Environmental Self-Auditing: Setting the Proper Incentives for Discovery and Correction of Environmental Harm

Alexander S. P. Pfaff
Columbia University and CERC

Chris William Sanchirico
University of Virginia, School of Law

Many firms conduct “environmental audits” to test compliance with a complex array of environmental regulations. Commentators suggest, however, that self-auditing is not as common as it should be, because firms fear that what they find will be used against them. This article analyzes self-auditing as a two-tiered incentive problem involving incentives both to test for and to effect compliance. After demonstrating the inadequacy of conventional remedies, we show that incentives can be properly aligned by conditioning fines on firms’ investigative effort. In practice, however, the regulator may not be able to observe such effort. Accordingly, we propose and evaluate the use of three observable proxies for self-investigation: the manner in which the regulator detected the violation; the firm’s own disclosure of violations; and the firm’s observed corrective actions. Each method has its own efficiency benefits and informational requirements, and each is distinct from EPA’s current audit policy.

1. Introduction

Since the 1970s, the amount of environmental regulation at all levels of government has increased significantly. Major federal statutes include the Clean Air Act, the Clean Water Act, the Resource Conservation and Recovery Act, the Toxic Substances Control Act, and the Comprehensive Environmental Response, Compensation, and Liability Act. In addition, most states have their own environmental laws and regulations. The resulting web of often highly technical requirements makes it difficult for the regulated enterprise itself to know whether it is in compliance with applicable law.1 In response, many firms have insti-

---

1 We thank the Center for International Business Education at Columbia University for financial support. For helpful comments, we thank Barry Adler, Jennifer Arlen, Jon Cannon, Louis Kaplow, Reinier Kraakman, James Sanchirico, Eric Talley, two anonymous referees, and seminar participants at Iowa State, the American Law and Economics Association’s 1998 meetings, and the NBER 1999 Summer Institute Workshop on Public Policy and the Environment. Thanks also to Benjamin Cohen for able research assistance.

© 2000 Oxford University Press
tuted a policy of conducting their own “environmental audits” [see, e.g., Kirsch and Viers (1996)].

Even with the increased incidence of such audits, however, many commentators suggest that the practice remains far rarer than it should be. Self-auditing is thought to be more thorough and efficient than periodic audits by the regulator. However, firms reportedly are reluctant to conduct such audits for fear that the information they gather on compliance problems will wind up in the hands of either government or adverse private parties, to be used against the firm in subsequent proceedings. By carefully investigating their own compliance status, so the argument goes, firms are effectively aiding in their own prosecution or adverse suit.2

The Colorado–Coors case illustrates both the necessity for self-investigation and the basis of firms’ reluctance to uncover their own violations. During a costly environmental review of its Golden, Colorado, brewing operations, the Coors Brewing Company discovered that the spillage of beer in the course of manufacturing was producing harmful—and illegal—ethanol emissions. Subsequently, the Colorado Department of Public Health and Environment sought to obtain more than $1 million in fines from Coors, based on Coors’ own disclosure of its internal review—even though the agency never would have known of the violations but for Coors’ own investigative effort [see Feeley (1995)].

Since 1986, the EPA has taken a series of tentative steps to address the self-auditing incentive problem, culminating in its December 1995 “Audit Policy.”3 In essence, the policy4 is to lower fines for violations

---

1. Hawks (1998) notes: “Occidental Petroleum estimated that in just four Texas facilities the company had to monitor 140,000 points for fugitive emissions, resulting in 4 to 7 million pieces of data.”

2. See, for example, Murray (1995): “Corporate America is worried that the audit provides the regulatory agency, the Department of Justice, or interested citizen groups with a paper trail that leads to expensive cleanup, fines, and possible criminal culpability for the corporation, its officers, directors, and even its employees”; Reed (1983): “auditing makes good sense for companies interested in managing their activities … [but an audit report] … could make an enforcement case for EPA or a state or local agency or a private group”; Moore and Newkirk (1995): “Substantial disincentives to self evaluation have existed because of enforcement risk associated with penalties imposed when non-compliance is found and reported and the fear that the information will be used against the company”; Feeley (1995): “A self-audit can become a ‘prosecutorial road map,’ allowing disclosure to be used as an enforcement tool”; Cooney et al. (1995c): “Environmental auditing is not used as frequently as it should be, however, due in part to governmental reluctance to give formal protection to internal documents generated during the audit process.”

3. EPA (December 22, 1995), as amplified by EPA (January 15, 1997) and EPA (March 1998).

4. The policy is a “guideline” for prosecutorial discretion and settlement negotiation.
discovered through self-auditing, so long as the firm (1) voluntarily discloses these violations, and then (2) promptly deals with the problem. Industry sentiment, however, is that the EPA policy does not go far enough. And legislation that would provide more sweeping protections is introduced perennially in both houses of Congress. Indeed, stronger protections already have been passed in several states, with Colorado notably among them. In all, 21 states have enacted into law some form of evidentiary “audit privilege” for the information revealed by self-audit; some also offer forms of immunity for harms discovered and corrected.

In this article we analyze the self-auditing dilemma as a two-tiered incentive problem, one involving incentives both to test for and to effect compliance. As in the classic torts problem, the ultimate goal is to induce firms to effect specified levels of “environmental care” (in terms of both precaution and cleanup) through a system of fines or charges probabilistically applied via random and imperfect auditing. In the classic torts problem, however, the issue for the firm is only whether, and not how, to exercise care; whereas in a technically and legally complicated regulatory regime, firms cannot exercise care until they first determine how care effort can and should be applied. Accordingly, the regulator must induce firms to test, or audit, their compliance status. The crux of the problem, however, is that such self-investigation inadvertently improves the regulator’s own auditing system, and thereby increases the frequency with which fines are assessed.

