Greenwash: Corporate Environmental Disclosure under Threat of Audit

Thomas P. Lyon

Erb Institute for Global Sustainable Enterprise
University of Michigan
701 Tappan St.
Ann Arbor, MI 48109
tplyon@umich.edu.

John W. Maxwell

Kelley School of Business
Indiana University
Bloomington, IN 47405
jwmax@indiana.edu

We develop an economic model of "greenwash," in which a firm strategically discloses environmental information and an activist may audit and penalize the firm for disclosing positive but not negative aspects of its environmental profile. We fully characterize the model's equilibria, and derive a variety of predictions about disclosure behavior. We rationalize conflicting results in the empirical literature, finding a nonmonotonic relationship between a firm's expected environmental performance and its environmental disclosures. Greater activist pressure deters greenwash, but induces some firms to disclose less about their environmental performance. Environmental management systems discourage firms with poor expected environmental performance from greenwashing, which may justify public policies encouraging firms to adopt them.

1. Introduction

Environmental issues have been on the corporate radar screen for years. Thousands of firms participate in the Environmental Protection Agency's partnership programs, and many others participate in industry-led environmental programs such as those of the World

We thank the coeditor and two referees for constructive suggestions that greatly improved the paper. We also thank Mike Baye, Rick Harbaugh, Charlie Kolstad, John Morgan, Michael Rauh, and participants in seminars at the American Economic Association meetings, Canadian Resource and Environmental Economics meetings, Berkeley, Dartmouth, Florida, Indiana, Northwestern, Santa Barbara, Toronto and Western Ontario for their helpful comments.

Business Council for Sustainable Development, the Chicago Climate Exchange, and the American Chemistry Council's "Responsible Care" program.[1] Despite these efforts, large portions of the public continue to view business as an enemy of the environment. Furthermore, although many companies publicize their environmentally friendly actions, others are surprisingly hesitant to promote their environmental successes or to issue detailed environmental reports. Part of the reason appears to be that activists react more angrily to firms that lay claim to being virtuous, and then are discovered to have feet of clay, than to firms that never make such claims. For example, BP makes frequent public claims about its efforts to reduce global warming yet was denounced at the Johannesburg Earth Summit, whereas Exxon has for years been among the loudest skeptics about climate change yet attracts less attention from activists. Indeed, based on his interviews with managers in charge of corporate social responsibility, Peloza (2005) finds that "Many managers worry that by overtly promoting their participation stakeholders might view the activity as self-serving. In fact, many respondents reported minimal or no attempts of self-promotion."[2]

Part of the reason managers hesitate to promote their good environmental deeds is that many such actions are attacked as "greenwash" by activists. Often these activists attempt to punish companies they view as greenwashers by embarrassing them in the media, and encouraging consumers to boycott them.[3] At the 2002 Earth Summmit in Johannesburg, a group of activists held a Greenwash Academy Awards event to criticize companies that falsely promote themselves as environmentally responsible and to "recognize these companies for what they are: hypocrites."[4] Winner for Best Greenwash was "BP for their Beyond Petroleum rebranding campaign," which highlights the company's investments in renewable energy without mentioning their major efforts in petroleum exploration.[5] Among the other awards, South African electricity firm Eskom was Runner up for Best Picture "for being a key member of Business Action for Sustainable Development while generating electricity from coal and nukes."[6] Ralph Nader reveals a similar skepticism regarding corporate participation in the U.N.'s Global Compact: "Companies are able to sign on to the compact and

1. For an introduction to corporate environmental strategy and its relation to public policy, see Lyon and Maxwell (2004b).
2. For example, one of his survey respondents commented that "We're pretty sensitive. We don't want to go out thumping our chests saying 'oh, aren't we wonderful and here's all the great things we do!' We want people to see for themselves and they can draw their own conclusions." See Peloza (2005), p. 16.
3. See "Greenwash 101" at http://www.thegreenlife.org/greenwash101.html.
4. For details, see http://www.corpwatch.org/article.php?id=3648.
5. Ibid.
6. Ibid.

'bluewash' themselves, as critics at the Transnational Research and Action Center in San Francisco have labeled the effort by image-impaired corporations to repair public perceptions by hooking up with the UN."[7]

In this paper, we present what is to our knowledge the first economic analysis of greenwash. For clarity of exposition we focus on environmental greenwash, but it should be clear that our analysis applies equally well to social issues, or any area where incomplete communications might be labeled hypocritical. Because public discussion of greenwash is often polemical and imprecise, we devote Section 2 to developing a clear formal definition of greenwash, and distinguishing it from other "disinformation" strategies. We conclude that greenwash can be characterized as the selective disclosure of positive information about a company's environmental or social performance, while withholding negative information on these dimensions.

Our definition of greenwash suggests that we model the phenomenon as a "persuasion game" along the lines of Milgrom and Roberts (1986), in which one player has certain pieces of verifiable information that he can either disclose or not. The economic literature on such games has shown that a stakeholder can induce full disclosure if verifiable disclosure is costless.[8] However, full disclosure may fail to occur if disclosure is costly, as shown by Verecchia (1983), or if the stakeholder is uncertain whether the firm possesses a given piece of information, as shown by Shin (2003). Sinclair-Desgagne and Gozlan (2003) assume disclosure is costly; they model a firm that is fully informed about its environmental performance and discloses information to an activist who punishes the firm if he believes its environmental performance will be poor. Although the existing disclosure literature offers valuable insights about partial disclosure, it does not address the possibility that stakeholders elect to punish partial disclosure itself, rather than punishing firms for being dangerous or dirty. Yet, as we discussed above, activists often attempt to punish greenwashers, while giving other firms with less stellar environmental records a pass. We explore this notion in a model based on Shin (2003), where there are multiple dimensions to the firm's environmental performance, and stakeholders are uncertain whether the firm actually possesses information about any given dimension. We extend the model by adding

7. Ralph Nader, "Corporations and The UN: Nike And Others 'Bluewash' Their Images," *San Francisco Bay Guardian*, September 18, 2000. Available at http://www.commondreams.org/views/091900-103.htm

8. See, for example, Milgrom and Roberts (1986), who show that with a single informed firm, the stakeholder can induce full disclosure if it adopts a skeptical posture, that is, it assumes the worst when full information is not presented.

a nonstrategic activist that increases its scrutiny of firms that selectively report good news, and punishes greenwash when it is discovered. Thus, in contrast to Sinclair-Desgagne and Gozlan (2003), we model a firm that is imperfectly informed about its environmental performance and discloses information to an activist who punishes the firm for engaging in greenwash, rather than for having poor environmental performance.

Our focus is on the optimal firm response to activist campaigns against greenwash, rather than on optimal activist behavior in itself; thus, the activist in our model is not a strategic agent, but rather an agent that is simply assumed to act according to a specified rule. The literature suggests a variety of motivations for activists, and there is no accepted model comparable to the model of maximizing profits for firms or maximizing re-election probability for politicians. Some economic models assume activists attempt to maximize environmental benefits (Innes, 2006), and others assume they have an objective that deviates from social welfare in some linear fashion (Grossman and Helpman, 2001). Baron (2003) argues that some activists may be "intransigent" types that do not behave as rational actors. Here we opt to model activist behavior according to a rule that appears to capture their actual behavior, rather than postulating a particular objective function and deriving optimal activist behavior. As we argued above, the evidence shows that many activists attack firms they see as engaging in greenwash, and that is the behavior we incorporate into our model.

We provide a full characterization of how disclosure varies with the firm's expected environmental performance, and with the quality of the firm's own internal information at the time it makes a disclosure. We show there is a nonmonotonic relationship between expected environmental performance and the number of expected disclosures. Moreover, increased activist pressure encourages firms with low performance to increase disclosures, but deters high performance firms from disclosing at all. In addition, we find that corporate adoption of a high-quality environmental management system (EMS) reduces incentives for greenwash, thereby providing a new rationale for public policies encouraging firms to adopt EMSs.

Incorporating concerns about greenwash into a model of environmental disclosure helps to explain some of the sharply conflicting results in the empirical literature on environmental disclosure regarding the relationship between environmental performance and environmental disclosures. For example, Patten (2002) and Cho and Patten (2007) find that firms with *worse* environmental records (as measured by higher ratios of toxic chemical emissions to sales) have higher levels of environmental disclosures, while Clarkson et al. (2008) find that firms with *better* environmental records (again measured by toxic emissions) have higher

levels of environmental disclosures. These conflicting results reflect conflicting theoretical approaches to understanding disclosure. Standard economic models of disclosure, such as Verecchia (1983) or Shin (2003), imply that firms with higher probabilities of success have more good news to convey, and hence engage in more disclosure. However, these models typically ignore the role of stakeholder pressure or regulatory threats in shaping disclosure behavior. In fact, Cho and Patten (2007) argue that disclosure is more likely by firms in environmentally sensitive industries because such firms "face greater exposure to the public policy process than companies from nonenvironmentally sensitive industries." Our model, by incorporating both firm characteristics and activist pressure, generates a rich set of testable predictions that shed light on the conflicting results in the existing literature and suggest more precise hypotheses for study in future work.

The remainder of the paper is organized as follows. Section 2 presents our definition of greenwash, and distinguishes it from other "disinformation" strategies. In Section 3, we present the basic disclosure model without activist auditing. Section 4 adds the activist to the model, and provides a full characterization of the model's equilibria. In Section 5, we use our analysis to draw out testable hypotheses for empirical study. Section 6 concludes.

2. What is Greenwash?

Formal analysis of greenwash requires a clear definition of the phenomenon. Unfortunately, popular usage of the term tends to be broad and vague; indeed, in their book on greenwash, Greer and Bruno (1996) never actually define the term.[9] On the first page of the Introduction, however, they complain that transnational corporations "are preserving and expanding their markets by posing as friends of the environment and leaders in the struggle to eradicate poverty."

9. Even academic discussions can be surprisingly broad. Laufer (2003), for example, presents a set of elements of greenwashing that include "confusion," "fronting," and "posturing." Confusion (p. 257) is achieved through "careful document control and strict limits on the flow of information made available to regulators and prosecutors." Fronting (p. 257) "is realized by subordinate scapegoating or reverse whistle blowing," and may involve such strategies as "cast doubt on the severity of the problem" or "emphasize uncertainty associated with the problem." Posturing (p. 256) involves the use of "front groups" to influence legislation or suggest that particular policies enjoy widespread "grassroots" support. Although we find these distinctions interesting, in our view these activities differ too much to be viewed as a single phenomenon; indeed, we have already modeled the use of "astroturf lobbying" through "front groups" in Lyon and Maxwell (2004a). Astroturf lobbying involves the provision of soft information targeted at a public decisionmaker to influence policy decisions. Greenwash involves public disclosure of hard information targeted to influence shareholder value.

Webster's New Millenium Dictionary of English defines greenwash as "The practice of promoting environmentally friendly programs to deflect attention from an organization's environmentally unfriendly or less savory activities." The *Concise Oxford English Dictionary (10th Edition)* defines it as: "Disinformation disseminated by an organization so as to present an environmentally responsible public image; a public image of environmental responsibility promulgated by or for an organization etc. but perceived as being unfounded or intentionally misleading." Both of these definitions emphasize the idea that the public has limited information about corporate environmental performance, and that corporations therefore can manipulate the dissemination of information to mislead the public.

The term "disinformation" implies the provision of deliberately false or fraudulent messages. To us, however, corporate greenwashing does not seem to fit this definition. Instead, the typical concerns raised by activists are that companies present positive information out of context in a way that could be misleading to individuals who lack background information about the company's full portfolio of activities. Consider the following example, taken from *Don't Be Fooled: The Ten Worst Greenwashers of 2003*:

> "Royal Caribbean points to its advanced wastewater treatment systems as a sign of environmental progressiveness, yet they are installed on just 3 of the company's 26 cruise ships. The advanced systems are only found on its Alaskan fleet, which due to Alaskan law are subject to the strictest environmental standards in the industry. Royal Caribbean deems them unnecessary on cruise ships that travel other routes."[10]

Terrachoice, an environmental marketing firm, defines the verb "to greenwash" in a slightly differen way: "the act of misleading consumers regarding the environmental practices of a company or the environmental benefits of a product or service." The company released a report in 2007 that studied the environmental claims of 1,018 products sold in "big box" retailers in the United States and Canada. The report concluded that all but one of the products made claims that were demonstrably false or risked misleading consumers. Of the "Six Sins of Greenwashing" identified by the firm, by far the most common was the "Sin of the Hidden Tradeoff," which was committed by 57% of the products.

