Rules and Standards: The Games Lawmakers Play

Andrew T. Hayashi*

March 10, 2019

Abstract

I analyze the punishment of harmful conduct when the optimal punishments depend on private information of the actors. I characterize the efficient single, strict liability, punishment if actors cannot credibly signal that information by producing favorable evidence of their type and then analyze the efficient punishment regime when there exists evidentiary signals of the actors' types. The efficiency of the punishment regime depends on whether the the law is articulated by the lawmaker as a rule or as a standard, which I argue corresponds to the difference between screening and signaling games of asymmetric information in this context. These games may have different equilibria, which introduces another dimension to the economic analysis of rules and standards.

Keywords: Law, Legal Institutions, Rules and Standards

JEL Codes: K00, K10, K42

*Professor of Law, University of Virginia School of Law. I'm grateful for helpful comments from Rich Hynes and Yehonatan Givati.

Contents

1 Introduction

Legal consequences often depend not only on individuals' actions, but also on facts that only they know and which are not directly verifiable by a court. In such cases, courts look to the facts and circumstances surrounding the actor's conduct to come to conclusions about the nature of this private information. For example, criminal liability often depends on mens rea requirements that a court can only draw inferences about from the surrounding evidence. Hate crime statutes punish people for acting on the basis of reasons that only they can know, such as whether they caused a bodily injury to a person because of that person's race, color, religion, or national origin.[1] Even tax law, which is sometimes misunderstood to be "strict liability" in the sense that tax consequences follow only from observable transactions, is rife with mental state requirements. The very question of whether an expense is deductible or not depends on whether it was made with the purpose to earn income rather than to generate a personal consumption benefit.[2] Under the common law doctrines of economic substance and business purpose, whether the tax consequences of a transaction will be respected depend on whether the taxpayer has a substantial non-tax purpose for entering into the transaction.[3]

The subject of my analysis is the public regulation of harmful conduct when actors have normatively relevant private information (about, for example, their purpose or intentions) that the lawmaker would like to use as the basis for assigning punishments. I focus specifically on the lawmaker's choice about whether to adopt a legal rule or legal standard for regulating that conduct. I begin by describing two different settings in which the first-best punishments depend on private information. In the first setting, actors derive an unobservable private benefit from an action that is generally, but not always, less than the harm created by that action. In this setting it is optimal for the lawmaker to deter only those actions for which the private benefit is less than the harm, but the first-best cannot be achieved because that benefit is unobservable. For example, consider the common law defense of necessity, known

[1] See, e.g., 18 U.S. Code §249.
[2] Compare I.R.C. §162 with I.R.C. §262.
[3] See I.R.C. §7701(o)(1)(B).

as "choice of evils" under the Model Penal Code. Several factors generally must be present to successfully invoke this justification for committing a crime, but one of those factors is that the harm or evil sought to be avoided by the criminal act is greater than that sought to be prevented by the law that is broken.[4] An actor may have committed a crime in order to avert some greater harm to herself, but she cannot directly reveal the subjective magnitude of that harm to the court. Or consider the economic substance and business purpose doctrines under federal income tax law noted above. The taxpayer's expectation of a significant pre-tax profit (the private benefit) from a transaction will weigh in favor of the transaction having economic substance and being respected, but the taxpayer's profit expectations are private information that cannot be directly revealed to a court.

In the second setting, the harmful action is socially inefficient for all actors because the social welfare weight attached to actors' private benefits are always less than the harm created by the act. This setting covers activities that, as an empirical matter, create a greater harm than the maximal benefit derived by any person in the population, as well as conduct that results in private benefits that are not given equal weight (in the extreme case, they are given zero weight) in the social welfare function. Utility that is disfavored in this way is known as "illicit" utility in the extant literature. Because the conduct in this setting is socially inefficient for all actors, a punishment that deters all actors would be optimal because no inefficient actions would be taken and no punishment would be imposed in equilibrium. I assume that no such punishment exists, so that some actors in this setting cannot be deterred by the maximal punishment available to the lawmaker. This both avoids a corner solution to problem of finding the optimal single punishment in this setting, and also reflects what I think is the empirical reality, given technological and legal constraints on the maximal punishment. As a result, in this setting it is optimal for the lawmaker to deter all actions that can be deterred but not to punish people who are undeterrable. In addition to the obvious welfarist justification for not punishing those who cannot be deterred, there are

[4]Model Penal Code §3.02(1)(a).

non-consequentialist justifications as well (Sunstein and Vermeule, 2005). Nevertheless, the first-best punishment regime in this setting does not have a natural correspondence to current criminal law doctrine. Perhaps the closest examples are the variety of defenses and excuses that arise from some notion of a lack of self-control.[5] For example, in many jurisdictions the insanity defense requires that the actor have acted from an "irresistible impulse." However, without more, the mere fact that an actor is undeterrable does not generally provide an excuse under current criminal law doctrine.

These two settings are only illustrative. The key feature of both, for the purpose of this paper, is that the first-best punishments are dependent on the regulated actors' types, which can generically represented by the preference parameter θ. Depending on the application, the actor's type may be the private benefit that they derive from committing the harmful act, the weight that they attach to the well-being of other people when engaging in potentially injurious behavior, or the desire to harm another person, just as examples. I begin the paper by describing the most efficient single punishment, which is imposed on a strict liability basis against anyone who commits the harmful act and is therefore, of course, independent of the actor's type. An example of such a punishment might be a period of incarceration for causing bodily injury to another person. This single, strict-liability, punishment for the conduct serves as a benchmark against which I compare punishment regimes that attempt to vary punishments on the basis of actors' types, such as whether an act resulting in bodily harm to another was done with the purpose of avoiding a greater harm or out of malice.

Because actors' types cannot be observed or directly verified by factfinders, laws that purport to assign consequences on the basis of types in fact assign consequences on the basis of *inferences* about types drawn by factfinders from the observable and verifiable facts and circumstances attending the actor's conduct.[6] Thus, a law that excuses the infliction of bodily harm if the actor's purpose was to avoid a greater harm is in fact a law that excuses

[5]For a summary of the importance of individual self-control in criminal law, see Hollander-Blumoff (2011).

[6]Hayashi (2017) provides an economic argument for how such "facts and circumstances" inquiries should be structured.

the harmful conduct if the factfinder believes that the actor had the mitigating purpose. There will typically be many verifiable facts that affect the factfinder's inference about the actor's type, and some of them will be within the control of the actor herself. I will refer to choice variables of the actor that are used by the factfinder to draw an inference about the actor's type sometimes as "signals," to be consistent with the economics literature's convention, but I sometimes also refer to these choice variables as "evidence," since that is the function they are serving in the legal context I am exploring. Thus, my project is to compare the welfare properties of a single, strict liability, punishment with punishment regimes that depend on signals of the actors' types.

An important decision to be made when basing punishments on these signals is whether the lawmaker specifies the complete mapping of signals to punishments, ex ante, in the relevant statute or regulations, or whether it instead delegates to courts to the responsibility to interpret actors' conduct and draw inferences about their types after the fact. In the first case, the lawmaker usurps some of the court's role in judging whether the facts and circumstances surrounding a harmful act, including the evidence generated by the actor, compel the belief that the actor is of a particular type. It does this by specifying the relevance and significance of certain facts and circumstances in the law when it is promulgated. In the latter case, judges are left free to form beliefs about the actor's type based on their priors and the inferences that they draw from the facts and circumstances in an open-ended inquiry. I argue that the distinction between these two approaches, which is a central difference between a legal rule and a legal standard, corresponds to the difference between a screening game and a signaling game of asymmetric information played between the legislature, courts and regulated actors. Thus, the choice between a rule and a standard is a choice which game to play with regulated actors. In section 7, I describe the factors that lawmakers should consider in choosing which game to play, or whether to play any game at all. An important factor in this choice is the efficiency of the separating equilibrium under a legal standard. Focusing on this factor sheds light on the efficiency properties of graduated punishments and

standards of proof, which operate through how they restrict judges' discretion.[7]

Economic analysis of the rules/standards choice in legal design has focused on cases in which all relevant information is public and verifiable. Consider traffic regulation. A rule might assign a punishment for driving more than 70 miles per hour. A standard might assign a punishment for driving negligently, where negligence is defined by the Hand rule. In either case, whether the factual predicate of the punishment has been satisfied is observable or ascertainable by any appropriately situated person. A driver's speed can just as easily be observed by a police officer with a radar gun as by the driver herself. Whether a driver has behaved negligently is an objective determination that depends on empirical questions about the relationship between circumstances and actions, such as the effect of driving at such and such a speed under such and such weather conditions. The driver has no special access to knowledge about these empirical questions. The driver may of course have better access to details about her manner of driving that are hard for police or judges to discover and that affect the probability of an accident, such as whether she is distracted when she drives, but such facts are, at least in principle, verifiable by objective measures and observable by anyone who is appropriately situated, such as a passenger.

A lot of regulation covers cases where all normatively relevant information is observable and verifiable. On the other hand, there is a also large class of activities where the normative objectives of the law depend on private information of regulated actors. To illustrate one such choice, consider Internal Revenue Code §183, which helps police the boundary between expenditures that should be deductible for federal income tax purposes, because the expenditures are incurred for the production of income, and expenditures that yield personal consumption benefits. The general rule of that section provides that "[i]n the case of an activity engaged in by an individual..if such activity is not engaged in for profit, no deduction attributable to such activity shall be allowed under this chapter." Whether an activity

[7]Raskolnikov (2015) discusses factors that typically provide a basis for graduating sanctions. None of these six factors is obviously implicated in my setup, although one might interpret the actors' private benefits from taking the action as being related to their culpability.

is engaged in for profit is a question about the motives of the actor and about which the factfinder can only try and draw inferences based on the surrounding circumstances. Section 183 demands an inquiry into the taxpayer's actual, subjective, expectations about profit. The accompanying Treasury Regulations are explicit that "[a]lthough a reasonable expectation of profit is not required, the facts and circumstances must indicate that the taxpayer entered into the activity, or continued the activity, with the objective of making a profit." [8] This is a standard under most scholars' definitions, including under the definition I provide in Section 4.

