MICHIGAN
ROSS SCHOOL OF BUSINESS

ERB INSTITUTE
FOR GLOBAL SUSTAINABLE ENTERPRISE
UNIVERSITY OF MICHIGAN

Working Paper

Greenwash: Corporate Environmental Disclosure under Threat of Audit

Thomas P. Lyon
Stephen M. Ross School of Business
at the University of Michigan

John W. Maxwell
Kelley School of Business,
Indiana University

Ross School of Business Working Paper Series
Working Paper No. 1055
March 2006

Greenwash: Corporate Environmental Disclosure under Threat of Audit*

Thomas P. Lyon† and John W. Maxwell‡

March 24, 2006

Abstract

We develop an economic model of "greenwash," in which a firm strategically discloses environmental information and a non-governmental organization (NGO) may audit and penalize the firm for failing to fully disclose its environmental impacts. We identify conditions under which NGO punishment of greenwash backfires, inducing the firm to become less rather than more forthcoming about its environmental performance. We show that complementarities with NGO auditing may justify public policies encouraging firms to adopt environmental management systems. Mandatory disclosure rules offer the potential for better performance than NGO auditing, but the necessary penalties may be so large as to be politically unpalatable. If so, a mix of mandatory disclosure rules, NGO auditing and environmental management systems may be needed to induce full environmental disclosure.

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 more participate in industry-led environmental programs such as those of the World 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 companies naturally want to publicize their environmentally-friendly actions, they are often surprisingly hesitant to promote their environmental successes or to

*We would like to thank Mike Baye, Rick Harbaugh, Charlie Kolstad, John Morgan, and participants in seminars at the American Economic Association meetings, Dartmouth, Indiana University, UC Berkeley, and UC Santa Barbara for their helpful comments.

†Ross School of Business, University of Michigan.

‡Kelley School of Business, Indiana University.

[1]For an introduction to corporate environmental strategy and its relation to public policy, see Lyon and Maxwell (2004b).

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, while 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." 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."[2] Similar concerns surround business efforts to be socially responsible.

Part of the reason managers hesitate to promote their good environmental deeds is that many such actions are attacked as "greenwash" by non-governmental organizations (NGOs). Often these NGOs 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 NGOs 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." 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. 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." Monsanto was Runner Up for the Lifetime Achievement Award for its "tireless promotion of Roundup Ready GM [genetically modified] crops as a solution to world hunger."[4] Ralph Nader reveals a similar skepticism regarding corporate social efforts:

> "One recent misstep is the U.N.'s 'Global Compact.' With the disappointing support of some international human rights and environmental organizations, the U.N. has asked multinational corporations to sign on to the compact's unenforceable and overly vague code of conduct. Companies are able to sign on to the compact and '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

[2] See Peloza (2005), p. 16.

[3] See the instructions for "How to Stop It" under Greenwash 101 at http://www.thegreenlife.org/greenwash101.html.

[4] For details, see "Greenwash Academy Awards Announced at Earth Summit," http://www.corpwatch.org/article.php?id=3648.

with the U.N..."[5]

The activists criticizing firms like BP and Monsanto are undoubtedly trying to press these firms to improve their environmental and social performance. Nevertheless, if companies fear being publicly "smeared" for their environmental and social initiatives, this outcome may not be achieved. Perhaps unfortunately, activists often react more angrily to firms that lay claim to being virtuous, and then are discovered to have feet of clay, than it does to firms that never make such claims. For example, BP makes frequent public claims about its efforts to reduce global warming yet is denounced at the Johannesburg Earth Summit, while Exxon has for years been among the loudest skeptics about climate change yet attracts less attention from activists. Unfortunately, popular usage of the term "greenwash" tends to be both strident and vague. For example, in their book on greenwash, Greer and Bruno (1996) never actually define the term. 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." The implication is clearly that companies are misrepresenting their environmental performance. However, our reading about greenwash indicates that firms' environmental claims are typically not false, although they often fail to present the full picture. It turns out this distinction has important implications for the effectiveness of NGO campaigns against greenwashers, and also suggests some novel public policy measures.

