $\tilde{\theta}^{1*} < \theta^{1*}$ (i.e., there will be more law-breaking on the first dimension).*

This proposition thus shows that the imposition of law sufficiently similar to prevailing laws on the second dimension increases law-abiding behavior on the first dimension, whereas a very strict law on the second dimension encourages law-breaking on the first. The first result reflects the option value of behaving on one law to be able to whistle-blow on the other. It suggests that introducing a well-designed second law can increase the value of law enforcement to individual citizens, encouraging them to comply with existing laws.

The second result is particularly important because it suggests that a *poorly designed* law on the second dimension can decrease the power of the first law. This is because a law that is too strict naturally leads to extensive law-breaking on the second dimension. But then those agents cease to be able to whistle-blow on the first dimension, the anticipation of which discourages law-abiding behavior on the first dimension. This reasoning may explain why laws that are viewed as unfair or unjust

lead to the erosion of respect for other laws and regulations in society, as discussed in the Introduction. This proposition also clarifies what a poorly design law is in our context. If the law imposed on the second dimension of behavior is moderate, so that we are in the first part of the proposition, it is not only effective but also improves law-abiding behavior on the first dimension. If it is excessively tight, so that we are in the second part of the proposition, then it backfires and leads to more law-breaking on the first dimension.

In light of this extension, we can return to our discussion of the broken windows theory of Kelling and Wilson (1982) in the Introduction. Consistent with the broken windows theory, pervasive law-breaking on the second dimension encourages law-breaking on the first dimension (which may be more costly for society, e.g., if ζ^1 is very high). Then, we might conclude that greater resources have to be devoted to prevent law-breaking behavior on the second dimension. Though this would not be entirely incorrect in our model, the real problem is not that there is high behavior on the second dimension, but that the law on the second dimension is too strict (badly designed). Decriminalizing moderate behaviors on the second dimension would be much more effective in preventing the "culture of law-breaking," and for reducing law-breaking on the first dimension.

5.2. Assortative Matching and the Power of Laws

We have so far assumed uniform random matching. In many situations, those engaged in illegal activities may be able to partner with others also doing so. A simple way of flexibly allowing for this possibility is to introduce some degree of assortative matching. To analyze this possibility, let us modify the model in one other dimension, by assuming that there is a continuum of agents.[29] Then, suppose that with probability q, an individual matches with another agent with exactly the same type as herself, and with probability $1 - q$, she matches with somebody else uniformly at random. As $q \to 0$, we converge to our baseline model. As $q \to 1$, we converge to fully assortative matching. Equilibria are again similar to the baseline, but both the threshold for law-breaking, θ^*, and the strategies of law-abiding and law-breaking agents are now modified as outlined in the next proposition.

PROPOSITION 7. *Suppose that in the baseline static model we have probability* $q \in [0, 1]$ *of assortative matching. Then Propositions* 1 *and* 3 *apply, and Proposition* 4 *is modified such that*

$$\beta(\theta_i) = \min\left[(a + (1-a)q)\theta_i + (1-a)(1-q)x, L\right] \text{ if } \theta_i < \theta^*$$

and

$$\beta(\theta_i) = (a + (1-a)q)\theta_i + (1-a)(1-q)\mathbb{E}[\theta|\theta > \theta^*] \text{ if } \theta_i > \theta^*,$$

29. Though this requires some care in the definition of a matching, we omit these details.

where x is the unique point satisfying

$$x = \mathbb{E}\left[\min[(a + (1-a)q)\theta + (1-a)(1-q)x, L]\right].$$

In addition, θ^* is decreasing in q.

The intuition for the change in the law-abiding and law-breaking strategies is straightforward (as they both reflect the increased likelihood of a match with somebody very close to one's own type and thus reduced mismatch). Though both law-abiding and law-breaking agents benefit from a higher probability of assortative matching, the gain is greater for law-breakers, since it enables them to avoid facing whistle-blowing and punishment. Hence, higher q encourages greater law-breaking. As a consequence, enforcing laws is more difficult in activities where law-breakers can more easily find and operate with other law-breakers, thus insulating them from whistle-blowing. This suggests that laws relying on private enforcement (whistle-blowing) can be more effective in certain activities, such as tax evasion, where the degree of assortative matching is likely to be more limited, than others such as smuggling.

5.3. Costly Whistle-Blowing

In this subsection, we introduce a cost of whistle-blowing, ε, and to simplify the discussion we set the probability of public enforcement to zero, that is, $\eta = 0$. Now agent i's payoff is

$$u_i(B_i, B_{-i}) = -a(B_i - \theta_i)^2 - (1-a)\left(B_i - B_{m(i)}\right)^2 - \zeta \bar{B}_i - w_{m(i)} \boldsymbol{I}_{\{b_i > L\}} \varphi - w_i \boldsymbol{I}_{\{b_{m(i)} > L\}} \varepsilon. \tag{6}$$

Let us start by considering the decision to whistle-blow (since it is no longer optimal to whistle-blow on all law-breakers). Since whistle-blowing reduces the action of one's partner to L, the gain to doing so for an agent who has chosen behavior B_i is

$$(1-a)\left[\left(B_i - B_{m(i)}\right)^2 - (B_i - L)^2\right] = (1-a)\left[B_{m(i)}^2 - L^2 + 2LB_i - 2B_{m(i)}B_i\right],$$

and whistle-blowing is a strict best response when this quantity exceeds ε. It is straightforward to verify that it will do so when $B_{m(i)} > w_\varepsilon(b_i)$ (where the switch from B_i to b_i reflects the fact that agent i is law-abiding and hence $B_i = b_i$). Incorporating this behavior, we can again reduce our multistage game to a static one in which each agent chooses $\beta(\theta_i)$, and define an equilibrium as a pure-strategy Bayesian equilibrium. Then the existence of equilibrium and the monotonicity of strategies again follow straightforwardly using the same arguments as Proposition 1.

The monotonicity of strategies allows us to simplify the whistle-blowing strategy by writing w_ε as a function of θ_i rather than b_i (by simply substituting for $b_i = \beta(\theta_i)$), so that whistle-blowing happens when the behavior of one's partner is above a

threshold $\overline{W}_{\varepsilon}(\theta_i)$. It is also straightforward to see that every equilibrium is characterized by a threshold θ^* such that types above this threshold break the law. Though the exact characterization in Proposition 4 no longer applies, it can be shown that any equilibrium with costly whistle-blowing converges to an equilibrium of the form given in Proposition 4 as $\varepsilon \to 0$.

PROPOSITION 8. *As $\varepsilon \to 0$, an equilibrium with costly whistle-blowing converges (weak*) to an equilibrium of the baseline model with a zero cost of whistle-blowing.*

An interesting implication of costly whistle-blowing (with $\eta = 0$) is the possibility of a distinction between "laws on the book" and laws that are actually enforced. In particular, for any positive cost of whistle-blowing, ε, there exists a range of behaviors on which nobody will blow the whistle. So even though these are banned behaviors (i.e., they are above L), everybody understands that they will be tolerated by society. There then exists another threshold above L, say $\hat{L}$, beyond which whistle-blowing begins (but not all behaviors above $\hat{L}$ will be whistle-blown upon as this will depend on the behavior of the partner).

Another result that follows immediately from the analysis of costly whistle-blowing is that the government can also increase the enforcement of laws by rewarding whistle-blowing. This can offset the reduction in whistle-blowing due to the cost, and in fact, it can induce whistle-blowing in situations (such as those highlighted in Proposition 5) where the unwillingness of agents to whistle-blow makes laws ineffective.

5.4. Whistle-Blowing by Law-Breakers

In the baseline model, we assumed that only law-abiding agents can whistle-blow. The analysis is similar when this assumption is relaxed. It can be verified that if the fine, φ, is sufficiently large (though less than the threshold $\bar{\varphi}$ defined in Proposition 3, above which the unique equilibrium involves all agents obeying the law), then no agent who has herself broken the law will whistle-blow. But when φ is sufficiently small, an agent who has broken the law by a small amount may prefer to whistle-blow when matched against somebody who breaks the law by a very large amount, because this will get both of them to coordinate at L (and thus avoid the disutility resulting from a significant mismatch). Nevertheless, we can utilize the same strategy as in the previous subsection, where we first solve for whistle-blowing, and then substituting for this, we look for an equilibrium in terms of $\beta(\theta_i)$s for each agent. An equilibrium can then be defined in the same fashion and is once again characterized by a threshold θ^*.

This extension also enables us to see how amnesty laws, which partially wave sanctions against law-breakers who turn whistle-blower, affect law-breaking. In our model, when law-breakers can whistle-blow and avoid punishment because of an amnesty, whistle-blowing greatly increases.[30]

30. For more on amnesty and incentives, see Spagnolo (2008).

6. Welfare

We conclude our formal analysis with a brief discussion of welfare properties. The (simple) laws that we have examined in this paper are not always fully optimal. If a social planner could influence ex post behavior, then she would do so in a pair-dependent manner, since there is a mismatch (miscoordination) externality within a pair. The laws we have studied in this paper are nevertheless relevant because they are much simpler to implement than fully pair-dependent taxes or subsidies conditional on the exact behavior of each agent. Moreover, they are optimal in some cases as we show next.

6.1. A Case in Which a Law and Full Compliance are Optimal

In the case where agents place full weight on their own type and do not care about coordinating with their match ($a = 1$), but still experience externalities, the welfare analysis is transparent and full compliance to simple laws becomes optimal.