10. See Johnson (2003).

> "The Sin of the Hidden Tradeoff is committed by suggesting a product is 'green' based on a single environmental attribute (the recycled content of paper, for example) or an unreasonably narrow set of attributes (recycled content and chlorine free bleaching) without attention to other important, or perhaps more important, environmental issues (such as energy, global warming, water, and forestry impacts of paper). Such claims are not usually false, but are used to paint a "greener" picture of the product than a more complete environmental analysis would support." (Terrachoice, 2007, p. 2)[11]

A follow-up study in 2009 found many more products that make environmental claims, and found that 98% of the 2, 219 products making such claims committed at least one of the "Six Sins of Greenwashing"; it also identified a new, seventh sin, which involves creating false suggestions of third-party endorsements (Terrachoice, 2009).

The above examples all suggest that greenwash is fundamentally about misleading consumers and investors by telling the truth, but not the whole truth. This suggests a model in which the firm discloses verifiable information, but may choose to withhold facts that do not reflect favorably on it, thereby persuading outsiders that the firm's performance is better than it is in reality. The "persuasion games" introduced by Milgrom and Roberts (1986) fit this situation well. This notion of disclosure captures a variety of types of business communications, including corporate annual reports, 10-Ks, sustainability reports, web sites, and advertising that is fact-based.[12] Taking this approach, we define greenwash as the *selective disclosure of positive information about a company's environmental or social performance, without full disclosure of negative information on these dimensions, so as to create an overly positive corporate image*.[13]

11. The other sins are: the Sin of No Proof, the Sin of Vagueness, the Sin of Irrelevance, the Sin of Fibbing, the Sin of the Lesser of Two Evils. (Terrachoice, 2007, p. 1)

12. We readily acknowledge that there are dimensions of corporate environmental strategy and communications that are not captured in our formulation. For example, like the rest of the disclosure literature, we focus only on the firm's disclosure decisions, and leave for future research the important challenge of integrating a disclosure model with a model of corporate choice of environmental projects. In addition, we do not analyze corporate communications that involve pure image advertising, for example, BP's selection of the green flower image for its logo or General Electric's use of scantily clad miners in a television ad about coal-fired power generation. Nevertheless, all of these types of corporate communications result in overly favorable corporate images, and that fundamental phenomenon is captured in our model.

13. Empirical research in accounting suggests that this is a common practice for firms that choose to engage in corporate environmental disclosure; see, for example, Deegan and Rankin (1996).

An interesting concrete example of selective disclosure comes from the Department of Energy's Voluntary Greenhouse Gas Reporting program, created by section 1605b of the Energy Policy Act of 1992. Kim and Lyon (forthcoming) show that electric utility participants in the 1605(b) program reported reductions in their greenhouse gas emissions during the period 1995–2003, but their actual emissions rose. Furthermore, during the same period, nonparticipant utilities reduced their emissions. This misleading reporting behavior is not illegal, for the program allows participants great flexibility in how they choose to report emissions reductions. In particular, firms can choose to report at the "project level" or the "entity level." The former allows a firm to report only on the outcomes of successful projects, while remaining silent about its aggregate performance. This is precisely what we mean by the term greenwash.

In the model that follows, we capture this notion of greenwash by assuming there are multiple dimensions of environmental performance, and that a firm can selectively choose which of these will be part of its corporate communications. When firms with mixed records engage in this type of selective disclosure, the result is that the receiver of the communication develops a view of the company that is better than its actual record. Although such distortions are familiar in modern advertising, they typically concern private goods, with advertising policed by the Federal Trade Commission. What sets greenwash apart is that there are externalities associated with the distortion of the company's image, which provides an explanation for why activists attack greenwash. Our model captures the image distortion created by greenwash in a clear and tractable fashion that is consistent with economic rationality. We do not assume that consumers fail to update information correctly, nor do we focus on the type of image-based advertising in which, for example, scantily clad men and women are deployed in coal mines.[14] Nevertheless, all of these types of corporate communications result in overly favorable corporate images, and that fundamental phenomenon is captured in our model.

3. The Basic Disclosure Game

Our model focuses on a single firm, whose stock is traded publicly, and an activist. The firm has two different activities that each has some potential environmental impact; the nonenvironmental aspects of the firm's operations are assumed to be already incorporated into the firm's

14. To view General Electric's advertisement featuring scantily clad men and women in a coal mine, visit http://www.youtube.com/watch?v=J1A146sANdg.

market value.[15] However, the firm's environmental performance is not known at the outset of the model. We assume the market sets the firm's value at its actuarily fair level.[16]

There are three periods. Let V_t represent the expected value of the firm in period t. At period 0, there is common knowledge about the likelihood there will be a liability associated with any given activity. During period 1, which may be of indeterminately long duration, each activity generates for the firm a "success" of value u (e.g., an outcome that improves the firm's public image) with probability $r \in (0, 1)$, and a "failure" of value $d < u$ with probability $1 - r$. Thus, the firm's expected number of environmental successes is simply $2r$. Its market value in period 0 is

$$V_0 = 2(ru + (1 - r)d) + \tilde{V}, \tag{1}$$

where $\tilde{V}$ is the total value created by the firm aside from its environmental impacts. Throughout the remainder of the paper, we will simplify notation by normalizing $\tilde{V}$ to 0. At period 2, all information about environmental impacts is revealed and becomes common knowledge, and is incorporated into stock prices. The important action in the model takes place in the interim period 1, during which the manager attempts to influence the firm's stock price through the information he discloses.[17]

We assume there is a probability $\theta \in (0, 1)$ that the manager actually learns the environmental impact of the activity by period 1.[18] Thus, at the interim period, the expected number of activities for which the manager has information on environmental outcomes is 2θ. The expected number of activities known to have environmental liabilities at the interim period is $2\theta(1 - r)$. The manager has the ability to disclose publicly the number of activities that are known to be successes or failures. We assume that all such disclosures are verifiable by outside parties. Thus, the manager is free to selectively withhold information,

15. We model two activities because this is the minimum number that allows us to capture the notion of greenwash as partial disclosure. We refer to environmental impacts for concreteness, but could just as easily refer to social and environmental impacts more generally.

16. The model draws upon the work of Shin (2003), but departs from it by using an additive rather than a multiplicative structure for payoffs, and by incorporating monitoring and punishment of greenwash.

17. There are many reasons a manager might want to influence the stock price. One obvious way the manager could exploit the stock price is by selling stock short in period 1 and buying it back in period 2, although this may be viewed as an illegal form of insider trading and subject to legal penalties. It is not necessary that the manager engage in possibly illegal short sales, however; all that is needed is that his salary is in some way tied to the performance of the firm's share price, for example, through bonuses, stock options, etc. For further details, see Milgrom and Roberts (1992).

18. We would expect θ to be greater for firms that have created an EMS. We return to this issue in Section 6.

but he cannot actually lie to outsiders. We assume the manager adopts a disclosure strategy that maximizes the value of the firm.

The firm's type is defined *ex ante* at period 0 by the pair (r, θ). The state of the world is defined *ex post* at period 1 by the pair (s, f), where s is the number of successes and f the number of failures about which the manager is informed. Let n be the total number of activities whose outcomes are known at the interim period, so that $n = s + f$. Let the manager's disclosures of the number of successes and failures be given by $\hat{s}$ and $\hat{f}$. We assume $V_1 = E(V_2)$. If the market knows s and f, as would be the case if the manager fully disclosed his information in period 1, then

$$V_1 = E(V_2) = us + df + (2 - s - f)(ru + (1 - r)d). \tag{2}$$

If the manager discloses a success, but fails to disclose both outcomes, then the activist investigates the manager's report for the possibility of greenwash (i.e., that the manager has a bad outcome that he failed to disclose). With probability α the activist obtains hard (verifiable) information about the true values of s and f at the interim period and, if greenwash is detected, mounts a successful campaign against the firm that imposes a punishment of cost P on the firm; with probability $1 - \alpha$ it learns nothing and takes no action against the firm. The punishment might come about because the activist triggers a consumer boycott, because it creates an advertising campaign that damages the firm's value, or through some other channel that the firm finds costly.[19] We will use the notation $\eta = \alpha P/(u - d)$ to indicate the "cost/benefit ratio" for greenwash, where αP is the expected penalty for greenwash, and $u - d$ represents the maximum value the firm could possibly obtain from successful greenwash. When the expected penalty for greenwash is zero, then $\eta = 0$, and we expect the firm to engage in greenwash. When the expected penalty for greenwash rises to $u - d$, then $\eta = 1$, and any potential benefits of greenwash are outweighed by the expected penalty; in this case we expect the firm to avoid greenwashing.

We are interested in Perfect Bayesian Equilibria (PBE), which involve specifying a disclosure strategy for the manager, a market valuation, and a set of market beliefs for each time t such that (a) the disclosure strategy $(\hat{s}, \hat{f})$ is a best response mapping for a firm with actual environmental outcomes (s, f), given the market's pricing policy and the beliefs of the market, (b) $V_1 = E(V_2)$ given the market's beliefs

19. Baron and Diermeier (2007) and Sinclair-Desgagne and Gozlan (2003) present models of interactions between firms and activists in which firms are punished for bad social outcomes, rather than being punished for greenwashing.

at period 1 and the manager's disclosure strategy, and (c) at period 0 the market believes the expected number of environmental failures is $2r$, and at period 1 it computes the expected number of environmental failures using Bayes' rule, conditional on any environmental reports. We will focus on pure strategy equilibria in section 3, and provide a full characterization of the model's equilibria in Section 4.

It is easy to see that if the market believed the manager always truthfully disclosed all successes and failures, then the manager would have incentives to report $f = 0$. The expected value of an activity whose social impact is unknown is greater than the value of a failure, that is, $ru + (1 - r)d > d$. As a result, the manager always prefers to minimize the number of failures reported, and report only the successes; if there is no penalty for withholding bad news, then full disclosure is not an equilibrium strategy.[20] However, we follow Shin (2003) and assume off-equilibrium beliefs on the part of the market that if the manager ever reports $\hat{f} > 0$, then all undisclosed outcomes are failures;[21] with these beliefs, then there exists a "partial disclosure" equilibrium in which the manager follows a strategy of disclosing only successes.

4. Disclosure with Activist Auditing

In this section, we assess how auditing by an activist and punishment for greenwash affects the manager's incentives to make environmental disclosures. We fully characterize the set of Perfect Bayesian Equilibria that can emerge in the model, and show how they are related to the underlying parameters of the model.

It is important to be clear about the nature of auditing and punishment in this model. The activist does not simply punish the firm for bad outcomes. Instead, the activist focuses on penalizing the firm for greenwashing, that is, disclosing a success while withholding information about a failure. It is natural to ask whether the activist can effectively punish partial disclosure without auditing, for example, by penalizing the firm retroactively based on the ultimate outcomes in period 2. It turns out this is not possible. As we noted in the Introduction, punishing partial disclosure is distinct from simply punishing the firm for bad social outcomes. Punishing partial disclosure involves punishing firms that were aware of, but failed to disclose, a failure,

20. Shin (2003) refers to the strategy of not disclosing any failures as "sanitization," but does not distinguish situations where the firm has positive as well as negative news to report, which are the sorts of situations in which greenwash may become a problem.

21. This rules out the possibility that the manager can benefit by employing the implausible strategy of disclosing (0,1) in the state (1,1). See Shin (2003), Appendix A, for further discussion of the role of off-equilibrium beliefs in this sort of disclosure model.

whereas at the same time disclosing a success. At period 2, however, all the activist knows is the ultimate number of failures, *not* the number that were known but not disclosed at the interim period. Thus, it is impossible to punish partial disclosure *per se* by only observing period 2 outcomes. Instead, it is essential to have some sort of independent auditing structure in period 1.

In our model, a penalty for greenwashing can only occur when the state is (1,1) and the firm discloses (1,0). From the firm's perspective, greenwashing creates a risk of being detected and punished. This will only happen if the activist decides to audit, learns that the firm is actually in state (1,1), and successfully imposes a penalty P. Clearly the activist will not accomplish all these steps with probability 1. We interpret α as the overall probability that the firm is audited, detected and successfully punished when it engages in greenwash. Thus, when a firm engages in greenwash, it faces an expected penalty of αP.

