The Economics of Environmental Monitoring and Enforcement: 
A Review

Jay P. Shimshack
email: jshimsha@tulane.edu

March 2014

Author Contact Information:

Department of Economics, Tulane University
206 Tilton Hall, 6823 St Charles Avenue
New Orleans, LA 70118
jshimsha@tulane.edu
(phone) 504-862-8353
(fax) 504-865-5869
http://econ.tulane.edu/shimshack/
The Economics of Environmental Monitoring and Enforcement: A Review

Abstract

Without monitoring and enforcement, environmental laws are largely non-binding guidance. While economists and philosophers have thought seriously about the broader public enforcement of law since at least the 18th century, environmental monitoring and enforcement remains both understudied and controversial. This paper reviews what we do and do not know about the subject. I review common environmental enforcement institutions, prescriptive and descriptive theories, empirical evidence on regulator behavior, and empirical evidence on deterrence effects.

Keywords

inspections, sanctions, penalties, deterrence, pollution
1. AIMS AND SCOPE

Without monitoring and enforcement, environmental laws are largely non-binding guidance. Scholars regularly cite traditional enforceable regulation as a leading motivator for environmental performance (Kagan et al 2003; Doonan et al. 2005; May 2005; Delmas and Toffel 2008). Government agencies spend billions monitoring and enforcing pollution regulations each year, and firms spend billions complying and paying penalties.

Economists and philosophers have thought seriously about the public enforcement of law since, at the latest, Bentham (1789). Yet, environmental monitoring and enforcement remains both understudied and controversial. Environmental economists most often ignore or assume away monitoring and enforcement issues. Other scholars and policymakers regularly call for transitions away from traditional enforcement towards cooperative, voluntary, or information-based approaches.¹

This paper reviews what we know about environmental monitoring and enforcement and concludes with a discussion of what we do not yet know. The paper builds on, and draws from, existing surveys (Cohen 1998; Heyes 2000; Gray and Shimshack 2011; Stranlund 2013). It contributes in two ways. First, the paper updates Cohen (1998). Second, it addresses a broader set of issues than Heyes (2000), Gray and Shimshack (2011), and Stranlund (2013).

The paper proceeds as follows. Section 2 reviews common environmental enforcement institutions in the US. Section 3 surveys both prescriptive and descriptive theory. Section 4 reviews empirical evidence on how environmental regulators behave, and how regulated entities respond to regulator actions. Section 5 concludes with key lessons and knowledge gaps.

¹ Harrison (1995) reviews early arguments.
2. UNDERSTANDING MONITORING AND ENFORCEMENT IN PRACTICE

How does environmental monitoring and enforcement work? How common are inspections and sanctions in the United States? How is environmental monitoring and enforcement changing over time? This section takes up these issues.

The discussion below emphasizes activity falling under the umbrella of the Environmental Protection Agency (EPA), and especially the regulatory setting under the three acts receiving the bulk of US monitoring and enforcement resource allocations. These acts are the Clean Air Act (CAA), the Clean Water Act (CWA), and the Resource Conservation and Recovery Act (RCRA). The discussion below also loosely describes institutions under the Safe Drinking Water Act (SDWA), the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), the Toxic Substances Control Act (TSCA), the Emergency Planning and Community Right to Know Act (EPCRA), and the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA). Institutions for environmental rules managed by agencies other than EPA, as well as institutions in other developed countries, are often similar in some respects and different in others.

2.1 Basic Structure

Legislation guiding environmental policy in the US is largely set at the federal level. In contrast, primary monitoring and enforcement responsibility is typically decentralized to states’ departments of environmental protection and local authorities. When states or localities have primary authority, dubbed ‘primacy,’ they are still required to regularly provide key activity metrics to EPA regional and federal offices. EPA offices regularly review state operations, and regional and national authorities also conduct their own inspections and issue their own

---

2 Scarlett et al. (2011) and Deily and Gray (2007) compare institutions across agencies. Harrison (1995), Rousseau (2007), Almer and Goeschl (2010), Billiet and Rousseau (2012), and Telle (2013) discuss institutions in selected developed countries outside of the US.
sanctions. EPA actions most often occur when and where decentralized enforcement is perceived as insufficiently rigorous, or when and where potential environmental impacts from specific violations are unusually large. EPA retains the right to revoke a state’s primacy under any given specific environmental statue, although such revocations are rare historically. States may decline primacy for resource or political economic reasons. EPA typically maintains primacy for newer and smaller environmental regulatory programs.

Executive oversight of environmental monitoring and enforcement is primarily indirect. The executive branch appoints the EPA administrator, the deputy administrators, and the assistant administrator for enforcement and compliance assurance. These individuals strongly influence enforcement intensity via budget negotiations with Congress, via within-agency resource allocations, and via impacts on bureaucratic culture. While direct executive office interventions on specific enforcement cases has been rare, presidential administration staffers have periodically been heavily involved in crafting policy initiatives that can impact monitoring and enforcement (Bressman and Vandenbergh 2006; Mintz 2012, pg. 182-183). A high-profile example from the mid-2000s was executive efforts to modify new source review (NSR) policies for power stations while several EPA cases on the matter were ongoing.

Legislative oversight of environmental monitoring and enforcement is also mostly indirect. Congress designs and passes the overarching environmental laws that specify enforcement authority and process. Legislatures authorize EPA budgets and can indirectly influence all EPA activities via budgeting choices. Congress must approve executive nominations for EPA leadership positions. Legislative authorities can also publicly express their collective preferences over enforcement intensity. For example, Congress publicly advocated for more aggressive aggregate environmental enforcement in the early 1980s and, as part of a
growing anti-regulatory platform, advocated for less aggressive aggregate environmental enforcement in the early 1990s. According to Mintz (2012)’s extensive interviews, however, the prevailing wisdom among enforcement officials is that individual Congressional committees or members are rarely directly involved in specific enforcement cases beyond routine requests for information.

2.2 Personnel and Instruments

Almost all state, regional, and federal monitoring and enforcement staff members are engineers and attorneys. Managers at both state and federal agencies are commonly internally-promoted attorneys or engineers. Because enforcement resources are consistently scarce, because environmental conditions at facilities are often complex, and because regulations can be vague and flexible, authorities at all levels have high levels of discretion regarding the frequency and severity of interventions. This discretion, in turn, can result in contentious interactions between enforcement authorities at different levels of government (Mintz 2012).