To see this, consider a utilitarian social planner's objective, which is to maximize the sum of the utility of all individuals in society, or equivalently

$$-\sum_i \left((B_i - \theta_i)^2 + (\zeta_m + (n-1)\zeta_0)B_i\right).$$

This is equivalent to minimizing

$$\sum_i \left(B_i - \theta_i - \frac{(\zeta_m + (n-1)\zeta_0)}{2}\right)^2 + \text{Constant},$$

where the constant at the end is independent of the B_is.

This is a well-understood incentive problem, where the planner and the agent's objectives differ by the (externality) term $((\zeta_m + (n-1)\zeta_0))/2$, and where the planner cannot observe θ_i. In this case, the optimal mechanism from the planner's perspective is to put in place a cap on behavior (e.g., see Holmstrom 1984), which corresponds to a fully enforced law in our setup.

Thus, in cases in which coordination incentives are absent, but externalities are present, fully enforced laws are optimal. Once coordination enters preferences, however, the optimal structure becomes more complicated.

6.2. Partial Compliance

Once coordination incentives are present it is no longer optimal to choose laws and fines that force full compliance. In particular, note that for any law $L > 0$, there exists some level of fine, defined as $\bar{\varphi}$ in Proposition 3, which will ensure full compliance. The next proposition shows that provided that the externality on others is not too large (compared to the complementarity), full compliance is not optimal. To establish this

result, we focus on the choice of a utilitarian social planner (and treat fines paid in equilibrium as pure transfers). To keep the notation uncluttered, we consider a case in which $\zeta_m = \zeta_o = \zeta$, but the extension is straightforward (with the cutoff being based on a weighted average of ζ_m and ζ_o).[31]

PROPOSITION 9. *Fix a distribution of types F and law $L \in (0, 1)$. Then for any partial compliance equilibrium (associated with some fine $\varphi < \bar{\varphi}$), there exists $\bar{\zeta}$ such that for any $\zeta < \bar{\zeta}$, utilitarian social welfare is greater under this partial compliance equilibrium than under full compliance. Moreover, there exist distributions and levels of externality for which the total expected utility maximizing law and fine involve partial compliance.*

This proposition implies that for sufficiently small externalities, any partial compliance is better than full compliance—and thus, the utilitarian social planner would not like to choose a very high fine $\varphi > \bar{\varphi}$. Naturally, this result may not be true if externalities are very large, but still points out that there are natural reasons for allowing some law-breaking in society.

The intuition for this proposition is instructive. Partial compliance leads to law-breaking only by high types. When a high type matches with a low type, the latter can whistle-blow and reduce the behavior of her partner back down to L. This insulates low types from the mismatch consequences of law-breaking by high types, and thus low-types (those below the law-breaking threshold θ^*) are only worse off in a partial compliance equilibrium relative to full compliance because of the negative externalities from high behavior. This reasoning also implies that under partial compliance, there will only be behavior above L when two high types (two types above θ^*) match. But it is costly for a utilitarian social planner to force two high-type agents down to L. In fact, by revealed preference, these high types strictly prefer partial compliance to full compliance (as they could have chosen law-abiding behavior even under $\varphi < \bar{\varphi}$). Consequently, if the externality that law-breakers impose on society is not too large (i.e., provided that $\zeta < \bar{\zeta}$), it is optimal to permit high-type agents to break the law when they are matched to each other. Notably, this conclusion holds for any $\varphi < \bar{\varphi}$, and thus for any partial compliance equilibrium (though of course the relevant threshold $\bar{\zeta}$ for the size of externalities does vary across partial compliance equilibria).

7. Conclusion

This paper has examined the interplay between social norms and the enforcement of laws. The main motivation for our approach comes from the fact that many laws are

31. ζ_m is the term that appears in all of the equilibrium calculations, since an agent's actions can only affect their partner's behavior. In contrast, when working through a welfare analysis, both ζ_m and ζ_o would appear, as an agent's ultimate utility is affected by both their partner's behavior and other agents' behaviors, even if those other behaviors are beyond the agent's influence—each agent is influenced by one other via ζ_m and $n - 2$ others via ζ_o, and so a given agent's behavior has an average externality of $(\zeta_m + (n - 2)\zeta_o)/(n - 1)$ on society, and so this would be the relevant term that would replace the ζ expressed here.

ineffective, in part because they conflict with prevailing social norms, making private agents unwilling to cooperate with law enforcement (for example, by whistle-blowing), while at the same time effective laws successfully change social norms, significantly increasing their potency.

In our model, agents choose a behavior (e.g., tax evasion, production of low-quality products, corruption, substance abuse, etc.), and then are matched uniformly with another agent. Utility depends negatively on the average behavior of other agents and on the mismatch between the behaviors of the two partners. A law is a cap (upper bound), L, on behavior and a law-breaker, when detected, pays a fine and has her behavior forced down to L. Incentives to break the law depend on social norms because detection has to rely, at least in part, on private cooperation and whistle-blowing. Law-abiding agents have an incentive to whistle-blow because this will reduce their partner's behavior, ameliorating the mismatch.

When laws are in conflict with norms so that many others are breaking the law, anticipating little whistle-blowing, each agent has further incentives to also break the law. We show that all equilibria are characterized in terms of a threshold for law-breaking, and a range of comparative statics of the lowest compliance equilibrium are presented. For example, greater fines for law-breaking reduce behavior and law-breaking among law-abiding agents, but also increase behavior among law-breakers (because law-breakers choose their behavior in the hope of matching with other law-breakers, and in this case, the composition of law-breakers has shifted toward higher type agents). A tighter law (banning more behaviors) leads to greater law-breaking, but also reduces behavior among law-breakers.

We further show that laws that are in strong conflict with prevailing social norms may backfire and lead to a significant decline in law-abiding behavior in society. In contrast, gradual imposition of moderately tight laws can be effective in changing social norms and can thus alter behavior without leading to pervasive lawlessness. We also show that excessively strict (or badly designed) laws concerning some dimensions of behavior encourage broader law-breaking in society.

We view our paper as a step toward a systematic analysis of the interaction between laws and norms. Important next steps in this research program could, inter alia, include the following:

- integrating the two-way interaction between laws and norms with collective decision-making in society (as laws are often determined by voting or elected legislatures that reflect societal preferences);[32]
- enriching the types of laws and fines (e.g., allowing fines that depend on the exact level of behavior) or uncertainty in the interpretation and enforcement of laws;
- allowing for collusion and bribes among agents (see, e.g., Acemoglu and Verdier 2000);

32. Endogenous constitutions and laws without social norms are discussed in, among others, Barbera and Jackson (2004) and Acemoglu, Egorov, and Sonin (2012).

- studying different types of social norms in shaping whether individuals dare whistle-blow;
- examining the impact of laws on the inferences that agents draw about the preferences and intentions of law-breaking and law-abiding agents (see, e.g., Benabou and Tirole 2011);
- investigating substitutability and complementarity between private and public law enforcement;
- studying how laws and norms influence peoples' preferences (endogenizing the θs);
- considering informal collective sanctions instead of laws, such as ostracism or group punishments;[33]
- empirically investigating these interactions, and the role of history and prevailing social norms on law-breaking behavior and law enforcement.

Appendix: Proofs

We start with a lemma that establishes the monotonicity and structure of best replies to (symmetric) strategies of the other players.

LEMMA A.1. *Let $\beta(\cdot)$ be a best response for some agent i to any strategy $\beta'(\cdot)$ played by the other players. Then $\beta(\theta_i) \geq \beta(\theta_i')$ whenever $\theta_i > \theta_i'$. Furthermore, either $\beta(\theta_i) \leq L$ in which case*

$$\beta(\theta_i) = \min\left[L, a\theta_i + (1-a)\mathbb{E}[\min[L, \beta'(\theta_{m(i)})]]\right];$$

or $\beta(\theta_i) > L$ and then

$$\beta(\theta_i) = a\theta_i + (1-a)\mathbb{E}\left[\beta'(\theta_{m(i)})|\beta'(\theta_{m(i)}) > L\right].$$

Proof of Lemma A.1. Consider any measurable $\beta'(\cdot)$ used by the other agents. Following (1) we can write agent i's expected payoff, including only the external effects that i's behavior affects, as

$$\begin{aligned}\mathbb{E}u_i(b_i, \theta_i, \beta') = &-\zeta_m \mathbb{E}[\min\ (\beta'(\theta_{m(i)}), L)] \qquad \text{if } b_i \leq L;\\ &-a\left(b_i - \theta_i\right)^2 - (1-a)\mathbb{E}\left[\left(b_i - \min\ [\beta'(\theta_{m(i)}), L]\right)^2\right]\end{aligned} \tag{A.1}$$

33. For various game theoretic explorations of such behaviors see Kandori (1992), Lippert and Spagnolo (2011), Jackson, Rodriguez-Barraquer, and Tan (2012), Ali and Miller (2013, 2015), and Acemoglu and Wolitzky (2015).