In this paper, we present what is to our knowledge the first economic analysis of greenwash. Since public discussion of greenwash is often polemical and imprecise, we begin in section 2 by developing a clear formal definition of greenwash, and distinguishing it from other "disinformation" strategies. In section 3, we build a simple model in which a company conducts multiple projects with environmental impacts that may turn out well or turn out poorly. Good results, if publicly known, produce rewards (which may come about through the market or through political or social forces) while bad results are damaging. The firm has the option whether or not to reveal its performance on any activities. In these respects, the model follows the disclosure literature in finance and accounting.

Where we depart from the disclosure literature is in section 4, where we add the phenomena of increased public scrutiny of firms that selectively report good news, and public backlash against perceived greenwash. It would seem that the purpose of the punishment is to prompt full disclosure, but we find that there are many circumstances in which this does not happen. We characterize fully how the possibility of NGO punishment influences the firm's disclosure decisions, and show how these effects depend upon underlying parameters reflecting the firm's probability of success in its environmental activities, and the probability

[5]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

the firm is informed about the outcome of its activities at the time it makes a disclosure. We find that punishing greenwash is more likely to motivate full disclosure in settings where the probability of success is low and the likelihood that the firm is informed is high, that is, for firms in dirty industries that are well informed about their own environmental impacts. In section 5, we consider potential complementarities between NGO auditing of greenwash and corporate adoption of an environmental management system (EMS), and show that these complementarities may justify public policies encouraging firms to adopt EMSs. Our analysis points to a new rationale for encouraging firms to adopt EMSs: with an EMS in place, the firm is more likely to be well informed about its own environmental impact, and more importantly, *the market knows that the firm is more likely to be well informed*. As a result the firm is unable to hide behind the veil of ignorance when it fails to fully disclose the impacts of its actions, and is thereby pressured to fully disclose. In section 6, we study the effects of mandatory disclosure rules (such as those created by the Public Company Accounting and Reform Act of 2002, commonly know as Sarbanes-Oxley). We show that mandatory disclosure rules offer the potential for better performance than NGO auditing, but that the necessary penalties may be so large as to be politically unpalatable. Finally, we consider the interaction between mandatory disclosure rules, NGO auditing, and the adoption of EMSs, and offer some tentative suggestions regarding how these mechanisms can best be combined. Section 7 concludes.

2 What is Greenwash?

Formal analysis of greenwash requires a clear definition of the phenomenon. Unfortunately, popular usage of the term, and even academic discussion of it, tends to be broad and vague. As mentioned above, in their book on greenwash, Greer and Bruno (1996) never actually define the term. 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 actions 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. While we find these distinctions useful, in our view, these activities differ too much to be as a single phenomenon; indeed, we have already modeled the use of "astroturf lobbying" through "front groups" in Lyon and Maxwell (2004a).[6]

Turning to the dictionary, we find that *Webster's New Millenium Dictionary*

[6] Astroturf lobbying involves the provision of soft information targeted at a public decision-maker to influence policy decisions. Greenwash involves public disclosure of hard information targeted to influence stock prices.

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 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. These ideas are consistent with what Laufer refers to as "confusion."

The term "disinformation" goes somewhat further than mere "confusion," and 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 NGOs 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*:[7]

> "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."

This example, like those outlined in the Introduction, depicts a company making a statement that is true, yet not the whole truth. We view this as paradigmatic of greenwash. In fact, *Don't Be Fooled* implicitly agrees with this perspective. Consider its discussion of BP's "On the Street" campaign:

> "On the Street" is only selectively honest. The ads mention BP's solar power and clean fuel initiatives, but fail to mention other important initiatives. For example, during 2003, BP made an "ultra-deep" petroleum discovery off the coast of Angola, launched an oil products terminal in the expanding market of Guangdong, China, and acquired a 50% stake in Russia's third-largest oil and gas business. Contrary to the focus of "On the Street", BP's innovations and investments are by no means limited to environmental endeavors."[8]

To us, it seems absurd that the public would believe all BP investments are environmental. Nevertheless, the BP example is of interest because it supports

[7] See Johnson (2003).
[8] Johnson (2003), page 14.

our view that greenwash can be defined as the *selective disclosure of positive information about a company's environmental or social performance, without full disclosure of negative information on these dimensions.*[9]

An excellent example of selective disclosure comes the Department of Energy's Voluntary Greenhouse Gas Reporting program, created by section 1605b of the Energy Policy Act of 1992. Kim and Lyon (2006) 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, non-participant 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.

