3-2014

Innovation and Optimal Punishment, with Antitrust Applications

Keith N. Hylton
Boston University School of Law

Haizhen Lin
Indiana University - Kelley School of Business

Journal of Competition Law & Economics, *10(1)*, *1–25*
doi:10.1093/joclec/nht027
Advance Access publication 12 September 2013

INNOVATION AND OPTIMAL PUNISHMENT, WITH ANTITRUST APPLICATIONS

*Keith N. Hylton** *& Haizhen Lin*†

ABSTRACT

This article modifies the optimal punishment analysis by incorporating investment incentives with external benefits. In the models examined, the recommendation that the optimal penalty should internalize the marginal social harm is no longer valid. We focus on antitrust applications. In light of the benefits from innovation, the optimal policy will punish monopolizing firms more leniently than suggested in the standard static model. It may be optimal not to punish the monopolizing firm at all, or to reward the firm rather than punish it. We examine the precise balance between penalty and reward in the optimal punishment scheme.

JEL: D42; K14; K21; K42; L41; L43

I. INTRODUCTION

The literature on the economic theory of punishment consists mostly of models in which offenders are held strictly liable and sanctioned with case-specific optimal penalties. The law, in contrast, applies case-specific cost-benefit tests to determine liability and uses standardized penalties.[1]

In spite of this difference between the theory of punishment and the practice in courts, the theory remains useful as a guide for policy. This article follows tradition by examining optimal penalties as a source of guidelines for legal policy. The focus here is innovation and punishment, especially in the context of antitrust.

The prevailing analysis of optimal antitrust penalties holds that in order to avoid inefficient overdeterrence, the optimal penalty should internalize the social losses generated by potentially anticompetitive conduct.[2] If enforcement

* Professor of Law, Boston University School of Law. Email: knhylton@bu.edu.

† Assistant Professor of Economics, Kelley School of Business, Indiana University. Email: hzlin@indiana.edu. The authors thank the editors of this journal for helpful suggestions.

[1] For example, in American antitrust law, the main application for this article's model, courts use rule of reason analysis to exempt efficient conduct for the most part, while applying statutorily set penalties or treble damages to the violations. In tort law, courts use the negligence test, or the risk-utility test in products liability cases, to determine liability. Each of these legal tests is a type of cost-benefit test.

[2] Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169 (1968).

were perfect and costless, such a penalty would internalize the transfer from consumers and the deadweight loss.[3] Thus, if a monopolizing firm introduces efficiencies, it might still have an incentive to carry out its monopolizing conduct, as long as the efficiency gain exceeds the deadweight loss.

But this analysis does not incorporate plausible social benefits from monopolization. The prevailing analysis is based on a static model in which the gains from monopolization accrue to the firm and the losses are suffered by society. In contrast, a dynamic perspective[4] would take into account the social gains from investments made by the firm in its quest to become a monopoly. Investments in market-creating or market-expanding innovation should be incorporated into the analysis of punishment.

We modify the optimal punishment analysis by incorporating investment that has external benefits—such as the creation of surplus for consumers. In the models examined here, the recommendation that the optimal antitrust penalty should internalize the marginal social harm—as measured by the sum of the consumer surplus transfer, the deadweight loss, and the cost of enforcement—is no longer valid. We explore the recommendations from two simple versions of the dynamic story, one in which the offender invests in market-expanding innovation before committing the offense (monopolization), and another in which the victim invests before suffering an offense.

In the offender-investment model, which is the core of this article, the optimal penalty for monopolization is a function of the consumer harm, the residual consumer surplus (after monopolization), the cost of enforcement, and the relative responsiveness of innovation and monopolization to changes in the penalty. Specifically, the optimal penalty is a weighted average of the static penalty (internalizing consumer harm and enforcement cost) and an innovation subsidy (internalizing consumer benefit and enforcement cost), with the weights determined by the relative sensitivities of investment and monopolization to punishment. This has implications for law and punishment policy.

The most obvious implication is that in light of the benefits from innovation, the optimal policy will punish monopolizing firms more leniently than suggested by the static model. It may be optimal not to punish the firm at all, or to reward the firm rather than punish it. In this sense, the model provides a Schumpeterian perspective on punishment,[5] as well as the groundwork for a

[3] William M. Landes, *Optimal Sanctions for Antitrust Violations*, 50 U. CHI. L. REV. 652 (1983).

[4] On the difference between static and dynamic punishment models, *see* Siu Fai Leung, *How to Make the Fine Fit the Corporate Crime? An Analysis of Static and Dynamic Optimal Punishment Theories*, 45 J. PUB. ECON. 243 (1991). Our framework is simpler than Leung's. We use the term dynamic here in the sense common in the antitrust literature to refer to an analysis that takes the investment effects of enforcement into account. *See* J. Gregory Sidak & David F. Teece, *Dynamic Competition in Antitrust Law*, 5 J. COMPETITION L. & ECON. 581 (2009).

[5] JOSEPH SCHUMPETER, CAPITALISM, SOCIALISM, AND DEMOCRACY ch. 7–8 (1943). Schumpeter famously criticized the static model of competition for ignoring the social benefits of innovation, and the need for firms to gain monopoly power in order to earn a positive return on innovation.

positive theory of monopolization law, which has been puzzlingly lenient over its history.[6]

In addition, comparative statics results for the optimal penalty differ from the static analysis. The optimal penalty does not increase monotonically with consumer harm, and a fall in the probability of apprehension does not necessarily imply an increase in the optimal penalty. A combination of a high fine and a low probability of punishment may not be optimal.

The connection between innovation and punishment is a concern in both antitrust and products liability. In *United States v. Microsoft*,[7] the D.C. Circuit Court of Appeals refused to apply a *per se* liability rule to the firm's technological integration of the internet browser and operating systems because of its fear that such a rule would discourage innovation.[8] In *Trinko v. Verizon*,[9] the Supreme Court cited the negative innovation effect as a basis for refusing to adopt the essential facilities theory of monopolization.[10] Products liability lawsuits against drug manufacturers have been met with the criticism that their success will deter the development and marketing of new drugs.[11] The unexplored issue in these cases is the precise relationship between the deterrence of offensive conduct and the encouragement of innovation in an optimal punishment scheme.[12]

Parts II.A and II.B set up the static model, which replicates the standard optimal penalty recommendation. Parts II.C.2 and II.C.3 explore punishment in a setting in which the offender makes an investment that provides social

For a review, *see* Edward Mason, *Schumpeter on Monopoly and the Large Firm*, 33 REV. ECON. & STAT. 139 (1951).

6 On the perceived leniency of monopolization law and its explanation, *see* David S. Evans & Keith N. Hylton, *The Lawful Acquisition and Exercise of Monopoly Power and Its Implications for the Objectives of Antitrust*, 4 COMPETITION POL'Y INT'L 203 (2008).

7 United States v. Microsoft, 253 F.3d 34 (D.C. Cir. 2001).

8 *Id.* at 89–90. In addition, much of the commentary about *Microsoft* focuses on the implications of antitrust enforcement for innovation in the technology industries; *see* David S. Evans & Richard Schmalensee, *Some Economic Aspects of Antitrust Analysis in Dynamically Competitive Industries*, *in* 2 INNOVATION POLICY AND THE ECONOMY 1 (Adam B. Jaffe, Josh Lerner & Scott Stern eds., MIT Press 2002).

9 Verizon Commc'n Inc. v. Law Offices of Curtis V. Trinko, 540 U.S. 398 (2004).

10 *Id.* at 407.

11 *See, e.g.*, Steven Shavell & A. Mitchell Polinsky, *Vioxx Verdict's Dark Side*, BOSTON GLOBE, Aug. 23, 2005, at A15, *available at* http://www.boston.com/news/globe/editorial_opinion/oped/articles/2005/08/23/vioxx_verdicts_dark_side/.

12 One question generated by this model is whether the optimal penalty results derived here could be implemented. There are two ways to approach implementation. First, the variables that the optimal penalty incorporates indicate the matters that courts should take into account in the relevant law on liability. Second, the optimal penalty could be estimated by acquiring information on the probability of detection (or apprehension), the harm to victims, the sensitivity of investment to cash flow, and the residual consumer surplus. Although courts do not attempt to estimate optimal case-specific penalties, a move to such a regime, as originally advocated in Becker, would not be infeasible. *See* Becker, *supra* note 2.

benefits, the returns to which are a function of an offense that he may commit in the future as well as the punishment for that offense. Part II.C.4 examines an extension in which the victims make investments, the expected returns to which are reduced by an offense that may be committed by the offender. Part III discusses implications for the law.

II. MODEL

A. Basic Assumptions

All actors are risk neutral and victims are the only parties who suffer loss. The state apprehends an offender after an injury has occurred. The state does not attempt to apprehend the offender in each instance of an offense, and therefore the probability of apprehension (equivalently, detection) given an injury is less than one.