Nearly all pollution monitoring and enforcement activities in the US are conducted on a media-specific or statute-specific basis. Technical and legal complexities make other strategies challenging. The pulp and paper industry’s integrated “cluster rule” is a notable exception, as it was designed to coordinate water and air regulation for more efficient facility compliance through process modifications. Other exceptions include multiple-media monitoring and enforcement activities targeted towards industries or sources periodically identified as EPA “national priorities” or “national enforcement initiatives.” 2011 to 2013 priority industries include animal feeding, minerals processing, energy extraction, coal-fired utilities, and cement sectors.
Common environmental instruments include self-reporting and continuous emissions monitoring. For large facilities regulated under the CWA, several provisions of the CAA, and EPCRA, self-reported pollution data are the primary source of compliance monitoring information. Self-reporting is common under many other statutes as well. In most cases, facilities self-report pollution snapshots or longer-term pollution summary measures at the pollutant-point source level. Continuous emissions monitoring systems, applied on the largest scale under the CAA’s Title IV Acid Rain program for power stations, achieve approximately the same goals with real time measurements and with automatic reporting.

Regulator inspections help confirm the accuracy of self-reported data. Also, for statutes or facilities without extensive self-monitoring requirements, inspections serve as the dominant source of compliance monitoring information. Evaluations vary substantially in scope and scale across facilities, industries, statutes, states, and time. Low-intensity inspections may involve visual inspections of emissions and abatement equipment. Medium-intensity inspections may involve reviews of facility operations, maintenance, sampling, and reporting procedures. High-intensity inspections typically may involve extensive sampling by the regulator.

Inspections are conducted at specific facilities ‘for cause’ or, more commonly, for administrative reasons under the auspices of ‘neutral selection.’ ‘For cause’ inspections are typically associated with compliance history, citizen complaints, anonymous employee complaints, or facility characteristics correlated with frequent violations or significant damages. ‘Neutral selection’ inspections are based on time since last inspection and regulator cost factors, such as geographic proximity to other facilities scheduled to be inspected. Monitoring guidelines set inspection frequency targets for facilities, but these targets are generally not legally binding. Three features of the environmental inspection process are noteworthy to many economists.
First, purely random inspections do not fall under most definitions of “neutral selection” or “for cause.” Second, the agency increasingly uses environmental justice as a targeting consideration, as the natural vulnerability of populations near environmental justice facilities indicates the potential for significant environmental harm. Third, facilities are typically notified by authorities in advance of impending inspections, so on-site inspections are not a surprise to facilities.

When pollution violations are disclosed or detected, authorities have several enforcement options. Informal sanctions include warning letters, telephone calls, and notices of violation. These actions are most frequently carried out by the lowest-level authority with primacy. Formal sanctions may include field citations, but the vast majority are administrative orders issued by state authorities or by the administrative law judges associated with state and regional offices. Significant enforcement actions may also include state, regional, or national civil litigation. A small number of egregious violations are referred to state Attorneys General or the Department of Justice for criminal prosecution.3

Enforcement guidelines vary by statute. However, as a general rule, the frequency and severity of environmental sanctions is a function of one or more of the following: the extent of damages from the violation, the penalized facility’s financial gain from the violation, the facility’s compliance and enforcement history, the economic impact of the penalty on the violator, and the violator’s intent. Fairness and the strength of the legal evidence can also influence sanction magnitudes. In practice, authorities typically pursue the minimum sanction necessary to achieve a return to compliance and some longer-run deterrence objective. Maximum

---

3 Facilities may appeal all types of sanctions. Penalized facilities may appeal administrative penalties to state administrative appeals boards or to the independent Environmental Appeals Board (EAB) in Washington, DC. Facilities facing civil or criminal penalties may appeal to state or federal courts.
penalties allowable under the law are rarely assessed, if ever.\textsuperscript{4} Administrative penalties are strongly prioritized over civil penalties. All else equal, enforcement attorneys try to avoid time consuming and costly litigation. Criminal referrals are especially rare, and occur for cases with demonstrable attempts at falsification and evasion, deliberate attempts to sidestep the regulatory environment altogether, or extreme environmental damages (Uhlmann 2009).

Environmental authorities’ general approach to pollution monitoring and enforcement has remained largely constant for decades. However, the EPA is now investigating, promoting, and beginning to implement programmatic changes dubbed ‘next generation compliance.’ Next generation environmental compliance exploits recent technological advances and more holistic rulemaking to increase the efficacy and efficiency of monitoring and enforcement. The paradigm has five key pillars: rules where compliance is the default; advanced pollution monitoring; electronic reporting; enhanced information disclosure; and innovative enforcement strategies, including 3\textsuperscript{rd} party certifications (Giles 2013).

\textbf{2.3 Levels and Trends}

Oversight of federal environmental monitoring and enforcement was consolidated in 1994 into the EPA’s Office of Enforcement and Compliance Assurance (OECA). In the first year of operation, OECA received a ~$690 million budget (~$436 million in 1994 dollars, adjusted to 2013 dollars). Figure 1 shows that real budgets, however, steadily declined for the next several years. Real OECA budgets then leveled off in the ~$600-$620 million range, where they remain. Note that budgets in Figure 1 represent lower bound estimates of federal environmental enforcement expenditures, since OECA regularly partners with other offices and agencies that have their own resources. Moreover, while enforcement budgets in Figure 1 include federal

\textsuperscript{4} The most common statutory maximum is $25,000 per day in violation. A high-profile environmental defense attorney referred to these maximums as “off the charts” (Mintz 2012, pg. 16).
grants to states for enforcement and compliance assurance, they do not include resource allocations from the states themselves.\(^5\)

Between 1994 and 2011, the EPA conducted approximately 19,850 inspections per year. This estimate does not include state-led inspections, and thus significantly understates total inspection counts.\(^6\) Total EPA inspections steadily increased from around 14,500 in 1995 to 23,200 in 1998, and then generally leveled off to a steady state level of between 18,000 and 22,000 per year. Figure 2 shows that inspection levels and trends vary significantly across statute. After a sharp increase in the mid-1990s, CWA inspections generally fluctuated in the 3,500 to 4,500 range. The EPA conducted about 3,100 RCRA inspections between 1994 and 2001 on average, but levels sharply increased and then sharply decreased over the period. EPA carried out approximately 2,800 CAA inspections per year, but levels fell below 1,000 in 2002 and almost reached 4,000 in 2006. The greatest number of federal inspections between 1994 and 2001, about 7,200 per year, were conducted under the auspices of the Safe Drinking Water Act.