We will use the notation $V_1(\hat{s}, \hat{f})$ to indicate the market's valuation of the firm at period 1 when the manager discloses $(\hat{s}, \hat{f})$. Note that when $\hat{n} \equiv \hat{s} + \hat{f} = 2$ the market has no problem inferring the firm's true value because information disclosures are verifiable. These values are easily seen to be $V_1(0, 2) = 2d$, $V_1(2, 0) = 2u$, and $V_1(1, 1) = u + d$. It is only in states where $\hat{n} \equiv \hat{s} + \hat{f} < 2$ that we must carefully analyze the market's inference problem.

Note that the optimal disclosure strategy for the firm in all states except (1,1) is trivial. A firm in state (0,0) has nothing to disclose. A firm in state (1,0) or (2,0) will disclose all its information because failing to do so would not maximize its value. A firm in state (0,1) or (0,2) has nothing positive to disclose, and will choose not to disclose any negative information. Partial disclosure consists of reporting (1,0) in state (1,1), which we label greenwash. Full disclosure in that state would consist of reporting (1,1) and nondisclosure would consist of reporting (0,0) in the same state. The firm receives no punishment for any report except when the state is (1,1) and the manager reports (1,0). Hence our focus is on what the manager will report when $(s, f) = (1, 1)$. There are four reporting possibilities: $(\hat{s}, \hat{f}) \in \{(0, 0), (1, 0), (0, 1), (1, 1)\}$. Given the arguments we have made above, however, it is clear that the manager will never report $(\hat{s}, \hat{f}) = (0, 1)$.

In order to understand the manager's reporting incentives, we must know how the market will interpret each of the three possible reports. Given any report, the probability that the state is actually (1,1) can then be computed via Bayes' Rule. Table 1 below presents the prior probability of each state at the interim period, along with the value the market attaches to that state. It is easy to see that greenwash can

Table I.
Interim Period States, Probabilities, and Values

State	Probability	$V_1(s, f)$
(2,0)	$r^2\theta^2$	$2u$
(1,0)	$2r\theta(1-\theta)$	$u + (ru + (1-r)d)$
(1,1)	$2r(1-r)\theta^2$	$u + d$
(0,0)	$(1-\theta)^2$	$2(ru + (1-r)d)$
(0, 1)	$2(1-r)\theta(1-\theta)$	$d + (ru + (1-r)d)$
(0,2)	$(1-r)^2\theta^2$	$2d$

pay—reporting (1,0) earns the firm a better value than does reporting (1,1).

We will use the notation $\mu(\hat{s}, \hat{f}; s, f)$ to indicate the probability the market assigns to the manager playing reporting strategy $(\hat{s}, \hat{f})$ when the state is (s, f).[22] Thus market beliefs μ constitute a set of values $\mu(\hat{s}, \hat{f}; s, f)$ for all $(s, f) \in \{(0, 0), (0, 1), (0, 2), (1, 0), (2, 0), (1, 1)\}$. We will denote the firm's expected value from a particular disclosure strategy by the notation $EV[\hat{s}, \hat{f} \mid s, f, \mu]$, where μ identifies the beliefs of the market and activist regarding the firm's behavior.

We define $\Psi(\hat{s}, \hat{f})$ as the probability the market assigns to observing a report $(\hat{s}, \hat{f})$; this is the sum of the probabilities of each interim state multiplied by the probability that the firm reports $(\hat{s}, \hat{f})$ in that state. For example,

$$\Psi(0, 0) = (1-\theta)^2\mu(0, 0 \mid 0, 0) + 2(1-r)\theta(1-\theta)\mu(0, 0 \mid 0, 1) + (1-r)^2\theta^2\mu(0, 0 \mid 0, 2) + 2r(1-r)\theta^2\mu(0, 0 \mid 1, 1).$$

We turn now to the expected value the firm obtains in state (1,1) from alternative possible disclosure strategies. If the firm reports (1,1), then the market knows the state with certainty, and the firm has market value

$$EV[1, 1 \mid 1, 1, \mu] = u + d. \tag{3}$$

If the firm in state (1,1) reports (1,0), then the market believes the state is either (1,0) and the firm is revealing truthfully, or (1,1) and the firm is engaging in greenwash. Thus, $\Psi(1, 0) = 2r\theta(1-\theta)\mu(1, 0 \mid 1, 0) + 2r(1-r)\theta^2\mu(1, 0 \mid 1, 1)$. If the activist audits, and finds that the state is really (1,1) but the firm engaged in greenwash, then the activist

22. In equilibrium, of course, we must have $\mu(\hat{s}, \hat{f}; s, f)$ equal to the firm's true probability of playing a given strategy.

launches a campaign against the firm that imposes a penalty P. The firm's expected value from reporting (1,0) is

$$EV[1,0\,|\,1,1,\mu] = [u + (ru + (1-r)d)]\frac{2r\theta(1-\theta)\mu(1,0\,|\,1,0)}{\Psi(1,0)} + [u+d]\frac{2r(1-r)\theta^2\mu(1,0\,|\,1,1)}{\Psi(1,0)} - \alpha P. \quad (4)$$

If the firm in state (1,1) reports (0,0), the market recognizes that the state may be (0,0), (0,1), (0,2) or (1,1).[23] Note that there is no possibility of a punishment in this case because a report of (0,0) does not constitute greenwash, as it does not claim any positive outcomes. The firm's expected value is

$$\begin{aligned} EV[0,0\,|\,1,1,\mu] = {} & [ru + (1-r)d]\frac{(1-\theta)^2 2\mu(0,0\,|\,0,0)}{\Psi(0,0)} \\ & + [d + (ru + (1-r)d)]\frac{2(1-r)\theta(1-\theta)\mu(0,0\,|\,0,1)}{\Psi(0,0)} \\ & + 2d\,\frac{(1-r)^2\theta^2\mu(0,0\,|\,0,2)}{\Psi(0,0)} \\ & + [u+d]\frac{2r(1-r)\theta^2\mu(0,0\,|\,1,1)}{\Psi(0,0)}. \end{aligned} \quad (5)$$

There are three types of pure-strategy equilibria that can emerge in this model in state (1,1): a) the firm fully discloses the state, b) the firm engages in partial disclosure, or c) the firm does not disclose at all. We label these the "full disclosure equilibrium," the "partial disclosure equilibrium," and the "nondisclosure equilibrium," respectively. Of course, this does not mean the firm will behave according to these labels if it is not in state (1,1). For example, a firm that would fully disclose in state (1,1) would still fail to disclose anything if it finds itself in state (0,2); as we pointed out earlier, all equilibria involve withholding of information in states (0,1) and (0,2). Furthermore, mixed strategy equilibria are possible, as well. For example, a firm may mix between full disclosure and partial disclosure, or between nondisclosure and partial disclosure; however, we note that a firm will never mix between full disclosure and no disclosure. The following proposition is our main result, and fully characterizes the set of possible equilibria.

Proposition 1: *In state* (1,1), *the firm's behavior is as follows. (a) If* $\eta = 0$, *then for all* (r, θ), *the firm engages in partial disclosure. (b) If* $\eta \in (0, 1/2)$,

23. A firm in state (1,0) or (2,0) has no incentive to report (0,0).

then for any $\theta \in (0, 1)$, *there exists a series of nonnegative values* $r_{FP} < r_{PF} < r_{PN} < r_{NP} < 1$ *such that for* $r \in (0, r_{FP})$ *the firm engages in full disclosure, for* $r \in [r_{FP}, r_{PF})$ *the firm mixes between full disclosure and partial disclosure, for* $r \in [r_{PF}, r_{PN})$, *the firm engages in partial disclosure, for* $r \in [r_{PN}, r_{NP})$ *the firm mixes between partial disclosure and nondisclosure, and for* $r \in [r_{NP}, 1)$ *the firm discloses nothing. (c) If* $\eta \in (1/2, 1)$, *then for* $r < \min(r_{FN}, r_{FP})$ *the firm fully discloses, for* $r \in (\min\{r_{FN}, r_{FP}\}, r_{FN})$ *the firm mixes between full disclosure and partial disclosure, for* $r \in [r_{FN}, \max(r_{FN}, r_{NP}))$ *the firm mixes between partial disclosure and nondisclosure, and for* $r \in [\max(r_{NP}, r_{FN}), 1)$ *the firm discloses nothing. (d) If* $\eta \geq 1$, *then for* $r < r_{FN}$ *the firm fully discloses, and for* $r > r_{FN}$ *the firm discloses nothing.*

Proof. See the Appendix.

The Proposition establishes the existence and nature of the optimal disclosure strategy for each type of firm (r, θ) and for every possible level of greenwash penalty η. Our key finding is that in contrast to Shin (2003), now the firm does not always pursue the partial disclosure strategy. Indeed, we are able to establish in detail exactly how the firm's strategy changes with the underlying parameters of the model. The following section provides a discussion of the nature of the equilibria and their relationship to the parameters r, θ, and η. In doing so, we explain in graphical form all of the content of Proposition 1.

5. Key Implications of the Model

To this point, our analysis has been strictly theoretical. In this section we use it to shed light on disclosure behavior in a variety of practical situations. Our emphasis is on the three key underlying parameters of the model, namely η, the cost-benefit ratio for greenwash, r, the probability that a given activity has a positive environmental outcome, and θ, the probability the manager is informed about an activity's outcome at the time he makes a report.

5.1 Changing Penalties for Greenwash

Our most important results have to do with how increasing activist pressure affects firms' disclosure behavior. In this section, we first provide an intuitive discussion of our main results in Proposition 1 using Figures 1–3. We then proceed to establish formally that an increase in activist pressure leads to a reduction in greenwashing, leads "uninformed green firms" to increase the likelihood of nondisclosure

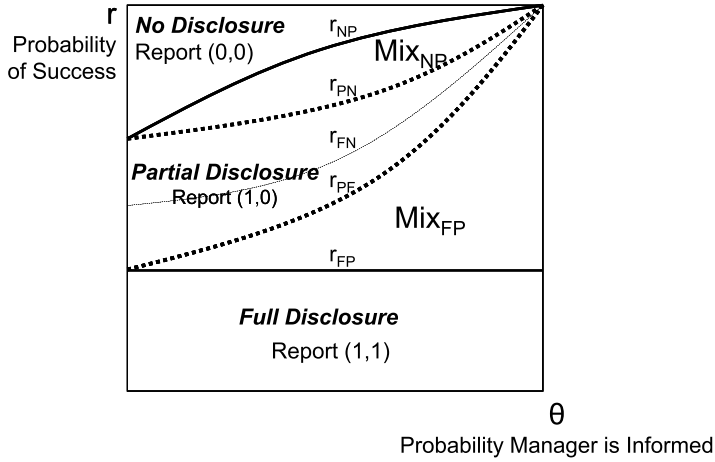

FIGURE 1. DISCLOSURE EQUILIBRIA WHEN GREENWASH PENALTIES ARE LOW ($\eta < .5$)

and leads " informed brown firms" to increase the likeihood of full disclosure.

As pointed out earlier, when $\eta = 0$, partial disclosure is the only equilibrium strategy for a firm in state (1,1)[24], that is the firm greenwashes with certainty Disclosing (1,0) produces a positive effect on external beliefs about the firm, and carries with it no penalty. Thus, partial disclosure dominates either full disclosure or no disclosure.

For $\eta \in (0, 1/2)$, each of the three types of pure strategies is employed in equilibrium for at least some value of (r, θ). However, it is also true that there is no equilibrium in pure-strategies in state (1,1) for some (r, θ) pairs. This is illustrated in Figure 1, which is divided by a series of curves labeled, from bottom to top, $r_{FP} < r_{PF} < r_{FN} < r_{PN} < r_{NP}$. In each case, for a given value of θ, r_{ij} indicates the value of r for which a firm that is believed by the market to be playing strategy i would find it equally profitable to play strategy j. There are three regions with pure-strategy equilibria: the full-disclosure region lies below the lower solid line, the nondisclosure region lies above the upper solid curve, and the partial-disclosure region lies between the two dotted curves . (The two thick dotted curves converge toward the thin curve lying between them as η approaches $1/2$, at which point greenwash is eliminated as an equilibrium pure strategy.) There are also two regions with

24. For the remainder of this section, we focus solely on the firm's behavior in state (1,1).

mixed-strategy equilibria, where the notation "Mix$_{ij}$" indicates the firm is mixing between strategies i and j.