Let z = the probability of apprehension, $0 < z' < z \leq 1$;[13] c = the cost to the state of apprehending the offender, $c > 0$; v = the loss suffered by a victim, $v > 0$; F = the fine imposed on apprehended offender. M = the gross gain to the offender from committing an offense, and is governed by the probability distribution function H with corresponding density function h, where $h(M) > 0$ for $0 \leq M < M_u$, and $h(M) = 0$ otherwise.

The offender cannot satisfy his preferences through the market; thus in order to enjoy the gain M he must commit an offense. Because the offender will commit the offense when $M > zF$, the probability of an offense is $1–H(zF)$. If $M_u \leq zF$ no crimes will be committed, so that $F = M_u/z$ is the minimum level of the fine that achieves complete deterrence.

The timeline of events is as follows: the offense occurs (probability $1–H(zF)$) causing a loss of v; enforcement occurs with probability z; the offender is apprehended at cost c, and then punished with a fine F.[14]

B. Optimal Punishment Policy: Static Case

The optimal punishment policy is the combination of the fine and the probability of apprehension that minimizes the cost of offenses and the cost of

[13] We assume a lower bound z' on the probability of apprehension because no state ever chooses to go without any law enforcement at all. The minimal "night watchman" state envisioned by philosophers limits itself to enforcing criminal prohibitions and (maybe) contracts. The minimal apprehension probability is the level consistent with the minimum expenditure necessary (staffing, equipment) to enforce criminal prohibitions.

[14] Once apprehended, punishment occurs with certainty, given the assumption of strict liability. In a later part, we consider a "rule of reason" standard under which punishment depends on whether the offender violated the rule of reason.

avoiding offenses:[15]

$$C = (1 - H(zF))(v + zc) + H(zF)E[M|M < zF]. \tag{1}$$

The optimal policy can be stated as follows:

Proposition 1: If $M_u > v + z'c$, then the optimal punishment policy is to set the fine so that it satisfies $F = F^* = v/z' + c$, and the probability of apprehension at the minimum level z' (internalization rule). If $M_u \leq v + z'c$, then the optimal policy is to set the fine and probability of apprehension so that $zF \geq M_u$ (complete deterrence rule).

This result replicates Becker, and Polinsky and Shavell (1992)[16] with minor modifications. In intuitive terms, if the offender's activity is potentially efficient, in the sense that the gain to at least some offenders exceeds the marginal social cost of the offense, then the optimal penalty internalizes the marginal social cost of the offense. In order to minimize enforcement costs, the state sets the probability of apprehension at its minimum level. However, if the offender's activity is not potentially efficient, the optimal policy is to completely deter it by eliminating the offender's gain.

It follows that an antitrust punishment authority should distinguish conduct that is potentially efficient from conduct that is not, and apply the internalization policy in the potentially efficient category and the complete deterrence policy in the inefficient category. Suppose that the monopolizing firm takes an act that allows it to extract surplus from consumers and may also generate an efficiency gain, as shown in Figure 1—such as an exclusive dealing contract that forecloses competition and at the same time reduces supply costs. Because $M = T + E$,[17] and $v = T + D$, the optimal static penalty is $F^* = (T + D) / z' + c$. If no efficiency gain were possible, the optimal policy would seek to eliminate expected profits. F^* would serve the gain-elimination purpose, as would any other penalty greater than T/z'.

C. Optimal Punishment Policy: Dynamic Setting

In the previous part we examined the enforcement model in the context of antitrust, replicating the static enforcement policy (internalize transfer and deadweight loss, plus enforcement cost). In this part, we extend the model to incorporate investment by the offender.

15 Equivalent objective functions would require maximizing the net benefit from offenses NB *(offenses)* $= (1 - H(zF))E[M - v - zc \mid M > zF]$, or the difference between net deterrence benefits and the costs of enforcement $H(zF)(v - E[M \mid M < zF]) - (1 - H(zF))zc$.

16 Becker, *supra* note 2; A. Mitchell Polinsky & Steven Shavell, *Enforcement Costs and the Optimal Magnitude and Probability of Fines*, 35 J.L. & ECON. 133 (1992).

17 Because M is a random variable and T is fixed, E, the efficiency gain, is a random variable. The efficiency gain associated with the offense determines the uncertainty associated with the gross gain from the offense.

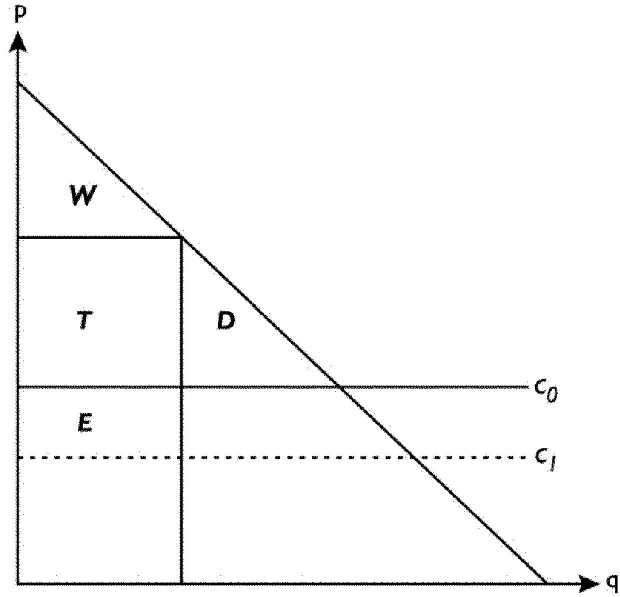

Figure 1. Monopolization with an efficiency gain

The reason for taking investment by the offender into account is that it is important in the antitrust setting, especially in monopolization cases. Suppose that the monopolizing firm has undertaken investments that benefit consumers, such as creating a new product market. The static enforcement policy may be socially excessive because it might prevent the firm from earning a break-even return on its investments in market creation. This is distinguishable from the static case without investment, where the product market was already in existence and the monopolizing firm takes an action that permits it to gain monopoly power, perhaps also with an efficiency gain. In the model below, the firm creates the market and then monopolizes it.

1. Assumptions

Using Figure 1, the firm invests in the first period creating the market. In the second period, the firm takes an action that monopolizes the market. As in the previous model, the second period action could generate an efficiency gain. When the monopolizing firm creates the market, it generates $S = T + D + W$. If the firm is deterred from monopolization, consumers get S. If the firm is not deterred from monopolization, consumers get W.

For example, suppose that in the first period the firm invests in the design and production of a new artificial tooth that will be ready to market in the second period. The tooth design can be copied by rivals easily, so the second period market has the potential to be perfectly competitive. However, the firm can reduce competitive pressure, and thereby appropriate part of the innovation return, by engaging in an exclusionary act at the start of the second period.[18] The ideal exclusionary act would be the attainment of a legal barrier to entry, such as a patent, but such options may not be available to the firm or may not be effective.[19] Suppose the firm's best option for appropriating some of the surplus from innovation is to enter into an exclusive dealing contract with a key resource supplier,[20] and that in addition to excluding competition the contract reduces supply costs (say, by permitting the resource supplier to better predict demand).[21] The returns from the creation of the new artificial tooth depend on the firm's later success in excluding competition.[22] It will have an incentive to monopolize if the gains from monopolization exceed expected antitrust penalties.

[18] The exclusionary or monopolizing act could take an infinite number of forms, many of them seemingly innocent. For example, even the decision to keep rival firms from learning about a new product under development has an exclusionary effect, because it allows the innovator to take advantage of lead-time. *See, e.g.*, Berkey Photo, Inc. v. Eastman Kodak Co., 603 F.2d 263 (2d Cir. 1979) (in which the defendant (Kodak) was sued under the Sherman Act for not pre-disclosing information on a new type of camera film it had created. The court held in favor of Kodak on the grounds that a pre-disclosure duty would weaken incentives to innovate.).

[19] If a patent or exclusive license were available (and effective), the analysis here would still apply—the optimal penalty should then be understood as the optimal fee for the patent or the license. Thus, whether or not legal barriers can be used, the optimal penalty result still provides a useful instrument for regulating the firm's conduct. As for the effectiveness of patents, Wesley M. Cohen, Richard R. Nelson, and John P. Walsh find that patents are "least emphasized" among manufacturing firms in comparison to other methods of appropriating the returns from innovation. The general reasons for this could be (1) legal restrictions on availability (novelty, nonobviousness, naturally occurring substances, and so forth), and (2) difficulty in enforcement. Wesley M. Cohen, Richard R. Nelson & John P. Walsh, *Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)* (NBER, Working Paper No. 7552, 2000).

[20] This hypothetical is based on the facts of United States v. Dentsply Int'l, Inc., 399 F.3d 181 (3d Cir. 2005). However, the exclusive dealing contract in *Dentsply* was with dealers rather than suppliers. An example involving an exclusivity contract with a supplier was Alcoa's contract to purchase electricity on the condition that the seller refrain from selling electric power to any other producer of aluminum. United States v. Aluminum Co. of Am. (Alcoa), 148 F.2d 416, 422 (2d Cir. 1945).