Note that greenwash is *not* the same as having a poor record of environmental performance. A firm can have a poor record without presenting any positive information about itself, or can have a relatively good record while simultaneously promoting its positive actions publicly and failing to discuss its (few) negative environmental impacts. Note also that greenwash is not the same as simply failing to report negative information; greenwash involves the additional step of selectively choosing to report positive information. These distinctions will turn out to have important implications as we develop our formal model below.

3 Basic Model

Our model focuses on a single firm, whose stock is traded publicly, and a non-governmental organization (NGO). The firm has N different activities that each have some potential environmental impact.[10] The magnitude of N is assumed to be common knowledge, e.g., available on the firm's web site or Annual Report; the non-environmental aspects of the firm's operations are assumed to be already incorporated into the firm's market value. However, the firm's environmental profile, i.e., the impact of the firm's portfolio of "environmentally friendly" activities, is not known at the outset of the model. We assume the market sets the firm's value at its actuarily fair level.[11]

There are 3 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

[9] 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).

[10] We refer to environmental impacts for concreteness, but could just as easily refer to corporate social responsibility more generally.

[11] 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 hypocrisy.

will be a liability associated with any given activity. 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, and a "failure" of value $d < u$ with probability $1 - r$. Thus, the expected number of environmental failures the firm faces is simply $(1 - r)N$. Its market value in period 0 is

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

where $\widetilde{V}$ 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 $\widetilde{V}$ 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.[12]

We assume there is a probability θ that the manager actually learns the social impact of the activity by period 1.[13] Thus, at the interim period, the expected number of activities for which the manager has information on social outcomes is θN. The expected number of activities known to have social liabilities at the interim period is $\theta(1 - r)N$. The manager has the ability to disclose publicly the number of activities that are known to be successes. We assume that all such disclosures are verifiable by outside parties. Thus, the manager is free to selectively withhold information, but he cannot actually lie to outsiders. We assume the manager adopts a disclosure strategy that maximizes the value of the firm.

Let n be the actual number of activities whose liabilities are known at the interim period, s be the number of successes and f the number of failures, so that $n = s + f$. Let the manager's disclosures of the number of successes and failures be given by $\widehat{s}$ and $\widehat{f}$. We assume $V_1 = E(V_2)$. If the market knows s and f, as would be the case if the manager fully disclosed its information in period 1, then

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

where $u =$ the additive impact of a success on the firm's value and $d =$ the additive impact of a failure on the firm's value. This formula is quite intuitive, since u and d are the values of successes and failures, respectively, and $(ru+(1-r)d)$ is the expected value of an activity whose social impact remains unknown.

If the manager discloses $\widehat{s} > 0$, and the total number of disclosures $\widehat{s} + \widehat{f}$ is less than N, the NGO may investigate the manager's report for the possibility of

[12] There are many reasons a manager wants to influence the stock price, e.g. compensation packages that are linked to stock price performance. For further details, see Milgrom and Roberts (1992).

[13] It is worth noting that we would expect θ to be greater for firms that have created an environmental management system. We return to this issue in section 6.

greenwash (i.e., that the manager has a bad outcome that he failed to disclose).[14] With probability α the NGO obtains hard (verifiable) information about the true values of s and f at the interim period and mounts a 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 NGO 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.[15]

We are interested in Perfect Bayesian Equilibria (PBE), which involve specifying a disclosure strategy for the manager, a market valuation, and a set of beliefs for each time t such that (a) the disclosure strategy $(\widehat{s}, \widehat{f})$ is a best response mapping for a firm with actual social profile (s, f), given the market's pricing policy and the beliefs of the market and the NGO, (b) $V_1 = E(V_2)$ given the market's beliefs at period 1 and the manager's disclosure strategy, and (c) at period 0 the market believes the expected number of social liabilities is rN, and at period 1 it computes the expected number of social liabilities using Bayes' rule, conditional on any social reports. We will focus on pure strategy equilibria.