After explaining why conventional tort remedies are incapable of solving the self-auditing incentive problem, we provide a “benchmark” solution, one that goes beyond conventional remedies by conditioning fines on firms’ investigative efforts. While this solution achieves efficiency, it requires information about firms’ self-auditing efforts that a regulator is unlikely to possess. Accordingly, we provide three additional solutions in which the regulator conditions fines on observable proxies for investigative effort. The proxies that we identify are the manner in

5. The policy stipulates that the gravity component of fines will be reduced by 100% for a firm that discovers a violation through a systematic auditing program, voluntarily discloses the violation within 10 days after discovery without prompting from the government or a third-party plaintiff, corrects the violation within 60 days, takes steps to prevent recurrence, and cooperates with the EPA throughout. Further, the violation cannot be part of a pattern of repeated violations and cannot be one that has caused or may cause “serious harm.” Gravity fines are reduced by 75% if the violation was not discovered as part of a systematic program, but all the other conditions listed are fulfilled.

6. For the 105th Congress, see S. 860 and S. 1332. For the 104th Congress, see S. 582 and H.R. 1047.
which the EPA detected the violation, the firm’s own disclosure of its violations, and the firm’s observed corrective actions.  
Section 2 presents a simple model of the self-auditing problem and explains why conventional remedies are incapable of producing the socially efficient outcome. In this section and in the rest of the article, the formal mathematics is preceded by a motivating discussion and followed by an explanatory discussion. Some readers may wish to skip the mathematical portions and focus solely on these discussions. Section 3 begins with a presentation of the benchmark solution of conditioning fines on firm investigation. It then sets forth three proposals for conditioning on investigation via observable proxy. Section 4 considers implementability. Section 5 concludes.

2. Model and Failure of Conventional Remedies
A firm is uncertain about the magnitude and the nature of environmental harm $h$ that its operation will cause. Without understanding the nature of the harm, the firm cannot take corrective action. Its beliefs over magnitude are given by the probability measure $P$. In order to

---

7. Shavell (1992) considers incentives to gather information on the presence of “risk” and the effectiveness of precaution, but investigation does not increase the chance of being sued or fined in his model. See note 9. Kaplow and Shavell (1994) show that self-reporting (i.e., disclosure) can lower both enforcement costs and risk-bearing costs; but agents in their model know the nature and magnitude of violations without having to investigate. See, our discussion of disclosure in Section 3.3. Arlen (1994) analyses the problem of vicarious corporation criminal liability and the incentive to monitor employees. Her solution of conditioning fines on monitoring effort is closely related to our benchmark solution. Written simultaneously and independently of our article, Arlen and Kraakman (1997) [A & K] extends Arlen’s model to the problem of inducing firms to monitor employee activities that may have harmful environmental consequences. A & K also consider duty-based regimes and self-reporting. A & K analyze self-auditing as a type of agency problem, while we analyze it as a self-knowledge problem faced by the firm as a unit. Notwithstanding the ubiquity of agency issues, we think the later is more in line with how those involved in the problem describe it. See note 3, supra. Further, A & K focus on enumerating the practical difficulties of directly conditioning fines on investigation of employees, while the bulk of our article concerns proposals for getting at investigation indirectly by conditioning on its observable proxies. Lastly, we provide important caveats to the supposed benefits of a disclosure-based approach, which do not appear in A & K. Dana (1996) argues against a policy of “audit immunity” because it would effectively weaken firms’ incentives to prevent and fix such violations. Yet our benchmark solution shows that it is possible to mitigate the self-investigation disincentive without fouling the incentive to prevent, by raising fines for firms that do not investigate. Indeed, Dana ultimately advocates increased fines for externally detected violations. Orts and Murray (1997) propose an evidentiary self-evaluative privilege for businesses conducting audits under an EPA-supervised disclosure system. They append fine reduction to this privilege “to encourage self-reporting and self-policing.”
discover the magnitude and nature of the harm, the firm must conduct an environmental self-audit (henceforth an “investigation”).

For analytical convenience, we assume that the firm decides only whether or not to investigate, not how much or how to investigate. We also assume that if the firm investigates, it learns with certainty both the magnitude and the nature of the harm. It costs the firm \( i \) to investigate. If the firm chooses to investigate, it then decides whether or not to fix the problem(s) that it finds. Again for simplicity, the firm decides only whether or not to fix, not how much or how to fix, and if the firm fixes the problem there is no residual harm. The cost of alleviating the harm depends on the magnitude of the harm and is denoted \( c_h \).

Left to itself, the firm will neither investigate nor fix since these are costly actions and it does not care about the harm it is causing. The second actor in our model, a regulatory agency, takes into account not only private investigation and fixing costs, but also the cost of the harm, as well as the cost of enforcement. The agency’s choice of enforcement effort affects the probability that the agency will learn the magnitude and nature of the harm caused by the firm. In accordance with rationales presented in the popular press for special treatment of self-auditing (see note 2), we assume that for any given amount of enforcement effort, this probability is always larger if the firm has investigated.

Two additional simplifications serve to minimize technical detail. First, we fix the agency’s enforcement effort (which determines the probability of detection as a function of firm investigation) and consider the agency’s problem of setting fines to encourage efficient investigation and fixing. With enforcement effort in the background, we specify that if the firm has investigated, the probability that the government learns the harm is \( \beta_h \). If it has not, the probability is \( \tilde{\beta}_h < \beta_h \). Second, we assume that at this fixed level of enforcement effort, upper bounds on fines are not binding in the agency’s problem.