[21] The tying contract in Int'l Salt Co. v. United States, 332 U.S. 392 (1947) was explained by John Peterman as a device that permitted the company to better predict, and thereby take advantage of economies in, the distribution of salt. John Peterman, *The International Salt Case*, 22 J.L. & ECON. 351 (1979).

[22] Again, this is true whether the firm uses a legal barrier or self help. We treat the innovation concept as something given to the innovator and examine the incentive to carry it out. This is distinguishable from the innovation race in which firms attempt to develop a new technology at the same time, in which case perpetual rivalry might, or might not, enhance innovation. *See* Glenn C. Loury, *Market Structure and Innovation*, 93 Q. J. ECON. 395 (1979); Tom Lee & Louis L. Wilde, *Market Structure and Innovation: A Reformulation*, 94 Q. J. ECON. 429 (1980); Philippe Aghion, Nicholas Bloom, Richard Blundell, Rachel Griffith & Peter Howitt, *Competition and Innovation: An Inverted-U Relationship*, 120 Q. J. ECON. 701 (2005). One can distinguish the incentive to search for innovations and the incentive to carry a particular innovation concept out. Competition may reduce the incentive to carry out a particular concept and increase the incentive to search for an innovation.

In the remainder we will focus on the optimality condition for the penalty. Becker's suggestion that the probability of enforcement should be set at its minimum level (under reasonable assumptions) has not been disturbed in the subsequent literature (for example, that by Polinsky and Shavell),[23] so there is no need to reexamine this issue.

2. Optimal Punishment: Offender Investment

The potential offender will invest if the expected return from monopolization, net of the penalty, exceeds his investment cost. Let the investment cost, k_o, be governed by the probability distribution Ψ with corresponding density ψ. The potential offender invests when $k_o < \bar{k}_o = (1-H(zF))[E(M \mid M > zF) - zF]$, and the probability of investment is $\Psi(\bar{k}_o)$.

The problem for the social planner is to choose the optimal fine to maximize the net social benefit:

$$\begin{aligned} NB = \Psi(\bar{k}_o)\{&[(1 - H(zF))E(M|M > zF) - E(k_o|k_o < \bar{k}_o)] \\ &+ (1 - H(zF))(S - v - zc) + H(zF)S\}, \end{aligned} \tag{2}$$

where the first term (bracketed) is the net gain from investment to the would-be monopolist, the second term is the net gain to society if investment is followed by monopolization ($S - v = S - T - D = W$), and the third term is the net gain to society if investment is not followed by monopolization. The first order condition with respect to the fine is

$$\begin{aligned} \frac{\partial NB}{\partial F} = {}&(v + zc - S - zF)\{\Psi(\bar{k}_o)h(zF) + \psi(\bar{k}_o)(1 - H(zF))^2\} \\ &- S\{\psi(\bar{k}_o)H(zF)(1 - H(zF)) - \Psi(\bar{k}_o)h(zF)\} = 0. \end{aligned} \tag{3}$$

Thus, the optimal fine satisfies

$$F^* = \frac{v}{z} + c - \frac{S}{z} + \frac{S}{z}\left(\frac{\Psi(\bar{k}_o^*)h(zF^*) - \psi(\bar{k}_o^*)H(zF^*)(1 - H(zF^*))}{\Psi(\bar{k}_o^*)h(zF^*) + \psi(\bar{k}_o^*)(1 - H(zF^*))^2}\right), \tag{4}$$

where the first two terms are the familiar static penalty,[24] which internalizes consumer harm and the enforcement cost. The remaining terms involve penalties or subsidies. The firm gets a subsidy of S/z (third term) for the potential

[23] If there is a fixed cost in setting the probability of apprehension, the optimal probability may be positive for sufficiently small values of the marginal enforcement cost. *See* Polinsky & Shavell, *supra* note 16. In addition, if there is a fixed upper limit on the fine, the optimal probability of apprehension will not necessarily be the minimum level. Risk aversion and the social interest in raising penalties in order to give prosecutors bargaining leverage are additional factors that could be considered. *See* Wallace P. Mullin & Christopher M. Snyder, *Should Firms Be Allowed To Indemnify Their Employees for Sanctions?*, 26 J.L. ECON. & ORG. 30 (2010). We do not explore these variations on the Becker framework here.

[24] Landes, *supra* note 3.

surplus it delivers to consumers when it creates a new product market. The surplus is divided by the probability of apprehension because the firm gets no subsidy from the state in cases in which it is not apprehended. The last term targets the investment incentive, and it is a combination of subsidy and penalty. The first component of the numerator reflects a penalty to discourage monopolization, the second a subsidy to encourage investment.

Overall, the subsidy effect dominates in the last two terms of (4). Simplifying, we have

$$F^* = \frac{v}{z} + c - \frac{S}{z}\left(\frac{\psi(\bar{k}_o^*)(1 - H(zF^*))}{\Psi(\bar{k}_o^*)h(zF^*) + \psi(\bar{k}_o^*)(1 - H(zF^*))^2}\right). \tag{5}$$

Substituting terms from Figure 1, the optimal penalty can be expressed as

$$F^* = (1 - \theta)\left(\frac{T + D}{z}\right) + \theta\left(\frac{-W}{z}\right) + c, \tag{6}$$

where θ is the last bracketed term in (5), and is a discontinuous function of F with the properties $\theta > 0$; $\theta = 1$ for $F^* \leq 0$; and $\theta'(F^*) > 0$ for $F^* > 0$.[25] This implies the following policy:

Proposition 2: Under the optimal enforcement policy, the penalty is the sum of two components: one, a weighted average of the penalty that internalizes consumer harm, $(T + D)/z$, and the investment subsidy $-W/z$, and two, the enforcement cost c, where the probability of enforcement z is set at the minimum level.

First, unlike the static scenario examined previously, the complete deterrence policy is no longer sensible. When the monopolist's investment makes the product market available, its conduct always provides some benefit to society, and it would therefore be suboptimal to set the penalty with the aim of completely eliminating the firm's gain.

Second, because the subsidy weight $0 < \theta \leq 1$, the optimal penalty is a two-part fine that (1), internalizes the cost of enforcement to the offender, and (2), regulates the monopolization and investment decisions with a component that is a convex combination of the static penalty and an investment subsidy. Because the subsidy weight is greater than zero, the *optimal penalty is unambiguously less than the static penalty that internalizes consumer harm*.

Third, comparative statics for the optimal penalty differ from the static case. Consider the behavior of the optimal penalty as consumer harm increases. Unlike the static case, *the optimal penalty does not go to infinity as the consumer*

[25] The properties of the optimal penalty are examined in the appendix, specifically in the proof of Proposition 2. The optimal penalty ($F^* = v/z - (S/z)\theta + c$) is equal to the net marginal social harm from the offender's conduct, which is the difference between the harm to consumers and the marginal innovation benefit. The innovation benefit is itself a function of the size of the penalty.

harm goes to infinity (see Proposition 2a in the appendix). The optimal penalty increases with consumer harm to a limit consistent with sustaining investment. Moreover, *a decline in the probability of apprehension does not necessarily cause the optimal fine to increase* (see Proposition 2b in the appendix). The reason is that, as the probability of apprehension decreases, the optimal subsidy weight could increase, because it may be socially preferable to encourage investment more. It is not necessarily desirable to combine an extremely low probability of apprehension with an extremely high fine.

The subsidy weight θ captures the relative sensitivities of the firm's innovation-investment and monopolization decisions with respect to the penalty. When

$$\frac{h(zF^*)}{1-H(zF^*)} > \frac{\psi(\bar{k}_o^*)H(zF^*)}{\Psi(\bar{k}_o^*)}, \tag{7}$$

the monopolization elasticity (with respect to the penalty) exceeds the investment elasticity, and the subsidy weight $\theta < 1$. The penalty is relatively large because its shadow price, the discouragement of innovation, is relatively low. When the inequality in (7) is reversed or replaced with a strict equality, the subsidy weight is equal to one, yielding a pure subsidy in place of a penalty.

Note that the optimal penalty formula can be expressed as $F^* = (1-\theta)[(T+D)\ /\ z + c] + \theta(-W/z + c)$, which is a weighted average of the optimal static penalty (internalizing consumer harm and enforcement cost) and the optimal innovation subsidy (internalizing residual consumer surplus and administrative cost of award process). When the marginal social harm, $v + zc$, exceeds the surplus S (or, equivalently, when the expected enforcement cost, zc, exceeds the residual surplus W) the optimal penalty is unambiguously positive. This makes sense because the loss from discouraging investment is relatively small under these conditions, though the enforcement policy is still lenient relative to the static model even in this worst-case scenario. When the surplus is greater than the marginal harm (equivalently, residual surplus exceeds expected enforcement cost), the optimal penalty can be positive (a penalty) or negative (a subsidy), depending on whether the surplus gained from deterring monopolization is greater than the wealth generated from investment (see Proposition 2 the appendix). Punishment becomes more severe, holding other factors the same, as the residual surplus falls and as the relative sensitivity of innovation to the penalty falls.