Administrative penalty orders are formal requirements to return to compliance that are accompanied by financial penalties. Federal and regional EPA offices levied nearly 1,800 administrative penalty orders per year, on average, between 1994 and 2011. This estimate, however, does not include state administrative penalty orders or EPA complaints unaccompanied by monetary sanctions. It therefore understates total administrative action counts. Figure 3 shows that EPA administrative penalty orders varied significantly across time and statute. EPA-led administrative penalties under the CWA increased between 1991 and 2005, and administrative penalties under RCRA and the CAA increased between 1998 and 2008. Penalty magnitudes also

\(^5\) I am unable to locate systematic estimates of state expenditures on environmental enforcement and compliance assurance. Informal estimates tend to suggest that incremental state enforcement expenditures across all states may be of similar magnitude to total federal enforcement expenditures.

\(^6\) Systematic data on inspection type and inspection intensity over time is not immediately available.
varied significantly across statute and time. The mean (nominal) EPA administrative penalty over all statutes between 1991 and 2005 was approximately $17,500. Median penalties are far lower.\footnote{Gray and Shimshack (2011) reported that median administrative penalties between 2001-2008 were approximately $7,850 under CERCLA, $3,000 under CWA, $7,200 under EPCRA, $600 under RCRA, and $3,600 under TSCA. Gray and Shimshack (2011) median penalties are calculated from data that include state administrative penalties.}

As noted earlier, civil and criminal referrals to the Department of Justice (DOJ) for pollution violations are infrequent. Between 1994 and 2011, EPA referred 288 cases per year to DOJ for civil litigation on average. The bulk of these referrals were for CERCLA, CAA, and CWA violations. Civil judicial case numbers trended downward over time, with the total number of EPA civil referrals generally varying between 300 and 400 in the mid-1990s and between 200 and 300 in the late 2000s. Total penalty amounts from civil judicial cases averaged about $95 million per year. Between 1998 and 2011, EPA referred 420 cases per year to DOJ for criminal litigation on average. Criminal judicial case numbers also trended downward over time, with the total number of EPA criminal referrals generally varying between 400 and 500 in the late 1990s and between 300 and 400 in the late 2000s. Financial penalties from criminal cases averaged approximately $85 million per year, with accompanying average annual jail sentences totaling approximately 134 years.

Fully characterizing levels and trends of state-led pollution monitoring and enforcement is beyond the scope of this paper. Figure 4, however, highlights significant heterogeneity across states. Figure 4 presents basic data on state-led compliance evaluations of CAA facilities, state-led formal enforcement actions levied against CAA facilities, state-led inspections of CWA major facilities, and state-led enforcement actions levied against CWA major facilities. To highlight some differences, consider that in 2011, NC led compliance evaluations at over 95% of its CAA facilities while NY led compliance evaluations at just over 10% of its CAA facilities.
PA led inspections at over 90% of its CWA majors while NY led inspections at less than 15% of its CWA majors. TX issued enforcement actions at over 50% of its CWA majors while PA issued enforcement actions at under 10% of its CWA majors. Readers are cautioned not to over-interpret Figure 4, as cross-state differences are attributable to many possible factors. Such factors include, but are not limited to, differences in industrial composition, facility characteristics, regulatory severity, regulatory action history, and data quality.

2.4 Private Monitoring and Enforcement

Nearly every major environmental statute allows for citizen suit enforcement (Naysnerski and Tietenberg 1992). Here, private groups like the Natural Resources Defense Council and Riverkeeper initiate lawsuits against polluters or government authorities to promote compliance and reduce pollution. Eight Citizen suits are typically permitted only when public regulators fail to pursue noncompliance. Further, statutes require that cases targeting polluters are preceded by a 60 day notice of intent to both authorities and polluters (Langpap and Shimshack 2010). During this period, public agencies are often able to preempt the suit with their own actions.

During the 1980s and 1990s, approximately 50-150 citizen suits were prosecuted per year. Citizen suits related to Clean Water Act violations were, and continue to be, far more common than suits under other statutes. For example, Smith (2004) suggests approximately 88 percent of suits between 1995 and 2000 were filed under the CWA provisions. Extensive self-reporting under the CWA allows citizen groups to regularly observe compliance at the facility level.

3. ECONOMIC THEORIES OF MONITORING AND ENFORCEMENT

---

8 Community organizations, and especially watershed groups, also regularly informally alert authorities of unusual discharges, raise community awareness of local pollution, and work with facilities to improve environmental performance. See Grant and Grooms (2012) and Grant and Langpap (2013) for more discussion of this informal monitoring and enforcement sector.
What does optimal environmental monitoring and enforcement look like? Why might monitoring and enforcement in the real world differ from optimal monitoring and enforcement? This section takes up the theoretical account. The reader should note that much of the theory of environmental monitoring and enforcement is developed as part of the broader theory of the public enforcement of law. Polinsky and Shavell (2000) and Shavell (2004) survey this literature in depth, so the present article focuses on key insights and influential citations.

Justifications for public enforcement, in lieu of private tort and contract law, are the natural theoretical point of departure. Becker and Stigler (1974), Landes and Posner (1975), and Polinsky and Shavell (2000) review the key arguments. Three reasons favor the public enforcement of law, and perhaps especially so in the context of environmental enforcement. First, information asymmetries may make it difficult for victims to identify the source of the harm. Second, there may be economies of scale and natural monopolies in monitoring and enforcement technologies. Third, positive externalities may arise from harm reduction or negative externalities may arise from private monitoring.