What investigation and fixing decisions would the regulatory agency make if it controlled the firm? Working backward, we consider first the decision to fix, conditional upon having investigated, and next the decision to investigate. Having investigated and found harm of magnitude \( h \), fixing is a best choice if and only if \( c_h \leq h \). This produces an optimal fixing decision contingent on \( h \). Given this best contingent plan,

---

8. Certainly there exist harms whose correction does not require investigation. But these are not, by definition, prey to the incentive problem that we study in this article and the EPA has attempted to address in its Audit Policy. Further, though investigation always precedes fixing, fixing may occur before or after occurrence of the harm. The variable \( h \) may represent actual harm (e.g., an ongoing emission) or expected future harm.
investigation imposes expected social costs of

\[
i + \int_{c_h \leq h} c_h \, dP + \int_{c_h > h} h \, dP.
\]

harms that are best fixed  harms that are best left unfixed

On the other hand, not investigating imposes expected social costs of \( i h dP \). Combining these terms, it is best for society to investigate if and only if

\[
i \leq \int_{c_h \leq h} h - c_h \, dP. \tag{1}
\]

The social costs of investigation are the direct costs \( i \) of conducting the search for harm. The social benefits are the potential cost savings from uncovering and fixing a harm whose magnitude exceeds its fixing cost. In order for our firm to make the optimal decision for society, we must employ a fine structure that induces the firm to make the same decisions as society would: that is, to make the decision to fix, having investigated, according to \( c_h \leq h \), and the decision to investigate according to Equation (1).

2.1 Failure of Conventional Remedies

Conventional tort remedies based on harm and/or “care” (i.e., strict liability and negligence, with punitive add-ons since not all violators are caught) are incapable of simultaneously producing the proper incentives for both investigating and fixing. Under strict liability, fines are based on harm caused, with no regard to whether the firm exercised “due care.” To align fixing incentives, the total fine \( F_h \) (both compensatory and punitive elements) for harm \( h \) must be such that \( \bar{\beta}_h F_h = h \iff F_h = h / \bar{\beta}_h \). Then investigation imposes expected costs on the firm of

\[
i + \int_{c_h \leq \bar{\beta}_h F_h} c_h \, dP + \int_{c_h > \bar{\beta}_h F_h} \bar{\beta}_h F_h \, dP,
\]

harms fixed harms not fixed

while not investigating imposes expected costs of \( \int \beta_h F_h \, dP \). Combining these terms, we see that the firm investigates if and only if

\[
i + \int_{c_h > \bar{\beta}_h F_h} \left( \bar{\beta}_h - \beta_h \right) F_h \, dP \leq \int_{c_h \leq \bar{\beta}_h F_h} \beta_h F_h - c_h \, dP. \tag{2}
\]
Selectively substituting $\bar{\beta}_h F_h = h$ yields

$$i + \int_{c_h \geq h} (\bar{\beta}_h - \beta_h) F_h \, dP \leq \int_{c_h \leq h} \beta_h F_h - c_h \, dP.$$  

This criterion differs from society’s investigation decision in Equation 1 on two scores, both of which imply that the firm will investigate too little. Term A in Equation 3 concerns cases in which the harm would not be fixed if found. Its presence in Equation 3 and not in Equation 1 reflects the fact that investigation exposes the firm to a higher probability of being caught. The second discrepancy, reflected in the difference between term B in Equation 3 and the right-hand side of Equation 1, concerns harms that would be fixed once revealed by investigation. In these cases, the firm compares the cost of fixing ($c_h$) to the expected fine conditional on not investigating. The discrepancy arises because this expected fine is calculated using $\beta_h$, the probability of being caught having not investigated. Since the fine $F_h$ was set so that $\bar{\beta}_h F_h = h$, it must be that $\beta_h F_h < h$.

To induce the firm that has self-investigated to make the right fixing decisions, we have set fines so that the firm’s expected fine equals harm. Since fixing occurs after investigation, these fines must be set based on the probability of detection that the firm faces after having investigated. This properly aligns fixing incentives, but simultaneously produces a divergence between the social and private net benefits of investigation. It is helpful to parse this divergence into cost and benefit components. Again, the social cost of investigation is the direct cost of self-investigating. The private cost, however, contains an additional component. The firm knows that its investigation may uncover harms that are not worth fixing. In those cases, its investigative effort will have served simply to increase the probability that it will be caught and charged for such unfixed violations. Society views such charges as transfers between agents, and thus ignores this effect. Individual firms, however, take the resulting increase in expected fines into account in deciding how much to self-investigate. Thus the firm’s (marginal) cost of investigation is higher than society’s, pushing the firm toward less than socially optimal investigation.9

While this divergence in the private and social costs of investigation has been the exclusive focus of public comment, a systematic analysis reveals that there is also a divergence in the social and private benefits of investigation. The social benefit to investigation is the expected net

---

gain from fixing environmental harms that are uncovered and found to be “worth fixing”; that is, those that cause more damage to society than they cost the firm to fix. Firms will undervalue these benefits. The social value of investigation and fixing arises out of alleviating the harm, while the firm’s gain from investigating and fixing comes from avoiding the expected fine it faces if it does not investigate. Thus the firm views the benefits of fixing in terms of the noninvestigation probability of detection. Recall, however, that to correctly align social and private fixing incentives, we began by setting fines so that the firm’s expected fine calculated with respect to the postinvestigation probability of detection equals harm. This implies that the expected fine calculated with respect to the lower, noninvestigation probability of detection falls short of harm. As a result, the firm’s gain from investigating and fixing is less than the harm avoided, and so less than society’s gain. Thus the firm’s (marginal) benefit of investigation is lower than society’s, also pushing the firm toward less investigation than is socially optimal. All told, higher private costs and lower private benefits act in concert to produce too little investigation.