The key policy implication is that it is not necessarily optimal to impose a penalty for monopolization that internalizes consumer harm—such a penalty may excessively deter innovation. Indeed, it may be optimal to subsidize rather than punish the monopolizing firm.[26] One function of the fine is to align

[26] It follows also that it may not be optimal to punish the firm even when there are no static efficiencies resulting from the monopolizing conduct. The model assumes the existence of

private and social incentives for innovation. If the firm monopolizes the market in the second period with probability 1, the social gain from the firm's first-period investment would be the sum of the residual consumer surplus and the monopoly transfer. The private gain, however, would be the monopoly transfer. Optimal punishment policy trades off deterrence of monopolization with equalizing the private and social gains from investment, and the latter goal requires a bounty based on the residual surplus.

Although we have focused on monopolizing conduct that might violate the antitrust laws, the model applies equally to patents (or to any method of rent appropriation). The penalty result here provides the optimal fee that should be charged to a patentee, which could be a fine or a prize, depending on the factors shown in (6).

3. *General Applications: Products Liability and Efficiency Defenses*

The previous part examined the offender investment model in the antitrust setting, under strict liability. In this part, we consider other applications, and the rule of reason test for liability.

Consider product safety regulation. The firm invests in the first period, creating the market. In the second period, it decides whether to take care to avoid imposing an injury v on the consumer.[27] It takes care only if the cost of taking care M is less than the expected fine zF. The firm's investment in the first period is a function of anticipated profit in the second, which is determined in part by the relationship between M and zF. The optimal penalty, $F^* = v/z - (S / z)\theta + c$, where θ depends on the relative elasticities of precaution and innovation, compromises internalization in order to encourage innovation. This is not a weighted average of consumer welfare components, as in the monopolization case examined earlier, because v does not have any necessary relationship to S. The optimal product safety penalty is unambiguously less than the one that internalizes the consumer injury, and it could be a pure subsidy.

Suppose that, instead of strict liability, the punishment authority operates under a rule of reason, which imposes punishment only when $M < v$.[28] The rule of reason test permits defendants to be exempted from punishment on the

---

static efficiencies. But the result that it may be optimal to reduce the penalty in order to encourage innovation applies just as well to the case where there are no static efficiencies. Thus, even if an exclusivity contract offers no cost advantage whatsoever, and serves the sole purpose of excluding competitors, it still may not be optimal to punish the firm for monopolization.

27 This discrete choice assumption simplifies matters and is consistent with the products liability case law. Every products liability design-defect case involves an allegation concerning a discrete decision to adopt a precaution, such as installing a safety bar on a hazardous machine.

28 In antitrust, this test is equivalent to a rule of reason test that permits efficiency defenses ($M > v$ is equivalent to the case where the efficiency gain exceeds the deadweight loss, $E > D$). Under a perfect-information rule of reason antitrust regime, there would be no need for antitrust enforcement, because all inefficient cases of monopolization would be deterred.

basis of efficiency defenses. This case has not been considered in the economic treatments of punishment.[29]

Under the reasonableness rule, the social objective would be to set the fine to maximize

$$\begin{aligned} NB = \Psi(\bar{k}_o)\{(1 - H(v))E(M|M > v) + [(H(v) \\ - H(zF))E(M|v > M > zF) - E(k_o|k_o < \bar{k}_o)] \\ + (1 - H(zF))(S - v - zc) + H(zF)S\}, \end{aligned} \tag{8}$$

where $\bar{k}_o = (1 - H(v))E(M \mid M > v) + (H(v) - H(zF))[E(M \mid v > M > zF) - zF]$. This objective function incorporates the facts that when $M > v$, the punishment authority will apprehend but not punish the offender; when $v > M > zF$, the offender will be punished but will not be deterred from the offense; and when $zF > M$, the offender will be deterred from committing the offense. The optimality condition for (8) implies that the optimal fine is smaller (greater) than the static penalty ($v/z + c$) when the surplus is large (small) relative to the marginal social harm.[30] The reason for enhancing the penalty (beyond the static level) is to discourage some investment, given that some offenders will not be punished under the reasonableness rule even though they have forced society to bear the costs of apprehending them. The subsidy provided by the optimal penalty is smaller because the reasonableness rule already exempts some potential offenders from any punishment. The static penalty comes closer to being socially optimal when the regulator punishes according to the reasonableness standard than when he punishes under the strict liability rule.

4. *Victim Investment*

Potential victims invest in some activity, at cost k_v. The gross gain from investment to the victim, if there is no offense, is B. However, because of the risk that an offense will destroy the value of the investment, the victim's expected gain is $[1\text{–}prob(offense)]B$. The victim suffers a direct loss v in the event an offense occurs. We allow v to differ from B because it is possible that the offense both destroys the value of the victim's investment and imposes a different direct loss on the victim.

Returning to the terms introduced in earlier, the expected private gain from investment is $H(zF)B$. Suppose k_v has probability distribution R, with corresponding density function r, $r(k_v) > 0$ for $k_v > 0$ and $r(k_v) = 0$ otherwise. The potential victim invests whenever $k_v < \bar{k}_v = H(zF)B$, so the probability he

[29] Becker, *supra* note 2; Polinsky & Shavell, *supra* note 16. In the static scenario, the most interesting feature of the rule of reason is that it enables the enforcement authority to avoid any enforcement expenditure at all. The offenders for whom $M > v$ would not be deterred by the threat of punishment (because they are exempted by the test), and those for whom $M < v$ would be deterred. There would never be a need for enforcers to act.

[30] $zF^* = v + zc - \frac{\psi(\bar{k}_o^*)(H(v) - H(zF^*))(S - (1 - H(v))(v + zc))}{\psi(\bar{k}_o^*)(H(v) - H(zF^*))^2 + \Psi(\bar{k}_o^*)h(zF^*)}$.

invests is $R(\bar{k}_v)$. The expected net benefit from investment is therefore

$$R(\bar{k}_v)E(\bar{k}_v - k_v | k_v < \bar{k}_v). \tag{9}$$

The investment benefit from increasing the fine is equal to the product of the marginal reduction in the probability of an offense, $zh(zF)$, and the expected gain among the pool of potential investors RB. Increasing the fine, through its deterrent effect, allows more of those who invest to realize their returns without seeing them destroyed by the offender. The expected net benefit from investment is maximized when the penalty is set at a level that eliminates the offender's gain, $F \geq M_u / z$, and minimized when the penalty is set at zero.

Of course, a social planner would not set out solely to maximize the expected net benefit from investment. The social objective is to maximize the net benefits from enforcement, which is the sum of the net benefit from investment and the net benefit from offenses, given enforcement:

$$\begin{aligned} NB &= NB(\mathit{investment}) + NB(\mathit{offenses}) \\ &= R(\bar{k}_v)(\bar{k}_v - E(k|k < \bar{k}_v)) \\ &\quad + (1 - H(zF))[E(M|M > zF) - (v + zc)]. \end{aligned} \tag{10}$$

Proposition 3: Let NB^* represent the value of the net benefit from enforcement under the optimal policy, and let $\overline{NB}$ represent the value of the net benefit from enforcement when offenses are completely deterred. Let $\bar{M} > v + z'c + R(\bar{k}_v^*)B$ denote the value of M_u such that $NB^* = \overline{NB}$. Then if $M_u > \bar{M}$, the optimal punishment policy is to set the penalty and probability of apprehension so that $F^* = \frac{v}{z'} + c + \frac{R(\bar{k}_v^*)B}{z'}$. If $M_u \leq \bar{M}$, then the optimal policy is to set the fine and the probability of apprehension so that $zF \geq M_u$.

The optimal penalty internalizes the direct loss, the enforcement cost, and the investment return forgone due to the fear of offenses. Here the circumstances under which a complete deterrence policy is optimal are broader than in the static case. If the maximum gain to offenders is less than the marginal social cost of an offense, complete deterrence is optimal, as in the static case. However, even if the maximum gain exceeds the marginal social cost of an offense, complete deterrence may be optimal, because the gain is insufficient to compensate for the cost of reduced investment.

The case in which offensive conduct discourages investment by potential victims was first considered by Jeremy Bentham, who referred to the "secondary effects" of criminal behavior.[31] Bentham noted that offensive conduct led to primary and secondary harms to society; where primary harms are the direct and derivative losses, as well as enforcement costs, and secondary harms are the costs that result from discouraged investment and extend "either over the

[31] Jeremy Bentham, The Principles of Morals and Legislation 153 (Prometheus Books 1988) (1789).

whole community, or over some multitude of unassignable individuals."[32] Secondary effects could include a range of costs generated by changes in behavior resulting from fear of crime. Bentham argued that punishment should be enhanced to internalize secondary costs. The penalty formula derived here formalizes Bentham's recommendation.