The intuition for the pure-strategy equilibria is straightforward. A firm with a low value of r fully discloses: it gains a lot from trumpeting a success, and loses little by withholding information about a failure (because it is already expected to fail); thus, there is little value in risking public backlash by refusing to disclose. At the other extreme, a firm with a high value of r does not disclose anything: it gains little by disclosing information about successes (because it is already expected to succeed), and loses a lot by disclosing a failure; thus, there is little value in risking public backlash by disclosing a success. For a firm with a moderate value of r, partial disclosure is attractive: disclosing a success can produce a significant improvement in public perception, and withholding information about a failure can prevent a significant negative public perception; thus, it is willing to risk public backlash by disclosing only partially.

There are also two regions in which there are no pure-strategy equilibria: between the upper solid curve and the upper dotted curve, the region labeled "Mix$_{NP}$," firms employ a mixed strategy that involves mixing between nondisclosure and partial disclosure; between the lower dotted curve and the lower solid line, the region labeled "Mix$_{FP}$," firms mix between full disclosure and partial disclosure.

At the point where $\eta = 1/2$, partial disclosure is just eliminated as an equilibrium pure strategy in state (1,1). Two types of pure-strategy equilibria continue to exist, but there are again regions in which pure-strategy equilibria do not exist. This is illustrated in Figure 2. Again, the nondisclosure region lies above the upper solid curve, and the full-disclosure region lies below the lower solid line. Now there is no pure-strategy partial disclosure region, because the penalty is large enough to eliminate it as an equilibrium. From a graphical perspective, the two former dotted curves bounding the partial disclosure region have collapsed together into the dashed curve in the middle of the graph. Once again, there are two regions in which there are no pure-strategy equilibria: the region labeled "Mix$_{NP}$," in which firms employ a mixed strategy that involves mixing between nondisclosure and partial disclosure, and the second, labeled "Mix$_{FP}$," in which firms mix between full disclosure and partial disclosure. Thus, even though partial disclosure is not a pure strategy equilibrium for any (r, θ) pairs, it is still part of the mixed strategies in the aforementioned regions. Note that at $\eta = 1/2$, the left intercept is at $r = 1/2$ for all three curves defining the full disclosure, nondisclosure and mixing regions.

For $\eta \in (1/2, 1)$, the mixed strategy regions shrink as η increases, as can be seen by comparing Figures 3–2. For $\eta > 1/2$, there is a critical

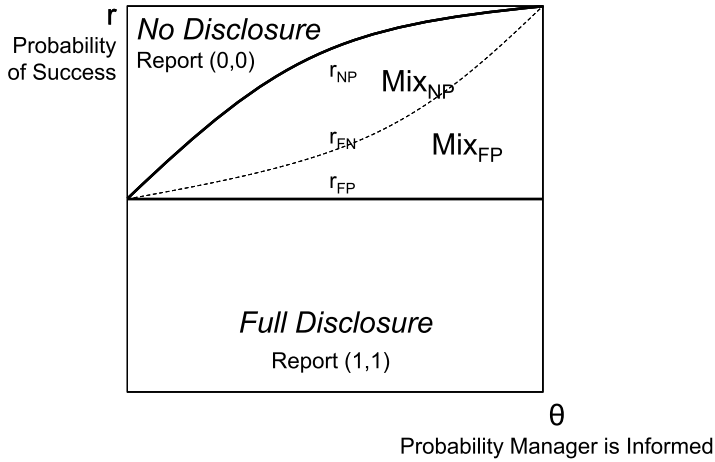

FIGURE 2. DISCLOSURE EQUILIBRIA WHEN GREENWASH PENALTIES ARE MODERATE ($\eta = .5$)

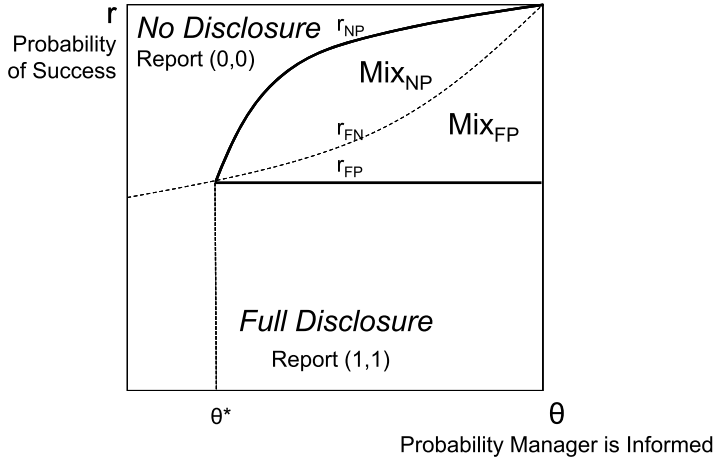

FIGURE 3. DISCLOSURE EQUILIBRIA WHEN GREENWASH PENALTIES ARE HIGH ($.5 < \eta < 1$)

value $\theta^* = 2 - 1/\eta$ where $r_{FN} = r_{NP} = r_{FP} = \eta$. For $\theta < \theta^*$, the unique equilibria are full disclosure for $r < r_{FN}$ and nondisclosure for $r > r_{FN}$. For $\theta > \theta^*$, we know that $r_{FP} < r_{FN}$, and the unique equilibrium is full disclosure for $r < r_{FP}$. For $r \in (r_{FP}, r_{FN})$ the equilibrium involves mixing between full disclosure and partial disclosure. For $r \in (r_{FN}, r_{NP})$ the equilibrium involves mixing between nondisclosure and partial

disclosure. Finally, for $r > r_{NP}$, the unique equilibrium is nondisclosure. In the limit, as η goes to 1, θ^* goes to 1 as well; then nondisclosure is the unique pure-strategy equilibrium for all $r > r_{FN}$ and full disclosure is the unique pure-strategy equilibrium for $r < r_{FN}$.

The fact that the greenwash regions shrink as the expected penalty grows is intuitive. The maximum benefit the firm can possibly obtain from greenwash is $u - d$. This occurs if the firm has a very high value of r, so the market grants the firm an expected value of close to u for undisclosed outcomes, whereas it would have gotten a d if it revealed the failure. If the penalty is large enough to outweigh this maximum possible benefit to partial disclosure, then it will deter the firm from using this strategy. Thus, if $\alpha P \geq (u - d)$, the firm in state (1,1) simply chooses between full disclosure or nondisclosure. We provide a formal proof below in Proposition 2.

Our analysis shows that when the activist attacks greenwash, this always reduces the incidence of greenwashing, but can lead to either more or less environmental disclosure. There is a real possibility that the threat of public backlash for greenwash will cause some firms to "clam up" rather than become more open and transparent. In particular, such a response is likely from firms with a high probability of successful projects, yet who are not fully informed about the environmental impacts of their actions. We call such firms " uninformed greens." (They are defined formally by $r > r_{FN} = 1/(2 - \theta)$. This expression is depicted in Figure 1 as the dotted curve that lies in the center of the partial disclosure region. For high greenwash penalties, that is for $\eta \geq 1$, a firm in this region opts not to disclose in the state (1,1).) Such a firm may shift from partial disclosure to nondisclosure, thereby reducing the amount of information received by the public. It gains little by disclosing information about successes (because it is already expected to succeed), and now loses more than before when it discloses a failure; thus, it becomes less willing to risk public backlash by disclosing a success. On the other hand, firms with a low probability of environmental success are more likely to respond by increasing their environmental disclosures. We call such firms "informed browns"; they are defined formally by $r < r_{FN} = 1/(2 - \theta)$. Such a firm may respond to an increase in expected penalties by switching from partial disclosure to full disclosure.[25] It gains a lot from trumpeting a success, and loses little by withholding information about a failure (because it is already expected to fail); thus, there is little value in risking public backlash by

25. Our discussion has touched only upon the state (1,1), because greenwash penalties do not affect disclosure behavior in any of the other states. Hence, if an increase in greenwash penalties leads to an increase in disclosures in state (1,1), this also implies an increased in total expected disclosures, as disclosure strategies in other states are unaffected.

refusing to disclose, and an increase in the penalty for partial disclosure makes it even more willing to fully disclose. We present these ideas formally in the following Proposition

Proposition 2: *An increase in the expected penalty for greenwash: (a) reduces the incidence of greenwashing, (b) causes uninformed greens to weakly increase their likelihood of nondisclosure, and (c) causes informed browns to weakly increase their likelihood of full disclosure.*

Proof. See the Appendix.

Empirical support for these predictions comes from data on participation in the Carbon Disclosure Project (CDP), an effort by institutional investors to pressure firms to disclose their carbon footprints and strategies for managing them. Reid and Toffel (2009) examine which firms are more likely to disclose to the CDP as a function of the pressures they face from activists and from regulatory threats. They find that firms in environmentally sensitive industries (such as oil, utilities, or transportation) are more likely to disclose to the CDP in response to shareholder resolutions against them than were other firms. This is precisely the effect we predict.

Our analysis also provides insight into how disclosure behavior changes when environmental incidents provoke an increase in scrutiny from activists. Consider the case of the Exxon Valdez, which struck a reef in Prince William Sound, Alaska, on 24 March 1989. The 11 million gallon spill attracted an enormous amount of attention from activists and the public, and caused oil company stakeholders, including citizens and shareholders, to re-evaluate the environmental riskiness of oil company operations. As one might expect, the incident also had a strong negative impact on the company's finances: within a year of the accident, Exxon had already spent over $2 billion to clean up the spill. Patten (1992) studied the disclosure behavior of major oil companies in the wake of the Valdez accident, and found that on average they doubled the amount of space in their annual reports devoted to environmental issues. This increase in disclosures was virtually required for Exxon, which had to describe the event to shareholders. The interesting finding was that other major firms in the industry also increased their disclosure behavior. Similar results were found by Deegan et al. (2000), who examined five major environmental incidents and the resulting changes in disclosure behavior. They find a common pattern in all five cases, with firms in the industry increasing disclosure after an environmental disaster. This was true for firms listed on U.S. stock markets but also for firms on other markets too. In our framework, this behavior can be seen as the result of an increase in scrutiny (higher α and hence higher

η) from activists across the entire industry, which motivated firms with substantial environmental risks to move toward greater disclosure.

5.2 Do Greener Firms Disclose More?

As discussed in Section 1, the empirical literature on environmental disclosure contains sharply conflicting results regarding the relationship between environmental performance and environmental disclosures. Our model, by incorporating both firm characteristics and activist pressures, generates a rich set of testable predictions that shed new light on the conflicting results in the existing literature and suggest more precise hypotheses for study in future work. We identify two distinct drivers of disclosure behavior. First, a "greener" firm (that is, a firm with higher r) is *more* likely to generate a strictly positive record that it is wants to disclose. For firms with a "pure" record—that is, either good or bad outcomes but not both—the unique equilibrium strategy for firms of all types is to disclose positive outcomes and withhold negative ones. For these states of the world, higher values of r are associated with better expected outcomes and hence a larger expected number of voluntary disclosures. Second, in the face of activist pressure, a greener firm is *less* likely to make any disclosures at all when it has a mixed record—both positive and negative outcomes, that is the state (1,1)—to report. This effect becomes increasingly pronounced as activist pressure intensifies. Thus, there is no monotonic relationship between environmental performance and environmental disclosure, and empirical researchers should not expect to find one in the data. Instead, empirical work should focus on including variables that capture the full set of drivers of disclosure behavior, including the intensity of activist pressure.

To connect performance and disclosure, we first need a proper definition of environmental performance. In our model, environmental performance is of interest from both the perspective of period 0 (*ex ante*) and period 1 (the interim period).[26] At the outset, the firm's expected environmental performance *ex ante* is captured by the parameter r, which expresses the firm's likelihood of positive environmental outcomes. At the interim period, performance is imperfectly known, even to the firm itself. The best measure of performance is the firm's state, for example, (1,1) or (0,0), although even then one can imagine ranking (1,1) either better or worse than (0,0); furthermore, we must be clear that this is really a measure only of the firm's *known* performance at that point.