One way to illustrate the importance of this second divergence is to consider negligence in lieu of strict liability. At first glance, negligence looks like a desirable alternative. A negligence standard, with the due care threshold keyed to the social costs and benefits of fixing, correctly aligns fixing incentives for self-investigating firms. Moreover, unlike strictly liability, it imposes no fines for those harms not worth fixing. Thus the divergence between private and social costs identified above is not present. But negligence imposes the same fines as strict liability for harms that are worth fixing, and thus produces the same divergence between the social and private benefits of investigation as above. The result is again a difference between the social and private net benefits from investigation—though there is now one source rather than two.  

3. A Benchmark and Three Proxy-based Solutions

3.1 A Benchmark Solution

Having demonstrated the inadequacy of conventional remedies, we propose a “benchmark solution,” whose applicability is limited by its informational requirements, but whose incentive structure will help to...
unify and illuminate the rest of our proposals. The basic idea is that fines are raised, relative to strict liability, for firms that have not self-investigated. Note that in contrast to EPA policy, firms that investigate will avoid these enhanced fines regardless of whether they fix or disclose what they find.

Formally, let $\bar{F}_h$ be the fine in the case in which the firm has investigated and let $\bar{F}_h$ be the fine when the firm has not investigated. Set $\beta_h \bar{F}_h = \beta_h \bar{E}_h = h$, implying that $\bar{F}_h > \bar{F}_h$ (so $\bar{F}_h$ from Section 2). Incentives to fix are then properly aligned. The cost of investigating is $\int_{c_h \geq \beta_h \bar{F}_h} c_h \, dP + \int_{c_h > \beta_h \bar{F}_h} \beta_h \bar{F}_h \, dP$. The cost of not investigating is $\int \beta_h \bar{F}_h \, dP$. Combining yields the analogy to Equation (2):

$$i + \int_{c_h > \beta_h \bar{F}_h} \left( \beta_h \bar{F}_h - \beta_h \bar{F}_h \right) \, dP \leq \int_{c_h \leq \beta_h \bar{F}_h} \beta_h \bar{F}_h - c_h \, dP. \quad (4)$$

Since $\beta_h \bar{F}_h = \beta_h \bar{E}_h = h$, Equation (4) reduces to Equation (1) and investigation incentives are also aligned.

The benchmark solution entails setting one fine level for firms that have investigated and another for those that have not, in such a manner that expected fines equal harm in both cases. Since detection probabilities increase as a result of self-investigation, actual fines will be higher for firms that have not self-investigated.

Bifurcating fines in this manner alleviates the two discrepancies discussed in Section 2, between the social and private costs and benefits of self-investigation. The additional private cost vanishes because even though investigation increases detection probabilities, it no longer alters the firm’s expected fine; rather the higher detection probability arising out of self-investigation is exactly offset by the lower fine. In terms of the discrepancy in benefits, when firms use noninvestigative detection probabilities in calculating the expected fines avoided, these expected fines now will equal the harm avoided by society.

The applicability of the benchmark solution is limited by the fact that the regulator is unlikely to possess information on firms’ investigative efforts. As the following three solutions show, however, it is possible to condition on investigation indirectly, through actions that the regulator does observe.

3.2 Proposal 1: Conditioning on Mode of Detection

As is well-known, a principal can effectively condition on an agent’s hidden action if she receives an informative signal of that action; thus contingent compensation is a means by which an employer can effectively condition on work effort by observing level of production. The issue then is whether an informative signal of investigative effort exists. We propose that one such signal is inherent in the nature of the self-auditing problem.
The special problem of self-auditing—according to industry, academia, and government—is that the act of investigating makes it more likely that the EPA will discover the firm’s unfixed violations. And the reason given for this is that investigation produces a “paper trail,” provides a “prosecutorial road map,” and/or inspires “whistle-blowing.” In short, the reason that we need to condition fines on investigation in the first place is that the by-products of investigation may be, at least in part, observable to the EPA. From a contrapositive perspective, if nothing about investigation were observable to the EPA, it is difficult to understand how investigation would increase the probability of detection, and it is then not clear why there is an incentive problem in the first place.

We propose that the same observable by-products of investigation that lie at the root of the self-auditing problem are the natural starting point for constructing an informative signal of self-investigation. Specifically we suppose that there may be several modes of detection, that is, manners in which the EPA detects a firm’s violation. We define these modes by what the detector (i.e., the EPA) knows about how it detected a violation. Thus, unlike investigative effort itself, such modes are (by definition) observable to the EPA.

We show that so long as investigative effort has any effect on the relative probability of detection by any one of these modes, then it will be possible to condition fines on such modes in a way that solves the self-auditing problem. Roughly speaking, the solution entails lowering fines when violations are detected by modes that are relatively more likely following investigation. Again, in contrast to EPA policy, this is regardless of whether the firm has fixed or disclosed what its investigation uncovers.

Formally, suppose that there are two modes of detection, $N$ and $D$ [for instance, discovery without ($N$) and with ($D$) help of a revealing document]. Given harm $h$, let $\beta^D_h$ and $\beta^N_h$ be the probability of detection via $D$ and $N$, respectively, when the firm has investigated. The probability that the investigating firm is detected is then $\beta^*_h = \beta^D_h + \beta^N_h$. Let $\beta_h$, $\beta^D_h$ and $\beta^N_h$ be the same probabilities when the firm has not investigated.