D. Headline Effects and Penalties

News headlines alter the investment decisions of offenders or victims. The headlines may lead the offender to believe that the likelihood of apprehension is greater than it is. To model headline effects, let the perceived probability of apprehension differ from the real probability of apprehension: $\tilde{z} = z(1+\mu)$. The optimal policy is now

$$F^* = \frac{1}{1+\mu}\left\{\frac{v}{z'} + c - \frac{S}{z'}\left(\frac{\psi(\bar{k}_o^*)(1 - H(z'(1+\mu)F^*))}{\Psi(\bar{k}_o^*)h(z'(1+\mu)F^*) + \psi(\bar{k}_o^*)(1 - H(z'(1+\mu)F^*))^2}\right)\right\}, \tag{11}$$

which suggests that the fine should be reduced to compensate for the offender's overestimate of the likelihood of punishment ($\mu > 0$ case). However, because the sign of the portion of the penalty regulating the investment decision is ambiguous, offender overestimation of the likelihood of apprehension could raise or lower the optimal penalty. An overestimate of the probability of punishment makes the punishment for monopolization seem more likely, but the associated change in investment incentives could lead to a reduction in the optimal subsidy component.

Now suppose the victim invests while relying on news headlines to predict the likelihood that he will reap the rewards. The break-even cost level for the victim is $\bar{k}_v = (1+\eta)H(zF)B$, where $\eta > 0$ means that the victim underestimates the likelihood of an offense that destroys his investment. The optimal penalty satisfies

$$F^* = \frac{v}{z} + c + \frac{R(\bar{k}_v^*)B - \eta(1+\eta)B^2 r(\bar{k}_v^*)H(zF)}{z}, \tag{12}$$

implying that the penalty should be reduced when the victim underestimates the likelihood of an offense, and increased when the victim overestimates the likelihood. The reason for reducing the penalty when the victim underestimates is to align private and social incentives to invest. If the victim thinks that there will not be an offense, he will invest too much in light of the return. The penalty is reduced in order to indirectly diminish the investment incentive.[33]

[32] *Id.*

[33] A similar problem is encountered in the context of crime and victim precaution. The optimal fine varies in order to control the incentives of both offender and victim; *see* Keith N. Hylton, *Optimal Law Enforcement and Victim Precaution*, 27 RAND J. ECON. 197 (1996).

The parameter η is an index of the alarm-to-danger ratio identified by Bentham.[34] If η is equal to zero, the danger and alarm caused by offensive activity are the same; the impressions potential victims get from reading news headlines are accurate indicators of the likelihood of an offense. If η is positive, the alarm is less than the danger, and if η is negative the alarm is greater than the danger. The optimal penalty, consistent with Bentham, implies that the fine should increase as the alarm increases relative to the danger, in order to internalize the negative investment effect.

III. DISCUSSION

A. Antitrust

Because the normative implications of this model have been mentioned in the course of its presentation, we will focus on positive implications for antitrust law here. The offender investment model suggests that the social payoff from innovation antecedent to or associated with monopolization should be part of rule of reason analysis under the Sherman Act. The law appears to reflect this recommendation already. Monopolies are not illegal *per se*.[35] Antitrust law immunizes firms from liability when they have acted merely as profit-maximizing monopolists (for example, by setting the monopoly price).[36] Liability is imposed under section 2 for predatory conduct and efforts to exclude rivals. Exploitative conduct is distinguished from exclusionary conduct.

Although antitrust law is underinclusive in comparison with the optimal penalty model presented here, the exemption provided to firms that merely exploit their market power rather than exclude rivals can be understood as an attempt by the law to accommodate the welfare gains from innovation. One paradox of antitrust, stressed in Judge Hand's *Alcoa* opinion, is that cartel pricing is *per se* illegal, whereas monopoly pricing is *per se* lawful.[37] These basic rules are not contradictory under our model.

[34] BENTHAM, *supra* note 31.

[35] The legal standard is described as a balancing test that compares anticompetitive harms with procompetitive benefits; *see* United States v. Microsoft, 253 F.3d 34, 59 (D.C. Cir. 2001).

[36] In *Alcoa*, Judge Learned Hand distinguished monopolies that are passively acquired, and monopolies acquired through superior skill, foresight, and industry, from monopolies that are actively acquired. *See* United States v. Aluminum Co. of Am. (Alcoa), 148 F.2d 416, 429–31 (2d Cir. 1945); Dennis W. Carlton & Ken Heyer, *Extraction vs. Extension: The Basis for Formulating Antitrust Policy Towards Single-Firm Conduct*, 4 COMPETITION POL'Y INT'L 285 (2008). Carlton and Heyer propose the distinction between extraction and extension as a normative guideline for monopolization law. This model distinguishes innovation (creation) from extraction. The distinctions are not the same. There are cases that could be described as extension (tying, exclusive dealing) where punishment would have to be moderated in light of investment incentive effects under this model.

[37] The paradox served as a key justification for Judge Learned Hand's reform of monopolization law in *Alcoa*. Hand argued that monopolies (monopoly pricing) should be viewed in the same way as cartels (cartel pricing). *Alcoa*, 148 F.2d. at 427–29. For a discussion of innovation and its implications for Hand's *Alcoa* argument, *see* Evans & Hylton, *supra* note 6.

For firms that engage in exclusionary conduct, the law is overinclusive, in the sense that it does not reduce expected penalties to compensate for the creation or expansion of markets through innovation. One exception is *United States v. Jerrold Electronics Corp.*,[38] where the court held that rule of reason rather than *per se* liability applied to tying policies that were instituted at a time when the market for antennae systems was in its infancy. Jerrold had played a central role in the creation of the market. The court exempted Jerrold from antitrust liability for the infancy period of its business, effectively a penalty reduction to compensate for innovation. Our model implies that the doctrine of *Jerrold Electronics* should be incorporated generally into monopolization law.[39]

Of course, a symmetrical and opposing conclusion applies where the dominant firm's conduct discourages innovation by rivals. The rule of reason should take into account evidence that the monopolist's conduct reduced innovation by potential competitors.[40]

B. Torts

Now consider implications for tort law. The damage multiplier approach[41] suggests that the optimal tort damage award will divide the harm by the probability of liability. However, this is inadequate as a method of internalizing social costs when there are negative investment effects from offenses. Conversely, the multiplier approach inefficiently overdeters when the offender's investment yields a positive externality and is dependent on profits from a later action that may cause harm.

[38] United States v. Jerrold Electronics Corp., 187 F. Supp. 545 (E.D. Pa. 1960), *aff'd per curiam*, 365 U.S. 567 (1961).

[39] Jonathan Baker, *Beyond Schumpeter vs. Arrow: How Antitrust Fosters Innovation*, 74 ANTITRUST L.J. 575 (2007). Baker argues that enforcement authorities are sensitive to innovation effects in their targeting decisions. If enforcement authorities targeted only those cases in which the monopolization effect is substantially greater than the investment effect (θ close to zero), then antitrust enforcement would be virtually optimal, as Baker contends. This argument assumes a best-case scenario for enforcement priorities, which seems implausible in view of the fact that enforcement priorities vary according to the type of administration in office (anti-business versus pro-business).

[40] The offender-investment and victim-investment models can be combined to yield an optimal penalty formula that balances opposing externalities. In the combined model, the penalty would be reduced relative to the benchmark static penalty to the extent that punishment reduced the social gain from investment by the offender, and enhanced to the extent that punishment of the offender encouraged potential victims to invest. The net direction would be ambiguous *a priori*, but could be simulated under parameter assumptions. Ilya Segal & Michael D. Whinston, *Antitrust in Innovative Industries*, 97 AM. ECON. REV. 1703 (2007). Segal and Whinston suggest that a policy that protects victims would be preferable because it would frontload profits to new innovators (entrants). These considerations complicate the analysis and raise questions about the ability of an enforcement authority to implement a policy free from error.

[41] A. Mitchell Polinsky & Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 HARV. L. REV. 869 (1998).

Suppose that the firm invests in a new vaccine in the first period. In the second period, when the product is on the market, it can adopt some precaution to reduce the likelihood of harm to consumers. The precaution could take many different forms: enhancing the warning label, or better monitoring of the production process. The firm decides in the second period whether to take the precaution by comparing the cost of precaution to the expected fine. This description of the vaccine marketing is analogous to the investment-monopolization model examined earlier. It follows that the optimal penalty will depend on several factors: the harm to the consumer, the positive externality to society (consumer surplus from innovation and externalities from vaccination), and the degree to which an increase in the penalty affects the investment incentive versus the precaution incentive.