However, because the taxpayer's purposes cannot themselves be observed, when a court applies §183 they draw conclusions about the presence of a profit motive from the facts and circumstances surrounding the taxpayer's conduct. Functionally, Section 183 is therefore a mapping from those facts and circumstances to the availability or denial of a deduction through an inference about what those facts tell us about the taxpayer's intent. An alternative way for the lawmaker to implement the purposes of Section 183 would be specify *which* facts will result in a conclusion about profit motive and therefore the availability of a deduction. The Treasury Regulations under §183 in fact provide a list of factors to be considered in coming to a conclusion about motive, but these factors are non-determinative and non-exhaustive. These factors include the time and effort expended by the taxpayer in carrying on the activity, the taxpayer's history of income or losses with respect to the activity, and whether the taxpayer carries on the activity in a businesslike manner and maintains complete and accurate books and records. These factors make §183 more rule-like than the statute itself appears, but they stop short of making §183 a rule because the factors are not determinative and also vague. This article is about the choice between rules and standards for laws like §183 that have private information (such as the existence of a particular motive) as a factual predicate.

In Section 2, I provide a brief overview of some of the relevant literature. Section 3

[8] Treas. Regs. /S 1.183-2(a).

describes two settings in which optimal punishments depend on private information of the regulated parties: one setting in which actors have heterogeneous private gains from a harmful act such that the act is only socially efficient for some actors, and a second setting in which all acts are inefficient, but where some actors are undeterrable and so it is not efficient to impose the maximal punishment. I characterize the efficient strict liability punishment in these settings when types are unobservable. Section 4 defines rules and standards and argues that they map to screening and signaling games when legal consequences depend on private information. In Sections 5 and 6, I assume that there exists observable and verifiable evidence (signals) of individual types and analyze the separating equilibria of the screening game and signaling games that are created by the adoption of legal rules and standards, respectively. Section 7 summarizes the choice between a strict liability punishment, a fault-based rule, and a fault-based standard. Section 8 concludes.

2 Literature

This paper explores optimal deterrence regimes for actors with heterogeneous preferences for the regulated act. I consider only punishments (and not damages) for the harm-creating activity.[9] I also distinguish between preferences that are incorporated on an equal basis into a utilitarian social welfare function with all other preferences, and preferences for illicit utility, which are not included in the social welfare function. I say that such actors have "bad intentions."

An actor intends a consequence of an action when that consequence is motivationally significant for her. That is, the consequence provides her with positive utility. The word "intent" is used in different scholarly literatures to mean different things.[10] My analysis is

[9] In Hayashi (2019), I consider damages and other mechanisms for inducing efficient conduct.

[10] Some readers may prefer the label "motive" or "purpose" for motivationally significant consequences, and reserve a more inclusive definition for "intent" that includes, for example, awareness of the natural and probable consequences of an act, regardless of whether they are desired or not. Disputes about the proper scope of the term "intent" are generally motivated by whether the definition captures all of the mental states associated with conduct that is blameworthy. I am not concerned with providing a definition that captures

limited to just that definition of intent that I have defined here, a definition that is consistent with usage in the extant literature within law and economics.[11] Intent is generally not the focus of economic analysis of law, although there are example of analyses in specific areas.[12] Focusing on the case of antitrust law, Cass and Hylton (2000) argue that legal rules that rely on intent can be useful in reducing errors costs from under and over inclusiveness in the application of rules. On the role of intent in criminal law, see Posner (1985), Shavell (2009), and Curry (2017). In contract law there is Bar-Gill and Ben-Shahar (2008), and in tort law there is Landes and Posner (1981), Ellis Jr (1983) and Hylton (2009). More generally, see also Hamdani (2007), Parker (1993) and Becker (1968).

I say that an actor has bad intentions when the utility they derive from a harmful act is not included in the social welfare function. I do not argue that the law *should* disfavor certain utility, but suggest that the law seems to already do this. I also note that differential social welfare weights need not be derived from philosophical or extra-welfarist constraints. Certain forms of utility may be devalued (or more precisely, be offset) endogenously in even a utilitarian social welfare function if others may derive disutility from seeing certain preferences vindicated. For example, there is evidence that people do not think that envy utility should be rewarded (Weinzierl, 2018). I refer to utility that is disfavored in this way as "illicit", following the literature.[13]

Regardless of the social weight given to individual preferences, individual heterogeneity in the settings I analyze implies that the optimal punishments vary with individual types. If there exists evidence that can signal individual types, there may be a schedule of evidence-dependent punishments that achieve a more efficient outcome than that under a single punishment.[14] I argue that when punishments depend on evidence of private informa-

all of the blameworthy conduct that we might want to punish.

[11] Cooter (1982), Polinsky and Shavell (1997).

[12] Cass and Hylton (2000, p.660)("Economic analyses of law have tended to ignore intent doctrines, focusing on rules framed in terms of the actor's conduct.")

[13] Baker (1977), Schwartz (1979), Cooter (1982), Shavell (1985), Shavell (2009), Ellis Jr (1982), Polinsky and Shavell (1997)

[14] The problem of assigning different taxes to individuals of different types is the mechanism problem that is central to the optimal taxation approach, and examples applying the same screening logic to tax law

tion of the regulated parties, the choice to specify the correspondence between evidence and punishment either ex ante or ex post is the choice between a rule and a standard. The literature on rules and standards is voluminous. Perhaps the most influential recent paper in law and economics on rules and standards from within a public regulation framework is Kaplow (1992), wherein the key difference between a rule and a standard is whether exogenous uncertainty about harm created by a regulated activity is resolved before or after actors choose whether to engage in the activity. Responding to Kaplow, Posner (1997) raises the question of whether legal standards cause people to coalesce around certain behaviors that signal that they have desirable character traits that entitle them to better treatment by the government or courts.[15] A more recent contribution in the context of regulation is Givati (2015), which explicitly allows for type-dependent punishments rather than manipulable signals of actor's types.[16]

There is a robust literature on the choice between specific and vague terms in the optimal contracting literature. A common assumption in in contract design scholarship is that parties' contractual obligations are conditioned only on information that can be verified by a court. see, e.g., Grossman and Hart (1986), Hart and Moore (1988), Hart (1995). Thus, contract law theory generally assumes that the fulfillment of contractual obligations is either costlessly verifiable, or non-nonverifiable and therefore non-contractable. However, Choi and Triantis (2008) and Choi and Triantis (2009) show that parties to a contract may prefer to incorporate vague standards that can only be verified at a cost because the threat of costly litigation provides an additional incentive for the promisor through the threat of litigation.[17]

There close similarities between the economics of specific and vague terms in contract

include Raskolnikov (2009) and Osofsky (2013).

[15]Posner's speculation is relevant to this project, although it is unclear why he considers only the possibility of pooling outcomes.

[16]See also Korobkin (2000) and Johnston (1995) for other examples of the economics of rules and standards.

[17]citetchoi2008completing introduces costly verification that results in both type I and type II errors, in contrast to the prior literature on costly verification, which assumes that verification is error-free. See,e.g., Gale and Hellwig (1985), Khalil (1997), Townsend (1979). There is also a large literature on incomplete contracting, which includes some empirical tests of models of why contracts might be deliberately incomplete. See, e.g., Sanga (2018)

theory and rules and standards in public law. They traditional justification for vague standards in the contracts literature is that they economize on upfront negotiation costs. An important paper in this literature is Scott and Triantis (2005), which argues that the choice between precise and vague terms in a contract is a choice between rules and standards, with the choice being motivated in large part by the tradeoff from reducing upfront transactions costs and back-end litigation costs. This tracks closely the key tradeoff identified by Kaplow (1992) between the costs of promulgating and adjudicating rules and standards. In contrast to both Kaplow (1992) and this contracts literature, the public regulation framework I adopt in this paper is concerned with private information that is nonverifiable at any cost, so that legal consequences can only attached to manipulable signals of that private information.

Finally, the analysis in this framework highlights the costs to parties under a fault-based legal regime of proving that they are of the type that is entitled to more favorable treatment, and to the strategics considerations that are central when the law tried to distinguish between parties on the baseis of evidence that they have some control over. Therefore, the results are also relevant to the economics of evidence law. For early work in this area see, e.g., Posner (2001) and Stein (2014), as well as Lempert (2001) and Posner (1998).

3 Setup

I consider two settings in which an uninformed lawmaker faces a heterogeneous population of actors who can engage in some harm-producing conduct. In the first setting, the private benefits to only some actors exceed the the social harm of the conduct. In the second setting, all actions are socially inefficient but the punishment is bounded above so some actors cannot be deterred. In both settings the first-best punishment varies by type. The analysis in the remainder of the paper is applicable to either setting.

3.1 Conduct that is Socially Efficient for Only Some Actors

An individual i can take an action that generates private benefit θ_i but imposes harm h on a third party. The individual's type θ is distributed on the interval $[0, \bar{\theta}]$ according to distribution function $F(\theta)$, which has density $f(\theta)$ that is strictly positive everywhere. The population is normalized to equal unity. Assume that $h < \bar{\theta}$ but that $\int_0^{\bar{\theta}} \theta f(\theta) d\theta < h$ so that the conduct is socially efficient for some individuals but there is net harm from the conduct in expectation.

All individuals have the same reservation value of 0 and an individual who commits the act receives punishment P with certainty. No individual who does not commit the act is punished. Thus, the individual commits the act if and only if $\theta_i > P$. I assume that the regulator must follow through with the punishment if the act is committed.[18] I follow Becker (1968) in assuming that the social cost of punishment is linear in the punishment to the actor, so that the cost is given by αP where $\alpha \geq 0$. Modeling the social cost of punishment as αP allows that the social cost may be less than the harm to the individual if, for example, there is retribution utility, or bigger than the harm if people dislike punishing others (Polinsky and Shavell, 2000b). I ignore any such preferences for retribution or regret, so that $\alpha \geq 1$ if P is hard treatment, such as imprisonment, and $\alpha = 0$ if P is a fine.[19] The lawmaker's problem is to choose punishments, which are imposed on anyone who commits the act, to minimize social costs from the activity and its regulation.