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 $f = 0$. Obviously a success is more valuable than a failure, since $u > d$. Thus, 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; full disclosure is not an equilibrium strategy.[16]

If the manager follows a strategy of partial disclosure in equilibrium, and the market knows this, then the firm's expected value at the interim stage is

$$V_{PD} = us + (N - s)(qu + (1 - q)d), \tag{3}$$

where

$$q = \frac{r - \theta r}{1 - \theta r}$$

is the probability of success of an activity conditional on the fact that the manager has not disclosed information about that activity.[17] Note that this expression has the same structure as equation (2), except that r (the ex ante probability that an activity succeeds) in (2) is replaced by q (the conditional probability that an undisclosed activity succeeds) in (3). The probability an undisclosed project succeeds is

[14] To simplify the analysis, we will assume the NGO commits *ex ante* to audit with fixed probability whenever $\widehat{n} < N$.

[15] Baron and Diermeier (2005) present a model of strategic NGO activism in which firms are punished for bad social outcomes, rather than being punished for hypocrisy.

[16] 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 hypocrisy may become a problem.

[17] Recall that by Bayes' Rule, the probability an undisclosed project succeeds is $q = \Pr(success|undisclosed) = \Pr(success\&undisclosed)/\Pr(undisclosed) = r(1-\theta)/(1-r\theta)$.

$$
\begin{aligned}
q &= \Pr(success|undisclosed) = \Pr(success\&undisclosed)/\Pr(undisclosed) \\
&= \frac{r(1-\theta)}{1-r\theta}.
\end{aligned}
$$

The partial disclosure equilibrium can be supported by a set of off-equilibrium beliefs on the part of the market that if the manager ever reports $f > 0$, then all undisclosed outcomes are failures.[18]

It is natural to ask whether the NGO can effectively punish partial disclosure without auditing, e.g. by penalizing the firm retroactively based on the ultimate outcomes in period 2. It turns out this is not possible. As we noted in section 2 above, punishing partial disclosure is distinct from simply punishing the firm for bad social outcomes.[19] Punishing partial disclosure involves punishing firms that were aware of, but failed to disclose, a failure. At period 2, however, all the NGO knows is the ultimate number of failures, *not* the number that were known 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. This is the issue to which we now turn.

4 Equilibria with Monitoring and Punishment

In this section we assess how auditing by an NGO affects the manager's incentives to make social 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. This analysis prepares us for a detailed examination in section 5 of how changes in expected penalties for greenwash change the nature of equilibria in the model.

4.1 Equilibria with Auditing by an NGO

In order to keep the analysis tractable and focused, we present it in the context of a model with $N = 2$. (Even with this simplication, some derivations of formulae are complicated enough that we relegate them to the Appendix.) This is the simplest setting in which partial disclosure can emerge as an equilibrium outcome. Furthermore, conducting the analysis for general N would significantly complicate the notation, but is unlikely to yield qualitatively new insights. Table 1 presents the firm's value for each set of possible reports the manager can make at period 1. In each box, the value consists of two components, each of which is indexed by the number of successes and failures reported by the manager at period 1. The first component is the firm's value as assessed by the market, and the second is the penalty imposed by the NGO. We will

[18] While this is not the only set of off-equilibrium beliefs that support the sanitization strategy, it is the simplest.

[19] Baron and Diermeier (2005) study the latter situation.

use the notation $V_1(\widehat{s}, \widehat{f})$ to indicate the market's valuation of the firm when it makes the disclosure $(\widehat{s}, \widehat{f})$. Note that when $\widehat{n} \equiv \widehat{s} + \widehat{f} = 2$ the market has no problem inferring the firm's true value, since information disclosures are verifiable. These values are easily seen to be $V_1(0,2) = 2d$, $V_1(2,0) = 2u$, and $V_1(1,1) = u + d$, and are presented on the diagonal in Table 1. It is only in states where $\widehat{n} \equiv \widehat{s} + \widehat{f} < 2$ that we must carefully analyze the market's inference problem. (It is also worth noting that if the firm faced no penalties it would always pursue the strategy of partial disclosure, because it raises the firm's value, giving it a false appearance of virtue; this is precisely the case treated above in section 3.)