Implicit in explanations of the self-auditing problem are two assertions about firms that have self-investigated: (1) they are more likely to get caught; and (2) when caught, they are likely to be caught by different means. The first is our assumption that $\beta^*_h > \beta_h$. The second is the statement that investigation has an effect on the probability of

---

12. For example, the EPA might detect a violation during a routine onsite inspection, or its inspection might uncover incriminating internal documents, or it might follow a tip from a disgruntled employee, who might have acquired damaging information on the line or via participation in an internal review, etc.
modes of detection conditional on the event that the firm is caught:

\[
\frac{\beta_h^D}{\beta_h^D + \beta_h^N} \neq \frac{\beta_h^N}{\beta_h^D + \beta_h^N} \iff \frac{\beta_h^D}{\beta_h^D + \beta_h^N} \neq \frac{\beta_h^N}{\beta_h^D + \beta_h^N}.
\] (5)

As is clear from the benchmark solution, the condition for achieving efficiency is that the firm’s expected fine is \( h \), whether or not it has investigated. Let \( F_h^D \) be the fine for harm \( h \) detected via mode \( D \). Let \( F_h^N \) be the same for detection via mode \( N \). The expected fine having investigated (not investigated) is \( \beta_h^D F_h^D + \beta_h^N F_h^N (\beta_h^D F_h^D + \beta_h^N F_h^N) \).

Thus efficiency is achieved if for, almost every \( h \), the following system has a solution:

\[
\frac{\beta_h^D F_h^D + \beta_h^N F_h^N}{\beta_h^D F_h^D + \beta_h^N F_h^N} = h
\]

A solution exists if the ratios \( \frac{\beta_h^D}{\beta_h^D} \) and \( \frac{\beta_h^N}{\beta_h^N} \) are not equal (i.e., if the lines describing each equation are not parallel). This is in turn equivalent to Equation (5):

\[
\frac{\beta_h^D}{\beta_h^D} \neq \frac{\beta_h^N}{\beta_h^N} \iff \frac{\beta_h^D}{\beta_h^D} \neq \frac{\beta_h^N}{\beta_h^N} \iff \frac{\beta_h^D}{\beta_h^D} \neq \frac{\beta_h^N}{\beta_h^N} \iff \frac{\beta_h^D}{\beta_h^D} \neq \frac{\beta_h^N}{\beta_h^N}.
\]

Thus equation (5) is sufficient\(^{13}\) for the existence of a solution to Equation (6).

Perhaps the best way to clarify how this solution operates is with a numerical example. Suppose that the chance of detection for a given harm, \( h \), is 30% when the firm has investigated and 15% when the firm has not. Further, detection might or might not involve the discovery of a revealing document. Among firms that self-investigate, 70% of those caught are caught with revealing documents, whereas among firms that

\(^{13}\) Two technical notes: (1) The condition is not necessary: \( 0.1 F_h^D + 0.2 F_h^N = h \), \( 0.1 F_h^D + 0.2 F_h^N = h \) is soluble but violates the condition. More generally, a solution exists if investigation does not affect some mode’s probability. (2) With more than two modes, it is sufficient that the system’s coefficient matrix has full rank.
do not self-investigate, this number is only 20%. Suppose then that the EPA charges the firm only one-third of the harm caused when it detects the violation via revealing document, but a full eight times harm when it detects without the revealing document. The EPA is thus rewarding the firm for having investigated, by lowering fines when its detection is aided by the likely, observable by-product of investigation. Given this policy, let us calculate the expected fine for a firm that has investigated. The chance of detection with documents for a self-investigating firm is \(0.3 \times 0.7 = 0.21\), whereas the chance without documents is \(0.3 \times 0.3 = 0.09\). Thus the expected fine is \(0.21(h) + 0.09(8h) = h\). For noninvestigating firms, the chance of detection with documents is \(0.15 \times 0.2 = 0.03\) and the chance without is \(0.15 \times 0.8 = 0.12\); a similar calculation shows that the expected fine for the noninvestigating firm is also \(h\). Because expected fines equal harm when the firm has investigated, fixing incentives are aligned. Because expected fines are invariant with respect to investigation, the private and social costs of investigation are equated. And because the expected fine from having not investigated equals harm, the social and private benefits of investigation are also the same.

3.3 Proposal 2: Conditioning Fines on Disclosure

As noted, the EPA requires disclosure before fines will be reduced, and this has been endorsed in the literature. In partial deference to this consensus, our second proposal makes fines contingent on whether the firm reports the violations it discovers. In contrast to EPA policy, this is regardless of whether the firm fixes what it finds. First, we show that under certain conditions disclosure can be used as a proxy for investigative effort. Next, we evaluate the advantages and disadvantages of relying on firm disclosure.

Disclosure as a Proxy for Investigative Effort. We modify our model to account for the firm’s decision to disclose. Now, after having investigated, the firm chooses not only whether to correct the harm but also whether to report the results of its investigation to the agency. We assume that this report is verifiable. Thus if the firm finds harm of \(h\), and decides to make disclosure to the agency, it is compelled to report \(h\). In particular, if it has not investigated, it has nothing to disclose. Thus disclosure, like fixing, arises only after investigation. We also assume that disclosure has no direct costs for the firm.