Before proceeding, it may be useful to review the typical utilitarian view of individuals’ compliance motivations in the presence of enforcement. This framework has been understood since at least Bentham (1789), although the theory was formalized in the seminal work of Becker (1968). Here, a rational risk-neutral agent considers a privately beneficial action that may also generate harm to third parties. The agent compares their expected benefits of the action with the expected cost of that action. In a setting with public enforcement of law, the expected cost of the action is related to the probability of detection and the magnitude of the sanction if detected. Under strict liability, where a penalty is imposed for realized harm regardless of intent or care, a risk-neutral agent commits the harmful act if expected benefits exceed expected penalties. Under
fault-based liability, where a penalty is imposed for realized harm only if the act is socially undesirable, a risk-neutral agent commits the harmful act when expected gains are high enough to avoid fault.\(^9\) Obvious extensions follow for act-based analogs to both strict and fault-based liability, where penalties are based on ex-ante expected harm rather than ex-post realized harm. Act-based liability is an especially appropriate framework when harm is treated as stochastic.\(^10\)

### 3.1 Prescriptive Theory

The most fundamental law and economic theories of enforcement explore various features of a social-welfare maximizing enforcement system. The most common baseline models assume costly monitoring and apprehension, costless imposition of monetary sanctions, and certain imposition of monetary sanctions conditional on detected violation. Common baseline models also assume Becker-style rational agents with no wealth constraints, who make binary decisions over a single course of action.

The simplest starting point is a model assuming the probability of detection is fixed at one. With strict liability or its act-based analog, an optimizing enforcement agency imposes an expected sanction equal to the expected harm. With fault-based liability or its act-based analog, an optimizing enforcement agency imposes an expected sanction equal to the expected harm provided the action is socially undesirable. In all cases, socially undesirable acts are deterred and socially desirable acts are left undeterred. A natural extension to the simplest model assumes the probability of detection is fixed at some probability greater than zero but less than one. Optimal sanctions are now based on expected harm divided by the probability of detection.

Beyond the identification of optimal sanction rules, two insights arise from simple models with fixed detection probabilities (Shavell 2004). First, when agents are risk-averse,

---

\(^9\) Socially undesirable in this context refers to a net or aggregate concept, i.e. when social harm exceeds private gain.

\(^10\) Shavell (2004) reviews advantages and disadvantages of the various liability rules, including those related to agency enforcement costs, agent risk bearing costs, and Coasian transaction costs (Coase 1960).
optimal penalties will be lower than the optimal sanctions when agents are risk-neutral. The intuition is that risk bearing is costly and that risk-averse agents are more easily deterred. Second, optimal sanctions may bear little resemblance to sanctions perceived as ‘fair’ or ‘reasonable.’ Sanctions for acts when the probability of detection is low can be very high, even when harm itself is relatively low. For example, a minor pollution violation that is unlikely to be detected is only deterred with optimal penalties that may seem extreme to the casual observer.

More nuanced models assume that a welfare-maximizing regulator chooses both the probability of detection and the magnitude of the sanction conditional of violation. Here, an optimizing enforcement agency chooses a low probability of detection and a high sanction. Indeed, one implication of Becker (1968) is that the optimal sanction with risk-neutral agents is often at or near its maximum allowable level. 11 The intuition is that the regulator saves monitoring resources without sacrificing deterrence. An optimizing enforcement agency will also choose a probability of detection that leads to some under-deterrence. In other words, the regulator chooses a monitoring probability that is lower than the monitoring probability necessary to equate the expected sanction with the expected harm. The intuition is that it is beneficial to allow some socially undesirable activities to save enforcement resources.

Some models allow for both monetary sanctions and non-monetary sanctions. Non-monetary sanctions naturally include incarceration, but may also include penalties like reputation damage. Under the assumption that non-monetary sanctions are costly to impose and monetary sanctions are costless to impose (or nearly so), the key insight is that an optimizing regulator will issue maximum monetary sanctions before issuing any non-monetary sanctions. Stated differently, an optimizing regulator will only resort to non-monetary sanctions when monetary

---

11 Maximum legal penalties may be defined in statutes, or penalties may be politically constrained. Marginal deterrence, discussed later, may also lead to constrained penalties for many actions with lower levels of harm.
sanctions are unlikely to achieve deterrence. In practice, these conditions typically arise when monetary penalties are constrained, when the probability of detection is limited, when agent wealth is constrained, and/or when the private gains from committing the harmful act are unusually high (Shavell 2004).

3.2 Extensions to Prescriptive Theory

Extensions to the basic prescriptive theory include models allowing errors in liability, imperfectly observable enforcement parameters, costly monetary sanctions, and multiple harmful actions. In models with errors in liability, the key insight is that all errors reduce deterrence. Type I errors reduce expected sanctions of noncompliance and Type II errors reduce the gap between expected benefits from compliance and expected benefits of noncompliance. Optimal sanctions are therefore higher with errors than optimal sanctions without errors. In models with imperfectly observable enforcement parameters, the key feature is that changes in perceived detection probabilities and perceived sanctions impact deterrence more than changes in actual detection probabilities and sections. In models with costly sanctioning, the key point is that optimal sanctions need to be higher than optimal sanctions with costless monitoring. The intuition is that expected penalties need to reflect both expected harm and expected imposition costs. In models where agents may choose multiple harmful acts, the key insight is the notion of ‘marginal deterrence.’ Marginal deterrence across acts with different levels of harm requires optimal sanctions that rise with damages. The key complication is that maintaining sufficiently steep penalty schedules may require artificially low sanctions and under-deterrence for lower-level harm (Mookerjee and Png 1994).

---

12 Segerson and Tietenberg (1992) consider more nuanced optimal agency trade-offs between monetary and non-monetary sanctions when principle-agent issues within the firm complicate compliance incentives.

13 Sanctioning may be costly for documentation, negotiation, litigation, political backlash, or other reasons.
One extension that has received particular attention in the environmental context involves the possibility of avoidance activities. Malik (1990a) models the implications of agents trying to lower the probability of being sanctioned by engaging in evasion, lobbying, or concealment efforts.\textsuperscript{14} The key point here is that larger penalties increase incentives for avoidance activities, so an optimizing enforcement agency may not choose the highest possible sanction. Heyes (1994) expands these ideas to a case where a regulator chooses between more frequent inspections and more thorough inspections. More frequent inspections provide incentives for concealment activities while more thorough inspections provide incentives for transparency. Heyes (1994) argues that an optimizing enforcement agency should emphasize thoroughness.