and

$$\begin{aligned}
&\mathbb{E}u_i(b_i,\theta_i,\beta') = \\
&\quad -\zeta_m\big(\eta\mathbb{E}[\min\ (\beta'(\theta_{m(i)}),L)] + (1-\eta)\mathbb{E}[\beta'(\theta_{m(i)})]\big) \qquad \text{if } b_i > L; \\
&\quad -\ \Pr(\beta'(\theta_{m(i)}) \le L)\big(a(L-\theta)^2 + (1-a)\mathbb{E}\big[(L-\beta'(\theta_{m(i)}))^2 | \beta'(\theta_{m(i)}) \le L\big] + \varphi\big) \\
&\quad -\eta\Pr(\beta'(\theta_{m(i)}) > L)(a(L-\theta)^2 + (1-a)\times 0 + \varphi) \\
&\quad -(1-\eta)\Pr(\beta'(\theta_{m(i)}) > L)\big(a(b-\theta)^2 \\
&\quad +(1-a)\mathbb{E}\big[(b-\beta'(\theta_{m(i)}))^2 | \beta'(\theta_{m(i)}) > L\big]\big). \qquad \text{(A.2)}
\end{aligned}$$

To understand (A.1), note that the first line is simply the expected externality from her match. The second line comes from the utility of the distance between action and own type and match's behavior; noting that the agent's match's behavior is always at most L, either due to whistle-blowing or enforcement if it starts above L. To understand (A.2), note that the first line is simply the expected externality from her match. The second line comes from noting that with probability $\Pr(\beta'(\theta_{m(i)}) \le L)$ the agent will meet a law-abiding agent, who will whistle-blow on her. In this case, her behavior will be reduced down to L and she will incur the fine φ. The third line comes from noting that with probability $\eta\Pr(\beta'(\theta_{m(i)}) > L)$, she will meet a fellow law-breaker but there will be public enforcement, and in this case she will again be subject to the fine, and her behavior will be reduced to L (but so will the behavior of her partner accounting for the 0 term). Finally, with probability $(1-\eta)\Pr(\beta'(\theta_{m(i)}) > L)$ the agent will meet another law-breaker and will not be subject to public enforcement. In this case, her utility is given by the convex combination of the distances between her behavior and her type, and her behavior and the behavior of the fellow law-breaker ($\beta'(\theta_{m(i)})$), and this gives the last line.

It then follows from the first-order necessary conditions that a best response must satisfy either $\beta(\theta_i) \le L$ and

$$\beta(\theta_i) = \min\big[L, a\theta_i + (1-a)\mathbb{E}[\min[L,\beta'(\theta_{m(i)})]]\big]; \qquad \text{(A.3)}$$

or $\beta(\theta_i) > L$ and

$$\beta(\theta_i) = a\theta_i + (1-a)\mathbb{E}\big[\beta'(\theta_{m(i)}) | \beta'(\theta_{m(i)}) > L\big]. \qquad \text{(A.4)}$$

Both of these functions are nondecreasing in θ_i, and (A.4) is always greater than (A.3). So, the only possible violation of the nondecreasing property would have to be a setting where the best response at θ_i is smaller than (or equal to) L, whereas at $\theta_i' < \theta_i$ the best responses are greater than L.

Consider any $\underline{b}_i$ and $\bar{b}_i$ such that $\underline{b}_i \leq L$ and $\bar{b}_i > L$ and let us evaluate

$$\mathbb{E}u_i(\bar{b}_i, \theta_i, \beta') - \mathbb{E}u_i(\underline{b}_i, \theta_i, \beta').$$

To rule out this last situation where the best reply at θ_i is no higher than L, whereas at $\theta_i' < \theta_i$ the best reply is higher than L, it is enough to show that $\mathbb{E}u_i(\bar{b}_i, \theta_i, \beta') - \mathbb{E}u_i(\underline{b}_i, \theta_i, \beta')$ is increasing in θ_i. From (A.1) and (A.2), it follows that

$$\begin{aligned}\mathbb{E}u_i(\bar{b}_i, \theta_i, \beta') - \mathbb{E}u_i(\underline{b}_i, \theta_i, \beta') = \\ \big(1 - (1-\eta)\Pr[\beta'(\theta_{m(i)}) > L]\big)a\big[\big(\underline{b}_i - \theta_i\big)^2 - \big(L - \theta_i\big)^2\big] \\ + (1-\eta)\Pr\big[\beta'(\theta_{m(i)}) > L\big]a\Big[\big(\underline{b}_i - \theta_i\big)^2 - \big(\bar{b}_i - \theta_i\big)^2\Big] + X\end{aligned}$$

where X is a term that is independent of θ_i. We simplify to get

$$\begin{aligned}\mathbb{E}u_i(\bar{b}_i, \theta_i, \beta') - \mathbb{E}u_i(\underline{b}_i, \theta_i, \beta') = \\ \big(1 - (1-\eta)\Pr[\beta'(\theta_{m(i)}) > L]\big)2a(L - \underline{b}_i)\theta_i \\ + (1-\eta)\Pr\big[\beta'(\theta_{m(i)}) > L\big]2a(\bar{b}_i - \underline{b}_i)\theta_i + Y\end{aligned}$$

where Y is a term that is independent of θ_i. This expression is increasing in θ_i whenever $\Pr[\beta'(\theta_{m(i)}) > L] > 0$. In the case in which $\Pr[\beta'(\theta_{m(i)}) > L] = 0$, a strategy of L offers a strictly higher payoff than any strategy above L, and so the best reply at $\theta_i' < \theta_i$ could not be higher than L. This concludes the proof.

Proof of Propositions 1 and 4. The monotonicity follows from Lemma A.1, and then the existence of a cutoff type describing any agent's best reply follows directly. Thus, by the lemma, any symmetric pure strategy equilibrium must be such that there is a threshold θ^* such that agents above this threshold break the law and those below do not (with full compliance corresponding to $\theta^* = 1$ and full law-breaking to $\theta^* = 0$). Moreover, from the characterization of strategies in the lemma it also follows that such an equilibrium $\beta(\cdot)$ must be such that when $\theta_i > \theta^*$,

$$\beta(\theta_i) = a\theta_i + (1-a)\mathbb{E}[\beta(\theta)|\theta > \theta^*].$$

Taking expectations conditional upon $\theta > \theta^*$ on both sides leads to

$$\mathbb{E}[\beta(\theta)|\theta > \theta^*] = \mathbb{E}[\theta|\theta > \theta^*],$$

and thus

$$\beta(\theta_i) = a\theta_i + (1-a)\mathbb{E}[\theta|\theta > \theta^*]. \tag{A.5}$$

When $\theta_i < \theta^*$, the Lemma A.1 implies that

$$\beta(\theta_i) = \min\left[L, a\theta_i + (1-a)\mathbb{E}[\min(\beta(\theta), L)]\right].$$

Thus,

$$\beta(\theta_i) = \min(L, a\theta_i + (1-a)x), \tag{A.6}$$

where $x \equiv \mathbb{E}[\min(\beta(\theta), L)]$.

Next, we show that for any θ^*, x is uniquely defined, as claimed.

We expand $x = \mathbb{E}[\min\ (\beta(\theta), L)]$ as

$$x = \Pr(\theta < \theta^*)\mathbb{E}[\min(\min(L, a\theta + (1-a)x), L)|\theta < \theta^*] + \Pr(\theta > \theta^*)L$$

or

$$x = \Pr(\theta < \theta^*)\mathbb{E}\left[\min\ (L, a\theta + (1-a)x)|\theta < \theta^*\right] + \Pr(\theta > \theta^*)L. \tag{A.7}$$

The right-hand side of this equation is a contraction (it is nondecreasing in x but always less than one-for-one as $1 - a < 1$). Thus, by the contraction mapping principle (A.7) has a solution and it is unique.

We have therefore established Proposition 4, subject to proving existence, which then completes the proof of Proposition 1.

Existence of an equilibrium follows straightforwardly as $\theta^* = 1$ is always an equilibrium: if all other agents abide by the law, then any action $b_i > L$ is dominated by $b_i = L$, and so then best replies must all be completely low-abiding. Then $\beta(\theta_i) = \min[a\theta_i + (1-a)x, L]$ is a best reply to itself when $\theta^* = 1$, providing one equilibrium. □

Proof of Proposition 2. Applying Lemma A.1 to the case in which $L = 1$, it follows that the unique best response of any agent i to the strategies of other agents is to set

$$\beta(\theta_i) = a\theta_i + (1-a)\mathbb{E}[b_{m(i)}].$$

Writing $\mathbb{E}[b]$ for the expected strategy of a randomly selected agent, the above expression implies that

$$\mathbb{E}[b] = \mathbb{E}[a\theta + (1-a)\mathbb{E}[b]],$$

which given $0 < a < 1$ has a unique solution of

$$\mathbb{E}[b] = \mathbb{E}[\theta].$$

Thus, the unique equilibrium without laws involves

$$\beta(\theta_i) = a\theta_i + (1-a)\mathbb{E}[\theta],$$

as claimed. □

Proof of Proposition 3. We have already shown that in any equilibrium, equations (2) and (3) have to hold given the law-breaking threshold θ^*. Clearly, the relevant thresholds are fixed points: if all other agents use threshold θ^*, then it is a best response for each to also use threshold θ^*. Given the monotonicity of best responses already

established in Lemma A.1, it is sufficient to look at the payoffs from law-breaking and law-abiding for the agent of type θ^*.

We first note that there is always an equilibrium with $\theta^* = 1$ given that $\varphi > 0$. This follows since, if all other agents abide by the law, then by breaking the law an agent's action will reduced down to L with certainty and the agent will pay the fine φ. The agent would have a strictly higher payoff from choosing L directly. Thus, it is a best response to obey the law when all other agents do, and so having $\theta^* = 1$ and actions as in equation (2) is an equilibrium. We next consider characterize equilibria for which $\theta^* < 1$.