This implies that strict products liability may not be optimal in the innovation setting, or in a setting in which the firm's product yields beneficial externalities (for example, vaccines). If the injuries caused by the product are not large in relation to the surplus created, the negligence standard may be preferable to strict liability.[42]

IV. CONCLUSION

Using a model of punishment with monetary penalties, we have examined the design of optimal penalties in settings where agents make investments. The key scenario examined is that in which the offender invests in an activity that benefits society, and the private return to that activity is a function of the offense he later commits, as well as the penalty. The optimal policy strikes a balance between internalizing the costs of the offender's conduct and subsidizing the offender's investment. In the monopolization setting, the optimal penalty is a weighted average of a penalty that internalizes the consumer harm and a subsidy that internalizes the consumer benefit created by investment, with the weights depending on the relative elasticities of investment and monopolization with respect to the penalty. Under certain conditions, the subsidy component may dominate the penalty, generating a reward for monopolization. Although rewarding a monopolizing firm seems counterintuitive and inconsistent with

[42] In the case of drugs, the law shows some signs of incorporating this implication, though the record is mixed. Comment k of Restatement (Second) of Torts, Section 402A, suggests that courts should exempt drug makers from strict liability for product defects (and defective designs). *See, e.g.,* Brown v. Superior Court, 44 Cal. 3d 1049 (1988) (in which the California Supreme Court held that comment k insulated all Food and Drug Administration-approved prescription drugs from strict liability for design defect. The Court reasoned that strict liability would deter innovation of new drugs. The California courts later applied the same reasoning to implanted prescription medical devices.). *See also* Artiglio v. Superior Court, 22 Cal. App. 4th 1388, 27 Cal. Rptr. 2d 589 (1994); Plenger v. Alza Corp., 11 Cal. App. 4th 349, 13 Cal. Rptr. 2d 811 (1992). However, *Brown* has been adopted in only a minority of U.S. states. To some extent, courts have moved in the direction of a negligence framework by embracing risk-utility analysis.

the static enforcement model, the law on monopolization—particularly its puzzling leniency—is best explained by a theory that views encouraging innovation as one of the implicit goals of antitrust policy.

APPENDIX

Proof of Proposition 1: The social planner's problem is to choose z and F to minimize

$$C = H(zF)(E(M|M \leq zF)) + (1 - H(zF))(v + zc)$$

$$= \int_0^{zF} Mh(M)dM + (1 - H(zF))(v + zc).$$

The first-order conditions are:

$$\frac{\partial C}{\partial F} = -zh(zF)(v + zc - zF)$$

$$\frac{\partial C}{\partial z} = -Fh(zF)(v + zc - zF) + (1 - H(zF))c.$$

Note that when z^* and F^* are chosen so that z^*F^* is greater than M_u, then the offender is completely deterred and the above equation equals $E(M)$. We discuss the optimal choice of enforcement rate and penalty below.

$$v + z'c \geq M_u \tag{1}$$

This is the case where the minimum cost from an offense is higher than the maximum benefit to the offender. In this case, it is optimal to eliminate offenses by setting z^*F^* greater than M_u. Here is the proof.

$$\begin{aligned} C &= H(zF)(E(M|M \leq zF)) + (1 - H(zF))(v + zc) \\ &\geq H(zF)(E(M|M \leq zF)) + (1 - H(zF))M_u \\ &> H(zF)(E(M|M \leq zF)) + (1 - H(zF)) \\ &\times (E(M|M \geq zF)) = E(M) \end{aligned}$$

In this scenario, $E(M)$ is the lower bound for social cost, as in Figure A1. Given a specific enforcement rate z, $F^* \in \{F : F \geq M_u/z\}$.

$$v + c < M_u \tag{2}$$

Given a specific value of z, let $F^* = \frac{v}{z} + c$ and $F^L = \frac{M_u}{z}$. Note here that $F^* < F^L$. When $F < F^*$; C is decreasing and when $F > F^*$, C is increasing

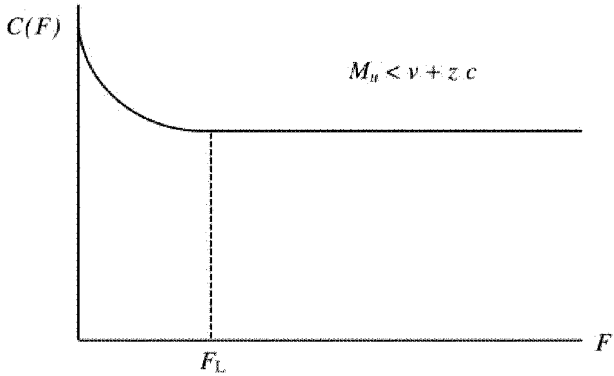

Figure A1. Social cost as a function of the penalty

until it reaches $E(M)$ for $\forall F \geq F^L$. Figure A2 shows the relationship. Thus for a given z, the optimal choice of F is F^*.

Moreover, given F^*, C is increasing in z so that it is optimal to set $z^* = z'$. Consequently, the optimal policy is $\{ z^* = z', F^* = \frac{v}{z'} + c \}$ and the corresponding social cost is $C^* = H(v + z'c)(E(M \mid M \leq v + z'c)) + (1 - H(v + z'c))(v + z'c)$. Note that C^* is smaller than $E(M)$.

$$v + z'c < M_u \leq v + c \tag{3}$$

Using the same logic, we know that given a specific value of z, the optimal F is $F^* = \frac{v}{z} + c$. Depending on the value of z, now F^* may be smaller or bigger than F^L. As for $\forall F \geq F^L$, C remains at the level $E(M)$ so it would be better to have $F^* < F^L$ so that we could reach the cost level less than $E(M)$. Thus in this scenario, the optimal policy is again reached when z is set at its minimal level and F equals to F^*, namely, $\{z^* = z', F^* = \frac{v}{z'} + c \}$. ■

Proof of Proposition 2: The proof follows in three steps.

First, we show that it is never optimal to set $z^* F^*$ greater than M_u. If $z^* F^* \geq M_u$, the offender will not make any investment and there will be no social gain. Instead, for any $z^* F^* < M_u$, we observe a positive social benefit as

$$\begin{aligned} NB &= \Psi(\bar{k}_o^*)\{(1 - H(zF^*))E(M|M > zF^*) \\ &\quad - \bar{k}_o^* + (1 - H(zF^*)(S - v - zc) + H(zF^*)S\} \\ &> \Psi(\bar{k}_o^*)\{(1 - H(zF^*))(zF^* + S - v - zc) + H(zF^*)S\} \\ &= \Psi(\bar{k}_o^*)\frac{\Psi(\bar{k}_o^*)h(zF^*)}{\Psi(\bar{k}_o^*)h(zF^*) + \psi(k^*)(1 - H(zF^*))^2}S > 0. \end{aligned}$$

Second, we need to prove that the optimal penalty z^* is the unique solution to the global maximum. The first order condition (equation 3 in the article) can

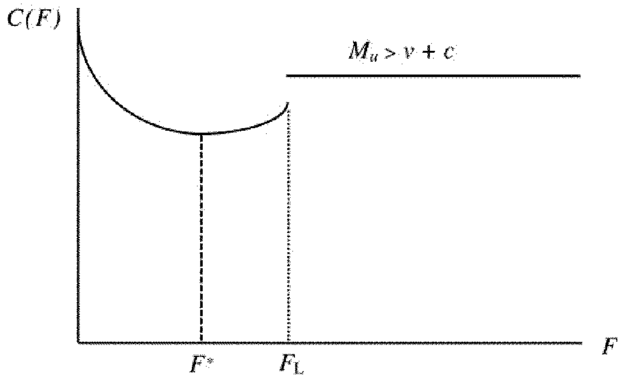

Figure A2. Social cost as a function of the penalty

be simplified as follows:

$$\frac{\partial NB}{\partial F} = (v + zc - zF)\{\Psi(\bar{k}_o)h(zF) + \psi(\bar{k}_o)(1 - H(zF))^2\} - S\{\psi(\bar{k}_o)(1 - H(zF))\} , \quad (2A)$$
$$= \frac{P}{z}\{\frac{v}{z} + c - F - \frac{S}{z}\theta\}$$

where $P = \Psi(\bar{k}_o)h(zF) + \psi(\bar{k}_o)(1 - H(zF))^2$ and $\theta = \frac{\psi(\bar{k}_o)(1 - H(zF))}{P}$. Because $P > 0$, if there exists F^* such that

$$F = \frac{v}{z} + c - \frac{S}{z}\theta(F), \quad (2B)$$

then we will have

$$\left.\frac{\partial NB}{\partial F}\right|_{F=F*} = 0.$$

To proceed, let us first take a look at θ, which is a function of F. When $F \leq 0$, $\theta = 1$; when $F > 0$, θ is strictly increasing in F because (1) $\frac{\Psi(\bar{k}_o)}{\psi(\bar{k}_o)}$ is increasing in $\bar{k}_o$ and $\bar{k}_o$ is decreasing in F; (2) $\frac{h(zF)}{(1-H(zF))}$ is decreasing in F; (3) $(1 - H(zF))$ is decreasing in F; four, both $\frac{\Psi(\bar{k}_o)}{\psi(\bar{k}_o)}$ and $\frac{h(zF)}{(1-H(zF))}$ are positive.