2	$V_1(0,2)$		
1	$V_1(0,1)$	$V_1(1,1)$	
0	$V_1(0,0)$	$V_1(1,0) - \alpha P$	$V_1(2,0)$
$\widehat{f} \diagup \widehat{s}$	0	1	2

Table 1: Value of the Firm for Possible Reports $(\widehat{s}, \widehat{f})$ in period 1

We focus on the case in which the true state is $(1,1)$, as this is the only possible case—for $N = 2$—when partial disclosure can occur. Specifically, partial disclosure would consist of claiming the state is $(1,0)$ when it is really $(1,1)$. This is the type of behavior that activists label greenwash, since it presents a false appearance of being better than one really is in truth. The firm receives no punishment for any situation except when it is a type $(1,1)$ and discloses $(1,0)$. Hence our focus is on what the manager will report when $(s,f) = (1,1)$. There are four reporting possibilities: $(\widehat{s}, \widehat{f}) \in \{(0,0),(1,0),(0,1),(1,1)\}$. Given the arguments we have made above, however, it is clear that the manager will never report $(\widehat{s}, \widehat{f}) = (0,1)$, so we focus on the other three cases in sequence.

In order to understand the manager's reporting incentives, we must know how the market will interpret each of the three possible reports. Consider them in turn. The probability that the state is actually $(1,1)$ can then be computed via Bayes' Rule. Table 2 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 reporting $(1,0)$ earns the firm a better value than does reporting $(1,1)$, since the expected value of a project, $ru + (1-r)d$, is greater than the known value of a failure, d.

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

Table 2: Interim Period States, Probabilities, and Values

We will use the notation $\mu(\widehat{s},\widehat{f};s,f)$ to indicate the probability the market assigns to the manager playing reporting strategy $(\widehat{s},\widehat{f})$ when the state is (s,f).[20] In addition, we will define $\Psi(\widehat{s},\widehat{f})$ as the probability the market assigns to observing a report $(\widehat{s},\widehat{f})$; this is the sum of the probabilities of each interim type of firm multiplied by the probability that type reports $(\widehat{s},\widehat{f})$. For example,

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

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

$$E[1,1|1,1]=u+d. \tag{4}$$

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; $(2,0)$ and the firm is failing to report a success; or $(1,1)$ and the firm is engaging in greenwash. Thus, $\Psi(1,0)=2r\theta(1-\theta)\mu(1,0|1,0)+r^2\theta^2\mu(1,0|2,0)+2r(1-r)\theta^2\mu(1,0|1,1)$. If the NGO audits, and finds that the state is really $(1,1)$ but the firm engaged in greenwash, then the NGO launches a campaign against the firm that imposes a penalty P. The firm's expected value from reporting $(1,0)$ is

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

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 $(1,1)$.[21] Note that there is no possibility of a punishment in this case, since a report of $(0,0)$ does not constitute greenwash, since it does not aver any positive outcomes. The firm's expected value is

$$\begin{aligned}E[0,0|1,1] &= [ru+(1-r)d]\frac{(1-\theta)^2 2\mu(0,0|0,0)}{\Psi(0,0)}\\ &\quad +[d+(ru+(1-r)d)]\frac{2(1-r)\theta(1-\theta)\mu(0,0|0,1)}{\Psi(0,0)}\\ &\quad +2d\frac{(1-r)^2\theta^2\mu(0,0|0,2)}{\Psi(0,0)}+[u+d]\frac{2r(1-r)\theta^2\mu(0,0|1,1)}{\Psi(0,0)}.\end{aligned} \tag{6}$$

Expressions (5) and (6) appear complicated, but are actually quite simple in equilibrium. For example, the manager never has incentives to hide a success,

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

[21] A firm in state $(1,0)$ or $(2,0)$ has no incentive to report $(0,0)$.

so a firm in state $(2,0)$ will never report $(1,0)$. Thus we know $\mu(1,0|2,0) = 0$. Furthermore, the NGO is assumed to only punish what it views as greenwash, that is, partial disclosure, which means there is no punishment for reporting $(0,0)$; thus, firms in states $(0,1)$ or $(0,2)$ always have incentives to report $(0,0)$, and $\mu(0,0|0,0) = \mu(0,0|0,1) = \mu(0,0|0,2) = 1$. Furthermore, when we solve for the truthful disclosure equilibrium, it must be the case that in equilibrium the manager truthfully reports the firm's state when it is a $(1,1)$, that is, $\mu(1,1;1,1) = 1$ and $\mu(0,0;1,1) = 0$, and the manager does not report falsely, that is, $\mu(1,0;1,1) = 0$. Substituting in these values of $\mu(\cdot)$ greatly simplifies equations (5) and (6).