Suppose that type θ^* decides to take a law-breaking action. Then, from (A.5) in the proof of Lemma A.1, the optimal law-breaking action will be $a\theta^* + (1-a)y$ (with $y \equiv \mathbb{E}[\theta|\theta > \theta^*]$), and since $\theta^* \geq L$ this behavior is above L. Her payoff can then be written based on (A.2) as

$$
\begin{aligned}
&-\Pr(\theta < \theta^*)\left(a(L-\theta^*)^2 + (1-a)\mathbb{E}[(L-a\theta-(1-a)x)^2|\theta < \theta^*] + \varphi\right) \\
&-\eta \Pr(\theta > \theta^*)(a(L-\theta^*)^2 + \varphi) \\
&-(1-\eta)\Pr(\theta > \theta^*)\left(a(a\theta^* + (1-a)y - \theta^*)^2\right. \\
&\left.+ (1-a)\mathbb{E}[(a\theta^* + (1-a)y - a\theta - (1-a)y)^2|\theta > \theta^*]\right) \\
&-\zeta_m\left(\eta\mathbb{E}[\min\ (\beta(\theta), L)] + (1-\eta)\mathbb{E}[\beta(\theta)]\right). \qquad \text{(A.8)}
\end{aligned}
$$

Suppose, instead, that type θ^* chooses to abide by the law, in which case she will set her behavior to $b = \min[L, a\theta^* + (1-a)x]$ and from (A.1) receive expected payoff

$$
\begin{aligned}
&-\Pr(\theta > \theta^*)\left(a(b-\theta^*)^2 + (1-a)(b-L)^2\right) \\
&-\Pr(\theta < \theta^*)(a(b-\theta^*)^2 + (1-a)\mathbb{E}[(b-a\theta-(1-a)x)^2|\theta < \theta^*]) \\
&\quad -\zeta_m\mathbb{E}[\min\ (\beta(\theta), L)]. \qquad \text{(A.9)}
\end{aligned}
$$

The threshold type θ^* is given by setting (A.8) equal to (A.9). We set the former on the left-hand side and the latter on the right-hand side and equal to each other, and then to help with the comparative statics analysis, we then transfer all terms involving φ and ζ_m to the right-hand side, and all other terms to the left-hand side; noting that that $\Pr(\theta > \theta^*) = 1 - \Pr(\theta < \theta^*)$; and dividing both sides by $(1-\eta)\Pr(\theta > \theta^*)$ (which is strictly positive in view of the fact that $\eta < 1$ and $\theta^* < 1$ as we see below). After

doing this we obtain

$$
\begin{aligned}
&-a(1-a)^2(\theta^*-y)^2-(1-a)a^2\mathbb{E}[(\theta-\theta^*)^2|\theta>\theta^*]\\
&-\left(\frac{1-(1-\eta)\Pr(\theta>\theta^*)}{(1-\eta)\Pr(\theta>\theta^*)}\right)a(L-\theta^*)^2\\
&+\left(\frac{1}{(1-\eta)\Pr(\theta>\theta^*)}\right)[\Pr(\theta>\theta^*)(1-a)(b-L)^2+a(b-\theta^*)^2]\\
&+\left(\frac{\Pr(\theta<\theta^*)}{(1-\eta)\Pr(\theta>\theta^*)}\right)(1-a)(\mathbb{E}[(b-a\theta-(1-a)x)^2|\theta<\theta^*]\\
&-\mathbb{E}[(L-a\theta-(1-a)x)^2|\theta<\theta^*])\\
=&\left(\frac{1-(1-\eta)\Pr(\theta>\theta^*)}{(1-\eta)\Pr(\theta>\theta^*)}\right)\varphi+\zeta_m\frac{(\mathbb{E}[\beta(\theta)]-\mathbb{E}[\min\ (\beta(\theta),L)])}{\Pr(\theta>\theta^*)}.
\end{aligned}
$$

Noting that

$$\mathbb{E}[\beta(\theta)]-\mathbb{E}[\min\ (\beta(\theta),L)]=\Pr(\theta>\theta^*)\mathbb{E}[\theta-L|\theta>\theta^*],$$

this simplifies to

$$
\begin{aligned}
&-a(1-a)^2(\theta^*-y)^2-(1-a)a^2\mathbb{E}[(\theta-\theta^*)^2|\theta>\theta^*]\\
&-\left(\frac{1-(1-\eta)\Pr(\theta>\theta^*)}{(1-\eta)\Pr(\theta>\theta^*)}\right)a(L-\theta^*)^2\\
&+\left(\frac{1}{(1-\eta)\Pr(\theta>\theta^*)}\right)\left[\Pr(\theta>\theta^*)(1-a)(b-L)^2+a(b-\theta^*)^2\right]\\
&+\left(\frac{\Pr(\theta<\theta^*)}{(1-\eta)\Pr(\theta>\theta^*)}\right)(1-a)\left(\mathbb{E}\left[(b-a\theta-(1-a)x)^2|\theta<\theta^*\right]\right.\\
&\left.-\mathbb{E}\left[(L-a\theta-(1-a)x)^2|\theta<\theta^*\right]\right)\\
=&\left(\frac{1-(1-\eta)\Pr(\theta>\theta^*)}{(1-\eta)\Pr(\theta>\theta^*)}\right)\varphi+\zeta_m\mathbb{E}[\theta-L|\theta>\theta^*]. \qquad \text{(A.10)}
\end{aligned}
$$

Note that since all of the transformations used to obtain (A.10) involve adding and subtracting numbers and dividing by the positive number, the left-hand side of (A.10) is greater than the right-hand side if and only if (A.8) is greater than (A.9).

We now consider the set of θ^* that satisfy (A.10).

First, note that $\theta^* \geq L$, since for type L choosing behavior L strictly dominates anything above it: L does strictly better against law-abiders, and gets maximal utility against law-breakers since all of their behaviors are reduced to exactly L, the most preferred point of an L type. So, let us consider the set of $\theta^* \in [L, 1)$ that satisfy

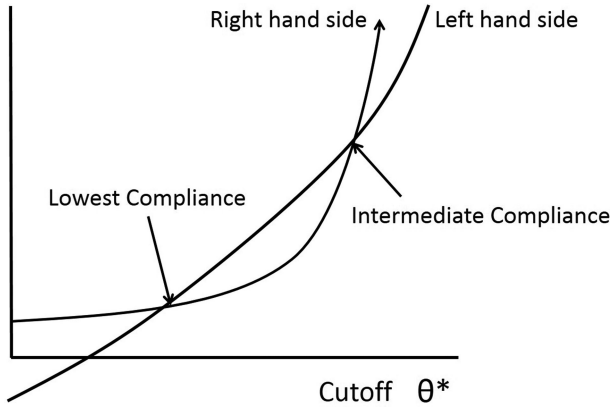

FIGURE A.1. Multiplicity of partial compliance equilibria.

(A.10). The fact that, for $\theta^* = L$, choosing behavior equal to L strictly dominates choosing any behavior above L, implies that the expression in (A.8) is strictly less than the expression in (A.9) when $\theta^* = L$. This implies that the left-hand side of (A.10) is less than its right-hand side at $\theta^* = L$.

Second, as shown above, if others are all abiding by the law, then it is a strict best response to abide by the law for an agent. Put differently, when $\theta^* = 1$ (A.8) is strictly less than the expression in (A.9). This implies that the left-hand side of (A.10) is strictly less than its right-hand side as $\theta^* \to 1$ (where this statement is for the limit $\theta^* \to 1$, since the expressions in (A.10) diverge at $\theta^* = 1$ due to the fact that they are both divided by $(1 - \eta)\Pr(\theta > \theta^*)$, though of course this does not affect their relative ranking as $\theta^* \to 1$).

Next, note that the right-hand side is continuous and increasing in θ^* and is always positive, and as just argued the right-hand side starts out and ends up strictly above the left-hand side. Note that the left-hand side is also continuous in θ^*. Thus, if there is an intersection and an interior equilibrium (rather than tangency which we discuss in the next paragraph), then the smallest intersection must involve the left-hand side cutting the right-hand side from below, and the greatest intersection must involve the reverse. This is pictured in Figure A.1, with the less curved line corresponding to the left-hand side and the more curved line to the right-hand side.[34] As Figure A.1 makes it clear, when there is an intersection between the two curves, there must be at least two of them and thus two equilibrium thresholds (all in the range $\theta^* \in [L, 1)$).

Now consider the case in which φ is large, and recall that the right hand side is increasing in φ whereas the left hand side is not. In that case the right-hand side still starts out at a higher intercept on the vertical axis, and for large enough φ will

34. The figures are drawn for $b = L$, in which case the left-hand side stays bounded whereas the right-hand side asymptotes to infinity as $\theta^* \to 1$ (since then $\Pr(\theta > \theta^*) \to 0$); but qualitatively similar pictures hold for the other case, though the left-hand side may also be unbounded.