If θ is observable then it is optimal for the lawmaker to set the type-dependent punishment $P^*(\theta) \geq \theta$ if $\theta < h$ and $P^*(\theta) = 0$ otherwise. This ensures that all inefficient acts are deterred and no punishment is imposed in equilibrium. However, if θ is unobservable, then the lawmaker sets P to minimize social costs by solving the following problem:

[18] The punishment may be socially costly and so imposing the punishment in the subgame in which the actor commits the act may not be subgame perfect in the one-shot game. Nevertheless, it may be necessary to impose the punishment to retain credibility for future iterations of this game.

[19] Of course, even for a fine α is probably positive because of transactions costs. We set it to zero for simplicity. I note that α could even be negative from incapacitation effects, but do not consider that possibility here.

$$\min_P \int_P^{\bar{\theta}} (\alpha P + h - \theta) f(\theta) d\theta$$

The lawmaker can always choose $P^* \geq \bar{\theta}$, in which case all actors are deterred and social costs are 0. However, this is not first-best, because those individuals for whom $\theta > h$ do not take the action. Note that $P^* = 0$ cannot be a solution, because in that case all individuals commit the act and the aggregate cost of all individuals committing the act is positive; the lawmaker would prefer to deter all acts resulting in social costs of 0. If there is an interior solution P^* to this problem, then the solution satisfies:

$$\alpha(1 - F(P^*)) = (h - P^*(1 - \alpha)) f(P^*) \tag{1}$$

For an interior solution, social costs are minimized where the marginal cost of increasing the punishment, which comes from imposing an additional unit of punishment on those who continue to commit the act, is equal to the benefits from deterring the marginal individual, which include the punishment on the individual that no longer needs to be imposed and the net harm averted by deterring them. Note that in any interior solution the aggregate social costs must be negative (i.e., there is a net social benefit), because the lawmaker can always deter all individuals and achieve social costs of 0. In the special case that $\alpha = 0$, such as for a fine, the tradeoff disappears and the optimal punishment is simply equal to the harm. In general, P^* can be more or less than h so that there are two sets of potential interior solutions to the problem: $P^* > h$, in which case some efficient acts are deterred, and $P^* < h$, in which case some inefficient acts are committed.

It is instructive to compare this setup with that of Polinsky and Shavell (2000a), who analyze both strict liability and fault-based sanction regimes. In a strict liability regime a sanction is imposed whenever harm occurs, and under a fault-based regime sanctions are imposed only if the act is "socially undesirable," by which they mean that the private gain

from the act is insufficient to outweigh the harm caused.[20] Their result that the optimal fine ($\alpha = 0$, in my setup) under a strict liability regime is equal to the harm is reproduced above. On the other hand, they note that there is "not a simple formula for defining the optimal imprisonment term" if the cost of punishment is positive, and that the optimal imprisonment term could result in underdeterrence or overdeterrence. Equation 1 provides the first order condition for an interior solution to the optimal punishment problem when $\alpha > 0$, but a closed form solution will not always be available. The key difference between their analysis and mine, at this stage, is that they assume that private gains are observable. When this is true, it is efficient to set the standard for fault equal to equal to the harm and induce compliance with a high fine. When the private gain is unobservable, a fault-based regime cannot be directly implemented. In my setup, the single punishment is imposed on a strict liability basis for anyone who commits the harmful act, and the punishment regimes analyzed in Sections 5 and 6 are quasi fault-based, in the sense that punishments are dependent on evidence of fault rather than fault itself (which is unobservable and unverifiable).

3.2 Bad Intentions

In this Section, I describe an alternative setup that serves as a baseline for the analysis of equilibria in Sections 5 and 6. Hard treatment is often reserved for crimes, and crimes are generally prohibited without regard for the private gains derived by the criminal. Suppose that, as before, actors derive unobservable utility θ from taking the action but that social costs from the action are simply h because the private benefits are not counted by lawmakers. This utility is labeled "illicit" in the literature.[21] The utility may be disvalued either because of some extra-welfarist constraint, or because other members of society get disutility from this kind of utility. Since lawmakers do not recognize any benefit from the activity, all actions are inefficient and should be deterred.[22] If the maximal penalty, such as the death penalty,

[20] Polinsky and Shavell (2000a, p.47).

[21] Cooter (1982), Ellis Jr (1982)Polinsky and Shavell (1997).

[22] Of course, differential weighting of utilities in the social welfare function is controversial because it is non-utilitarian. Parker (1993). I do not defend the philosophical decision to do so in this paper, but do

would deter all actors then this is the optimal punishment for the lawmaker.

However, suppose that that some actors cannot be deterred given the maximal punishment that can be imposed. The maximal punishment could be determined by technology (e.g., we may not be able to implement a punishment more severe than death), or it could be fixed by law, reflecting some extra-welfarist constraint. Denote this maximal punishment by $\bar{P}$. An actor will commit the act if $\theta > P$. Let $\hat{\theta}(P)$ be the actor that is indifferent between committing the act and not when threatened with punishment P, so that $\hat{\theta}(P) = P$.

The first-best punishment regime would set $P^*(\theta) = 0$ if $\theta > \hat{\theta}(\bar{P})$, because these types cannot be deterred in any event, and $P(\theta) \geq \theta$ otherwise which deters all $\theta \leq \hat{\theta}(\bar{P})$. However, if θ is unobservable then the lawmaker sets solves the following cost minimization problem:

$$\min_{P \leq \bar{P}} \int_{P}^{\bar{\theta}} (\alpha P + h) f(\theta) d\theta \tag{2}$$

By setting $P^* = 0$, all individuals commit the crime and the aggregate social cost is h. If $P^* = \bar{P}$, then all individuals for whom $\theta > \bar{P}$ will act and aggregate social cost will be $(\alpha \bar{P} + h)[1 - F(\bar{P})]$. Unlike the previous setting, it is now impossible to deter everyone and the social costs of the activity must be greater than zero. If there is an interior solution, social welfare is $(\alpha P^* + h)[1 - F(P^*)]$. As in the previous example, the interior optimum balances imposing greater punishment on those who commit the act with the benefits of deterring the marginal actor, by deterring them and therefore not having to punish them. The lone difference between the first order conditions is the utility θ of the marginal actor, which is missing when this utility is illicit.[23]

note that the decision to weight all utilities equally can also lead to some rather controversial conclusions. Finkelstein (2000, p.71).

[23]There is no necessary relationship between the optimal punishment in these two examples because it depends on the density $f(\theta)$, which may be nonlinear and nonmonotonic. Note again, however, that there is certain to be positive social costs in this setup, because none of the actions yield recognized benefits and all solutions yield positive social costs.

4 Rules and Standards, Screening and Signaling

The important feature of both settings is that the first-best punishment regime is type-dependent and (weakly) decreasing in the actors' types, but the first-best cannot be achieved because types are private information. In both settings, an actor i has utility from acting and being punished of $\theta_i - P$. Suppose now that the actor can take actions that create verifiable information that can be used in court as favorable evidence of the actor's type. For example, consider the factors listed in the regulations under I.R.C. §183 that are treated as evidence of a profit motive. The lawmaker may be able to achieve a more efficient outcome than that induced with a single punishment by making punishments dependent on the amount of that evidence. Let the quantity of evidence produced by the actor be given by s and assume that the cost of generating that evidence is given by the continuous and differentiable function $v(s;\theta)$ which satisfies the usual properties of being convex in the argument and that the indifference curves of different types cross only once in (s, P) space. Specifically, the partial derivatives $v_s > 0, v_{ss} > 0$ and $v_\theta < 0, v_{s\theta} < 0, \forall s > 0$. That is, any given evidence/punishment combination is less costly for higher types and the marginal disutility of generating evidence is also lower for higher types. To fix ideas, let $v(s;\theta) = s^2/\theta$ so that each actor i's utility from acting, producing evidence s, and receiving the evidence-dependent punishment $P(s)$ is given by $\theta_i - P(s) - s^2/\theta_i$.

I restrict attention to punishment regimes with two properties. First, I assume that the punishment regime must include a punishment for anyone who takes the action and does not produce any favorable evidence at all. The lawmaker must provide some outcome for actors who commit the harm but do not attempt to prove they deserve a lighter punishment.[24] Second, I consider punishment regimes involving only two evidence/punishment pairs, so that the lawmaker's problem is to specify a punishment for the act and a second, more lenient, punishment that depends on the favorable evidence of the individual's type. One might think of the second punishment as resulting from compelling evidence of an excuse,

[24] I intend to prove that the optimal regime has only two punishments in a subsequent draft.

justification, or mitigating circumstances.

I consider the implementation of punishment regimes under two different chronologies. In Section 5, the chronology is as follows: (1) the lawmaker specifies the punishment for committing the act, the lighter punishment available to any actor who is able to prove her justification/excuse, and the amount of evidence that is necessary to receive that lighter punishment; (2) actors decide whether to act and how much favorable evidence to generate; and (3) courts assign punishments on the basis of the lawmaker's instructions in the first stage. In Section 6, the sequence of play is: (1) the lawmaker specifies the punishment for committing the act and a lighter punishment available to any actor of a sufficiently high type; (2) actors decide whether to act and how much evidence to produce; and (3) courts impose the punishments created by the lawmaker on the basis of the their beliefs about the actors' types. Judges' beliefs about the actors' types are correct in equilibrium.

This difference in chronology is a key difference between screening and signaling games of asymmetric information. In screening games the uninformed parties (the lawmaker and courts, in this case) provides a menu of offers and the informed parties (the regulated actors) simply respond to these offers (Riley, 2001, p.443). Alternatively, if the informed parties move first and their actions are interpreted by the uninformed parties, then it is a signaling game. This seemingly innocuous difference in timing can result in very different equilbria in the two kinds of games. Signaling games typically have many equilibria, depending on the beliefs of the uninformed party about what happens off of the equilibrium path (Riley, 2001, p.444). In screening games the issue is generally the opposite; an equilbrium in pure strategies may not exist and if it does it is unique. The setup in this paper differs from the canonical models of competitive signaling and screening because the lawmaker is a monopolist. For that reason, the setup is similar to those in optimal tax theory (Mirrlees, 1971) and mechanism design approaches generally. The model also differs in that there are two uninformed parties, with one setting the evidence/punishment offers (the lawmaker) and the other imposing the punishments (the courts).