Another relevant extension addresses self-reporting. The key insight here is that enforcement systems with self-monitoring can be incentive compatible while lowering enforcement costs and enhancing regulator efficiency. Malik (1993) showed that regimes with self-reporting can be welfare improving if less frequent regulator inspections are coupled with more frequent regulator punishment. Kaplow and Shavell (1994) demonstrated that regimes with self-reporting can be implemented without affecting agents’ incentives. The intuition is that equivalent deterrence can simply be achieved with penalties for self-disclosed violations equal to (or, in some cases, somewhat less than) the certainty equivalent of expected penalties without self-reporting. Kaplow and Shavell (1994) also note that an additional advantage of self-reporting is a reduction in agents’ risk bearing costs. Innes (1999a, 1999b, 2001) shows that the welfare enhancing effects of self-reporting may be especially high if agents might be expected to engage in ex-ante avoidance activities and/or ex-post remediation effects.

\textsuperscript{14} Linder and McBride (1984) provide earlier attention to the possibility of concealment activities in an environmental context. However, that paper is ultimately concerned with enhancing enforcement performance in a decentralized regulatory setting.
A third pertinent extension in the environmental context involves the monitoring and enforcement of market-based pollution control instruments. The key point is that, under a pollution tax or cap-and-trade system, the marginal benefit of noncompliance is typically the per unit pollution price (Stranlund and Dhanna 1999). 15 Since the pollution price and therefore the marginal benefits of noncompliance do not vary across facilities, compliance decisions may be independent of abatement costs. Implications for enforcement targeting follow. See Stranlund (2013) for further discussion.

3.3 Descriptive Theory

In the environmental context, several features of observed regulator behavior are broadly consistent with prescriptions of the optimal enforcement models discussed above. Environmental sanctions are frequently a function of harm or damages. Under-deterrence may be widespread. Monetary penalties are common relative to incarceration, and criminal sentences are typically reserved for only those cases where deterrence with monetary sanctions may be especially difficult. Statutes have penalty schedules that often reflect the general notion of marginal deterrence. Self-reporting is common. Market-based regulatory systems are enforced differently than command-and-control regulatory systems.

However, other features of observed regulator behavior may be largely inconsistent with prescriptions of traditional law and economic theory. A broad descriptive literature, then, attempts to explain these ‘departures’ from economically optimal behavior in several ways. Many descriptive theories consider public enforcement agencies as captured regulators, in the spirit of Stigler (1971) and Peltzman (1976). Others conceptualize enforcement bureaucracies as

---

15 Under additional conditions described in Stranlund and Dhanna (1999), the expected penalty may have little direct influence on the true level of the harm and expected penalties only impact incentives for truthful reporting. Under a tradable permits system, the expected penalty may have only an indirect influence on the level of harm via impacts on permit prices. These indirect impacts, however, can have important implications for the stability and efficiency of transferable discharge permit markets (Malik 1990b). See Stranlund (2013) for greater discussion.
conflict minimizing or attention avoiding, along the lines of Hilton (1972), Joskow (1974), and Leaver (2009). Some theorists model enforcement agencies as budget maximizers, in the spirit of Niskanen (1971). Yet others consider enforcement agencies as maximizing compliance subject to strong institutional constraints, or as minimizing enforcement costs subject to target compliance rates. See, for example, Garvie and Keeler (1994).

An especially frequent motivation for descriptive theories of environmental enforcement is a common belief that static law and economic models would predict compliance rates that are well below those observed in real world pollution control contexts. Observed inspection rates, sanction probabilities, and penalty magnitudes seem too low to generate high compliance. In an attempt to rationalize this apparent puzzle, Harrington (1988) proposed a dynamic repeated-game between a regulated firm and an environmental enforcement authority facing highly constrained penalties. Harford (1991), Harford and Harrington (1991), Harford (1993), Friessen (2003), and others refined the basic model. The essence of regulator behavior in these studies is a state-dependent strategy, where the regulator adjusts inspection frequency and/or sanction intensity based on agents’ past performance. Generally compliant agents will face infrequent inspections and/or low sanctions for violations, and generally noncompliant agents will face frequent inspections and/or high sanctions for violations. The expected penalties of noncompliance therefore include both immediate sanctions and additional sanctions for violations that would have been otherwise undetected or lightly punished in future periods. This dynamic ‘enforcement leverage’ may be used to enhance compliance, in many cases cost effectively, beyond what might be achieved with static enforcement.

Other descriptive theories attempting to explain simultaneously high compliance rates and low enforcement intensity include Livernois and McKenna (1999) and Heyes and Rickman

---

16 Harrington (1988) is an application and extension of Landsberger and Mielijson (1982).
Livernois and McKenna (1999) study the sanctioning policy of a cost-minimizing regulator managing an enforcement regime with self-reporting and an exogenous goal of achieving a fixed compliance rate. They show that increasing sanctions for violations lowers the incentive for truthful reporting, and therefore lower penalties could possibly increase compliance in some circumstances. Heyes and Rickman (1999) study regulator-firm interactions in the presence of multiple compliance domains. Here, a regulator may enhance overall compliance by allowing some noncompliance in one domain in exchange for compliance in another domain. As noted in Heyes (2000), the ‘regulatory dealing’ of Heyes and Rickman (1999) can be thought of as a cross-sectional analog of dynamic enforcement leverage games.

Other prescriptive and descriptive theories going beyond simple Becker-style models emphasize jointness in pollution production. In the general spirit of Heyes and Rickman (1999), Shimshack and Ward (2008) illustrated that compliance for a given pollutant will be influenced by expected penalties for different pollutants generated in the same production process. Becker-style models typically consider pollutants individually. With pollution jointness, optimizing agencies should design and implement permitting, inspection targeting, and enforcement strategies holistically.