Let θ_{min} be the value of θ when F is approaching zero. It is straightforward that $\theta_{\min} = \lim_{F\to 0} \theta = \frac{\psi(\bar{k}_o)}{\Psi(\bar{k}_o)h(zF)+\psi(\bar{k}_o)} < 1$. [The value of $\theta_{\min}$ will depend on the shape of the distributional forms of $\psi(\cdot)$ and $H(\cdot)$. For example if we assume that both of them take an exponential form with mean of $1/\lambda^{\Psi}$ and $1/\lambda^{H}$, then $\theta_{\min} = \frac{1}{(\lambda^{H}/\lambda^{\psi})(e^{\lambda^{\psi}/\lambda^{H}}-1)+1} < \frac{1}{2}$.]

Let us then consider equation (2B). The left-hand side (*LHS*) of equation (2B) is F itself, which could be treated as a strictly increasing function of F. The right-hand side (*RHS*) then is a nonincreasing function where

$$RHS = \begin{cases} \dfrac{v}{z} + c - \dfrac{S}{z} & \text{if } F \leq 0 \\ \dfrac{v}{z} + c - \dfrac{S}{z}\theta & \text{if } F > 0, \end{cases}$$

and the *RHS* reaches its maximum at $\frac{v}{z} + c - \frac{S}{z}\theta_{\min}$ when $F \to 0$. The following three cases are explained for the equality of the *LHS* and *RHS*.

Case 1: $\frac{v}{z} + c - \frac{S}{z} \geq 0$

It is obvious that in this case, $\frac{v}{z} + c - \frac{S}{z}\theta_{\min} > 0$. There exists a unique solution such that *LHS* = *RHS* and the unique equilibrium is $F^* = \frac{v}{z} + c - \frac{S}{z}\theta^* > 0$. The proof of the global maximum is quite straightforward: when $F < F^*$, $\frac{\partial NB}{\partial F}\big|_{F<F^*} > 0$; and when $F > F^*$, $\frac{\partial NB}{\partial F}\big|_{F>F^*} < 0$. This case corresponds to the graph in Figure A3 below.

Before moving to Case 2, we will briefly discuss two properties of the Case 1 solution.

Proposition 2a: As v goes to infinity, the optimal penalty converges to a constant F^{**} that satisfies $\theta(F^{**}) = 1$.

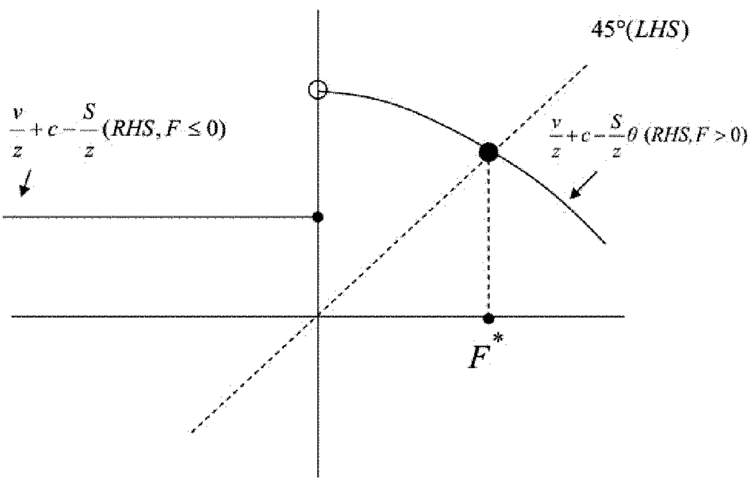

Figure A3. Case 1

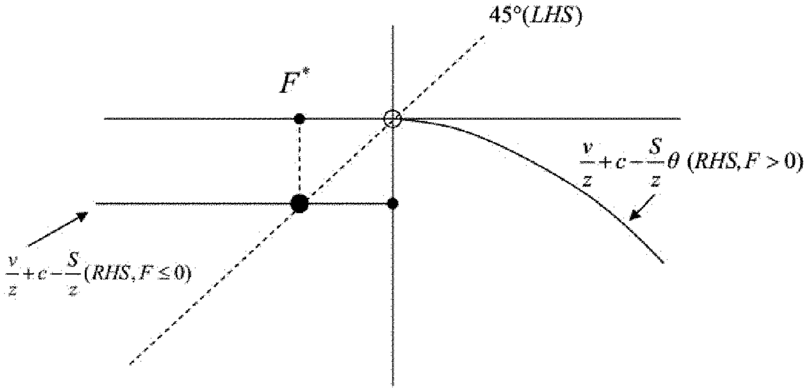

Figure A4. Case 2

Proof: Recall that the optimal penalty F^* satisfies:

$$RHS = \frac{v}{z} + c - \frac{S}{z}\theta(F^*) = c - \frac{w}{z} + \frac{(v+w)}{z}(1-\theta(F^*))$$
$$= F^* = LHS$$

First, it is easy to show that F^* is increasing in v. To see why, assume that v is increased to v' and the associated optimal penalty is $F^{*\prime}$. We need to show that $F^{*\prime} > F^*$. Assume otherwise that $F^{*\prime} \le F^*$, $\theta(F^{*\prime}) \le \theta(F^*)$, so that $RHS(F^{*\prime}) > RHS(F^*)$. However $LHS(F^{*\prime}) \le LHS(F^*)$ so that $RHS(F^{*\prime}) > LHS(F^{*\prime})$, a contradiction. Second, we need to show that when v goes to infinity, F^* does not goes to infinity. Otherwise if F^* goes to infinity, θ will also go to infinity. As a result, $RHS(F^*)$ goes to negative infinity, which contradicts the fact that LHS (F^*) goes to positive infinity. Last, because F^* is increasing in v and because F^* is finite as v goes to infinity, then it must be the case that θ goes to 1 so that the optimal penalty makes $RHS = LHS$. We have shown that θ is increasing in F, so that the optimal penalty is F^{**} which satisfies $\theta(F^{**}) = 1$.

Proposition 2b: When z is decreasing, F^* is ambiguous.

Proof: It is easy to show that when z is decreasing, zF^* is decreasing as well. To see why, because $v/z_0 + c - (S/z_0)\theta(z_0F_0^*) = F_0^*$, we have $v + z_0c - S\theta(z_0F_0^*) = z_0F_0^*$. Now z_0 is decreased to z_1 and $v + z_1c - S\theta(z_1F_1^*) = z_1F_1^*$. If $z_1F_1^* \ge z_0F_0^*$, we know that $S\theta(z_1F_1^*) \ge S\theta(z_0F_0^*)$ (as θ is increasing in zF) and so that $v + z_1c - S\theta(z_1F_1^*) < v + z_0c - S\theta(z_0F_0^*)$. However, if $z_1F_1^* \ge z_0F_0^*$, $v + z_1c - S\theta(z_1F_1^*) \ge v + z_0c - S\theta(z_0F_0^*)$ by the definition of the optimum. We see a contradiction here so that $z_1F_1^* < z_0F_0^*$. The remaining question is whether F^* is increasing or decreasing. Using the implicit function theorem, $\partial F^*/\partial z = [c - (1 + s\theta')F^*]/z(1+ s\theta')$, where $\theta' = \frac{\partial\theta}{\partial(zF)}$, and this depends on the comparison of c and $(1 + s\theta')F^*$, which further depends on the function θ, especially the value of θ' at F^*.

Case 2: $\frac{v}{z} + c - \frac{S}{z} < 0$ and $\frac{v}{z} + c - \frac{S}{z}\theta_{\min} < 0$

Again we obtain a unique optimum $F^* = \frac{v}{z} + c - \frac{S}{z} < 0$ for $LHS = RHS$. F^* is the global maximum as when $F < F^*$, $\frac{\partial NB}{\partial F}\big|_{F<F^*} > 0$; and when $F > F^*$, $\frac{\partial NB}{\partial F}\big|_{F>F^*} < 0$. This case corresponds to the graph in Figure A2.

Case 3: $\frac{v}{z} + c - \dfrac{S}{z} < 0$ and $\frac{v}{z} + c - \frac{S}{z}\theta_{\min} \geq 0$

There exist two F^* (one for $F^* < 0$ and the other for $F^* \geq 0$) such that $LHS = RHS$: $F_1^* = \frac{v}{z} + c - \frac{S}{z} < 0$ and $F_2^* = \frac{v}{z} + c - \frac{S}{z}\theta^* > 0$. It can be verified that both are local maximums because $\frac{\partial NB}{\partial F}\big|_{F<F_1^*} > 0$ and $\frac{\partial NB}{\partial F}\big|_{0 \geq F > F_1^*} < 0$; and $\frac{\partial NB}{\partial F}\big|_{0<F<F_2^*} > 0$ and $\frac{\partial NB}{\partial F}\big|_{F>F_2^*} < 0$. The solutions are illustrated in Figure A5. The global maximum is the F^* that provides the higher social benefit.