26. Environmental performance can also be characterized in period 2, when there are only three possible states: (2,0), (1,1), or (0,2). At this point, performance is common knowledge, so disclosure behavior is irrelevant.

Disclosure only occurs in the interim period, so it seems natural to assess the relationship between performance and disclosure from the perspective of the interim period. However, with multiple dimensions of performance, it is not obvious how to rank the various possible states. For example, is a firm in state (0,0) a better environmental performer than a firm in state (1,1)? The only way to evaluate state (0,0) is in terms of its expected performance, $ru + (1 - r)d$. But this evaluation requires adopting the *ex ante* notion of performance, which in turn requires that we focus on the expected number of disclosures. Thus, we shall focus on whether firms with better expected performance (e.g., higher values of r) have a greater number of expected disclosures.

To measure expected disclosures *ex ante*, we must account for the firm's disclosure behavior in each possible state and then weight this behavior by the probability of each relevant state. In fact, there are only three states in which the firm makes any disclosures at all: states (2,0),(1,0), and (1,1). In the state (2,0), which occurs with probability $r^2\theta^2$, the firm discloses two successes. In the state (1,0), which occurs with probability $2r\theta(1 - \theta)$, the firm discloses one success. Thus, the number of expected disclosures for a firm of type (r, θ) in the states (2,0) and (1,0) is

$$2r^2\theta^2 + 2r\theta(1 - \theta).$$

As our analysis has shown, the state (1,1) is more complicated. In this state, which occurs with probability $2r(1 - r)\theta^2$, the number of disclosures depends upon the nature of the equilibrium, and could be either zero, one, or two. For example, a firm with a high likelihood of good performance may choose to make no disclosures in state (1,1) whereas a firm with a lower value of r may engage in greenwash. In fact, Proposition 1 implies that when firms have a mixed record (i.e., one success and one failure) firms with lower r always disclose (weakly) more than firms with higher r. In Figure 1, for example, a firm with a high r makes no disclosures in state (1,1), whereas a firm with a middling level of r engages in greenwash, and a firm with a low r discloses fully. We record this observation in the following Corollary to Proposition 1.

Corollary 3: *For firms with mixed records, a firm with a lower probability of environmental success r makes more disclosures, in expectation, than a firm with the same θ but a higher value of r.*

The foregoing discussion shows why there is no monotonic relationship between r and the number of disclosures. High-r firms are more likely to have purely positive records to disclose, but if they end up with a mixed record, they are more likely to adopt a strategy of withholding

information. The balance of the two depends upon the nature of the equilibrium. If the firm adopts a strategy of making no disclosures in state (1,1), then total expected disclosures are $2r^2\theta^2 + 2r\theta(1-\theta)$. It is straightforward to show that this number is increasing in r. If the firm adopts a pure strategy of engaging in greenwash in state (1,1), then total expected disclosures are $2r^2\theta^2 + 2r\theta(1-\theta) + 2r(1-r)\theta^2 = 2r\theta$, and it is obvious this number is increasing in r. If the firm adopts a strategy of full disclosure in state (1,1), then total expected disclosures are $2r^2\theta^2 + 2r\theta(1-\theta) + 4r(1-r)\theta^2 = 2r\theta(1+\theta-r\theta)$. Differentiating with respect to r yields $2\theta(1+\theta-2r\theta)$, which can be shown to be positive for $r > 0$ and $\theta > 0$. Hence, total expected disclosures are once again increasing in r. If the firm adopts a mixed strategy, then the firm's expected number of disclosures when it greenwashes with probability p_{10} and fully discloses with probability p_{11} is

$$E(Disclosures) = 2r^2\theta^2 + 2r\theta(1-\theta) + 2r(1-r)\theta^2(p_{10} + 2p_{11}). \tag{6}$$

The above argument establishes that for two firms using exactly the same disclosure strategy in the state (1,1), the firm with the higher r has a larger number of expected disclosures. However, a firm with a higher value of r is more likely to adopt a strategy of making no disclosures in the state (1,1), which can lead it to have lower expected disclosures than a firm with a lower r. We establish these points in the following Proposition.

Proposition 4: *For two firms using exactly the same disclosure strategy in the state (1,1), the firm with the higher r has a larger number of expected disclosures. However, if a firm with a higher r would make a different disclosure than another firm in state (1,1), then it may have either more or fewer expected disclosures.*

Proof. See the Appendix.

The ambiguous effect of r on total expected disclosures should not be surprising in light of Corollary 3, which shows that for state (1,1), increases in r weakly decrease the number of disclosures. Hence there is a tradeoff between the effects of increasing r in states (2,0) and (1,0), and the effects of increasing r in state (1,1). As a result, when the firm faces the possibility of penalties for greenwashing, there is no general result linking r and the number of expected disclosures. It is interesting to note that this result contrasts with that of Sinclair-Desgagne and Gozlan (2003), in whose model an increase in the subjective likelihood of environmental damages always triggers more environmental disclosures. The difference is that in their model the firm knows its environmental outcomes with certainty at the outset of

the game, and the beliefs of other stakeholder are simply subjective probability estimates of the outcome. In our model, the firm is not perfectly informed about its own performance at the outset of the game, and r represents both the actual probability of a good outcome as well as the rational expectation of both the firm and the market about this probability. Nevertheless, their results parallel our results for the interim stage of the game, as presented in our Corollary 3 above.

This observation has important implications for empirical work, and offers an explanation for the contradictory findings in the empirical literature. Indeed, empirical researchers should not expect to find a simple monotonic relationship between environmental performance and environmental disclosures. It is essential to control for activist pressures as well as the firm's environmental performance. Furthermore, these two dimensions jointly influence the firm's disclosure strategy, and hence should be controlled for through an interaction term, as done, for example, by Reid and Toffel (2009).

5.3 Environmental Management Systems and Activist Auditing

In our model, θ measures the likelihood that a firm knows its environmental impacts at period 1. In practice, an increase in θ is likely to correspond to a firm's adoption of an EMS, such as ISO 14001. An EMS is a set of management processes and procedures that allows an organization to integrate environmental issues into day-to-day decisions. Of course, a necessary component of an EMS is a reliable system for measuring a firm's environmental impacts.[27] In this section, we show that the adoption of an EMS has important effects on a firm's disclosure strategy, and identify a new rationale for public policies that encourage firms to adopt EMSs.

As shown in Proposition 2, the threat of activist audits does not lead all firms to increase their disclosures. In particular, activist auditing leads to reduced disclosures from "uninformed greens" ($r > r_{FN}$), although it also leads to an increase in disclosures from "informed browns" ($r < r_{FN}$). As $\eta \to 1$, firms no longer engage in greenwash at all. At this point, firms separate themselves into two groups. One group, positioned below the dashed line in Figure 3, elects to fully disclose. The other group, positioned above the dashed line in Figure 3, elects not to disclose at all. As θ increases, more and more firms fall below the dashed line in Figure 3. This means that activist auditing is more likely to increase disclosures for firms that have a high value of θ.

27. For more details on EMSs, see Coglianese and Nash (2001) and Delmas (2000).

Thus, whether activist auditing is likely to increase disclosures depends in part on the presence of EMS within the audited firms.

We can make the above discussion more precise if we focus on high values of θ. Figures 1–3 all show that as θ, approaches unity, there are only two types of disclosure equilbria: full disclosure, and mixing between full and partial disclosure. Of particular interest is how the firm's mixing behavior changes with θ in the region labeled Mix_{FP} in the Figures. As shown earlier, we can derive a closed-form solution for the firm's probability of greenwashing in this region, which is given by

$$p_{10} = \left(\frac{r}{\eta} - 1\right) \frac{1-\theta}{(1-r)\theta}.$$

Differentiating with respect to θ reveals that

$$\frac{\partial p_{10}}{\partial \theta} = \frac{(\eta - r)}{(1-r)\eta\theta^2} < 0,$$

where the sign is determined by the fact that we are in a region where $r > \eta$. Thus, p_{10} is monotonically decreasing in θ. Furthermore, it is easy to see that as $\theta \to 1$, $p_{10} \to 0$. Thus, in the limit we obtain the result of the early literature, for example, Milgrom and Roberts (1986), which found that full disclosure is an equilibrium when it is common knowledge that the firm has complete information about its own environmental performance.[28] We record the foregoing discussion in the following Proposition.

Proposition 5: *As $\theta \to 1$, the equilibrium disclosure strategy for firms with mixed records converges to full disclosure.*

In our model, firms have no incentive to invest in an EMS. In period 0, as we can observe from equation (1), the firm's value is not enhanced by adopting an EMS because the EMS only affects θ, which does not enter the firm's *ex ante* expected value. Furthermore, in period 1, if the manager learns he has a failure, he would prefer not to have an EMS in place, because its existence makes it harder for him to convince the market that he does not know the true state. As the likelihood increases that the manager knows the environmental outcomes of the firm's activities, the market increasingly interprets nondisclosure as withheld negative information rather than as true uncertainty on the

28. In the limit as $\theta \to 1$, there are only three possible states: (2,0), (1,1), and (0,2). In the state (2,0), the firm prefers to disclose its successes. In the state (0,2), the market will infer that the firm has two failures if it does not disclose, so the firm may as well disclose fully. However, as long as $\theta < 1$, the firm will elect not to disclose information about failures because there is no penalty for nondisclosure. Thus, greenwashing behavior is more sensitive to changes in θ than is nondisclosure.

part of the manager. Adopting an EMS improves the manager's internal information, and thus makes the market increasingly skeptical when the manager does not fully disclose all possible environmental information. On the other hand, when the manager has no failures in period 1, he is indifferent to the presence of an EMS because he can verifiably disclose successes to the market anyway. Thus, the manager has no incentive to invest in an EMS.

Of course, our model does not incorporate the benefits of an EMS in terms of improved internal control and ability to comply with environmental regulations. Nevertheless, our analysis does identify a countervailing incentive that may deter firms from adopting EMSs. Furthermore, our observations are broadly consistent with the empirical results of Delmas (2000), who finds that many U.S. firms elect not to adopt ISO 14001 (a particular form of EMS) because they wish to limit public access to internal information about their environmental performance.

Our results suggest that public policy pressures may be required to induce a broad cross-section of firms to adopt EMSs. Interestingly, Coglianese and Nash (2001, p. 15) find that there has been "an explosion of programs in the United States that offer financial and regulatory incentives to firms that implement EMSs." These programs are being implemented at both the federal and state levels. Whether these programs are likely to achieve their objectives is unclear. Coglianese and Nash (2001, p. 16) point out that "[a]ll of these policy initiatives are premised on the assumption that EMSs make a difference in environmental performance. Yet this question merits research and evidence rather than untested optimism."

Our analysis points to a new rationale for encouraging firms to adopt EMSs, one that does not appear to have been recognized in prior literature, either by academics or practitioners. We do not presume that an EMS makes any difference in environmental performance, but instead simply assume an EMS improves the manager's internal information about the firm's environmental performance. In this capacity, an EMS operates as a complement to activist auditing of environmental disclosure and greenwash. With a strong EMS in place, when a manager discloses nothing about the firm's environmental performance, the market infers that the manager is hiding some negative information, and thus downgrades its rating of the firm's value. In turn, this means that an activist's threat to punish greenwash is more likely to drive the manager to disclose fully rather than to not disclose at all.

6. Conclusions

This paper has presented what is to our knowledge the first economic model of "greenwash," in which a firm strategically discloses environmental information and an activist may audit and penalize the firm for disclosing positive but not negative aspects of its environmental performance. We modeled this phenomenon using tools from the literature on financial disclosure. In our model, an activist can audit corporate environmental reports, and penalize firms caught engaging in greenwash, that is presenting good environmental news while hiding bad news. Our model is relatively simple, yet produces some interesting positive implications. We show that when faced with activist pressure, the types of firms most likely to engage in partial disclosure are those with an intermediate probability of producing positive environmental and social outcomes. For such firms, disclosing a success can produce a significant improvement in public perception, and withholding information about a failure can prevent a significant negative public perception; thus, they are willing to risk public backlash by disclosing only partially. Our results rationalize conflicting results in the empirical literature because we show that there exists a nonmonotonic relationship between a firm's expected environmental performance and its environmental disclosures. The reason is that high performers are more likely to have purely positive records to disclose, but if they end up with a mixed record, they are more likely to adopt a strategy of withholding information. In addition, we find that activist auditing of corporate disclosure behavior is more likely to induce a firm to become more open and transparent if the firm is likely to have socially or environmentally damaging impacts, and if the firm is relatively well informed about its environmental or social impacts. This description fits quite well with the broad types of firms typically singled out for scrutiny and outrage by activists.