How does this difference between screening and signaling models correspond to the difference between rules and standards? My answer has to do with the delegation of authority for interpreting evidence by the lawmaker to the courts, and the resulting uncertainty actors face in how courts will judge their conduct and the range of possible outcomes. The definitions I provide below of rules and standards capture only a part, but I think an important part, of the difference between the two ways of legislating.

Laws can assign punishments on the basis of three kinds of factual predicates: facts that are publicly observable, facts that are observable to only some people (privately observable), and facts that are not directly observable by anyone (unobservable). An example of a publicly observable fact is how fast someone is driving their car at a particular time. This fact can be directly measured or observed. A well-functioning radar gun or speedometer will report the speed of the car. This fact is publicly observable in the sense that no person has privileged access to observation of that fact by virtue of who they are (as opposed to where they happen to be situated at a place and time). Although, as a practical matter, the driver of the car may be better positioned to know her speed than another driver she passes on the road, the fact of her speed is equally observable by anyone who is similarly situated to the driver, such as a passenger.

By contrast, there are facts to which certain people have privileged access. Privately observable facts cannot be directly observed by anyone else, and the person who is in a position to observe the fact cannot directly reveal it to others. An example of this kind of fact is someone's intentions or purposes in committing a particular act. For example, whether a taxpayer spends her weekends taking photographs with the aim of selling them and earning a profit is something that she knows, but which cannot be directly observed by anyone else. Moreover, she cannot directly reveal to anyone what her intentions are; all she can do is make assertions or take other actions that she hopes will persuade them. For example, a savvy taxpayer would conduct her hobby in a businesslike fashion, including keeping books and records, if she wanted to have a chance at deducting the expenses of her

hobby after application of §183.

Finally, there are unobservable facts. Applications of the Hand formula have such facts as inputs. Whether the failure to take a precaution with cost B that can avert loss L is negligent depends on whether $B < PL$. This in turn depends on the effect of the precaution on the probability of loss P, and this probability cannot directly be observed by anyone. For example, consider applying this formula to a precaution that would reduce the emission of a carcinogen into a water supply. The proportion of the people who develop cancer because a particular carcinogen is present in their drinking water is an unobservable fact. Confronted with litigation about whether the emitter behaved negligently, a rational judge's beliefs about this proportion will be based in part on evidence they are provided about the effects of the carcinogen and the characteristics of the population exposed to the tainted drinking water. These beliefs will also depend on the judge's idiosyncratic prior beliefs about how the carcinogen operates and how she interprets the new evidence with which she is presented. Because of differences in prior beliefs and and differences in interpretation of new evidence, different judges may have come to different beliefs about P after reviewing the same evidence.

This is the difference between rules and standards that I emphasize in this paper: the factual predicates for rules are publicly observable whereas the factual predicates for standards are unobservable by judges, either because the facts are unobservable to everyone (as in the case of claims about causation) or because they are observable only by regulated actors (as in the case of intentions or purposes). The important consequence of this distinction is that judges will have a wider range of beliefs about whether a punishment is applicable in the case of a standard than in the case of a rule. As a result, regulated actors can plan their affairs with greater certainty of the consequences of their actions under a rule than a standard. This emphasis on uncertainty is also at the heart of the rule/standard distinction made by Kaplow (1992), which focuses on standards with unobservable information. My focus is on standards with privately observable information, where actors know whether the factual predicate of the punishment regime has been satisfied but the judges does not.

Thus, a rule specifies a mapping from publicly observable facts to legal consequences, whereas a standard assigns legal consequences to inferences from publicly observable facts. This is true both when the inference is about an unobservable fact, such as the causal effect of a precaution on the probability of harm, and when the inference is about something knowable with certainty only by regulated actors, such as their intentions. When regulated actors are able to produce favorable evidence about their intentions (or other private information), a rule specifies a mapping from this evidence to a punishment. Thus, for any given act/evidence combination, a rule specifies the punishment and the court's only function is to observe the evidence and assign the punishment that the lawmaker has specified for that evidence. This creates a screening game between the actor and the judge. By contrast, a standard in this setting specifies the punishments that are assigned to actors based on their types. Since types are not observable, judges interpret evidence for themselves and assign punishments on the basis of their beliefs about actors' types. This creates a signaling game between actors and judges.

This characterization of the difference between rules and standards is incomplete and also a little naïve. It is incomplete because it ignores other differences between rules and standards. For example, consider a standard that assigns liability for harms "created through negligence." If negligence is well defined, such as through the Hand formula, then the only uncertainty actors face is whether one has, in fact, acted negligently. However, if the Hand formula is not embedded in the definition of negligence then the legislature has also delegated to the courts the question of what it means to act negligently. This is an additional source of uncertainty that is important but outside the scope of this paper.

The characterization is also naïve in its account of facts. First, at bottom, it may be that no facts are directly observable; the speed of a car at a point in time is something that we can only measure imperfectly. Thus, different judges may have different beliefs about the speed that an individual was driving after reviewing the same radar and speedometer evidence (they may have different beliefs about the accuracy and bias of these measurement

tools or different priors), and so the difference between observable and unobservable facts is one of degree. Second, I have suggested that people are in some sense able to directly observe things like their own intentions. This is not so clear. Sometimes our own purposes are opaque to us. Attempting to justify these theoretical distinctions with any real rigor is a philosophical project also beyond the scope of this paper. For my purposes, it is sufficient to draw the distinction between the three categories of legal rules (rules, standards based on private information, and standards based on public information) along empirical lines. We can say that a law is a rule if there is little variance across judges in beliefs about whether the factual predicate is true; a law is a standard based on private information if there is much less variance for the regulated actor than across judges in whether the standard has been met; and a law is a standard based on public information if regulated actors have the same uncertainty about whether the standard has been met than judges do.

To simplify the presentation, in the next two sections I describe equilibria of the screening and signaling games associated with rules and standards using a graphical presentation. Both sections consider the case described in subsection 3.1, in which the action is socially beneficial for only some actors.

5 The Screening Game

Let P^* be the optimal punishment when there is no signaling evidence available and assume that $P^* < h$, so that there are some actors who commit an inefficient act and are punished for it. For the sake of exposition, I focus on a discrete type example with two types that satisfy the following inequalities: $\theta_H > h > \theta_L > P^*$.[25] As described in subsection 3.1, both actors $i \in L, H$ have utility from taking the action and producing evidence s and receiving punishment $P(s)$ of: $\theta_i - P(s) - s^2/\theta_i$. Social welfare for the lawmaker is given by:

[25]Although it is possible for P^* to be greater than h, I do not consider this case yet. Polinsky and Shavell (2000a) note that the "possible optimality of overdeterrence strikes us as more theoretical than real." Implicitly, to support $P^* > 0$, there must also be some $\theta < \theta_L$, which I ignore. A similar analysis follows, however, if $h > \theta_H > \theta_L$.

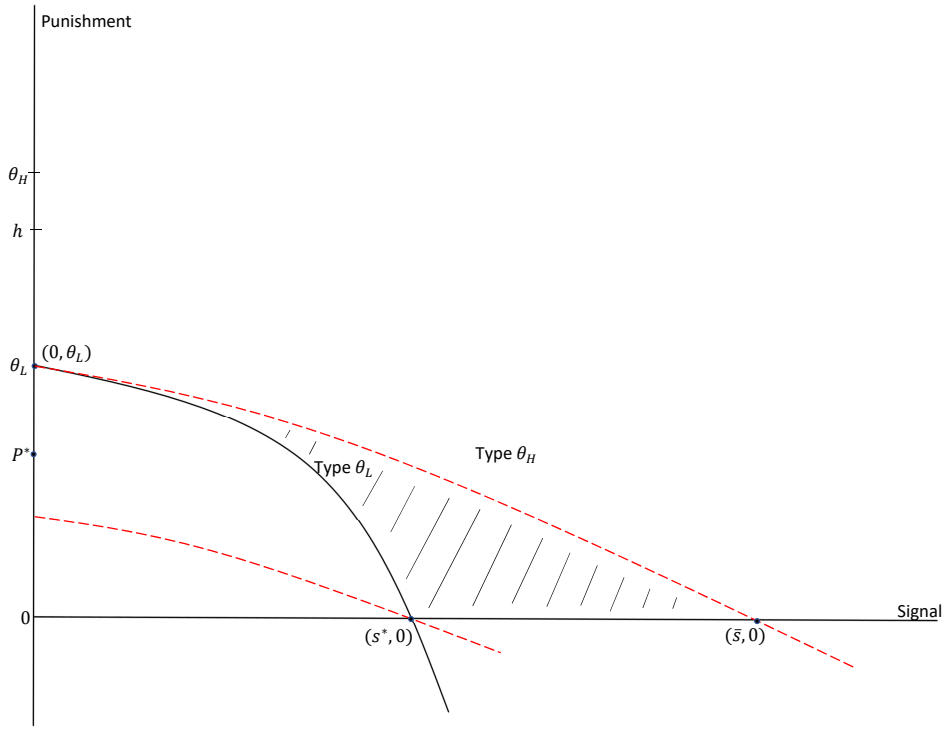

Figure 1: Separating Equilibria Under a Rule

$\sum_i (h - \theta_i - P(s) - s^2/\theta_i - c_r)$ over all actors i who take the action, where c_r is the cost of applying the legal test by the court.

In the strict liability equilibrium with p^*, both types commit the act. Now suppose that there are forms of evidence that serve as signals of the actors' types, i.e., having the properties described in Section 3. The lawmaker may be able to improve on the strict liability equilibrium by implementing a punishment regime with two punishments of the form (s, P), pairing evidentiary thresholds with punishments. Indifference curves for θ_H and θ_L in (s, P) space are shown in Figure 1. The red dashed lines represent indifference curves for θ_H and the solid black line is the indifference curve for θ_L. Utility is increasing to the southwest.