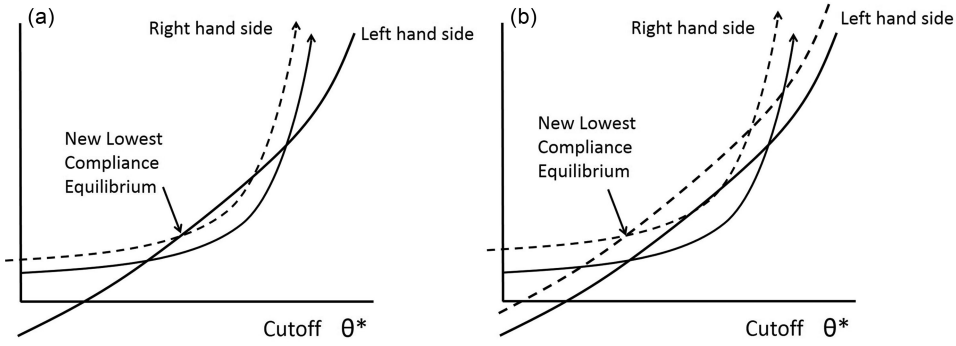

FIGURE A.2. Comparative statics in the lowest compliance equilibrium. (a) Comparative statics of the equilibrium thresholds due to an increase in fine, public enforcement, or externality. (b) comparative statics of the equilibrium thresholds due to a tighter law.

have no intersection with the left-hand side. The threshold value for there to be a tangency between the two curves is defined as $\bar{\varphi}$ (and when there are intersections for all values of φ, we set this as $\bar{\varphi} = 0$). So when $\varphi > \bar{\varphi}$, there are no interior equilibria and the unique equilibrium is full compliance. When $\bar{\varphi} > 0$ and $\varphi < \bar{\varphi}$, there are at least two intersections as drawn in Figure A.1, and the full compliance equilibrium always continues to exist. This completes the proofs of Propositions 3 and 4. □

Proof of Corollary 1. Recall that the lowest compliance equilibrium threshold for law-breaking, θ^*, is given by the smallest intersection of (A.8) and (A.9), or equivalently, by the left-hand side of (A.10) intersecting the right-hand side from below as shown in Figure A.1.

First note that the definition of "small" changes from footnote 20 is satisfied in Figure A.1 since the lowest compliance equilibrium is given by the intersection of the two curves (rather than a tangency point). Now all of the comparative statics follow by noting how the parameters in question shift the right-hand and left-hand sides of (A.10). In particular, the right-hand side shifts up and the left hand side is unchanged when φ, η, or ζ_m increases, as pictured in Figure A.2(a).

When L decreases (and laws become more strict), both sides of equation (A.10) shift as pictured in Figure A.2(b). The right-hand side shifts up, whereas the left-hand side's shift depends on parameters, and so which effect dominates depends on the level of the parameters. In particular, there is a critical ζ_m, $\bar{\zeta}_m$ (as a function of the other parameters), above which the equilibrium threshold increases, and below which it decreases (since ζ_m affects the magnitude of the change of the right-hand side equation).

Specifically, the derivative of the right hand side is $-\zeta_m$. The derivative of the left-hand side with respect to L depends on whether $b = L$ or differs from it. Let us first treat the case in which when L declines, $b = L$ in some small neighborhood. In that case, the derivative of the left-hand side is $-2a(\theta^* - L)$, and so the critical $\bar{\zeta}_m$

for which the two effects balance is

$$\bar{\zeta}_m = 2a(\theta^* - L).$$

Above this, the increase in the right-hand side is larger from a decrease in L (leading to an increase in θ^*), whereas below it, the shift in the left-hand side is greater from a decrease in L (leading to a decrease in θ^*).

Next, let us suppose that $b < L$ when L declines slightly from its starting level. In that case, the derivative of the left-hand side is

$$\begin{aligned}
&-2a\left(\theta^* - L\right) \\
&+\left(\frac{1}{(1-\eta)\Pr(\theta > \theta^*)}\right)\left(2\Pr(\theta > \theta^*)(1-a)(L-b) + 2\Pr(\theta < \theta^*)a(\theta^* - L)\right. \\
&\left.-2\Pr(\theta < \theta^*)(1-a)\mathbb{E}[(L - a\theta - (1-a)x)|\theta < \theta^*]\right)
\end{aligned}$$

and so

$$\begin{aligned}
\bar{\zeta}_m = 2a&\left(\theta^* - L\right) \\
&-\left(\frac{1}{(1-\eta)\Pr(\theta > \theta^*)}\right)\left(2\Pr(\theta > \theta^*)(1-a)(L-b)\right. \\
&+2\Pr(\theta < \theta^*)a(\theta^* - L) \\
&\left.-2\Pr(\theta < \theta^*)(1-a)\mathbb{E}[(L - a\theta - (1-a)x)|\theta < \theta^*]\right).
\end{aligned}$$

This latter expression could be 0 or less, in which case the change is then unambiguous.

The rest of the results in parts 1 and 3 of the corollary follow directly from the inspection of equations (2) and (3). Finally, part 2 follows by considering large changes defined as changes that shift one of the two curves so much that the interior equilibria disappear. This case clearly leads to full compliance as the unique equilibrium and reduces overall average behavior. □

Proof of Proposition 5. Given $\lambda = 0$, generation 0's behavior is described by

$$\beta_0(\theta_i) = \beta^{\text{no law}}(\theta_i) = a\theta_i + (1-a)\mathbb{E}[\theta]. \tag{A.11}$$

Consider an unanticipated tightening of the law to $L' < (1-a)\mathbb{E}[\theta] - a$. Let us start by analyzing the whistle-blowing behavior of type $\theta = 0$ of generation $t = 0$. Because society was previously in a full compliance steady-state equilibrium, all agents from generation $t = 0$, and in particular type $\theta = 0$, are law-abiding according to L and are hence grandfathered after the law change, and can whistle-blow. By whistle-blowing, type $\theta = 0$ can bring her partner with behavior $b > L'$ down to L'

(this period's law). From equation (4), the gain of doing so is

$$(b-\beta_0(0))^2-(L'-\beta_0(0))^2=(b-(1-a)E[\theta])^2-(L'-(1-a)E[\theta]))^2$$
$$<(b-(1-a)\mathbb{E}[\theta])^2-a^2$$

where the equality uses the fact that, from (A.11), $\beta_0(\theta=0)=(1-a)\mathbb{E}[\theta]$, and the inequality exploits the fact that $L'<(1-a)\mathbb{E}[\theta]-a$. Then for any behavior $L'<b\leq a+(1-a)\mathbb{E}[\theta]$, the best response for type $\theta=0$ is not to whistle-blow. Thus, it is, a fortiori, the unique best response for any type to not whistle-blow against such behavior. □

Next, consider behavior of generation $t=1$. Now note that (i) this generation cannot whistle-blow on the previous generation (that was law-abiding according to L); (ii) there will not be whistle-blowing from the previous generation provided that $b\leq a+(1-a)\mathbb{E}[\theta]$; (iii) behavior in the next period does not matter for their utility given $\lambda=0$. Then, for η sufficiently small (i.e., $\eta<\bar{\eta}$ with $\bar{\eta}>0$ suitably defined) this generation will have a unique best response of $\beta^{\text{nolaw}}(\theta)$ as specified in (A.11).

Next, consider the lowest compliance equilibrium β' and associated threshold $\theta^{*\prime}$ associated with L'. Note that $\mathbb{E}[\beta_1(\theta)]=\mathbb{E}[\theta]<\mathbb{E}[\beta'(\theta)|\theta>\theta^{*\prime}]$. Thus, the best response β_2 to β_1 has a lower cutoff $\theta_2^*\leq\theta^{*\prime}$ and also, with the same reasoning as in the proof of Lemma A.1, results in a strategy such that $\mathbb{E}[\beta_2(\theta)|\theta>\theta_2^*]\leq\mathbb{E}[\beta'(\theta)|\theta>\theta^{*\prime}]$. Iterating, the cutoff for law breaking of generation t, $\theta_t^*\leq\theta'$, and $\mathbb{E}[\beta_t(\theta)|\theta>\theta_t^*]\leq\mathbb{E}[\beta'(\theta)|\theta>\theta^{*\prime}]$ holds for every t, with an increasing sequence of cutoffs θ_t^*.

Note also, that in period 2 onward, types with $\theta\leq L'$ will wish to abide by the law, since they can force any partners' behavior to be at most L', as their partners are not grandfathered under the original law, but must obey L'. Thus, although $\theta_1^*=0$, thereafter $\theta_t^*>L'$. Since $\{\theta_t^*\}$ is an increasing sequence and bounded above by $\theta^{*\prime}$, it converges. Its limit point must then be the threshold for the lowest compliance steady-state equilibrium, and hence $\theta^{*\prime}$. Weak* convergence of the equilibrium actions to the steady-state then follows from the (uniquely-defined) form of the best responses described in the proof of Lemma A.1.

Finally, note that the equilibrium is unique as each generation's behavior is completely determined by the previous generation's.

(Gradual Tightening of Law). Next consider a sequence of gradual tightenings defined by

$$L_1=\beta_0(1)-\varepsilon=a+(1-a)\mathbb{E}[\theta]-\varepsilon$$

and

$$L_t=\max\ [L',L_{t-1}-\varepsilon],$$

with ε defined below.

We next show inductively that each generation follows the current law given that all previous generations have. As the induction step, suppose that for all previous generations $t' < t$, $\beta_{t'}$ conforms to the law $L_{t'}$ and is a best response to the previous generation's strategy. We will then show that this is true for t.