Other theories, typically drawing from non-economic disciplines or behavioral economics, explore both prescriptive and descriptive aspects of monitoring and enforcement when agents are motivated to comply for reasons beyond rational utilitarian calculations (Ayres and Braithwaite 1992, Thornton et al. 2005). One insight is that observed compliance may be motivated by moral or ideological beliefs and values (Burby and Paterson 1993). Under these conditions, an optimizing agency may use ‘carrots’ as well as ‘sticks’, regularly provide services, and impose sanctions widely seen as fair and prompt. A second insight is that observed
compliance may be motivated by social norms (Scholtz 1984; Winter and May 2000). Under these conditions, an optimizing agency may foster long-term personal relationships and work to promote a culture where compliance is seen as the norm. A third insight is that some observed noncompliance may emerge because agents don’t fully understand complex and changing requirements (Spence 2001; Stafford 2006). Here, an optimizing agency may offer extensive compliance assistance.

A final strand of the descriptive theory literature focuses on relationships between public enforcement and private citizen suit enforcement. Heyes and Rickman (1999) predict that citizen suits may crowd in or crowd out public monitoring and enforcement, depending on the presence and extent of regulatory dealing. Langpap (2007) explores more fully endogenous relationships in more detail. A key result is that citizen suits with high expected penalties are likely to crowd out public enforcement, and in many cases this crowd out will enhance the efficiency of public agency behavior.

4. THE EMPIRICAL ACCOUNT

How do environmental authorities determine inspection probabilities and sanction magnitudes in the real world? Do environmental monitoring and enforcement interventions enhance compliance? Do environmental inspections and sanctions reduce pollution? This section takes up the empirical account.

4.1 Empirical Determinants of Enforcement Agency Behavior

Empirical evidence suggests that environmental enforcement authorities regularly consider benefits and costs. Regarding benefits, higher emissions and damages are frequently associated with increased inspection probabilities (Dion et al. 1998; Stafford 2002; Gray and Shadbegian 2004). Administrative, civil, and criminal penalty magnitudes typically increase with
harm (Eppler and Visscher 1984; Cohen 1992; Kleit et al. 1998; Oljaca et al. 1998). CAA enforcement actions may be more common in nonattainment areas (Gray and Deily 1996). Regarding administrative costs, states with higher paid employees conduct lower intensity CWA inspections on average (Helland 1998). Facilities recently inspected under the CWA are less likely to be immediately inspected again (Helland 1998; Shimshack and Ward 2005). Storage tank inspections are more likely when nearby facilities are also inspected (Eckert and Eckert 2010). Regarding compliance costs, EPA and states direct fewer monitoring and enforcement actions towards facilities that are important local employers or have especially high probabilities of shutdown (Deily and Gray 1991; Gray and Deily 1996; Helland 1998).

Other observed determinants of enforcement intensity may be consistent with direct agency benefits and costs, but overall welfare effects are more ambiguous. First, facilities’ compliance history is typically an important determinant of monitoring and enforcement activity (Kleit et al. 1998; Oljaca et al. 1998; Stafford 2002; Eckert and Eckert 2010). The presumed intuition is that recent violators may be more likely to violate again. Second, while authorities are more likely to inspect facilities with a higher threat of private citizen suits, they are less likely to penalize these facilities (Langpapa and Shimshack 2010, Ashenmiller and Norman 2011). This enforcement crowd out effect is consistent with theory and may reflect shifting resource allocations away from settings where private interventions are already influencing deterrence. Third, regulators respond to enforcement conditions in other jurisdictions. Federal CWA inspections are more common after state CWA inspections, and vice versa (Earnhart 2004a). Surface mining regulators lower enforcement intensity in response to lower enforcement intensity in nearby states, a result that may be consistent with an enforcement race to the bottom.

17 Earnhart (2000) explored citizen suits and public enforcement in the Czech Republic, but did not explore the extent of crowd in or crowd out between them.
Somewhat similar strategic interactions are observed for CAA, CWA, and RCRA enforcement (Konisky 2007).

Regulator actions that are wholly inconsistent with direct benefit and cost comparisons are also readily observed. CWA and CAA inspection propensities are related to Congressional representatives’ voting scores and committee memberships, perhaps suggesting the importance of bureaucratic interest (Helland 1998; Innes and Mitra 2011). Highly corrupt states pursue more lax environmental oversight, relative to less corrupt states, after receiving enforcement primacy (Grooms 2012). Many authors find that both inspection and enforcement probabilities are closely related to community characteristics. Most notably, characteristics associated with political activism, such as income, education, voter turnout, and environmental group membership, appear especially influential for state-level interventions (Earnhart 2004a; Earnhart 2004c; Helland 1998).

4.2 Empirical Investigations of Deterrence

The evidence from environmental settings indicates that monitoring and enforcement actions get results. Most directly, requirements of administrative compliance orders and judicial resolutions help reduce immediate environmental harm. For example, the EPA asserts that its 2011 federal actions resulted in 3.6 billion pounds of hazardous waste treated and 1.6 billion pounds of air and water pollution reduced.\footnote{Source: EPA Compliance and Enforcement Annual Results: 2011 Fiscal Year. Online at: \url{http://www.epa.gov/compliance/resources/reports/endofyear/eoy2011/index.html}.}

More importantly, the empirical deterrence literature consistently finds that regulated facilities adjust subsequent environmental behavior following inspections, sanctions, or increased threats of inspections and sanctions. Environmental monitoring and enforcement actions generate \textit{specific deterrence}, meaning they improve future performance at the evaluated
or sanctioned facility. Environmental monitoring and enforcement actions generate *general deterrence*, meaning that they spillover to improve future performance at facilities other than the evaluated or sanctioned facility. Finally, environmental monitoring and enforcement actions appear to even generate *beyond compliance behavior*, meaning that they can induce facilities to reduce pollution well below legally allowable levels.

### 4.2.1 Empirical Investigations of Specific Deterrence

The empirical literature finds relatively consistent evidence for specific deterrence under air quality regulations. EPA enforcement actions were followed by enhanced compliance with CAA regulations in the steel industry during the 1970s and 1980s (Gray and Deily 1996; Deily and Gray 2007). EPA and state monitoring and enforcement actions reduced both the duration of noncompliance and the rate of noncompliance in the pulp and paper industry during the 1980s (Nadeau 1997; Gray and Shadbegian 2005). In the late 1990s and early 2000s, coal-fired power plants facing threats of New Source Review lawsuits reduced emissions relative to those plants not facing similar threats (Keohane et al. 2009). EPA air compliance evaluations reduced aggregate TRI-reported toxic emissions across several manufacturing industries in the 1980s, 1990s, and early 2000s (Hanna and Oliva 2010).