By offering the punishment $(0, \theta_L)$ and a second evidence/punishment pair (s, P) anywhere in the crosshatched region, the lawmaker can induce separation of the two types. θ_L will be deterred from taking the action, because she strictly prefers $(0, \theta_L)$ to (s, P) in the

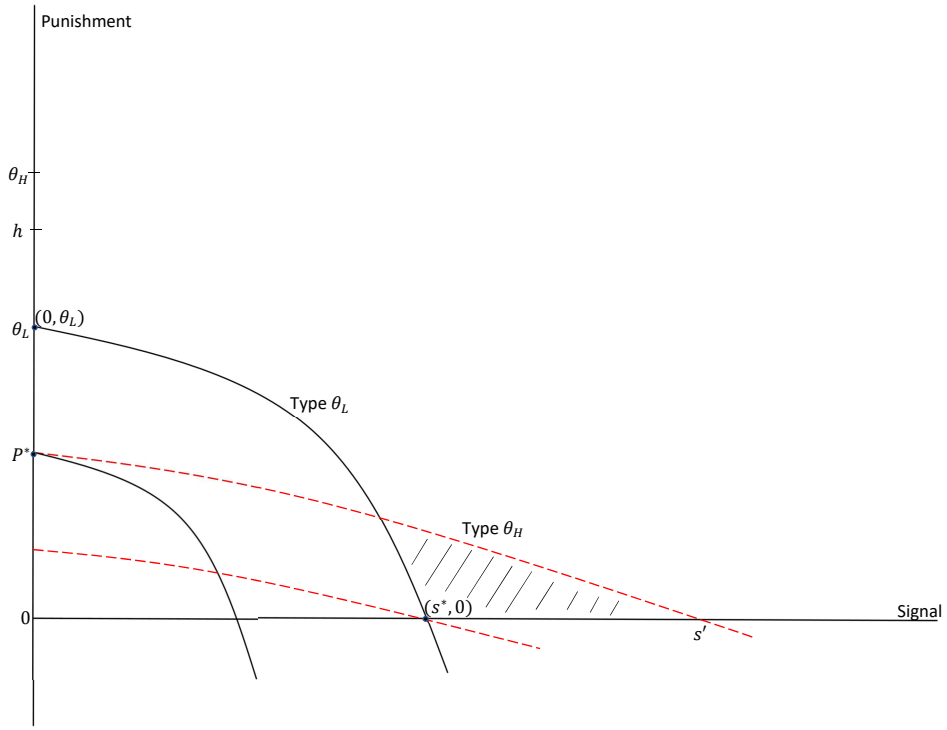

Figure 2: Separate Punishments That Make θ_H Better Off

crosshatched region, and facing the punishment θ_L makes her indifferent between taking the action and not. Deterring θ_L yields strictly positive benefits to the lawmaker, relative to the strict liability equilibrium, because of the net harm averted and the avoided cost of punishing θ_L for taking the act.

θ_H will commit the act, receiving a lighter punishment than θ_L in exchange for producing sufficient evidence of his type. Creating separation between the two types allows the regulator to increase the penalty and deter some inefficient acts, but use the signal/evidence to lower the burden on the high types. Note that some of evidence/punishment points in the crosshatched region lie on a higher (i.e., worse) indifference curve for θ_H than the curve that passes through P^*. Thus, not all separating equilibria will make θ_H better off. Figure 2 shows the subset of separating equilibria that would make θ_H better off, relative to the strict liability equilibrium.

Separating equilibria that leave θ_H better off may not exist, however. If preference maps do not vary quickly enough with the signal, or if the increase in punishment necessary to deter θ_L is too high, then the indifference curve for θ_L passing through $(0, \theta_L)$ and the indifference curve for θ_H passing through P^* may intersect below the x-axis. In that case, because punishments are bounded below by 0, the equilibrium punishment for θ_H may involve producing so much evidence that it leaves her worse off than under the single, strict liability punishment.

The welfare consequences of the new separating equilibrium relative to the strict liability equilibrium do not depend only on the welfare of the two types, however. This is obvious in that θ_L is made worse off in the separating equilibrium because she is deterred from doing something that would be in her (but not society's) interest. But it is also true that even if θ_H is not made worse off in the separating equilibrium, it is possible that society may be made worse off because the social costs of punishment may differ from the private costs.

To illustrate this, Figure 3 adds two different indifference curves of the *lawmaker* with respect to punishments imposed on θ_H. These curves appear as dotted blue lines in the figure. Their shape is different than the shape of the red dashed indifference curve for θ_H, because the marginal rate of substitution for the lawmaker between the costly evidence and the punishment differs from the rate of substitution for the actor. This is because the marginal cost of punishment for the lawmaker is α, which will not in general be equal to 1. The flatter blue dotted curve represents the lawmaker's preferences when $\alpha > 1$, and the steep blue dotted curve represents the lawmaker's preferences when $\alpha < 1$.

Unless $\alpha = 1$, any change in θ_H's welfare in moving from the strict liability equilibrium to the separating equilibrium will differ from the change in social welfare. If $\alpha > 1$ then the change in social welfare will be bigger than the private change, and if $\alpha < 1$ then the change in social welfare will be smaller than the change in θ_H's utility. In fact, it is possible for a separating equilibrium to leave θ_H better off than in the strict liability equilibrium but to *increase* the social costs of punishing θ_H. For example, suppose that in the separating

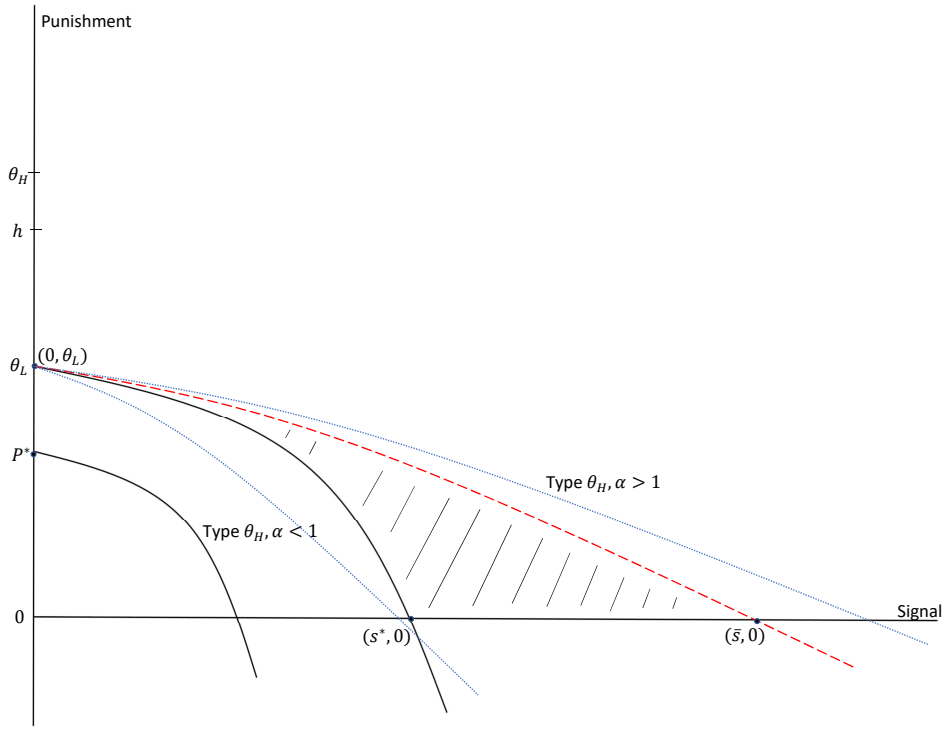

Figure 3: Separating Equilibria Under Different Costs of Punishment

equilibrium that θ_H chooses s' just a bit larger than s^* and that the lawmaker's preferences with respect to punishing θ_H are given by the steeper blue dotted line. In that case, θ_H is made better off by the move to a separating equilibrium, but the social costs associated with θ_H have increased. The substitution of the costly evidentiary burden s' for the punishment P^* is better for θ_H, but because the social cost of the punishment is so much less than the private cost, social costs have increased. When α is sufficiently small, θ_H may choose the separating punishment even though the social planner would prefer that she choose $(0, \theta_L)$.

The improvement in social welfare from moving from P^* to a separating equilibrium is equal to the benefits of deterring θ_L plus the benefits from reducing the punishment (but adding the evidentiary burden) to θ_H. Stated in terms of the marginal cost of punishment, the benefits from moving to a separating equilibrium will be positive if:

$$\alpha > \frac{s^2/\theta_H + \theta_L - h}{2P^*} \tag{3}$$

The benefits are increasing in the marginal cost of punishment and the benefits of deterring θ_L, and decreasing with the (punishment and evidence production) costs borne by θ_H in the new equilibrium. If the inequality in equation 3 is satisfied, then the optimal punishment regime is given by $(0, \theta_L)$ and $(s^*, 0)$, which deters θ_L and provides θ_H with a complete justification for his action if he provides evidence s^* of his type. The lawmaker can induce the constrained first-best equilibrium by formulating the law as a rule: $P = \theta_L$ if $s < s^*$ and $P = 0$ if $s \geq s^*$.[26] In equilibrium, θ_H and θ_L will separate into the two punishments, with θ_H not choosing any $s > s^*$, thereby achieving the constrained first-best outcome. This is known as the "Riley outcome" (Fudenberg and Tirole, 1991, p.451).

Although the lawmaker can select the constrained first-best equilibrium if it promulgates a rule, this comes at a cost. For the reasons give in Kaplow (1992), rules are costlier to promulgate than standards. On the other hand, standards based on private information can result in a range of outcomes, some of which may be worse than promulgating a law that ignores the private information altogether. I explain this claim in greater depth in the next section, but the intuition is that it may be in the interest of θ_H to generate a lot of costly evidence of his type to distinguish himself from θ_H when signaling evidence is available, but that θ_H would prefer to be under a single-punishment, strict liability regime.