Consider the following comparison of the two F^* solutions:

$$F_1^* = \frac{v}{z} + c - \frac{S}{z} < 0$$

so that

$$\begin{aligned}
\bar{k}_o^1 &= \bar{M} - zF_1^* = \bar{M} + S - v - zc \\
NB_1 &= \Psi(\bar{k}_o^1)\{\bar{M} - E(k|k < \bar{k}_o^1) + (S - v - zc)\} \\
&= \Psi(\bar{k}_o^1)\bar{k}_o^1 - \int_0^{\bar{k}_o^1} k\psi(k)dk
\end{aligned}$$

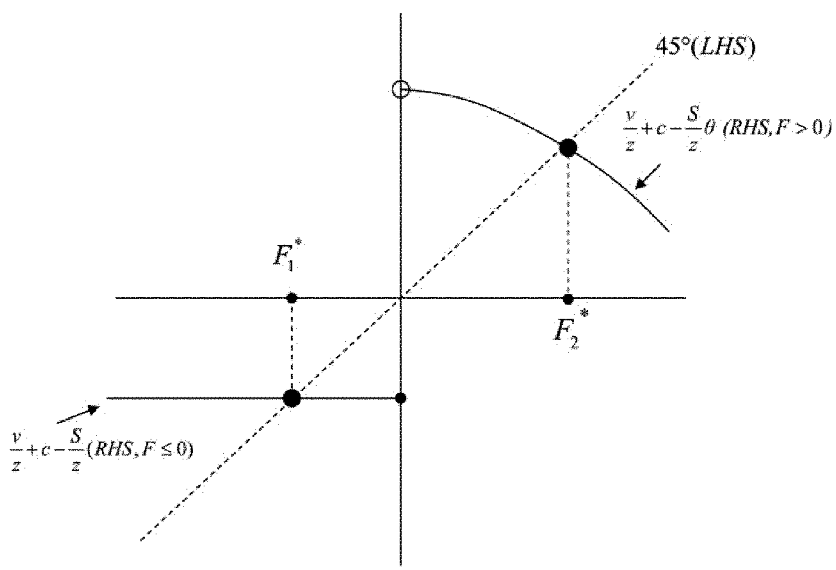

Figure A5. Case 3

$F_2^* = \frac{v}{z} + c - \frac{S}{z}\theta^* > 0$ so that $\bar{k}_o^2 = \int_{zF_2^*} Mh(M)dM - (1 - H(zF_2^*))zF_2^* < \bar{k}_o^1$

$$\begin{aligned} NB_2 &= \Psi(\bar{k}_o^2)\{(1 - H(zF_2^*))E(M|M > zF_2^*)\bar{M} - E(k|k < \bar{k}_o^2) \\ &\quad + (1 - H(zF_2^*))(S - v - zc) + H(zF_2^*)S \\ &= \Psi(\bar{k}_o^2)\{\bar{k}_o^2 + (1 - H(zF_2^*))zF_2^* - E(k|k < \bar{k}_o^2) \\ &\quad + (1 - H(zF_2^*))(S - v - zc) + H(zF_2^*)S \\ &= \Psi(\bar{k}_o^2)\bar{k}_o^2 - \int_0^{\bar{k}_o^2} k\psi(k)dk + S - S(1 - H(zF_2^*))\theta^*. \end{aligned}$$

Note that θ^* must be less than one because $F_1^* = \frac{v}{z} + c - \frac{S}{z} < 0$ and $F_2^* = \frac{v}{z} + c - \frac{S}{z}\theta^* > 0$. Also note that $\Psi(\bar{k}_o)\bar{k}_o - \int_0^{\bar{k}_o} k\psi(k)dk$ is strictly increasing in $\bar{k}_o$ and $\bar{k}_o^2 < \bar{k}_o^1$, we have $\Psi(\bar{k}_o^1)\bar{k}_o^1 - \int_0^{\bar{k}_o^1} k\psi(k)dk > \Psi(\bar{k}_o^2)\bar{k}_o^2 - \int_0^{\bar{k}_o^2} k\psi(k)dk$. Still, NB_2 could be greater than NB_1 given a large enough S. The net subsidy solution F_1^* is preferable to the penalty F_2^* if the enhanced wealth from additional investment is greater than the wealth gained from deterring monopolization with the penalty.

Last, we show that it is optimal to set $z = z'$. The first-order condition with respect to z is as follows:

$$\begin{aligned} \frac{\partial NB}{\partial z} &= F(v + zc - S - zF)\{\Psi(\bar{k}_o)h(zF) + \psi(\bar{k}_o) \\ &\quad \times (1 - H(zF))^2\} - \psi(\bar{k}_o)(1 - H(zF))c\mathrm{F} \\ &\quad - FS\{\psi(\bar{k}_o)H(zF)(1 - H(zF)) - \Psi(\bar{k}_o)h(zF)\}. \end{aligned}$$

Given that $F = F^*$, $\frac{\partial NB}{\partial z}\Big|_{F=F^*} = -\psi(\bar{k}_o)(1 - H(zF^*))c < 0$ so that it is optimal to choose the minimum enforcement rate. ■

Proof of Proposition 3: The social welfare function can be expressed as

$$NB = R(\bar{k}_v)(\bar{k}_v - E(k|k < \bar{k}_v)) + (1 - H(zF))\{E(M|M > zF) - (v + zc)\},$$

where $\bar{k}_v = H(zF)B$.

Assume that $zF < M_u$. The two first-order conditions for the welfare maximization problem are:

$$\frac{\partial NB}{\partial F} = R(\bar{k}_v)Bzh(zF) + zh(zF)(zc + v - zF),$$

where the first term denotes marginal gain from investment; and

$$\frac{\partial NB}{\partial z} = Fh(zF)(zc + v - zF + R(\bar{k}_v)B) - (1 - H(zF))c.$$

Using the same logic of the basic model (see Proposition 1), it is easy to show that for a given value of z, F is optimally set so that $F^* = \frac{v}{z} + c + \frac{R(\bar{k}_v^*)B}{z}$. And because NB is decreasing in z when $F^* = \frac{v}{z} + c + \frac{R(\bar{k}_v^*)B}{z}$, it is optimal to choose enforcement rate at its minimal level, namely $z^* = z'$.

Now let NB^* denote the social welfare evaluated at (z', F^*), namely,

$$NB^* = R(\bar{k}_v^*)(\bar{k}_v^* - E(k|k < \bar{k}_v^*)) + (1 - H(z'F^*))\{E(M|M > z'F^*) - (v + z'c)\}.$$

And let $\overline{NB}$ be the social welfare evaluated when $F^* \in \{F{:}F \geq M_u/z\}$. In this latter scenario, the offense is completely deterred and social welfare depends solely on the victim's net gain from investment, where $\overline{NB} = R(B)[B - E(k_v|k_v < B)]$.

$$M_u \leq v + z'c + R(\bar{k}_v^*)B \tag{3}$$

In this case, it is optimal to set z^* F^* greater than M_u as NB^* is smaller than $\overline{NB}$.

$$\begin{aligned} NB^* - \overline{NB} &< R(\bar{k}_v^*)(\bar{k}_v^* - E(k|k < \bar{k}_v^*)) + (1 - H(z'F^*)) \\ &\quad \times (M_u - (v + z'c)) - R(B)[B - E(k_v|k_v < B)] \\ &\leq R(\bar{k}_v^*)(\bar{k}_v^* - E(k|k < \bar{k}_v^*)) + (1 - H(z'F^*))R(\bar{k}_v^*)B \\ &\quad - R(B)[B - E(k_v|k_v < B)] \\ &= \int_{\bar{k}_v^*}^{B} kr(k)dk - B\int_{\bar{k}_v^*}^{B} r(k)dk \\ &< B\int_{\bar{k}_v^*}^{B} r(k)dk - B\int_{\bar{k}_v^*}^{B} r(k)dk = 0 \end{aligned}$$

$$M_u > v + z'c + R(\bar{k}_v^*)B \tag{4}$$

First, let $\Delta = NB^* - \overline{NB}$ and it is easy to show that Δ is increasing in M_u as

$$\frac{\partial \Delta}{\partial M_u} = M_u h(M_u).$$

Let $\bar{M}$ denote the value of M_u such that $NB^* - \overline{NB} = 0$. Based on case (3), it is obvious that $\bar{M}$ is bigger than $v + z'c + R(\bar{k}_v^*)B$. For $M_u \in \{M_u : M_u > \bar{M}\}$, $NB^* > \overline{NB}$ so that $\{z', F^*\}$ is the optimal policy; for $NB^* \leq \overline{NB}$, $NB^* \leq \overline{NB}$ so that $z^*F^* > M_u$ is the optimal policy. ∎