The model also has interesting normative implications. If the activist's goal is to increase firm disclosures, then it needs to be very careful in targeting suspected greenwashers. There is a real possibility that the threat of public backlash for greenwash will cause firms to "clam up" rather than become more open and transparent. In particular, such a response is likely from firms with a high probability of successful projects, yet who are not fully informed about the environmental impacts of their actions. For firms such as this, activist pressures designed to increase disclosure may backfire and produce exactly the opposite of the intended results. On the other hand, firms with a low probability of environmental success can be pressured into making

more environmental disclosures, and thus make better targets for anti-greenwash campaigns.

The likelihood that a firm responds to the threat of activist auditing by opting for nondisclosure is reduced if the firm has adopted an EMS, and the complementarity between EMSs and activist auditing of greenwash points to a benefit from public policies that mandate the adoption of EMSs. Indeed, our analysis points to a new rationale for encouraging firms to adopt EMSs. An EMS brings the market closer to a state of common knowledge, thereby increasing market efficiency. With an EMS in place, the manager is better informed about his firm's environmental impact, and the market knows that the manager is better informed. As a result the manager is unable to hide behind the veil of ignorance when he fails to fully disclose the impacts of his firm's actions, and is thereby pressured to fully disclose.

There are a number of areas in which further research would be valuable. One need is for more empirical work on greenwash, its effects on firm valuation, and its interaction with activist information campaigns. Ramus and Montiel (2005) and Kim and Lyon (forthcoming) represent useful steps in this direction, but more work is needed before we have a robust empirical understanding of the phenomenon. A second need is to explore more fully the motivations of activist groups that monitor and punish corporate hypocrites. Articulating their objective functions—maximizing membership, maximizing financial contributions, affecting change in the industry, or some mix of the above—would allow for a strategic analysis of activist behavior, and the equilibrium of such a model would produce further insights into corporate nonmarket strategy. Third, it would also be interesting to extend the model so that the firm's activities are heterogeneous in nature, varying in cost, likelihood of success, and social or environmental impact. This would allow for an analysis of firms' incentives to invest in projects known to have a high probability of success but low social or environmental value, an accusation leveled against some firms. In this case, partial disclosure may divert scarce funds from valuable risky projects to relatively certain but low-value projects.

Appendix

In this appendix, we present some derivations of formulae that appear in simplified form in the text.

Full Disclosure Equilibrium

$$\begin{aligned}\Psi(0,0) &= (1-\theta)^2 + 2(1-r)\theta(1-\theta) + (1-r)^2\theta^2 \\ &= (1-r\theta)^2\end{aligned}$$

$$
\begin{aligned}
&EV[0,0\,|\,1,1,\mu_F] \\
&\quad = \frac{(1-\theta)^2 2(ru+(1-r)d)\mu(0,0\,|\,0,0)}{\Psi(0,0)} \\
&\qquad + \frac{2(1-r)\theta(1-\theta)[d+(ru+(1-r)d)]\mu(0,0\,|\,0,1)}{\Psi(0,0)} \\
&\qquad + \frac{(1-r)^2\theta^2 2d\,\mu(0,0\,|\,0,2)+2r(1-r)\theta^2(u+d)\mu(0,0\,|\,1,1)}{\Psi(0,0)} \\
&\quad = \frac{(1-\theta)^2 2(ru+(1-r)d)+2(1-r)\theta(1-\theta)[d+(ru+(1-r)d)]}{\Psi(0,0)} \\
&\qquad + \frac{(1-r)^2\theta^2 2d}{\Psi(0,0)} \\
&\quad = \frac{2\,(d(1-r)+ru(1-\theta))}{(1-r\theta)}
\end{aligned}
$$

$$
\begin{aligned}
\Psi(1,0) &= 2r\theta(1-\theta)\mu(1,0\,|\,1,0)+r^2\theta^2\mu(1,0\,|\,2,0) \\
&\quad +2r(1-r)\theta^2\mu(1,0\,|\,1,1) \\
&= 2r\theta(1-\theta)
\end{aligned}
$$

$$
\begin{aligned}
EV[1,0\,|\,1,1,\mu_F] &= (u+(ru+(1-r)d))\frac{2r\theta(1-\theta)\mu(1,0\,|\,1,0)}{\Psi(1,0)} \\
&\quad +2u\frac{r^2\theta^2\mu(1,0\,|\,2,0)}{\Psi(1,0)} \\
&\quad +(u+d)\frac{2r(1-r)\theta^2\mu(1,0\,|\,1,1)}{\Psi(1,0)}-\alpha P \\
&= (u+(ru+(1-r)d))\frac{2r\theta(1-\theta)}{\Psi(1,0)}-\alpha P \\
&= u+(ru+(1-r)d)-\alpha P.
\end{aligned}
$$

Nondisclosure Equilibrium

$$
\begin{aligned}
\Psi(0,0) &= 1-\Pr(1,0)-\Pr(2,0) \\
&= 1-2\theta r(1-\theta)-r^2\theta^2 \\
&= 1-\theta r\,(2-(2-r)\theta)
\end{aligned}
$$

$$
\begin{aligned}
&EV[0,0\,|\,1,1,\mu_N]\\
&\quad=\frac{(1-\theta)^2 2(ru+(1-r)d)+2(1-r)\theta(1-\theta)(d+(ru+(1-r)d))}{1-\theta r\,(2-(2-r)\theta)}\\
&\qquad+\frac{(1-r)^2\theta^2 2d+2r(1-r)\theta^2(u+d)}{1-\theta r\,(2-(2-r)\theta)}.
\end{aligned}
$$

Partial Disclosure Equilibrium

$$
\begin{aligned}
\Psi(1,0)&=2r\theta(1-\theta)\mu(1,0\,|\,1,0)+2r(1-r)\theta^2\mu(1,0\,|\,1,1)\\
&=2r\theta(1-r\theta)
\end{aligned}
$$

$$
\begin{aligned}
EV[1,0\,|\,1,1,\mu_P]&=(u+(ru+(1-r)d))\frac{2r\theta(1-\theta)\mu(1,0\,|\,1,0)}{\Psi(1,0)}\\
&\quad+2u\frac{r^2\theta^2\mu(1,0\,|\,2,0)}{\Psi(1,0)}\\
&\quad+(u+d)\frac{2r(1-r)\theta^2\mu(1,0\,|\,1,1)}{\Psi(1,0)}-\alpha P\\
&=(u+(ru+(1-r)d))\frac{2r\theta(1-\theta)}{\Psi(1,0)}\\
&\quad+(u+d)\frac{2r(1-r)\theta^2}{\Psi(1,0)}-\alpha P\\
&=(u+(ru+(1-r)d))\frac{2r\theta(1-\theta)}{2r\theta(1-r\theta)}\\
&\quad+(u+d)\frac{2r(1-r)\theta^2}{2r\theta(1-r\theta)}-\alpha P\\
&=\frac{u(1+r(1-2\theta))+d(1-r)}{1-r\theta}-\alpha P.
\end{aligned}
$$

$$
\begin{aligned}
\Psi(0,0)&=(1-\theta)^2\mu(0,0\,|\,0,0)+2(1-r)\theta(1-\theta)\mu(0,0\,|\,0,1)\\
&\quad+(1-r)^2\theta^2\mu(0,0\,|\,0,2)+2r(1-r)\theta^2\mu(0,0\,|\,1,1)\\
&=(1-r\theta)^2
\end{aligned}
$$

$$
\begin{aligned}
&EV[0,0\,|\,1,1,\mu_P] \\
&\quad = \frac{(1-\theta)^2 2(ru+(1-r)d)\mu(0,0\,|\,0,0)}{\Psi(0,0)} \\
&\qquad + \frac{2(1-r)\theta(1-\theta)(d+(ru+(1-r)d))\mu(0,0\,|\,0,1)}{\Psi(0,0)} \\
&\qquad + \frac{(1-r)^2\theta^2 2d\,\mu(0,0\,|\,0,2)+2r(1-r)\theta^2(u+d)\mu(0,0\,|\,1,1)}{\Psi(0,0)} \\
&\quad = \frac{2\left(d(1-r)+ru(1-\theta)\right)}{(1-r\theta)}.
\end{aligned}
$$

We use the following Lemma in the proof of Proposition 1.

Lemma: *For* $r > \frac{1-\sqrt{2(1-\theta)}}{\theta}$, $N(r,\theta) \equiv \frac{(1-r)\left(r^2\theta^2+1\right)}{(1-\theta r(2-(2-r)\theta))} > G(r,\theta) \equiv \frac{1-r}{1-r\theta}$.

Proof. Define

$$
\begin{aligned}
\Delta(r,\theta) = N(r,\theta) - G(r,\theta) &= \frac{(1-r)\left(r^2\theta^2+1\right)}{(1-\theta r\,(2-(2-r)\theta))} - \frac{1-r}{1-r\theta} \\
&= \frac{\left(2\theta-2r\theta+r^2\theta^2-1\right)}{\left(2r\theta-2r\theta^2+r^2\theta^2-1\right)}\frac{(1-r)\,r\theta}{(1-r\theta)},
\end{aligned}
$$

the second term of which is always nonnegative. Hence the sign of $\Delta(r,\theta)$ is the same as the sign of

$$
\frac{\left(2\theta-2r\theta+r^2\theta^2-1\right)}{\left(2r\theta-2r\theta^2+r^2\theta^2-1\right)}.
$$

Differentiating the denominator shows that it is increasing in both r and θ for $r \in (0,1)$ and $\theta \in (0,1)$. Evaluating the denominator at $r=\theta=1$, we find that it is zero at that point; hence, it is negative for all for $r \in (0,1)$ and $\theta \in (0,1)$. Thus, we focus on signing the numerator. Solving for $2\theta(1-r)+(r\theta+1)(r\theta-1)=0$ we find there are two roots. The one that is defined for $r \in (0,1)$ and $\theta \in (0,1)$ is $r = \frac{1-\sqrt{2(1-\theta)}}{\theta}$. Hence the numerator is negative for $r > \frac{1-\sqrt{2(1-\theta)}}{\theta}$. Thus, we have shown that $\Delta(r,\theta) > 0$ for $r > \frac{1-\sqrt{2(1-\theta)}}{\theta}$. □

Proof of Proposition 1. The interested reader may benefit from noting that Figures 1–3 contain a series of curves that divide the unit square with axes θ and r into strategy regions. Ascending from the bottom of the figure, these curves are defined by a series of nonnegative values $r_{FP} < r_{PF} < r_{FN} < r_{PN} < r_{NP} < 1$. Note that in Figures 2 and

3 the curves defined by r_{PF} and r_{PN} disappear because the expected penalty for greenwash is so large as to eliminate partial disclosure as a pure strategy equilibrium. Note also that we will denote market beliefs corresponding to a particular pure-strategy equilibrium with the notation μ_i for $i \in \{F, N, P\}$ defining the full-disclosure, nondisclosure and partial-disclosure equilibria, respectively. To avoid unnecessary notational clutter, we will use this notation sparingly, reserving it for expressions representing the firm's expected values under different sets of beliefs.

(a) Consider the case where $\eta = 0$. Because $ru + (1-r)d > d$, the firm will always engage in partial disclosure if there is no penalty for doing so.

(b) Consider the case where $\eta \in (0, 1/2)$. Begin with $r < r_{FP}$. The term $r_{FP} \equiv \frac{\alpha P}{u-d} = \eta$ is defined as the value of r for which $EV[1,1 \mid 1,1,\mu_F] = EV[1,0 \mid 1,1,\mu_F]$, so for all $r < r_{FP}$, the firm prefers full disclosure to partial disclosure. In addition, the condition $EV[1,1 \mid 1,1,\mu_F] > EV[0,0 \mid 1,1,\mu_F]$ simplifies to $r < r_{FN} \equiv \frac{1}{2-\theta}$. We know that $r_{FP} \equiv \frac{\alpha P}{u-d} = \eta < r_{FN} \equiv \frac{1}{2-\theta}$, because $\eta < 1/2$. Hence, for $r < r_{FP}$ we know the firm prefers full disclosure to nondisclosure. Hence, the unique equilibrium for $r < r_{FP}$ is full disclosure.