We first show that for any type, choosing L_t is preferable to any behavior above L_t. The loss in utility from such conformity is bounded above by $a + (1 - a)\varepsilon^2$ (where the first term is the loss in not matching own type, and the latter is the furthest change between any match from the previous generation from abiding compared to playing any higher strategy). Breaking the law, on the other hand, leads to a loss of at least $\eta\varphi$ due to public enforcement. Thus, for $a + (1 - a)\varepsilon^2 < \eta\varphi$ the agents are better off abiding by the law than breaking it. Picking any $\bar{a} < \eta\varphi$ and then setting $\varepsilon < \sqrt{(\eta\varphi - a)/(1 - a)}$ suffices to establish the result.

Once $L_t = L'$ after a finite number of steps, all future generations will obey the law L', and thereafter the argument in Lemma A.1 implies that the behavior of all types is nonincreasing in time (and decreasing when it is above their steady-state behavior), and again the equilibrium strategy profile converges (weak*) to the full compliance steady-state equilibrium profile, as the limit point must be an equilibrium and it involves full compliance.

Proof of Proposition 6. To prove the first part, and note that with $L^2 = 1$ (i.e., no law on the second dimension), the equilibrium threshold θ^* is given by exactly the same characterization as in Proposition 4, and in particular, by the intersection of the left- and right-hand sides of equation (A.10). Next consider the imposition of a law $\tilde{L}^2 = L^1$. Then, in the lowest compliance equilibrium, all law-abiding agents (those with types less than θ^{1*}) in the first dimension would also abide by the law in the second dimension (since the two dimensions are symmetric). But then types just above θ^{1*} now have an additional reason to abide by the law (on both dimensions), since this will give them an option to whistle-blow on very high behaviors on the other dimension. This shifts up the benefit to law-abiding behavior and thus the right-hand side of equation (A.10), increasing θ^{1*} and reducing law-breaking. By continuity, the same argument applies to $\tilde{L}^2$ not too far from L^1, thus establishing that there exist $\bar{\delta}$ and $\underline{\delta}$ such that any $\tilde{L}^2 \in (L^1 - \underline{\delta}, L^1 + \bar{\delta})$ induces more law-abiding behavior on the first dimension. □

To prove the second part, consider $\tilde{L}^2 > 0$ but small. This implies that only types very close to zero will abide by the law on the second dimension and thus there will be very little whistle-blowing on the first dimension, reducing θ^{1*} and thus encouraging law-breaking on the first dimension.

Proof of Proposition 7. This proposition follows immediately from writing out the best responses of law-abiding and law-breaking agents as in the proofs of Propositions 3 and 4 with the assortative matching technology, and thus the details are omitted to save space. □

Proof of Proposition 8. To prove this result, note that the threshold for whistle-blowing $\overline{W}_{\varepsilon}(\theta_i)$ defined in the text as a function of the cost of whistle-blowing, ε, converges to

L (for all $\theta_i \leq \theta^*$) as ε converges to zero. This ensures that any sequence of equilibria with costly whistle-blowing converges to an equilibrium of the baseline model (weak*) using standard upper hemicontinuity arguments (e.g., the sort of argument behind Theorem 2 of Jackson et al. 2002). □

Proof of Proposition 9. Let us first suppose that $\zeta = 0$, so that there are no externalities. Then, for a given law L and distribution of types F, consider first the level of fine $\varphi_1 > \bar{\varphi}$ so that there is full compliance (from Proposition 3). We will now compare utilitarian welfare in this full compliance equilibrium to that under the level of fine $\varphi_2 < \bar{\varphi}$ leading to partial compliance. Under partial compliance there exists a law-breaking threshold $\theta^* < 1$ such that below this threshold, there is law-abiding behavior, and above this threshold, there is law-breaking (and thus law-breaking for a positive measure of agents).

First, note that for all $\theta' \leq \theta^*$, behavior under both scenarios (where the fine is φ_1 and φ_2) is given by

$$\beta(\theta') = \min\ [a\theta' + (1-a)x, L],$$

where x is the unique fixed point of $x = \mathbb{E}\,[\min\ [a\theta + (1-a)x, L]]$. This follows, from similar arguments as in our earlier proofs (cfr. Proposition 4; because law-abiding agents can whistle-blow on law-breakers when they match and reduce their behavior to L). Since there are no externalities, this implies that these agents have exactly the same expected utility under both scenarios, that is,[35]

$$U_2(\theta') = U_1(\theta') \text{ for all } \theta' \leq \theta^*,$$

where $U_k(\theta')$ denotes the expected utility of type θ' under scenario $k = 1, 2$.

Second, observe that all agents above $\theta' > \theta^*$ can still choose law-abiding behavior when the fine is $\varphi_2 < \bar{\varphi}$, which will give them exactly the same expected utility $U_1(\theta')$ as in the equilibrium with fine $\varphi_1 < \bar{\varphi}$. Since they choose to break the law, by revealed preference we have that $U_2(\theta') \geq U_1(\theta')$. But in fact we also know that only type θ^* is indifferent between law-breaking and law-abiding behavior, and thus we have

$$U_2(\theta') > U_1(\theta') \quad \text{for all } \theta' > \theta^*.$$

This implies that utilitarian social welfare is always strictly greater with partial compliance than full compliance when there are no negative externalities (i.e., $\zeta = 0$). But given this strict ordering, it also follows that for any level of fine $\varphi_2 < \bar{\varphi}$ (and thus for any partial compliance equilibrium), there exists $\bar{\zeta} > 0$ such that for $\zeta < \bar{\zeta}$,

35. Note that as before, $\theta^* \geq L$, and so $\beta(\theta^*) = L$ (and similarly for all higher θ_i). Therefore, the eventual behavior B for all agents with $\theta_i \geq \theta^*$ when matched with a law-abiding agent will be L. So under both scenarios, the equilibrium behavior of (law-abiding) agents with $\theta_i < \theta^*$ can be determined by using the fact that the behavior of any of their partners with $\theta_i \geq \theta^*$ will be L. Thus, the equilibrium calculations are identical in both cases.

utilitarian social welfare is greater under the (nearby)[36] partial compliance equilibrium than under full compliance.

We prove the last claim in the proposition for an extreme distribution with equal weight on just two types, as this is easily extended to a continuous distribution that approximates this distribution. Consider a distribution with equal weight on two types, $1 > \theta_2 > \theta_1 > 0$, and let $\bar{\theta} = (\theta_1 + \theta_2)/2$.

Consider any law L that is strict enough to have partial compliance, in particular such that

$$\bar{\theta} + \frac{a}{2}(\theta_2 - \theta_1) > L > \theta_1,$$

and a low enough φ so that there exists a partial/low compliance equilibrium (so that θ_1s comply and θ_2s do not). Then, the low compliance equilibrium is such that $\beta(\theta_2) = \theta_2$ and

$$\beta(\theta_1) = \frac{2a\theta_1 + (1-a)L}{1+a}.$$

The full compliance equilibrium has the same action for type θ_1, but $\theta_2 = L$ for type θ_2.

Let us thus check that there is a law $L < \bar{\theta} + a(\theta_2 - \theta_1)/2$ that leads to higher expected total utility than any $L \geq \bar{\theta} + a(\theta_2 - \theta_1)/2$. Given that the types θ_2 and θ_1 are both better off when meeting their own type under a partial compliance equilibrium than under no law, then taking externalities as sufficiently small, it is enough to show that the total utility of a type θ_1 meeting a θ_2 is better for at least some $L < \bar{\theta} + a(\theta_2 - \theta_1)/2$ than a law above that threshold. The total utility just accounting for the cross-type matching and the portion affected by the law (when the externality is 0) is

$$-a\left(\frac{2a\theta_1 + (1-a)L}{1+a} - \theta_1\right)^2 - a(L-\theta_2)^2 - 2(1-a)\left(\frac{2a\theta_1 + (1-a)L}{1+a} - L\right)^2.$$

The derivative of this with respect to L is

$$-2a\left(\frac{(1-a)}{1+a}\right)^2(L-\theta_1) - 2a(L-\theta_2) - 4(1-a)\left(\frac{2a}{1+a}\right)^2(L-\theta_1).$$

36. Noting that the equilibrium changes continuously in a neighborhood if the equilibrium was not a tangency point, and that corresponding utilities vary continuously.

When $L = \bar{\theta} + \frac{a}{2}(\theta_2 - \theta_1)$, this becomes

$$-a\frac{(1-a)^2}{1+a}(\theta_2-\theta_1) + a^2(1-a)(\theta_2-\theta_1) - (1-a)\frac{8a^2}{1+a}(\theta_2-\theta_1),$$

$$= a(1-a)(\theta_2-\theta_1)\left[\frac{-1+a+a+a^2-8a}{1+a}\right],$$

which is negative, implying that the total utility is increased as L is decreased below $L = \bar{\theta} + a(\theta_2 - \theta_1)/2$ and there is partial compliance. □

A.1 Dynamic and Steady-State Equilibria

Let us consider the equilibria of the dynamic economy introduced in the text. These dynamic equilibria can be characterized in a similar manner as our static equilibria. Moreover, as stated in the text, steady-state equilibria (under constant laws and fines) coincide with the equilibria of the static model. These results are summarized in the next proposition.