There are three types of pure-strategy equilibria that can emerge in this model in state $(1,1)$. The firm: a) fully discloses the state, b) engages in partial disclosure, or c) does not disclose at all. We now examine each of these three equilibria in turn.

4.2 Full Disclosure Equilibrium

In order for a firm in state $(1,1)$ to disclose fully, we require $E[1,1|1,1] > E[0,0|1,1]$ and $E[1,1|1,1] > E[1,0|1,1]$. In addition, if market participants believe the full disclosure equilibrium is being played, their beliefs must reflect the nature of this equilibrium, that is, they believe that with probability one a firm in state $(1,1)$ discloses fully rather than engaging in partial disclosure or not disclosing at all. Formally, this means that $\mu(0,0|1,1) = \mu(1,0|1,1) = 0$, and $\mu(1,1|1,1) = 1$.

Since disclosed information is verifiable, it is easy to see that

$$E[1,1|1,1] = u + d.$$

Understanding the payoff for non-disclosure is more complex. By definition, in the full disclosure equilibrium the market believes that a firm in state $(1,1)$ will fully disclose. Hence, when the market observes non-disclosure, it concludes the state is $(0,0), (0,1)$, or $(0,2)$. The market then assigns the firm an expected value that reflects the payoff of each of these three states, weighted by the probability of each one occurring, conditional on the observation that the firm disclosed nothing. Calculation details are in the Appendix, but some algebraic manipulation reveals that

$$E[0,0|1,1] = \frac{2\left(d(1-r) + ru(1-\theta)\right)}{(1-r\theta)}.$$

Finally, the expected value of partial disclosure is

$$E[1,0|1,1] = u + (ru + (1-r)d) - \alpha P.$$

The intuition for this value is simple: market participants believe the full disclosure equilibrium is being played, so the only time a firm would report $(1,0)$ is when the state is $(1,0)$. One can see immediately that if the expected penalty were $\alpha P = 0$, then the firm would always prefer to disclose $(1,0)$ rather than

$(1,1)$, since by so doing the firm creates an impression of being more socially responsible than it is in fact. The only thing that will prevent the firm in state $(1,1)$ from making such a disclosure is the threat of a punishment if it is found guilty of greenwash.

In order for full disclosure to be a Perfect Bayesian Equilibrium, the market's beliefs must be consistent with actual firm behavior, and it must be the case that $E[1,1|1,1] > E[0,0|1,1]$ and $E[1,1|1,1] > E[1,0|1,1]$. We have found it helpful to visualize the payoffs from the three disclosure strategies in a three-dimensional diagram such as Figure 1, which depicts the case where $u = 1.5, d = .5$, and $\alpha P = .3$. The dark-shaded surface represents profits from non-disclosure, $E[0,0|1,1]$; the medium-shaded surface shows the profits from partial disclosure, $E[1,0|1,1]$; and the lightly-shaded surface shows the profits from full disclosure, $E[1,1|1,1]$. For the full disclosure equilibrium, the relevant region in the diagram is the rectangular region for which the lightly-shaded surface is highest.