Next consider the range $r \in [r_{FP}, r_{PF})$. We define $r_{PF} \equiv \frac{\eta}{\theta\eta+(1-\theta)}$ to be the value of r for which $EV[1,0 \mid 1,1,\mu_P] > EV[1,1 \mid 1,1,\mu_P]$, and it is straightforward to calculate $r_{PF} - r_{FP} = \frac{\theta\eta(1-\eta)}{\theta\eta+(1-\theta)} > 0$, so we know $r_{PF} > r_{FP}$. On this range, $EV[1,1 \mid 1,1,\mu_F] < EV[1,0 \mid 1,1,\mu_F]$, so full disclosure is not a pure strategy equilibrium. However, $EV[1,0 \mid 1,1,\mu_P] > EV[1,1 \mid 1,1,\mu_P]$ reduces to $r > r_{PF} \equiv \frac{\eta}{\theta\eta+(1-\theta)}$, so for $r < r_{PF}$, partial disclosure is not a pure-strategy equilibrium either. Hence, for $r_{FP} < r < r_{PF}$, there is no pure-strategy equilibrium, and both full disclosure and partial disclosure are undominated strategies. We know that on this region the manager never chooses full nondisclosure We are interested in showing that there exists a mixed-strategy equilibrium. Define $\xi[1,1 \mid 1,1,r,\theta,\alpha P,\mu_{11}] = EV[1,1 \mid 1,1,r,\theta,\alpha P,\mu_{11}] - EV[1,0 \mid 1,1,r,\theta,\alpha P,\mu_{11}]$, using the simplifying notation $\mu_{sf} = \mu(\hat{s}, \hat{f}; s, f)$. Some algebra yields

$$\xi[1,1 \mid 1,1,r,\theta,\alpha P,\mu_{11}] = \alpha P - r(u-d) + \frac{\alpha P(1-r)\theta(1-\mu_{11})}{1-\theta}.$$

It is immediate that $\xi[1,1 \mid 1,1,r,\theta,\alpha P,\mu_{11}]$ is continuous and decreasing in μ_{11}. For a mixed strategy equilibrium to exist for given values $r, \theta, \alpha P$, the market must assign a probability μ_{11} to the manager disclosing fully when the state is (1,1) such that $\xi[1,1 \mid 1,1,r,\theta,\alpha P,\mu_{11}] = 0$. We are interested in the range $r_{FP} < r <$

r_{PF}, and by definition, for every r in this range, we know that $\xi[1,1\,|\,1,1,r,\theta,\alpha P,0] > 0$ and $\xi[1,1\,|\,1,1,r,\theta,\alpha P,1] < 0$. Hence, because $\xi[1,1\,|\,1,1,r,\theta,\alpha P,1]$ is continuous, there exists some value $\mu_{11} \in (0,1)$ such that $\xi[1,1\,|\,1,1,r,\theta,\alpha P,\mu_{11}] = 0$. In fact, on this range there exists a closed form solution with $\mu_{10} = 1 - \mu_{11} = (r/\eta - 1)(1-\theta)/((1-r)\theta) = p_{10}$.

Next consider the range $r \in [r_{PF}, r_{PN})$. We define r_{PN} as the value of r for which $EV[1,0\,|\,1,1,\mu_P] = EV[0,0\,|\,1,1,\mu_P]$, which can be shown to be $r_{PN} = (1-\eta)/(1-\theta\eta)$. Straightforward calculation shows that

$$r_{PN} - r_{PF} = \frac{(1-\theta)(u-d-2P\alpha)(u-d)}{((u-d)(1-\theta)+P\theta\alpha)(u-d-\theta P\alpha)}.$$

It is easy to see that for $\eta \in (0,1/2)$ we have $(u-d-\theta P\alpha) > (u-d-2P\alpha) = 2\eta > 0$, so $r_{PN} > r_{PF}$. For $r > r_{PF}$, we know $EV[1,0\,|\,1,1,\mu_P] > EV[1,1\,|\,1,1,\mu_P]$. Similarly, for $r < r_{PN}$, we know $EV[1,0\,|\,1,1,\mu_P] > EV[0,0\,|\,1,1,\mu_P]$. Hence, on this range, partial disclosure is a pure-strategy equilibrium.

Next consider the range $r \in [r_{PN}, r_{NP})$. We define r_{NP} such that for all $r > r_{NP}$, we have $EV[0,0\,|\,1,1,\mu_N] > EV[1,0\,|\,1,1,\mu_N]$. This can be shown to be equivalent to $(1-r_{NP})\left(r_{NP}^2\theta^2+1\right)/(1-\theta r_{NP}(2-(2-r_{NP})\theta)) = \alpha P/(u-d)$. On the range of interest here, $r < r_{NP}$, so nondisclosure cannot be a pure-strategy equilibrium. However, $r > r_{PN}$, so partial disclosure cannot be a pure-strategy equilibrium. We know that on this region the manager never engages in full disclosure. We are interested in showing that there exists a mixed-strategy equilibrium. Define $\xi[0,0\,|\,1,1,r,\theta,\alpha P,\mu_{00}] = EV[0,0\,|\,1,1,r,\theta,\alpha P,\mu_{00}] - EV[1,0\,|\,1,1,r,\theta,\alpha P,\mu_{00}]$. Some algebra yields

$$\begin{aligned}
&\xi[0,0\,|\,1,1,r,\theta,\alpha P,\mu_{00}] \\
&\quad= \frac{2(1-r\theta)((1-r)d+(1-\theta)ru)+(u+d)2r(1-r)\theta^2\mu_{00}}{(1-r\theta)^2+2r(1-r)\theta^2\mu_{00}} \\
&\qquad- \frac{(u+(ru+(1-r)d))2r\theta(1-\theta)+(u+d)2r(1-r)\theta^2(1-\mu_{00})}{2r\theta(1-\theta+(1-r)\theta(1-\mu_{00}))} \\
&\qquad+\alpha P.
\end{aligned}$$

It is easy to see that $\xi[0,0\,|\,1,1,r,\theta,\alpha P,\mu_{00}]$ is continuous in μ_{00} because neither the denominator of the first or the second fraction ever approaches zero. For a mixed strategy equilibrium to exist for given values $r,\theta,\alpha P$, the market must assign a probability μ_{00} to the manager withholding all information when the state is (1,1) such that $\xi[0,0\,|\,1,1,r,\theta,\alpha P,\mu_{00}] = 0$. We are interested in the range

$r_{PN} < r < r_{NP}$, and by definition, for every r in this range, we know that $\xi[0, 0 \mid 1, 1, r, \theta, \alpha P, 0] > 0$ and $\xi[0, 0 \mid 1, 1, r, \theta, \alpha P, 1] < 0$. Hence, because $\xi[0, 0 \mid 1, 1, r, \theta, \alpha P, \mu_{00}]$ is continuous, there exists some value $\mu_{00} \in (0, 1)$ such that $\xi[0, 0 \mid 1, 1, r, \theta, \alpha P, \mu_{00}] = 0$.

To show that $r_{NP} > r_{PN}$, we construct a proof by contradiction. Define $G(r, \theta) = \frac{1-r}{1-r\theta}$ and $N(r, \theta) = \frac{(1-r)\left(r^2\theta^2+1\right)}{(1-\theta r(2-(2-r)\theta))}$. Recall that r_{NP} is defined implicitly by the relation $N(r_{NP}, \theta) = \eta$ and r_{PN} is defined implicitly by $G(r_{PN}, \theta) = \eta$. Now suppose that for a given η it is the case that $r_{NP} < r_{PN}$. It is straightforward to show that $dG/dr = \frac{\theta-1}{(r\theta-1)^2} \leq 0$, so $G(r_{NP}, \theta) > G(r_{PN}, \theta) = \eta$. Because $N(r_{NP}, \theta) = \eta$ it follows directly that $G(r_{NP}, \theta) > N(r_{NP}, \theta)$. The Lemma shows that this can only be true for $r < \frac{1-\sqrt{2(1-\theta)}}{\theta}$. Solving for the minimimum of $N(r, \theta)$ subject to the constraints $r < \frac{1-\sqrt{2(1-\theta)}}{\theta}$, $\theta > 0$, and $\theta < 1$, we find the minimum occurs at $\theta^* = 2(\sqrt{2} - 1)$ and $r^* = (1 - \sqrt{6 - 4\sqrt{2}})/\theta^*$; substituting in r^* and θ^* yields a value of $N(r^*, \theta^*) \simeq .85355$. Because we are only concerned with $\eta < .5$, we know that $N(r_{NP}, \theta) > \eta$ on the range $r < \frac{1-\sqrt{2(1-\theta)}}{\theta}$ and there cannot exist a solution $r_{NP} < \frac{1-\sqrt{2(1-\theta)}}{\theta}$. Thus, it must not be true that $r_{NP} < r_{PN}$.

Finally, consider the range $r \in [r_{NP}, 1]$. For all $r > r_{NP}$, we know that $EV[0, 0 \mid 1, 1, \mu_N] > EV[1, 0 \mid 1, 1, \mu_N]$, so nondisclosure dominates partial disclosure. Furthermore, nondisclosure dominates full disclosure, that is, $EV[0, 0 \mid 1, 1, \mu_F] > EV[1, 1 \mid 1, 1, \mu_F]$, if $r > r_{FN} \equiv \frac{1}{2-\theta}$. We already know that $r > r_{NP} > r_{PN}$. We need to show that $r_{PN} > r_{FN}$. Recall that $r_{PN} = (1 - \eta)/(1 - \theta\eta)$ and $r_{FN} = \frac{1}{2-\theta}$. So $r_{PN} - r_{FN} = ((2 - \theta)(1 - \eta) - (1 - \theta\eta))/((1 - \theta\eta)(2 - \theta)) = ((1 - \theta)(1 - 2\eta))/((1 - \theta\eta)(2 - \theta)) > 0$ because $\eta < 1/2$. Hence, $r > r_{FN}$ and nondisclosure is a pure-strategy equilibrium. To show that $r_{NP} < 1$, note that $N(1, \theta) = 0$, so for any $\eta > 0$ it is impossible to have $N(r_{NP}, \theta) = \eta$ with $r_{NP} = 1$.

(c) Consider now the case where $\eta > 1/2$. Begin with the range $r < \min(r_{FN}, r_{FP})$. Because $EV[1, 1 \mid 1, 1, \mu_F] > EV[0, 0 \mid 1, 1, \mu_F]$ is equivalent to $r < r_{FN} \equiv \frac{1}{2-\theta}$, we know that full disclosure dominates nondisclosure. Because $EV[1, 1 \mid 1, 1, \mu_F] > EV[1, 0 \mid 1, 1, \mu_F]$ is equivalent to $r < r_{FP} \equiv \frac{\alpha P}{u-d} = \eta$, we know that full disclosure dominates partial disclosure. Hence, full disclosure is pure strategy equilibrium on this range.

Next consider the range $r \in [\min(r_{FN}, r_{FP}), r_{FN})$. For $r_{FP} < r < r_{FN}$, there is no pure-strategy equilibrium, and both full disclosure and partial disclosure are undominated strategies. As shown in part

(b) above for the case where $r \in [r_{FP}, r_{PF})$, there is a mixed strategy equilibrium with solution $\mu_{10} = (r/\eta - 1)(1-\theta)/((1-r)\theta) = p_{10}$.

Next consider the range $r \in [r_{FN}, \max(r_{FN}, r_{NP}))$. For $r_{FN} < r < r_{NP}$, there is no pure-strategy equilibrium, and both nondisclosure and partial disclosure are undominated strategies. We know that on this region the manager never engages in full disclosure. Using the same reasoning as in part (b) for the range $r \in [r_{PN}, r_{NP})$, we can show there exists a mixed strategy equilibrium.