6 The Signaling Game

Alternatively, the lawmaker could formulate the law as follows: $P = \theta_L$ if the actor is of type $\theta \leq \theta_L$ and $P = 0$ if the actor is of type $\theta \geq \theta_H$. This formulation makes the excuse dependent on the actor's type, not on her providing a specified amount of favorable evidence. In this case, the judge's beliefs about the actor's type, based on the evidence that the actor

[26] Alternatively, the regulator could even specify $P = 0$ if $s = s^*$ and $P = \theta_L$ otherwise.

provides, will determine the equilibrium. I assume that courts are concerned with minimizing error in the assignment of punishments to actor types, so that any pooling outcome in which θ_L receives $P = 0$ or where θ_H receives $P = \theta_L$ is not an equilibrium. Instead, courts will experiment until they demand a sufficiently high level of s to establish a separating equilibrium in which θ_L is deterred but θ_H takes the action. Any equilibrium in which θ_H chooses $s > s^*$ is inefficient, although it may still be a welfare improvement over the strict liability equilibrium. In the remainder of this section I discuss three ways that the lawmaker may steer the courts and actors to an equilibrium that is more efficient than $\{(0, \theta_L), (s, 0)\}$ in the case that $s > s^*$.

6.1 Restrictions on Judicial Beliefs

The problem of too many equilibria in signaling games has generated a large literature on "equilibrium refinements." Although judicial beliefs at the equilibrium punishments are pinned down by the actors' choices, out-of-equilibrium beliefs are not restricted in the same way. Suppose that a judge believes that actors are of type θ_L unless they produce evidence in quantity $s > s^*$. In this separating equilibrium, only types θ_H will choose s, so the judge's beliefs are correct in equilibrium. This equilibrium is inefficient, however, and the reason is that the judge has such pessimistic beliefs about the actor's type for $s' \in (s^*, s)$; believing that the actor must be of type θ_L if she produces a quantity of evidence in this range. Note, however, that θ_L would rather not commit the act than generate evidence s' even if she could be assured of receiving no punishment at all; she strictly prefers $(0, \theta_L)$ to $(s', 0)$. If the judge recognizes this, it seems like it would be unreasonable for her to assign any positive probability to the actor being of type θ_L when she observes evidence in this range. This is the reasoning behind the Cho-Kreps "intuitive criterion" (Cho and Kreps, 1987), which is a popular refinement that restricts the set of equilibria in this game by restricting the set of reasonable judicial beliefs. The only beliefs that survive application of this criterion are those that assign probability 1 to the actor being of type θ_H for any $s \geq s^*$. With this

restriction on judges' beliefs, the only equilibrium that survives is one in which θ_H chooses s^*; the constrained first-best equilibrium that is achieved under a rule.

One way that judicial beliefs might actuall come to be "reasonable" in this way is that they evolve over time through an iterative process to induce the efficient outcome. For example, consider the standard described in the introduction for the deductibility of "hobby losses" under §183 of the Internal Revenue Code. As the court within a jurisdiction passes judgment on whether particular fact patterns are sufficient evidence of whether an activity is engaged in for profit, taxpayers might be expected to push for lower and lower thresholds through more aggressive tax positions. If the court endorses some threshold $s < s^*$, then it will inadvertently create a pooling equilibrium and the jurisdiction will be flooded with taxpayers who do not in fact have a profit motive but claim tax deductions for their hobbies. If courts care about the accurate application of the law, they should be expected to adjust the s threshold back up to s^* to induce the separation of types.

6.2 Partial Excuses

If judges cannot be relied on to have reasonable out-of-equilibrium beliefs, either by recognizing and trying to comply with the intuitive criterion or some related refinement or through an iterative process of trial-and-error, then the lawmaker may consider two alternatives to steer actors and courts to a more efficient equilibrium. The first is to provide something less than a full excuse/justification to actors who provide favorable evidence of their type.

Figure 4 illustrates a separating equilibrium $(s, 0)$ that is inferior to $(s^*, 0)$. The blue dotted curves again represent the lawmaker's indifference curves for combinations of evidence and punishment, depending on whether the marginal cost of punishment is high or low (the flatter line reflects the higher marginal cost of punishment α). Suppose now that the lawmaker stipulates that high types are entitled to only a partial excuse or justification, so that instead of receiving no punishment at all if the judge believes they are of the high type they will receive some positive punishment given by $\tilde{P}$.

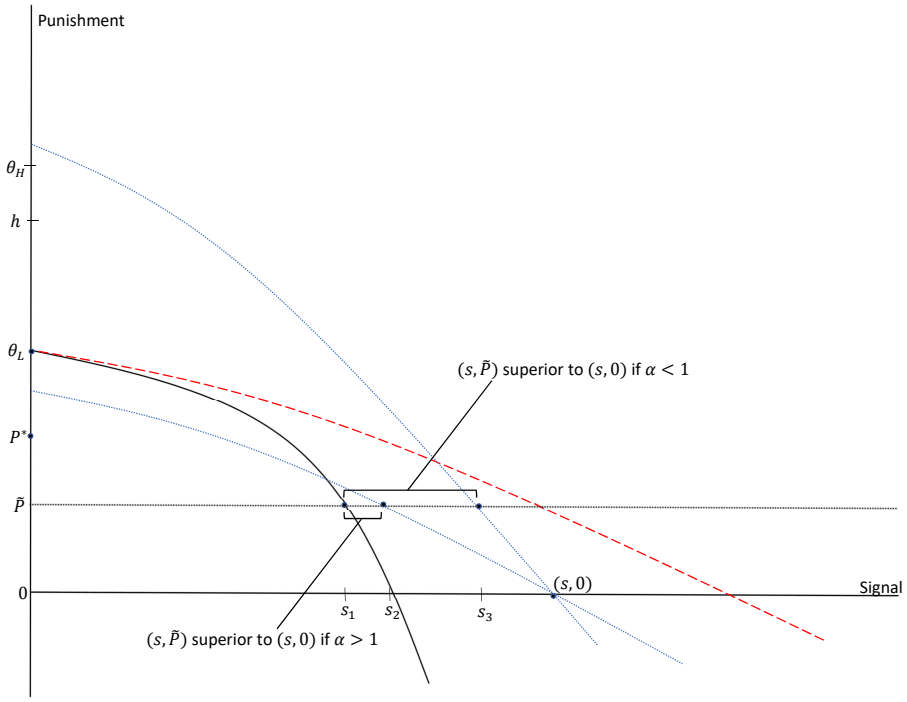

Figure 4: Partial Justification Equilibria

By providing only a partial excuse, the lawmaker changes the set of possible separating equilibria. Previously, the amount of evidence that θ_H produced in a separating equilibrium fell between the place that the solid black line intersects the horizontal axis at 0 and the place that the dashed red line crosses that same axis. When the horizontal axis is raised from 0 to $\tilde{P}$, the new evidentiary signals sent by θ_H fall between the places that the black and red lines cross that new axis.

Let the generic equilibrium for θ_H in this new partial excuse regime be $(s', \tilde{P})$. What is the likelihood that this equilibrium will be more efficient than the old equilibrium at $(s, 0)$? This depends on α and s. The greater the amount of evidence s required in the full-excuse equilibrium, i.e., the more pessimistic the beliefs of the judge about the actor's type, the more likely it is that the new partial excuse equilibrium $(s', \tilde{P})$ will be less socially costly. One can see this straightforwardly by moving the point $(s, 0)$ further to the right. As this equilibrium moves to the right (i.e., it requires more wasteful evidence production), the set

of equilibrium points on the new axis (at $\tilde{P}$) that fall below the social indifference curve passing through $(s, 0)$ becomes larger.

Whether the equilibrium under a partial excuse regime is more efficient than $(s, 0)$ also depends on the marginal social cost of punishment. If the marginal social cost of punishment is less than the marginal cost to the actor (i.e., $\alpha < 1$), such as in the case of a fine, then the new equilibrium on $\tilde{P}$ will tend to be more efficient than the full-excuse equilibrium. To see this, follow the steep blue dotted line, which represents a low marginal social cost of punishment, from the point $(s, 0)$ up through the axis at $\tilde{P}$ and note that the potential equilibrium points on the $\tilde{P}$ line between s_1 and s_3 lie on lower indifference curves than the line that passes through $(s, 0)$.

On the other hand, when the social cost of punishment is greater than the cost to the individual, such as in the case of hard treatment like imprisonment, the partial-excuse equilibrium is less likely to be socially preferred to the full-excuse equilibrium. This can be seen in Figure 4 by focusing on the flatter blue dotted line, which reflects a high marginal social cost of punishment, and noting that only the points on the $\tilde{P}$ line between s_1 and s_2 lie on lower (preferred) indifference curves to the curve that passes through $(s, 0)$. Thus, a punishment regime with partial excuses is most likely to result in a better outcome than a regime with full excuses when (1) judges are pessimistic about actors' types and demand high values of s to conclude that an actor is of the high type, or (2) when the marginal social cost of punishment is low.

6.3 Standards of Proof

In this subsection I consider one more way that the lawmaker might facilitate convergence on a more efficient equilibrium and lower the evidentiary production costs for high types: the burden of proof. Suppose that the judge believes that the actor is of type θ_H with probability 1 only if the actors produce evidence in quantity s that is at least as big as s_2, where $s_2 > s^*$. In the separating equilibrium, θ_H will choose s_2 and θ_L will choose $s = 0$.

Generically, the judge will assign some positive probability to the actor being of type θ_H if she observes intermediate values of s between 0 and s_2 (which are off the equilibrium path), and that probability is increasing in the amount of favorable evidence produced. The green solid curve in Figure 5 represents the judge's expectation about the actor's type. The judge has correct beliefs about the actor's type when the evidence produced is 0 or s_2, because her beliefs must be correct in equilibrium. For intermediate values of s, which are not chosen in equilibrium, her beliefs are largely unconstrained. As in the previous figures, the red dashed line is an indifference curve for θ_H and the solid black line is an indifference curve for θ_L.