The agency sets two fines: \(F_h^D\) if the firm discloses harm \(h\), and \(F_h^N\) if the agency finds out on its own that the harm is \(h\). Of importance, the agency does not condition directly on investigation and so \(F_h^N\) is assessed both in the case in which the firm has investigated, not fixed, and not disclosed and in the case in which the firm has not investigated (and so neither fixed\textsuperscript{14} nor disclosed).

\textsuperscript{14} Recall that if the firm fixes, the harm is alleviated. Recall also that fixing requires investigation.
Proceeding backwards as before, suppose that the firm has investigated. It has four choices. It can (1) fix and disclose, (2) fix and not disclose, (3) not fix and disclose, or (4) not fix and not disclose. Whether or not the firm discloses, fixing yields direct costs of \( c_h \) (as disclosure itself is costless). This is all of the costs for alternatives 1 and 2. If the firm chooses alternative 3, not fixing but disclosing, the firm is fined \( F_h^I \). Finally, alternative 4, not fixing and not disclosing, results in an expected fine of \( \beta_h F_h^N \).

In deciding whether to fix, the firm compares the uniform payoffs from fixing i.e., alternatives 1 and 2 to the better of the payoffs from not fixing (i.e., alternatives 3 and 4). The effective fine from not fixing is then the minimum of fines in 3 and 4, and the firm fixes if and only if

\[
c_h \leq \min \{ F_h^T, \beta_h F_h^N \} = m_h.
\]  

Stepping backward to the investigation decision, and making use of our definition of \( m_h \), investigation imposes on the firm expected costs of

\[
i + \int_{c_h > m_h} m_h - \beta_h F_h^N \, dP + \int_{c_h \leq m_h} m_h \, dP,
\]

while not investigating imposes costs of

\[
0 + \int \beta_h F_h^N \, dP.
\]

Combining these terms, we see that the firm will investigate if and only if

\[
i + \int_{c_h > m_h} m_h - \beta_h F_h^N \, dP \leq \int_{c_h \leq m_h} \beta_h F_h^N - c_h \, dP.
\]

Aligning fixing incentives requires setting fines so that \( m_h = h \). Investigation incentives are aligned as well if we also set \( \beta_h F_h^N = h \), for then term D disappears and term E is the same as the right-hand side of Equation (1). If \( m_h \) and \( F_h^N \) were independent variables, we could do this without further thought. However, \( m_h \) depends, in principle, on \( F_h^N \) and our analysis thus requires the additional step of showing that when \( \beta_h F_h^N = h \), we can find an \( F_h^T \) so that \( m_h(\min(F_h^T, \beta_h F_h^N)) = h \). It turns out that \( F_h^T = h \) does the trick: since \( \beta_h > \beta_h \), setting \( \beta_h F_h^N = h \) ensures that \( \beta_h F_h^N > h \) and so \( F_h^T = h \) will be the minimum fine defining \( m_h \).

Conditioning fines on disclosure correctly aligns the incentives for both investigating and fixing. Of importance, though, this has nothing to do with any characteristic inherent to disclosure. Rather it follows from the fact that disclosure, if assumed to be verifiable, may be used as a proxy for investigation per se.

To understand the intuition behind this finding, return to the case of fines conditioned neither on disclosure nor on investigation, as in our discussion of strict liability in Section 2. There we faced the following difficulty: equating the expected fine conditional on investigation with actual harm—in order to set proper postinvestigation fixing incentives—causes expected fines conditional on not investigating to fall...
short of what is necessary to align investigation incentives. Now focus on the mirror image view of this problem: setting the fine to get the noninvestigation case right. This causes expected fines conditional on having investigated to exceed the proper level. Getting the noninvestigation case right alleviates the divergence between the social and private benefits of investigation, but the resulting increase in postinvestigation expected fines means that the divergence in investigation costs remains and that a new divergence in the decision to fix is added.

Conditioning on disclosure provides a way of addressing this overshoot in all cases in which the firm has investigated. Essentially, the agency offers the firm who has investigated an “out”: if the firm tells what it has found, the agency lowers the fine below what the firm expects to pay if it keeps quiet. Of course, the firm prefers lower fines, and so takes the option. This effectively reduces the expected fine for investigating firms.

In sum, this policy effectively equates investigation and disclosure. Our verifiability assumption ensures that disclosure is possible only when investigation has occurred. Conversely, lowering the fine for disclosed violations ensures that firms that investigate always disclose. Hence the agency's “disclosure discount” effectively allows it to condition fines on investigation per se, which is the key to setting proper incentives.

Advantages and Disadvantages of Disclosure Proxy. One disadvantage of the disclosure solution is that it relies on the assumption that disclosure is truthful. In reality, credibility is a serious issue; firms will be motivated to shade their reports in any way that will reduce the fine. Firms might, for instance, disclose only the less severe harms among those that they actually uncover, reasoning that paying the lesser fine will buy them some reprieve from the EPA's detection fervor, and so help to insulate their truly egregious harms from detection. Similarly, firms that have not even investigated may have an incentive to fabricate reports of moderate harms to throw the EPA off the scent of any serious harms that the firm suspects may be ongoing. Thus the informational assumptions of the disclosure solution may rival those of the benchmark solution.

One potential advantage of conditioning on disclosure is that it enables savings in enforcement costs. Suppose, as proposed above, that the EPA sets fines so that the certain fine for disclosed violations is less than the expected fine for violations discovered by the EPA itself. Then firms will prefer to disclose their violations rather than play the “audit

---

15. Note that there is an important difference between faking detection mode and falsely disclosing: the former may be fashioned into informative costly signals, while the latter is purely “cheap talk.”
lottery.” And the EPA will therefore learn of all violations without ever having to audit, and so without ever incurring enforcement costs [see, e.g., Kaplow and Shavell (1994) and Arlen and Kraakman (1997)].