The empirical literature also consistently finds evidence of specific deterrence under water quality regulations. Both inspections and increased threats of inspections enhanced compliance with water quality regulations in the US and Canadian pulp and paper industries during the 1980s (Magat and Viscusi 1990; Laplante and Rilstone 1996). Sanctions, and especially federal fines, were associated with subsequent pollution reductions at wastewater treatment plants during the 1990s and at chemical facilities during the late 1990s and early 2000s (Earnhart 2004a, Earnhart 2004b; Glicksman and Earnhart 2007). Formal enforcement actions
with monetary penalties impacted compliance and pollution at pulp and paper plants during the 1990s and 2000s (Shimshack and Ward 2005; Shimshack and Ward 2008).

Evidence for specific deterrence arises in many other environmental settings as well. Three seminal papers found that some monitoring activities resulted in reduced oil spill frequency and oil spill size (Epple and Visscher 1984; Cohen 1987; Grau and Groves 1997). While some enforcement interventions failed to significantly impact gas and liquid pipeline operation during the late 2000s and early 2010s, federal cases initiated against operators may have improved many aspects of environmental performance (Stafford 2014). Deterrence effects of inspections under Canadian petroleum storage regulations appear small, but positive nonetheless (Eckert 2004). Rule changes increasing liability or penalties significantly reduced hazardous waste violations and toxic releases in the late 1980s and 1990s (Stafford 2002; Stafford 2003; Alberini and Austin 1999; Alberini and Austin 2002).19

Empirical evidence for specific deterrence is not restricted to North American contexts. Plants facing regulatory inspections in Mexico self-reported increased compliance during the 1990s (Dasgupta et al. 2000). While evidence on deterrence in European contexts is surprisingly rare (Tosun 2012), it appears that increased enforcement did reduce environmental crime and waste dumping in Germany during the 1990s and 2000s (Almer and Goeschl 2010; Almer and Goeschl 2013). The probability of detection strongly influenced Danish farmers’ compliance with agro-environmental regulations during the late 1990s (Winter and May 2001). Inspections reduced subsequent air and water pollution emissions at manufacturing facilities in China in the 1990s (Dasgupta et al. 2001).

---

19 Sigman (2009) demonstrated that liability rule changes for toxic pollution can have important consequences for property markets.
Despite a strong consensus in favor of specific deterrence, a handful of empirical studies find that some interventions reduce environmental performance. While this finding could be consistent with some non-economic theories of deterrence, it seems more likely that these studies failed to adequately address the non-random nature of environmental inspections and sanctions. Gray and Shimshack (2011) discuss endogeneity challenges arising in observational deterrence investigations, as well as possible empirical solutions to these challenges.

Observational endogeneity also suggests a promising and growing role for experimental evidence. Lab-based experiments of basic deterrence theory find that increasing the probability of detection or the sanction for noncompliance increases compliance (e.g. Anderson and Stafford 2003; Friesen 2012). Murphy and Stranlund (2006) demonstrated that enforcement pressure reduced emissions in laboratory permit markets, but only via changes in the allowance prices as predicted by theory. Telle (2013) provided compelling evidence from a natural experiment in Norway that inspections raised environmental compliance. Duflo et al. (2013) found that reformed incentives for third party auditors reduced false self-reports and reduced pollution at plants in the Indian state Gujarat.

4.2.1 Empirical Investigations of General Deterrence and Beyond Compliance Behavior

Scholars and regulators have long believed that monitoring and enforcement actions spillover to enhance compliance for agents other than the inspected or sanctioned agent. The intuition is that monitoring and enforcement actions enhance the regulator’s reputation for toughness, and plants update their beliefs in response to new perceptions of regulatory stringency. In a qualitative survey of regulated companies in Oregon, Carlough (2004) found that ten to forty percent of respondents reported making changes in response to hearing about

---

20 Stranlund (2011) and Friesen and Gangadharan (2013) more completely discuss the laboratory evidence on enforcement in environmental markets.
inspections or penalties at other Oregon facilities. Percentages were higher for larger facilities. Thornton et al. (2005) surveyed US industrial facilities and found that nearly ninety percent of respondents were aware of at least some enforcement actions at other firms. Sixty percent of respondents reported making changes in response to hearing about other facilities’ penalties. However, recall about specific ‘signal cases’ at other facilities in the same state and sector was often imperfect.

Shimshack and Ward (2005) provided early quantitative evidence for general deterrence. The deterrence effects of monetary penalties for violations in the pulp and paper industry during the 1990s were almost as strong for other facilities as they were for the sanctioned facility itself. Further, these general deterrence effects had significant effects on aggregate compliance. The statewide noncompliance rate for conventional water pollutants fell by nearly two-thirds in the year following a fine on any plant in the state. Gray and Shadbegian (2007) also found that interventions at manufacturing plants during the late 1990s enhanced compliance at both the evaluated facility and at facilities located nearby. Spillover effects, however, were limited by jurisdictional boundaries. Facilities located nearby, but across state lines, did not increase compliance in response to others’ inspections.

General deterrence may also stem from private enforcement actions. Langpap and Shimshack (2010) found that violations at all wastewater plants in a state fell significantly following citizen suits against a wastewater treatment plant in that state. This general deterrence effect, however, was attenuated by the net crowd out of public enforcement. In other words, citizen suits enhance compliance, but direct deterrence effects are overstated because citizen suits in a state and sector reduce public enforcement in that state and sector.
Economists typically view all types of enforcement as a means of achieving compliance. Recent evidence suggests that public enforcement actions may also increase beyond compliance or over-compliance behaviors. Shimshack and Ward (2008), using data from the pulp and paper industry during the 1990s and 2000s, found that increased expected penalties induced facilities that typically comply to lower pollution even further below legally permitted levels. Moreover, facilities that may have been noncompliant reduced discharges to well below permitted levels when expected penalties increased. These empirical results are puzzling when interpreted through the lens of a simple deterministic one-pollutant model of the firm. However, Shimshack and Ward (2008) demonstrated that beyond compliance behavior can be fully rationalized by economic theory. That paper found practically and economically significant evidence that stochastic pollution and jointness in pollution production drive the results. In periods of high perceived regulatory scrutiny, facilities lower target discharges to reduce the probability of accidental violations from randomness. Similarly, in periods of high regulatory scrutiny, facilities lower target discharges of one pollutant to reduce the probability of a violation for a jointly produced co-pollutant.