Finally, consider the range $r \in [\max(r_{NP}, r_{FN}), 1)$. Because $EV[0, 0 \mid 1, 1, \mu_N] > EV[1, 1 \mid 1, 1, \mu_N]$ is equivalent to $r > r_{FN} \equiv \frac{1}{2-\theta}$, we know that nondisclosure dominates full disclosure. Because $EV[0, 0 \mid 1, 1, \mu_N] > EV[1, 0 \mid 1, 1, \mu_N]$ is equivalent to $r > r_{NP}$, we know that nondisclosure dominates partial disclosure. Hence, nondisclosure is a pure-strategy equilibrium on this range.

(d) For $\eta \geq 1$, the maximum gain from greenwashing is less than $u - d$ and the firm never has incentives to engage in greenwash. Thus it simply compares full disclosure and nondisclosure. Because $EV[0, 0 \mid 1, 1, \mu_N] > EV[1, 1 \mid 1, 1, \mu_N]$ is equivalent to $r > r_{FN}$, nondisclosure is a pure-strategy equilibrium for $r > r_{FN}$ and full disclosure is a pure-strategy equilibrium for $r < r_{FN}$. □

Proof of Proposition 2: Part (a): We proceed in three steps. First, we show that the size of the region in which greenwash is a pure strategy is diminishing in αP. Second, we show that in the mixing regions, the probability of greenwashing falls with αP. Third, we show that the size of the lower and upper mixing regions does not increase when αP increases.

Step 1: For this step, it is more notationally convenient to work in terms of η rather than αP. As shown in the Proof of Proposition 1, the incentive compatibility conditions for a partial disclosure equilibrium imply that such an equilibrium exists for $r \in [r_{PF}, r_{PN}]$, where $r_{PF} \equiv \frac{\eta}{\theta\eta + (1-\theta)}$ and $r_{PN} = (1-\eta)/(1-\theta\eta)$. If $\alpha P = 0$, then $r_{PF} = 0$ and $r_{PN} = 1$, so partial disclosure is the unique equilibrium strategy for all r and θ. If $\eta = 1/2$ then $r_{PF} = r_{PN}$ and there is no region of partial disclosure. For $\eta \in (0, 1/2)$ the partial disclosure equilibrium exists. Let $R_P = r_{PN} - r_{PF}$ Some calculation shows that $R_P = [(1-\theta)(2\eta - 1)]/[(\theta - \theta\eta - 1)(1 - \theta\eta)]$ Differentiating with respect to η yields $\partial R_P/\partial\eta = -(1-\theta)\left(\theta^2(1-2\eta) + 2\eta^2\theta^2 + 2(1-\theta)\right)/[(\eta\theta - 1)^2(\eta\theta - \theta + 1)^2]$, which is strictly negative for all $\eta < 1/2$.

Step 2: For this step, it is more notationally convenient to work in terms of η rather than αP. In the lower mixing region, we can solve for

the mixed-strategy equilibrium in closed form, with the probability of greenwash being $p_{10} = (r/\eta - 1)(1-\theta)/((1-r)\theta)$. Differentiating with respect to η yields $\partial p_{10}/\partial \eta = -r(1-\theta)/[\eta^2\theta(1-r)] < 0$. Hence the probability of greenwashing decreases with η.

In the upper mixing region, we do not get a closed-form solution. Nevertheless, we can totally differentiate the equation $\nabla = EV[0,0 \mid 1,1,\mu] - EV[1,0 \mid 1,1,\mu] = 0$ defining the mixing probability to determine how the probability of greenwash, p_{10}, changes with changes in αP. We note that $d\nabla/dp_{10} = dEV[0,0 \mid 1,1,\mu]/dp_{10} - dEV[1,0 \mid 1,1,\mu]/dp_{10} > 0$, the sign of which can be established by straightforward differentiation, and $d\nabla/d\alpha P = -dEV[1,0 \mid 1,1,\mu]/d\alpha P = 1$. Hence $dp_{10}/d\alpha P = -[d\nabla/d\alpha P]/[d\nabla/dp_{10}] < 0$. Thus, we have shown that the probability of greenwashing in the upper mixing region decreases with αP, and by extension, with η.

Step 3: It remains to show that the size of the lower and upper mixing regions does not increase with αP. This is easy to do for the lower mixing region, which is bounded above by $\min\{r_{FN} = 1/(2-\theta), r_{FP} = \alpha P/(u-d)\}$. Hence, when αP increases, so does the size of the full disclosure region, shrinking the size of the lower mixing region. The upper mixing region is bounded above by r_{NP}, which is defined implicitly by the equation $(1-r_{NP})\left(r_{NP}^2\theta^2+1\right)/(1-\theta r_{NP}(2-(2-r_{NP})\theta)) = \alpha P/(u-d)$. Consider a point (r',θ') that is in the nondisclosure region for some given penalty $(\alpha P)'$. The fact that it is in this region means that $EV[0,0 \mid 1,1,\mu] > EV[1,0 \mid 1,1,\mu] = [u(1+r'(1-2\theta')) + d(1-r')]/[1-r'\theta'] - (\alpha P)'$. Recall that $EV[0,0 \mid 1,1,\mu]$ is independent of αP. Hence, for any $(\alpha P)'' > (\alpha P)'$, the point (r',θ') is still in the nondisclosure region. Thus, the upper mixing region does not grow in size when αP increases.

Part (b). Uninformed greens are firms with $r > r_{FN}$. These firms either opt for nondisclosure, greenwash, or they randomize between these two strategies. The proof of part (a) above shows that when αP increases, the size of the greenwash pure-strategy region shrinks, the probability of greenwashing in the upper mixing region falls, and the size of the nondisclosure region does not shrink. It follows immediately that uninformed greens increase their likelihood of nondisclosure.

Part (c). Informed browns are firms with $r < r_{FN}$. These firms either opt for full disclosure, greenwash, or they randomize between these two strategies. The proof of part (a) above shows that when αP increases, the size of the greenwash pure-strategy region shrinks, the probability of greenwashing in the lower mixing region falls, and the

size of the full disclosure region increases. It follows immediately that informed browns increase their likelihood of full disclosure. □

Proof of Proposition 4: The first part of the proposition has been proven in the text. For the second part of the proposition, it is sufficient to focus on the case where in state (1,1) the firm mixes between greenwash with probability p_{10} and full disclosure with probability $p_{11} = (1 - p_{10})$. Substituting this last expression into the number of expected disclosures gives

$$\begin{aligned} E(Disclosures) &= 2r^2\theta^2 + 2r\theta(1-\theta) + 2r(1-r)\theta^2(p_{10} + 2p_{11}) \\ &= 2r^2\theta^2 + 2r\theta(1-\theta) - 2r(1-r)\theta^2 p_{10}. \end{aligned}$$

Differentiating this equation with respect to r yields

$$\frac{\partial E(Disclosures)}{\partial r} = 4r\theta^2 + 2\theta(1-\theta) - 2(1-2r)\theta^2 \frac{\partial p_{10}}{\partial r}.$$

Recall from the proof of Proposition 1 that

$$p_{10} = (\frac{r}{\eta} - 1)\frac{1-\theta}{(1-r)\theta},$$

and therefore

$$\frac{\partial p_{10}}{\partial r} = \frac{(1-\theta)(1-\eta)}{(1-r)^2\eta\theta} > 0.$$

Note that $\partial E(Disclosures)/\partial r$ changes sign depending upon θ. For small θ, the first two terms approach zero and the third term grows large and negative. Hence, for small θ expected disclosures fall with increases in r. For $\theta \to 1$, the first two terms approach $4r\theta^2$ and the third term approaches zero. Hence, for large θ, expected disclosures increase with increase in r. Hence, within the lower mixing region, an increase in r has an ambiguous effect on expected disclosures. Similar reasoning can be applied to situations where a change in r causes the firm to switch from any one strategy to another, with the result that expected disclosures can increase or decrease depending upon r and θ. □

References

Baron, D.P., 2003, "Private Politics," *Journal of Economics and Management Strategy*, 12, 31–66.

—— and D. Diermeier, 2007, "Strategic Activism and Nonmarket Strategy," *Journal of Economics and Management Strategy*, 16, 599–634.

Cho, CH. and D.M. Patten, 2007, "The Role of Environmental Disclosures as Tools of Legitimacy: A Research Note," *Accounting, Organizations and Society*, 32, 639–647.

Clarkson, P.M., Y. Li, G.D. Richardson, and F.P. Vasvari, 2008, "Revisiting the Relation between Environmental Performance and Environmental Disclosure: An Empirical Analysis," *Accounting, Organizations and Society*, 33, 303–327.

Coglianese, C. and J. Nash, 2001, *Regulating from the Inside: Can Environmental Management Systems Achieve Policy Goals?* Resources for the Future Press.

—— and M. Rankin, 1996, "Do Australian Companies Report Environmental News Objectively? An Analysis of Environmental Disclosures by Firms Prosecuted Successfully by the Environmental Protection Authority," *Accounting, Auditing and Accountability Journal*, 9, 50–67.

——, ——, and P. Voght, 2000, "Firms' Disclosure Reactions to Major Social Incidents: Australian Evidence," *Accounting Forum*, 24, 101–130.

Delmas, M.A., 2000, "Barriers and Incentives to the Adoption of ISO 14001 by Firms in the United States," *Duke Environmental Law and Policy Forum*, 11, 1–38.

Greer, J. and K. Bruno, 1996, *Greenwash: The Reality Behind Corporate Environmentalism*, Penang, Malaysia: Third World Network and New York: Apex Press.

Grossman, G.M. and E. Helpman, 2001, *Special Interest Politics*, Cambridge, MA: The MIT Press.

Innes, R., 2006, "A Theory of Consumer Boycotts under Symmetric Information and Imperfect Competition," *Economic Journal*, 116, 355–381.

Johnson, G., 2003, *Don't Be Fooled: The Ten Worst Greenwashers of 2003*, Boston: Earthday Resources for Living Green. Available at http://www.thegreenlife.org/dontbefooled.html (accessed December 29, 2010).

Kim, E.H. and T.P. Lyon, "Strategic Environmental Disclosure: Evidence from the DOE's Voluntary Greenhouse Gas Registry," Journal of Environmental Economics and Management (forthcoming)..

Laufer, W.S., 2003, "Social Accountability and Corporate Greenwashing," *Journal of Business Ethics*, 43, 253–261.

Lyon, T.P. and J.W. Maxwell, 2004a, "Astroturf: Interest Group Lobbying and Corporate Strategy," *Journal of Economics and Management Strategy*, 13, 561–597.

—— and ——, 2004b. *Corporate Environmentalism and Public Policy*, Cambridge, UK: Cambridge University Press.

Milgrom, P.R. and J. Roberts, 1986, "Relying on the Information of Interested Parties," *RAND Journal of Economics*, 17, 18–32.

—— and ——, 1992, *Economics, Organization and Management*, Englewood Cliffs, NJ: Prentice Hall.

Patten, D.M., 1992, "Intra-Industry Environmental Disclosures in Response to the Alaskan Oil Spill: A Note on Legitimacy Theory," *Accounting, Organizations and Society*, 17, 471–475.

——, 2002, "The Relation between Environmental Performance and Environmental Disclosure: A Research Note," *Accounting, Organizations, and Society*, 27, 763–773.

Peloza, J., 2005, "Corporate Social Responsibility as Reputation Insurance," Working Paper 24, Center for Responsible Business, University of California, Berkeley.

Ramus, C. and I. Montiel, 2005, "When Are Corporate Environmental Policies a Form of Greenwashing?" *Business and Society*, 44, 377–414.

Reid, E.M. and M.W. Toffel, 2009, "Responding to Public and Private Politics: Corporate Disclosure of Climate Change Strategies," Working Paper 09-019, Harvard Business School.

Shin, H.S., 2003, "Disclosures and Asset Returns," *Econometrica*, 71, 105–133.

Sinclair-Desgagne, B. and E. Gozlan, 2003, "A Theory of Environmental Risk Disclosure," *Journal of Environmental Economics and Management*, 45, 377–393.

Terrachoice, 2007, The Six Sins of Greenwashing. Available at http://sinsofgreenwashing.org/findings/greenwashing-report-2007/ (accessed December 29, 2010).

——, 2009, The Seven Sins of Greenwashing. Available at http://sinsofgreenwashing.org/findings/greenwashing-report-2009/ (accessed December 29, 2010).

Verecchia, R.E., 1983, "Discretionary Disclosure," *Journal of Accounting and Economics*, 5, 179–194.