Suppose now that the lawmaker lowers the burden of proof, specifying that the excuse applies whenever it is more likely than not that the actor is of the high type. In Figure 5, the judge believes that the actor is more likely than not to be of the high type for any amount of evidence $s \geq s_1$. Under this standard of proof, θ_H knows that he will be fully excused if he produces evidence of at least s_1, which he will do.

Thus, the new separating equilibrium under a more-likely-than-not standard is $\{(0, \theta_L), (s_1, 0)\}$, which is more efficient than the old equilibrium of $\{(0, \theta_L), (s_2, 0)\}$. Note also that although the court has been instructed in this case to excuse the actor if it only more likely than not that she is type θ_H, in the new separating equilibrium *only* high types are actually excused. Lowering the standard of proof runs the risk, however, of destroying the separating equilibrium. The judge's beliefs may be such that s_1, the evidentiary threshold that persuades the judge that actor is more likely than not to be θ_H, is less than s^*. In that case, setting the standard of proof to 50% will induce both types to produce evidence in quantity s_1 and be fully excused, because the point $(s_1, 0)$ will be preferred by both types to $(0, \theta_L)$.

This outcome is especially problematic because it is uncertain whether courts and actors will adjust to re-establish a separating equilibrium. If the proportion of high types in the population is sufficiently high (i.e., $> 50\%$), then judges may be satisfied with this outcome, even if they have a preferences for accuracy in assigning punishments. After all, the lawmaker has instructed the judge that a full excuse is appropriate even if the judge is not certain that

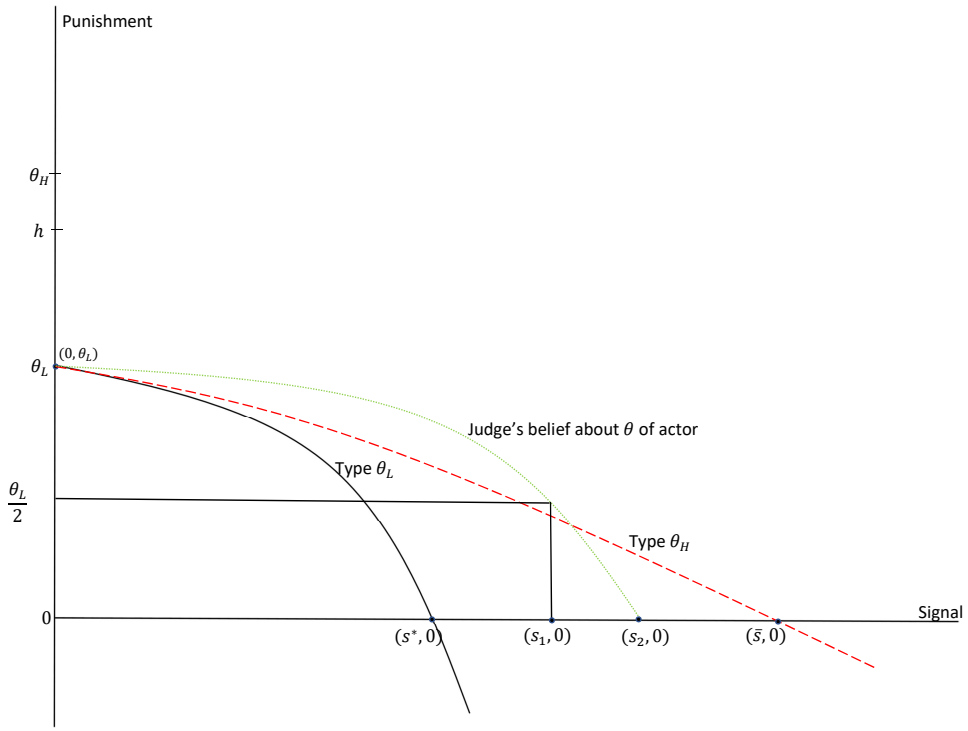

Figure 5: Preponderance of the Evidence Standard of Proof

the actor is a high type.[27] If, on the other hand, types θ_H make up less than half of the population then we should expect judges to increase the evidentiary threshold to excuse the defendant until more than 50% of those who are excused are in fact of type θ_H.

7 Choosing a Game (Or Whether to Play At All)

In this section I summarize the factors that decide whether it is more efficient to regulate conduct using a strict liability punishment, or a quasi fault-based punishment regime that excuses certain actors based on the amount of favorable evidence they produce. I proceed by first discussing whether a rule or standard is more efficient. I then describe the circumstances under which a quasi fault-based punishment regime is more efficient than strict liability.

7.1 The Choice Between a Rule and Standard

Let the costs of promulgating a standard and a rule be given by p_s and p_r, where $p_s < p_r$. Also let the costs of adjudicating these laws be given by c_s and c_r where $c_r < c_s$. I assume that both a rule and a standard induce separation of actors by type, deterring θ_L and fully excusing θ_H in exchange for producing a certain amount of favorable evidence. The only cost differences between a rule and a standard then are the costs of promulgation, adjudication, and the efficiency of the equilibrium under a standard relative to the equilibrium under a rule (which is assumed to be constrained first-best).

Note that the equilibrium evidence s' under a standard depends on the beliefs of the factfinder about what quantity of evidence supports an inference that an actor is of the high type. If the lawmaker is choosing the law for a large number of judges whose beliefs induce different amounts of evidence in equilibrium, or the lawmaker is making law for only one judge but the lawmaker is uncertain about that judge's beliefs, then the distribution of judicial beliefs must be taken into account. Let the distribution function for s', the equilibrium

[27] This outcome is a variant of the "paradox of the gatecrasher" in evidence law.

amount of evidence that exculpates the actor under a standard, be $G(s')$ and the associated continuous density function be $g(s')$, which is strictly positive everywhere on the interval $[s^*, \bar{s}]$ (which are the evidence threshholds that support a separating equilibrium). A rule will be more efficient than a standard if

$$p_r - p_s < \int_{s^*}^{\bar{s}} \int_{\theta_H}^{\bar{\theta}} \Big(v(s'; \theta) + c_s - v(s^*; \theta) - c_r \Big) f(\theta) g(s) d\theta ds \tag{4}$$

A rule is expected to be more efficient than a standard if the additional promulgation costs or a rule are less than the expected additional evidence production and adjudication costs over the range of possible separating equilibria under the standard.

7.2 Should the Law Ignore Actors' Types?

In this subsection I identify when it is preferable to impose a single, strict-liability, punishment rather than choose a rule or standard that imposes a punishment θ_L only on actors of a low type and fully excuses actors who produce favorable evidence s'. Let the costs of promulgating and adjudicating a strict liability punishment be p_{ss} and c_{ss}, and the costs of promulgating and adjudicating the most efficiencty quasi fault-based regime (either a rule or a standard) be given by p_x and c_x.

I assume that promulgating a fault-based standard costs as much as promulgating a strict liability punishment, but promulgating a fault-based rule costs much more. This is because a standard simply stipulates a single punishment with an excuse for actors "of a high type"; it does not specify what evidence must be taken as determinative of being a high type. I also assume that the adjudication costs of both a rule and a strict liability punishment are much less than the cost of adjudicating a standard. This is for the familiar reason that determining whether a set of facts entitle an actor to be excused is costly for the court, and that it is costlier than simply determining whether certain facts are present. The adjudication costs for the fault-based rule and strict liability are close because, like a rule, a

strict liability punishment only asks the court to inventory the verifiable findings of fact to determine whether the predicate for punishment has been met. A quasi fault-based rule or standard will be better than strict liability if

$$p_x - p_{ss} < \int_{P^*}^{\theta_L} \Big(h - \theta + \alpha P^* + c_{ss}\Big) f(\theta) d\theta + \int_{\theta_L}^{\bar{\theta}} \Big(\alpha P^* - v(s'; \theta) + c_{ss} - c_x\Big) f(\theta) d\theta \quad (5)$$

The left-hand side is the additional promulgation costs incurred by moving from strict liability to fault-based regime.The first term on the right-hand side is the strictly positive benefit from deterring inefficient acts by increasing the baseline punishment from P^* to θ_L. These benefits include the harm avoided net of the private benefit to the actor, and the punishment and adjudication costs that does not need to be incurred for these newly deterred actors. The second term is the change in social welfare that comes from reducing (to zero) the punishment on those who still commit the harmful act but also requiring them to produce costly favorable evidence to signal that they should be excused, as well as changing the adjudication costs from those associated with strict liability to those under a fault-based regime. This second term could be positive or negative.

7.3 An Example

In this section I consider a simple numerical example to illustrate the choice between a rule, a standard, and strict liability. Let the actors' types θ be distributed uniformly on the interval $[0, 1]$, so that $F(\theta) = \theta$ and $f(\theta) = 1$. Suppose that there is only one equilibrium that the lawmaker needs to worry about under a standard, in which the high type produces evidence s'. Let the cost of producing this evidence be given by $v(s; \theta) = s^2/\theta$. In this case, Equation 4 simplifies so that a rule is more efficient than a standard when:

$$s'^2 - s^{*2} > \frac{(p_r - p_s) - (1 + \theta_H)(c_s - c_r)}{\ln(\theta_H)} \quad (6)$$

The left hand side is the additional wasteful evidence production costs under the standard above the costs under a rule. The more inefficient is the equilibrium under a standard, the more attractive is the rule. Rules are also more likely to be desirable the smaller is the additional promulgation cost of a rule and the larger is the additional adjudication cost under a standard. Also, a standard becomes more attractive for higher values of θ_H, since higher values of θ_H reduce the welfare costs associated with producing favorable evidence.