4.2.3 Extensions to the Empirical Deterrence Literature

The environmental deterrence literature increasingly explores the heterogeneity of enforcement responses across facility characteristics. As a general rule, the impact of inspections and sanctions varies across facility size, firm size, firm financial status, industrial sector, community characteristics, permit conditions, and other characteristics (e.g. Alberini and Austin 2002; Carlough 2004; Gray and Shadbegian 2005; Deily and Gray 2007; Earnhart 2009; Hanna
and Oliva 2010; Earnhart and Segerson 2013). It is not yet clear, however, if and when deterrence heterogeneity is systematic or generalizable.21

The impact of environmental monitoring and enforcement activities also varies across instruments. State inspections, federal inspections, state administrative sanctions, federal administrative sanctions, civil penalties, and criminal penalties generate different deterrence effects on average. Laboratory and empirical evidence suggests that marginal changes in sanction magnitudes may impact deterrence more than marginal changes in the probability of detection (Anderson and Stafford 2003; Friesen 2012; Nadeau 1997; Earnhart 2004a; Shimshack and Ward 2005; Glicksman and Earnhart 2007). This result is not universally found, however. Federal actions seem to generate larger deterrence effects than state actions, and civil penalties seem to generate deterrence impacts at least as large as administrative penalties (Earnhart 2004a, Miller 2005, Earnhart 2009). The evidence comparing deterrence from criminal sanctions with deterrence from other sanctions is sparse, but Miller (2005) found that DOJ criminal penalties reduced environmental recidivism rates at US companies more than administrative or civil penalties.

5. DISCUSSION

What have we learned? First, environmental monitoring and enforcement actions are effective on average. Inspections and sanctions directly reduce pollution, deter future violations, and even encourage beyond compliance behavior at the monitored or sanctioned facilities. Inspections and sanctions spillover to deter violations and reduce pollution at facilities other than the monitored or sanctioned facility. Second, current environmental monitoring and enforcement practices are unlikely to be fully cost effective. Environmental regulators do not appear to strictly

---

21 Murphy and Stranlund (2007)’s experimental evidence showed that, as predicted by theory, enforcement responses don’t seem to vary significantly with agent characteristics when the regulation is price-based.
maximize social welfare, and observed regulator behavior often departs from optimal enforcement.

The above lessons notwithstanding, we have a lot left to learn about how - and how well - monitoring and enforcement works. First, how effective are monitoring and enforcement interventions in an international, and especially developing country, context? It remains unclear if and when empirical lessons from the North American experience translate to other contexts. Second, how do social norms, social dynamics, and economic psychology influence monitoring and enforcement? Environmental economists have been slow to investigate the causes and consequences of behavioral complications, social interactions, and compliance motivations other than simple utilitarian calculations. Third, how does the practice of monitoring and enforcement match the theory? My sense is that many worthy existing hypotheses remain untested.

We also have a lot left to learn about the social trade-offs involved in monitoring and enforcement. First, are traditional monitoring and enforcement instruments efficient? While the evidence indicates that enforcement actions get results, it is not clear how benefits compare to administrative and compliance costs. Second, are monitoring and enforcement instruments cost effective? Ideally, we would understand how the environmental ‘bang per buck’ from traditional inspections and sanctions compares to the environmental ‘bang per buck’ from voluntary, cooperative, and information-based alternatives. Similarly, when are alternative policy instruments complements to traditional enforcement, and when are alternative policy instruments substitutes to traditional enforcement? Voluntary programs and transparency programs can both leverage and undermine enforcement.22

---

22 I do not mean to suggest that the topics in the final two paragraphs have not been studied at all. Dasgupta et al. (2001) and Lo et al. (2009) study developing country contexts; May and Winter (2000), Winter and May (2001), and Muehlenbachs et al. (2013) explore social norms and social interactions; Helland (1998a) and Heyes and Rickman (1999) test prominent theories; Cohen (1986) and Magat and Viscusi (1990) perform benefit-cost analyses; Foulon
DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGEMENTS

I thank Michael Ward and Wayne Gray, as many of the ideas in this paper came out of joint work. Jon Silberman and Dinah Koehler provided helpful discussions of related work.
LITERATURE CITED


Almer C, Goeschl T. 2010. Environmental crime and punishment: empirical evidence from the German penal code. Land Econ. 86: 707-726


Billiet CM, Rousseau S. 2012. How real is the threat of imprisonment for environmental crime? Eur. J. Law Econ. 1-16


Coase, RH. 1960. The problem of social cost. J Law & Econ. 3: 1-44


Earnhart D. 2004c. The effects of community characteristics on polluter compliance levels. Land Econ. 80: 408-432


Eppele D, Visscher M. 1984. Environmental pollution: modeling occurrence, detection, and deterrence. J. Law and Econ. 27: 29-60


Hilton GW. 1972. The basic behavior of regulatory commissions. *Amer. Econ. Rev.* 62:47-54


Malik AS. 1990b. Markets for pollution control when firms are noncompliant. *J. Environ. Econ. and Manag.* 18: 97-106


Nadeau LW. 1997. EPA effectiveness at reducing the duration of plant-level noncompliance. J. Environ. Econ. Manag. 34: 54-78


Polinsky AM, Shavell S. 2000. The economic theory of public enforcement of law. J. Econ. Lit. 38:45-76


Scholz JT. 1984. Voluntary compliance and regulatory enforcement. Law & Policy. 6: 385-404


Figure 1. Inflation adjusted EPA Budget allocations for enforcement and compliance assurance, in 2013 dollars, fiscal years 1994 through 2013. Source: EPA Historical Planning, Budget, and Results Reports. Online at: http://www2.epa.gov/planandbudget/archive#BudgetSummary.
Figure 2. Inspections by major statute, fiscal years 1994 through 2013. Source: EPA National Enforcement Trends. Online at: http://www.epa.gov/compliance/data/results/nets.html.