PROPOSITION A.1. *Dynamic equilibria are in monotone strategies (and are characterized in the Appendix, where we also discuss the proof of existence).*

Moreover, let $\mathcal{B}$ be the set of equilibrium strategies of the static game, and $\mathcal{B}^*$ be the set of steady-state behaviors from the equilibria of the dynamic game (with the same parameter values as the static game and with $L_t = L$ and $\varphi_t = \varphi$ for all t). Then $\mathcal{B} = \mathcal{B}^*$, and every steady-state equilibrium is described by a strategy of the form

$$\beta^*(\theta_i) = \begin{cases} \beta_{abiding}(\theta_i) & \text{if } \theta_i < \theta^* \\ \beta_{breaking}(\theta_i) & \text{if } \theta_i > \theta^* \end{cases}$$

for some threshold θ^*, where $\beta_{abiding}(\theta_i)$ and $\beta_{breaking}(\theta_i)$ are as defined in equations (2) and (3) in Proposition 4, and then i whistle-blows occurs if and only if $\theta_i < \theta^*$ and a match breaks the law.[37]

In steady state, all generations have the same distribution of behavior, and thus the matching facing each agent with respect to the previous and the next generations is entirely analogous to that in the static model. Proposition A.1 then implies that the comparative statics from the static model generalize to steady-state equilibria.

Two points on dynamic equilibria are also worth noting. First, there will generally be multiple dynamic equilibria. One special case of interest is when $\lambda = 0$, whereby each generation best-responds to the behavior of the previous generation, which ultimately ties back to a starting condition, and thus at one extreme there is uniqueness. When λ is high, there can be a multiplicity of equilibria because of the expectation of

37. Again, we do not specify actions for the 0-probability event that $\theta = \theta^*$.

how the next generation will behave, which may vary as laws change with time, and so this is a source of multiplicity that is distinct from the multiplicity of steady states.

Second, the dynamic equilibria of this model also generate a type of "social multiplier" (e.g., Glaeser, Sacerdote, and Scheinkman 1996). In particular, if there is high law-breaking at time t, then these agents will not whistle-blow on their future partners, and anticipating this, there will be high law-breaking at time $t + 1$, and so on.

Proof of Proposition A.1. The characterization of equilibrium proceeds in a similar fashion to the proof of Proposition 4, except that the distribution of behavior in the previous and the next generations needs to be treated separately, explicitly taking into account both the type-dependent whistle-blowing behavior of one's partner and the probability of public enforcement. In particular, to characterize this behavior let us define an indicator variable for whether an agent of generation $t' \in \{t-1, t+1\}$ is either whistle-blown upon by her match from generation t or faces public enforcement. Let

$$I_{t,t'}(\theta_i, b_{t'}) = 1 \text{ if } b_{t'} > L_{t'} \text{ and either } b_{t'} > W_{t,t'}(\theta_i) \text{ or there is public enforcement,}$$

and

$$I_{t,t'}(\theta_i, b_{t'}) = 0 \text{ otherwise.}$$

□

With this notation, and also noting that a dynamic equilibrium involves a sequence of thresholds, $\{\theta_t^*\}_{t=0}^{\infty}$, above which the agent of generation t breaks the law, we can determine the behavior of each generation as

$$\beta_t(\theta_i) = \min[a\theta_i + (1-a)((1-\lambda)x_{t,t-1}(\theta_i) + \lambda x_{t,t+1}(\theta_i)), L] \text{ if } \theta_i < \theta_t^*,$$

whereas if $\theta_i > \theta_t^*$ then

$$\beta_t(\theta_i) = a\theta_i + (1-a) \times \frac{(1-\lambda)\Pr(I_{t-1,t}(\theta_{t-1}, \beta_t(\theta_i)) = 0) z_{t,t-1}(\theta_i) + \lambda \Pr(I_{t+1,t}(\theta_{t+1}, \beta_t(\theta_i)) = 0) z_{t,t+1}(\theta_i)}{(1-\lambda)\Pr(I_{t-1,t}(\theta_{t-1}, \beta_t(\theta_i)) = 0) + \lambda \Pr(I_{t+1,t}(\theta_{t+1}, \beta_t(\theta_i)) = 0)},$$

where, for $t' \in \{t-1, t+1\}$,

$$x_{t,t'}(\theta_t) = \mathbb{E}[\beta_{t'}(\theta_{t'}) \left(1 - I_{t,t'}(\theta_t, \beta_{t'}(\theta_{t'}))\right) + L_{t'} I_{t,t'}(\theta_t, \beta_{t'}(\theta_{t'}))]$$

and

$$z_{t,t'}(\theta_t) = \mathbb{E}[\beta_{t'}(\theta_{t'}) | I_{t',t}(\theta_{t'}, \beta_t(\theta_t)) = 0].$$

It follows from these expressions that all dynamic equilibria are in monotone strategies (i.e., $\beta_t(\theta_i)$, $w_{t,t-1}(\theta_i)$, $w_{t,t+1}(\theta_i)$ is nondecreasing in θ_i for each t).

Note that we have specified the existence of thresholds, $\{\theta_t^*\}_{t=0}^{\infty}$, determining behavior, but have not claimed the existence of thresholds for whistle-blowing behavior. Depending on the specific sequence of laws, whistle-blowing may take a more complex form. For example, if tomorrow's law is stricter than today's, then someone taking a low action today may wish to whistle-blow on all law-breakers tomorrow, whereas an agent who is exactly at today's law may prefer to whistle-blow only on those from the next generation whose behavior is sufficiently above her own behavior. Whistle-blowing still involves cutoffs above which an agent whistle-blows and which are monotone in type, but they may not depend on type.

Existence of equilibrium in this dynamic setting is also more challenging. We next sketch the argument for existence.

Existence of Equilibria in the Dynamic Game (sketch). Existence of equilibria is complicated by two issues: first, the game is discontinuous due to whistle-blowing, and lacks strategic complementarities, and second, it has an infinite horizon of overlapping generations. With discretionary whistle-blowing (whereas it was dominant in the static case), one needs to account for those actions explicitly. We provide a sketch of an argument to establish existence, at least of equilibria with randomizations over whistle-blowing. We begin by allowing for mixed strategies (even though for many cases the mixing would be present only in tie-breaking and then only for exceptional cases), where a symmetric mixed strategy for the agents of some generation in our dynamic game is a mixture of behaviors as a function of their types and cutoff for whistle-blowing as a function of their matches' realized behaviors (and specifying mixtures at the exact cutoffs). The most tractable way to represent this is as distributional strategies in the sense of Milgrom and Weber (1985). In particular, we could represent a strategy for an agent of a given generation as a joint distribution over that agent's possible types, behaviors, and cutoffs, for whistle-blowing on past and future generations (and mixtures at the cutoffs)—where the distribution's marginal on types agrees with the distribution on types F_t. Next, let us discuss the existence of equilibria for any finite horizon, taking as given the starting and ending strategies, as these can then be viewed as if they were just games with t players (the representative players of each generation). That is, for every specification of σ_0 and σ_{t+1}, we look for vectors of strategies $(\sigma_0, \sigma_1, \ldots, \sigma_t, \sigma_{t+1})$ for which each of $\sigma_1, \ldots, \sigma_t$ are best responses to all of the other strategies. The game has discontinuities due to the whistle-blowing, and so such existence does not follow standard arguments. Nonetheless, payoffs are continuous as a function of the outcomes: chosen behaviors and *mixtures* on whistle-blowing and fines. Thus, we then view the game as one satisfying the definition of a game with indeterminate outcomes from Jackson et al. (2002), for which there exist equilibria involving randomization over the outcomes as a function of types. In this game the whistle-blowing (and eventual behaviors) would be determined as part of the equilibrium endogenously as a limit of finite game approximations to the game. Here, those whistle-blowing behaviors will be given by pure strategies in cases where there is a strict incentive one way or the other, and thus will match with a best response, but may involve randomization in cases of indifference (all up to sets of measure 0). In cases of

indifference, given the structure of this game, the mixtures on whistle-blowing can be realized by mixtures by the players—as they happen at points of indifference. Thus, they can (almost surely) be translated back into an actual mixed strategy equilibrium of the game (while at the same time imposing that whistle-blowing is sequentially rational in the finite approximations to ensure subgame perfection). Finally, these equilibria can be embedded in the common space given by the set of all infinite sequences of strategies $(\sigma_0, \sigma_1, \ldots, \sigma_t, \ldots)$, by letting E_t be the set of all infinite sequences of strategies for which $\sigma_1, \ldots, \sigma_t$ are best responses to all of the other strategies. So, by establishing that E_t is nonempty and compact for each t, it then follows that $E_{t+1} \subset E_t$ for all t. Hence by Tychonoff's Theorem, these sets have a nonempty intersection, $\cap E_t = E_\infty$, which is the set of equilibria for the infinite horizon. □

Finally, the result about the equivalence of behaviors in the static model and in steady-state equilibria follows immediately by observing that in steady state, the maximization problem of the agents is similar to that in the static model.

References

Acemoglu, Daron (1995). "Reward Structures and the Allocation of Talent." *European Economic Review*, 39, 17–33.

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin (2012). "Dynamics and Stability of Constitutions, Coalitions, and Clubs." *American Economic Review*, 102(4), 1446–1476.

Acemoglu, Daron, Davide Cantoni, Simon Johnson, and James A. Robinson (2013). "The Consequences of Radical Reform: The French Revolution." *American Economic Review*, 101(7), 3286–3307.