When is a quasi fault-based regime better than strict liability? Using the first order condition given in Equation 1, the optimal strict liability punishment is $P^* = \frac{\alpha - h}{2\alpha - 1}$. Note that for this to be strictly positive then $\alpha > h \iff \alpha > 1/2$ and conversely $\alpha < h \iff \alpha < 1/2$. Limiting our attention to the case of underdeterrence in the case of strict liability (i.e., $P^* < h$), this implies that $h > 1/2 \iff \alpha > 1/2$ and $h > 1/2 \iff \alpha > 1/2$. Together, these restrictions imply that if $h > 1/2$ then $\alpha > h$ and if $h < 1/2$ then $\alpha < h$. Normalizing the cost of adjudicating and promulgating the strict liability law to be 0 and substituting into inequality 5, then the best quasi fault-based law is better than strict liability if:[28]

$$-\frac{(\alpha - h)^2}{2(2\alpha - 1)} < \theta_L h - c_x(1 - \theta_L) - \frac{\theta_L^2}{2} + s'^2 \ln(\theta_L) - p_x \tag{7}$$

Using this inequality, we can calculate some comparative statics of the effect of α, h, θ_L on the efficiency of a quasi fault-based law. Taking the derivative of the left-hand side of inequality 7 with respect to alpha gives

$$\partial lhs/\partial\alpha = -\frac{\alpha(\alpha - 1) - h(h - 1)}{(2\alpha - 1)^2} < 0$$

Thus, the attractiveness of a quasi fault-based law relative to strict liability is increasing the in marginal cost of punishment.[29] Moving the term θ_L to the left-hand side of equation 7, and taking the derivative with respect to h gives

[28]The RHS is equal to $\frac{P^*(h-\alpha)}{2}$.

[29]Note that the function $x(x-1)$ is decreasing for $x \in [0, 1/2)$ and increasing thereafter, and is negative valued for values of x less than 1. Using the facts that $h > 1/2$ implies $\alpha > h$ and $h < 1/2$ implies $\alpha < h$, $\partial/\partial\alpha < 0$.

$$\partial lhs/\partial h = \frac{\alpha - h}{2\alpha - 1} - \theta_L < 0$$

Note that the first term is simply equal to P^* which is less than θ_L. Thus, quasi fault-based laws become more attractive as the harm from the act increases. Finally, the relative efficiency of a quasi fault-based law depends on the low type. Taking the derivative of the right-hand side of equation 7 with respect to θ_L gives

$$\partial rhs/\partial \theta_L = h + c_x - \theta_L + \frac{s'^2}{\theta_L} > 0$$

Note that $\theta_L < h$, by assumption. This means that the efficiency of a quasi fault-based regime also increases with θ_L. Looking at the other parameters in the inequality above, predictably, the attractiveness of a quasi fault-based law declines as the costs of promulgating and adjudicating it increase.

8 Conclusions

In this paper I analyze the regulation of harmful conduct where the optimal punishments depend on private information of the regulated parties. The optimal strict liability punishment may be less than maximal if either some acts are socially efficient or some acts cannot be deterred at the maximal punishment.

If the actors can produce evidence that signals that they are of a favored type, then lawmakers are faced with an important question affecting the efficiency of the punishment regime: whether to specify the correspondence between evidence and punishments, or whether to specify the correspondence between actors' types and punishments and thereby leave it to courts to draw inferences about the relationship between evidence and types. I argue that this is the difference between a rule and a standard, and also the difference between whether courts and actors are engaged in a screening game or a signaling game of

asymmetric information. Although lawmakers can ensure that the most efficient equilibrium is achieved by adopting a rule and inducing a screening game, this comes at the cost of identifying, *ex ante*, the separating equilibrium in which there is the least amount of wasteful evidence production.

In the event that identifying this equilibrium is too costly to justify adopting a rule, I identify three ways that lawmakers can encourage convergence on more efficient equilibria under a standard. First, the lawmaker might be able to steer judges beliefs to be consistent with the intuitive criterion or other restrictions on out-of-equilibrium beliefs. Second, it may be helpful to provide only partial, rather than complete, excuses/justifications to actors who can demonstrate that they are eligible. Partial justifications are attractive when the social cost of the punishment is especially low or when judges have especially pessimistic beliefs about the evidence required to reveal that the actor is of the high type. Finally, lowering the burden of proof that the probability that the actor is of the high type can also induce a more efficient equilibrium under a standard.

References

Baker, C. E. (1977). Counting preferences in collective choice situations. *UCLA L. Rev.*, 25:381.

Bar-Gill, O. and Ben-Shahar, O. (2008). An information theory of willful breach. *Mich. L. Rev.*, 107:1479.

Becker, G. S. (1968). Crime and punishment: An economic approach. In *The economic dimensions of crime*, pages 13–68. Springer.

Cass, R. A. and Hylton, K. N. (2000). Antitrust intent. *S. Cal. L. Rev.*, 74:657.

Cho, I.-K. and Kreps, D. M. (1987). Signaling games and stable equilibria. *The Quarterly Journal of Economics*, 102(2):179–221.

Choi, A. and Triantis, G. (2008). Completing contracts in the shadow of costly verification. *The Journal of Legal Studies*, 37(2):503–534.

Choi, A. and Triantis, G. (2009). Strategic vagueness in contract design: The case of corporate acquisitions. *Yale LJ*, 119:848.

Cooter, R. D. (1982). Economic analysis of punitive damages. *S. Cal. L. Rev.*, 56:79.

Curry, P. A. (2017). Malice aforethought. *Review of Law & Economics*, 13(1).

Ellis Jr, D. D. (1982). Fairness and efficiency in the law of punitive damages. *S. Cal. L. Rev.*, 56:1.

Ellis Jr, D. D. (1983). An economic theory of intentional torts: A comment. *International Review of Law and Economics*, 3(1):45–57.

Finkelstein, C. (2000). The inefficiency of mens rea. *Cal. L. Rev.*, 88:895.

Fudenberg, D. and Tirole, J. (1991). Game theory, 1991. *Cambridge, Massachusetts*, 393(12):80.

Gale, D. and Hellwig, M. (1985). Incentive-compatible debt contracts: The one-period problem. *The Review of Economic Studies*, 52(4):647–663.

Givati, Y. (2015). An incomplete contracting approach to administrative law. *American Law and Economics Review*, 18(1):176–207.

Grossman, S. J. and Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of political economy*, 94(4):691–719.

Hamdani, A. (2007). Mens rea and the cost of ignorance. *Virginia Law Review*, pages 415–457.

Hart, O. (1995). *Firms, contracts, and financial structure*. Clarendon press.

Hart, O. and Moore, J. (1988). Incomplete contracts and renegotiation. *Econometrica: Journal of the Econometric Society*, pages 755–785.

Hayashi, A. T. (2017). A theory of facts and circumstances. *Ala. L. Rev.*, 69:289.

Hayashi, A. T. (2019). The law and economics of bad intentions. *Mimeograph*.

Hollander-Blumoff, R. (2011). Crime, punishment, and the psychology of self-control. *Emory LJ*, 61:501.

Hylton, K. N. (2009). 2009 monsanto lecture-intent in tort law. *Val. UL Rev.*, 44:1217.

Johnston, J. S. (1995). Bargaining under rules versus standards. *JL Econ. & Org.*, 11:256.

Kaplow, L. (1992). Rules versus standards: An economic analysis. *Duke Lj*, 42:557.

Khalil, F. (1997). Auditing without commitment. *The RAND Journal of Economics*, 28(4):629.

Korobkin, R. B. (2000). Beahvioral analysis and legal form: Rules vs. standards revisited.

Or. L. Rev., 79:23.

Landes, W. M. and Posner, R. A. (1981). An economic theory of intentional torts. *International Review of Law and Economics*, 1(2):127–154.

Lempert, R. (2001). The economic analysis of evidence law: Common sense on stilts. *Va. L. Rev.*, 87:1619.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The review of economic studies*, 38(2):175–208.

Osofsky, L. (2013). Who's naughty and who's nice-frictions, screening, and tax law design. *Buff. L. Rev.*, 61:1057.

Parker, J. S. (1993). The economics of mens rea. *Virginia Law Review*, pages 741–811.

Polinsky, A. M. and Shavell, S. (1997). Punitive damages: An economic analysis. *Harv. L. Rev.*, 111:869.

Polinsky, A. M. and Shavell, S. (2000a). The economic theory of public enforcement of law. *Journal of economic literature*, 38(1):45–76.

Polinsky, M. A. and Shavell, S. (2000b). The fairness of sanctions: some implications for optimal enforcement policy. *American Law and Economics Review*, 2(2):223–237.

Posner, E. A. (1997). Standards, rules, and social norms. *Harv. JL & Pub. Pol'y*, 21:101.

Posner, R. A. (1985). An economic theory of the criminal law. *Columbia law review*, 85(6):1193–1231.

Posner, R. A. (1998). An economic approach to the law of evidence. *Stan. L. Rev.*, 51:1477.

Posner, R. A. (2001). Comment on lempert on posner. *Virginia Law Review*, pages 1713–1721.

Raskolnikov, A. (2009). Revealing choices: Using taxpayer choice to target tax enforcement. *Colum. L. Rev.*, 109:689.

Raskolnikov, A. (2015). Six degrees of graduation: Law and economics of variable sanctions. *Fla. St. UL Rev.*, 43:1015.

Riley, J. G. (2001). Silver signals: Twenty-five years of screening and signaling. *Journal of Economic literature*, 39(2):432–478.

Sanga, S. (2018). Incomplete contracts: An empirical approach. *The Journal of Law, Economics, and Organization*, 34(4):650–679.

Schwartz, G. T. (1979). Economics, wealth distribution, and justice. *Wis. L. Rev.*, page 799.

Scott, R. E. and Triantis, G. G. (2005). Anticipating litigation in contract design. *Yale LJ*, 115:814.

Shavell, S. (1985). Criminal law and the optimal use of nonmonetary sanctions as a deterrent. *Columbia Law Review*, 85(6):1232–1262.

Shavell, S. (2009). *Foundations of economic analysis of law*. Harvard University Press.

Stein, A. (2014). Inefficient evidence. *Ala. L. Rev.*, 66:423.

Sunstein, C. R. and Vermeule, A. (2005). Is capital punishment morally required? acts, omissions, and life-life tradeoffs. *Stanford Law Review*, pages 703–750.

Townsend, R. M. (1979). Optimal contracts and competitive markets with costly state verification. *Journal of Economic theory*, 21(2):265–293.

Weinzierl, M. (2018). Welfarism's envy problem extends to popular judgments. In *AEA Papers and Proceedings*, volume 108, pages 28–32.