However, there are at least two reasons why such cost savings might be more limited than at first appears. First, the argument implicitly assumes that enforcement costs would be incurred only in the counterfactual circumstance in which the firm fails to disclose. Yet in reality, even if all firms disclosed, so that detection activities were never actually conducted, the EPA would still need to maintain a state of readiness (like the peacetime military) in order to make credible its threat that its detection capability could and would be employed against firms that do not disclose. These upfront costs seem to be ignored when disclosure’s cost savings are touted.

The second caveat is specific to the two-tiered incentive problem that we are studying, in which a firm cannot (verifiably) disclose unless it has investigated. Recall that it may be optimal for a firm not to investigate. Thus some firms cannot and will not disclose. Unlike in the simple disclosure model discussed above, the EPA may actually find itself in the position of having to follow up on its original threat to conduct costly random audits. To be sure, it may be in EPA’s myopic interest to go back on its threat, but this would likely undermine its effort to set proper incentives in future periods.

3.4 Proposal 3: Inverse Negligence

Our third and last solution entails imposing additional fines when the EPA detects a harm that the firm would have fixed had it discovered the harm on investigating. Thus this third proposal uses the firm’s fixing behavior—as opposed to what it discloses, or how it was caught—as a signal of its investigative effort. Since fines are now conditioned on whatever harms the EPA actually detects, fine reduction is, by definition, not conditioned on the firm’s hidden actions, as it is in the benchmark solution.

In contrast with EPA’s audit policy, disclosure is not a condition for fine reduction (i.e., avoiding the enhanced fines) under inverse negligence. And for those harms that an investigating firm would not have remedied, neither is fixing. The comparison with general EPA criminal policy is also interesting. There, under standard scienter provisions, the firm is punished more severely if it knew about the harm but did not fix.

16. “Inverse negligence” is really “upside down and backwards negligence.” In negligence, if the firm takes “due care,” it has no liability, if not, it pays an amount keyed to harm. The rule here is “backwards” from negligence because the firm pays the harm-based fine when it takes due care. The rule is “upside down” because the adjustment away from the harm-based fine is an increase rather than a decrease.

17. But the EPA must know the functional relationship between harm and fixing cost. See note 19.
it. Under inverse negligence, by contrast, the firm is punished for its 
ignorance—in particular, for its ignorance of harms that it would have 
fixed had it known about them.

Formally, start with conventional strict liability as in Section 2 (\( \bar{\beta}_h F_h = h \); no conditioning on investigation). Consider the "would-have-fixed" harms; those satisfying \( c_h \leq h = \bar{\beta}_h F_h \). Note that the incentive to fix these harms would remain intact if we were to raise the \( F_h \) on these harms (and these harms only) by \( K \) so that \( c_h \leq h < \beta_h(F_h + K) \). The public benefit of investigating would still be \( \int h \, dP - (\int_{h < c_h} h \, dP + \int_{h \geq c_h} c_h \, dP) \), while the private benefit becomes

\[
\int_{\beta_h F_h = h < c_h} \beta_h F_h \, dP + \int_{\beta_h F_h = h \geq c_h} \beta_h F_h(F_h + K) \, dP - \left( \int_{\beta_h F_h = h < c_h} \beta_h F_h \, dP + \int_{\beta_h F_h = h \geq c_h} c_h \, dP \right).
\]

The difference between these terms is

\[
\int_{\beta_h F_h \geq c_h} \left( (\beta_h - \bar{\beta}_h) F_h - \beta_h K \right) \, dP + \int_{\beta_h F_h < c_h} (\beta_h - \bar{\beta}_h) F_h \, dP.
\]

Efficiency obtains if this difference vanishes: that is, if

\[
K = \int \left( \frac{\beta_h - \bar{\beta}_h}{\bar{\beta}_h} \right) h \, dP / \left( E[\beta_h | h \geq c_h] P(h \geq c_h) \right).
\]

Thus we set \( K \) to the (harm-weighted) average increase in the probability of detection due to investigation, discounted by the chance \( (E[\beta_h | h \geq c_h] P(h \geq c_h) \) that \( K \) will be imposed.

The key to this proposal is that additional fines are imposed only for harms that self-investigating firms will choose to fix. This implies, first, that the additional fines do not foul fixing incentives; the increase in fines just gives investigating firms even more reason to fix what they are already fixing. Second, additional fines are paid only by those firms that choose not to investigate. Thus the additional fines increase what we have called in Section 2 the private benefit from investigation, namely the benefit of avoiding expected fines on harms that an investigating firm would fix. If we wished to, therefore, we could raise the private benefits from investigation to equal the social benefits. However, this would do nothing to mitigate the fact that the private costs of self-investigation exceed the social costs: recall that firms, but not society, take into account the increased probability of detection of harms that would not be fixed following self-investigation. Nonetheless, by increasing the private benefits from investigation over and above society’s benefit, we
may match firms’ higher investigation costs with correspondingly higher investigation benefits, in such a manner that the net gain from investigation is the same for both firm and society.

4. Implementability and Statistical Knowledge

The information necessary to implement actual policy may be divided into two types: (1) statistical knowledge of key functional relationships, such as between investigation and the probability of detection; and (2) particular knowledge of the level of relevant variables for a given firm, such as a given firm’s investigative effort. The three proxy-based solutions proposed above all address the practical problem of obtaining particular knowledge about a given firm’s investigative effort. In this section we discuss the issue of statistical knowledge.