Acemoglu, Daron and Matthew O. Jackson (2015). "History, Expectations, and Leadership in the Evolution of Social Norms." *Review of Economic Studies*, 82, 423–456.

Acemoglu, Daron and Thierry Verdier (2000). "The Choice between Corruption and Market Failures." *American Economic Review*, 90(1), 194–211.

Acemoglu, Daron and Alex Wolitzky (2015). "Sustaining Cooperation: Community Enforcement vs. Specialized Enforcement." NBER Working Paper No. 21457.

Akerlof, George and Janet L. Yellen (1994). *Gang Behavior, Law Enforcement, and Community Values.* Working Paper, Brookings Institution, Values and Public Policy Series.

Ali, S. Nageeb and David A. Miller (2013). "Enforcing Cooperation in Networked Societies." mimeo Penn State University.

Ali, S. Nageeb and David A. Miller (2015). "Ostracism and Forgiveness." forthcoming, American Economic Review.

Alonso, Ricardo, Wouter Dessein, and Niko Matouschek (2008). "When Thus Coordination Require Centralization?" *American Economic Review*, 98(1), 145–179.

Aldashev, Gani, Imane Chaara, Jean -Philippe Platteau, and Zaki Wahhaj (2012). "Formal Law as a Magnet to Reform Custom." *Economic Development and Cultural Change*, 60, 795–828.

Barbera, Salvador and Matthew O. Jackson (2004). "Choosing How to Choose: Self-Stable Majority Rules and Constitutions," *Quarterly Journal of Economics*, 119, 1011–1048.

Barfield, Thomas (2010). *Afghanistan: A Cultural and Political History.* Princeton University Press, Princeton, NJ.

Becker, Gary S. (1968). "Crime and Punishment: An Economic Approach," *Journal of Political Economy* 76, 169–217.

Becker, Gary S. and George J. Stigler (1974). "Law Enforcement, Malfeasance, and Compensation of Enforcers," *The Journal of Legal Studies* 3, 1–18.

Benabou, Roland and Jean Tirole (2011). "Laws and Norms." NBER Working Paper No. 17579.
Benoît, Jean-Pierre and Juan Dubra (2004). "Why do Good Cops Defend Bad Cops," *International Economic Review*, 45, 787–809.
Berkowitz, Daniel, Katharina Pistor, and Jean -Francois Richard (2003). "Economic Development, Legality, and the Transplant Effect." *European Economic Review*, 47, 165–195.
Bernstein, Lisa (1992). "Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry." *The Journal of Legal Studies*, 21, 115.
Bierstedt, Robert (1963). *The Social Order*, 2nd ed., McGraw-Hill, NY.
Bisin, Alberto and Thierry Verdier (2001). "The Economics of Cultural Transmission and the Dynamics of Preferences," *Journal of Economic Theory*, 97, 298–319.
Calvó-Armengol, A. and Yves Zenou (2004). "Social Networks And Crime Decisions: The Role of Social Structure In Facilitating Delinquent Behavior." *International Economic Review*, 45, 939–958.
Carbonara, Emmanuela, Francesca Parisi, and Georg von Wangenheim (2006). "Legal Innovation and the Compliance Problem." *Minnesota Journal of Law, Science and Technology*, 9, 837–860.
Cooter, Robert (1998). "Expressive Law and Economics." *The Journal of Legal Studies*, 27, S2: 585–607.
Doepke, Matthias and Fabrizio Zilibotti (2008). "Occupational Choice and the Spirit of Capitalism." *The Quarterly Journal of Economics*, 123, 747–793.
Dyck, Alexander, Adair Morse, and Luigi Zingales (2010). "Who Blows the Whistle on Corporate Fraud." *Journal of Finance*, 65, 2213–2253.
Ellickson, Robert (1991). *Order Without Law.* Harvard University Press, Cambridge.
Ferrer, Rosa (2010). "Breaking the Law when Others Do: A Model of Law Enforcement with Neighborhood Externalities." *European Economic Review* 54, 163–180.
Galor, Oded (2011). *Unified Growth Theory*. MIT Press, Cambridge, MA.
Gibbs, Jack P. (1965). "Norms: The Problem of Definition and Classification." *American Journal of Sociology*, 70, 586–594.
Glaeser, Edward L., Bruce Sacerdote, and Jose A. Scheinkman (1996). "Crime and Social Interactions." *Quarterly Journal of Economics* 111, 507–548.
Hay, Jonathan R., Andrei Shleifer, and Robert W. Vishny (1996). "Toward a Theory of Legal Reform," *European Economic Review*, 40, 559–567.
Hay, Jonathan R. and Andrei Shleifer (1998). "Private Enforcement of Public Laws: A Theory of Legal Reform." *American Economic Review*, 88(2), 398–403.
Hoffman, Martin L. (1977). "Moral Internalization: Current Theory and Research" *Advances in Experimental Social Psychology*, Vol. 10. Academic Press, New York.
Holmstrom, Bengt (1984). "On the Theory of Delegation." In *Bayesian Models in Economic Theory*, edited by M. Boyer and R. Kihlstrom. North-Holland, New York, pp. 115–141.
Jackson, Matthew O., Tomas Rodriguez-Barraquer, and Xu Tan (2012). "Social Capital and Social Quilts: Network Patterns of Favor Exchange." *American Economic Review*, 102(5), 1857–1897.
Jackson, Matthew O., Leo K. Simon, Jeroen M. Swinkels, and R. Zame William (2002). "Communication and Equilibrium in Discontinuous Games of Incomplete Information." *Econometrica*, 70, 1711–1740.
Kandori, Michihiro (1992). "Social Norms and Community Enforcement," *Review of Economic Studies*, 59, 63–80.
Lippert, Steffen and Giancarlo Spagnolo (2011). "Networks of Relations and Word-of-Mouth Communication," *Games and Economic Behavior*, 72, 202–217.
Lynn, John A. (1997). *The Giant of the Grand Siecle: French Armies, 1600–1715*. Cambridge University Press, Cambridge, UK.
McAdams, Richard H. and Eric B. Rasmusen (2007). "Norms in Law and Economics." In *Handbook of Law and Economics*, Vol. 2, edited by A. Mitchell Polinsky and Steven Shavell. North Holland Press
Mailath, George J. and Larry Samuelson (2006). *Repeated Games and Reputations*. Oxford University Press, New York.
Morris, Richard T. (1956). "A Typology of Norms." *American Sociological Review*, 21, 610–613.

Nisbett, Richard and Dov Cohen (1996). *Culture of Honor: The Psychology of Violence in the South.* Westview Press.

Parisi, Francesco and Georg von Wangenheim (2006). "Legislation and the Countervailing Effects from Social Norms." In *Evolution and the Design of Institutions*, edited by Christian Schubert and Georg von Wangenheim. Routledge.

Parsons, Timothy (2010). *The Rule of Empires: Those Who Built Them, Those Who Endured Them, and Why They Always Fall.* Oxford University Press.

Pinker, Steven (2011). *The Better Angels of Our Nature: Why Violence Has Declined.* Viking, New York.

Pistor, Katharina (1996). "Supply and Demand for Contract Enforcement in Russia: Courts, Arbitration, and Private Enforcement." *Review of Central and East European Law*, 22, 55–87.

Posner, Eric (2002). *Laws and Social Norms*. Harvard University Press, Cambridge.

Posner, Richard (1997). "Social Norms and the Law: An Economic Approach" *American Economic Review*, 87(2), 365–369.

Rasmusen, Eric Bennett (1996). "Stigma and Self-Fulfilling Expectations of Criminality." *Journal of Law and Economics*, 39, 519–544.

Sah, Raaj K. (1991). "Social Osmosis and Patterns of Crime." *Journal of Political Economy*, 99(6), 1272–1295.

Shavell, Steven M. (2004). *Foundations of Economic Analysis of Law.* Belknap Press of Harvard University Press, Cambridge, MA.

Simmel, Georg (1903). *The Metropolis and Mental Life*. Petermann, Dresden.

Simmel, Georg (1908). *Sociology: Investigations on the Forms of Association*. Duncker and Humblot, Leipzig.

Sorokin, Pitirim A. (1947). *Society, Culture, and Personality*. Harper and Brothers, New York, NY.

Spagnolo, Giancarlo (2008). "Leniency and Whistleblowers in Antitrust." In *Handbook of Antitrust Economics*, edited by P. Buccirossi, Chap. 12. MIT Press.

Tabellini, Guido (2008). "The Scope of Cooperation: Values and Incentives." *Quarterly Journal of Economics*, 123, 905–950.

Tyler, Tom R. (1990). *Why Do People Obey the Law*. Yale University Press, New Haven, CT.

Williams, Robin M. (1960). *American Society: A Sociological Interpretation*. Alfred A. Knopf, New York NY.

Wilson, James Q. and George Kelling (1982). "Broken Windows: The Police and Neighborhood Safety." *The Atlantic*, March, 1–9.

Woodward, Comer V. (1955). *The Strange Career of Jim Crow.* Oxford University Press, New York.

Wright, Gavin (2013). *Sharing the Prize: The Economics of the Civil Rights Revolution in the American South*, Harvard University Press, Cambridge, Massachusetts.

Wyatt-Brown, Bertram (1982). *Southern Honor: Ethics and Behavior in the Old South.* Oxford University Press, New York.

Young, Peyton H. (1998). *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press, Princeton, NJ.