As a preliminary matter, we note that it is common in the literature for statistical knowledge issues to take a back seat to problems of particular knowledge. For example, while the optimal tax problem typically assumes that the tax authority cannot observe the ability of any particular taxpayer, optimal taxes are a function of both the population distribution of ability and the effect of ability on earning power. And, even in the most basic principal/agent problem, the employer must have a statistical knowledge of how an employee’s unobservable effort affects the probability of different levels of output.

One reason for this regularity in the literature is that the exercise of gathering statistical knowledge is generally not prey to the serious strategic revelation problems that are inherent in obtaining particular knowledge. If we ask a particular firm whether it has investigated, and the firm knows that we will use this information to set its punishment or reward, then the firm has a compelling incentive to lie. On the other hand, if we ask each firm in an industry about the functional relationship between self-investigation and detection, each individual firm realizes that its answer is just one of many data points for use in the design of a system that it may never encounter.

Indeed, in the specific case of environmental self-auditing, there may be even more reason to be optimistic about the possibility of gathering the requisite statistical knowledge. Here, the impetus for change comes from the firms themselves, who complain that investigation increases the chance of detection. The EPA may be able to gain much statistical knowledge by asking firms to explain with supporting statistical evidence the precise nature of this causal relationship. And presumably the EPA has obtained some useful information in this regard, for as we have

---

18. Information is usually classified according to the extent to which it is common knowledge. The distinction here between statistical and particular knowledge is less precise, but more helpful in this setting.
noted, it has a program in place that conditions fines on investigative effort in certain circumstances.

Nonetheless, statistical knowledge requirements may still pose practical difficulties, and it is therefore worth clarifying the requirements of each of the solutions that we propose. For the benchmark solution the regulator needs to understand the relationship between investigation and the unconditional probability of detection, so that fines can be properly conditioned on investigative effort. The first proxy solution requires estimates of the relationship between investigation and detection by at least two separate modes. The second, disclosure-based solution requires the same statistical knowledge as the benchmark solution. It is thus less demanding on this score than conditioning on modes of detection; but, again, it rests on the assumption that self-reported particular knowledge is truthful. Inverse negligence requires knowledge of this same relationship between investigation and detection for (at least some) would-have-fixed harms, in order that enhanced fines may be properly calculated. Because it also relies on the identification of some would-have-fixed harms, inverse negligence also requires knowledge of the functional relationship between the size of the harm and the cost of fixing it.

The proper method for obtaining estimates of these relationships is beyond the scope of this article. We do note, however, that one major consideration will be the trade-off between cost and accuracy as it pertains to choosing the level of aggregation for EPA estimates. Quite possibly the statistical relationships that we have been discussing vary significantly across industries, regions, and firms. In deciding how finely to parse its data, the agency must trade-off the increased accuracy of more disaggregated estimates with the fact that obtaining any given level of confidence across all groups will require larger samples. In principle, there is an optimal level of disaggregation that balances the costs of obtaining additional data with the expected benefits of more specific estimates.19

This is no doubt a difficult problem to solve with precision. But the EPA faces sectoral and regional variation in all of its programs, and so is accustomed to the challenge of developing workable approaches to the problem. In 1994, for instance, the EPA launched a “Common Sense Initiative,” with a mandate to explore ways of moving beyond a “one-size-fits-all” approach to regulation. In line with this initiative, the EPA keeps “Industry Sector Notebooks” with specific information on

---

19. Note that whether a type of information is particular or statistical is endogenous to EPA’s choice of how far to disaggregate: the distinction is meaningful conditional on this choice. For example, information on fixing costs (as a function of harm), as is required in inverse negligence, is in principle firm specific. Yet the cost/accuracy trade-off may induce EPA to employ industry-wide estimates of this relationship.
major sectors, including information on sector-specific compliance issues. Based in part on this information, the EPA has developed a number of sector-based and/or geographically based initiatives.

5. Conclusion

This article offers several proposals for addressing firms' disincentive to conduct environmental audits. Each is a means of reducing punitive fines for self-investigating firms, regardless of whether the firm has fixed the harm it uncovers, and (with one exception) of whether the firm discloses the results of its investigation. In contrast, EPA policy not only requires both disclosure and fixing, but also in effect assesses additional, criminal fines against firms that knew of, but failed to address, noncompliance.

Our benchmark solution indicates that in principle the solution to the self-auditing problem lies in making fines contingent on whether the firm has investigated. Because the EPA likely cannot directly observe investigative effort, though we provide three ways of getting at investigation indirectly. The first proposal is to use the manner in which the EPA detects the violation as a signal of investigative effort. The second proposal makes fine reduction contingent on whether the firm reports the violations it discovers. The third solution uses fixing behavior as a signal of investigative effort: a firm pays additional fines for the harms that it would have fixed had it investigated.

Built in to the EPA's audit policy is a provision for a review of the program's success after 3 years. We believe that this review might well be more scientific and informative if it included comparisons among several variations on the existing policy. Indeed, the EPA might even put in place a controlled experiment of sorts, with different versions of the audit policy being enforced in different regions and/or industries. We are, of course, aware that the number of policies that could be tested is small and that the costs of testing any particular policy would be substantial, especially if that policy turned out to be seriously flawed. Thus it would be imperative that candidate policies for dealing with the self-auditing problem first pass muster under a systematic, theoretical analysis of the auditing problem. Perhaps the sort of analysis presented in this article could serve as a starting point for this preliminary, but pivotal, step in experimental design.

References