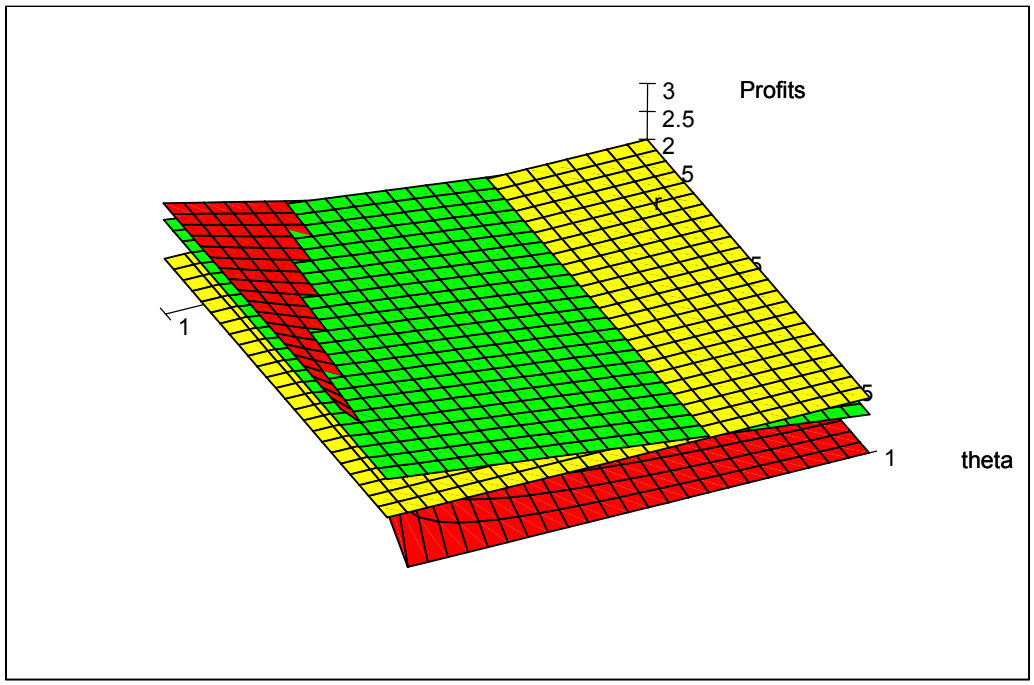

Figure 1: Full Disclosure Equilibrium

The condition that $E[1,1|1,1] > E[0,0|1,1]$ is equivalent to

$$u + d > \frac{2\left(d(1-r) + ru(1-\theta)\right)}{(1-r\theta)},$$

which simplifies dramatically to

$$r < \frac{1}{2-\theta}. \tag{7}$$

This expression turns out to be a very important determinant of firm behavior in the model, and it arises in later sections as well as here. One way to think about this inequality is that it determines the firm's disclosure strategy when the punishment for greenwash is so great as to eliminate partial disclosure as a viable strategy. In this case, a firm in state $(1,1)$ must choose between full disclosure or non-disclosure. Interestingly, it turns out that the market's beliefs about the firm's behavior do not affect the form of inequality (7). The reason is that at the point of intersection between surface $E[1,1|1,1]$ and surface $E[0,0|1,1]$, the firm is indifferent between disclosing fully or not at all. As a result, whether the market expects the firm to disclose fully or not has no impact on $E[0,0|1,1]$.

The condition $E[1,1|1,1] > E[1,0|1,1]$ simplifies to

$$r < \frac{\alpha P}{u-d}. \tag{8}$$

In Figure 1, $E[1,1|1,1] = E[1,0|1,1]$ represents the boundary between the dark-shaded and the medium-shaded surfaces, and $E[1,1|1,1] = E[0,0|1,1]$ represents the boundary between the dark and light surfaces (hidden in the diagram). Proposition 1 summarizes the foregoing analysis of the existence of a full-disclosure equilibrium

Proposition 1 *A full disclosure equilibrium exists for all* $r \leq \min\{1/(2-\theta), \alpha P/(u-d)\}$.

The basic intuition regarding full disclosure is that when the probability of success is low, there is little advantage to the firm in hiding a failure, since undisclosed activities will essentially be branded as failures by the market anyway.

4.3 Non-Disclosure Equilibrium

The formal requirements for a non-disclosure equilibrium are $E[0,0|1,1] > E[1,1|1,1]$ and $E[0,0|1,1] > E[1,0|1,1]$. The beliefs consistent with the equilibrium are $\mu(1,1|1,1) = \mu(1,0|1,1) = 0$, and $\mu(0,0|1,1) = 1$.

Once again, the payoff to full disclosure does not depend upon beliefs because disclosures are fully verifiable. As in the previous section, the payoff to full disclosure is

$$E[1,1|1,1] = u + d.$$

The payoff to partial disclosure is also unchanged from the previous section. Here, the beliefs associated with the equilibrium are that a firm in state $(1,1)$ chooses not to disclose any information. If the market sees a firm disclose $(1,0)$ then, it believes the firm is in state $(1,0)$. Thus, a firm in state $(1,1)$ can engage in greenwash, if it so desires, and obtain payoff

$$E[1,0|1,1] = u + (ru + (1-r)d) - \alpha P.$$

The payoff to non-disclosure is different than it was in the full disclosure equilibrium. Specifically, the market now believes there are four types that choose to not disclose: $(0,0), (0,1), (0,2)$, and $(1,1)$. The total probability a firm chooses to not disclose is

$$\Psi(0,0) = 1 - \theta r\left(2 - (2-r)\theta\right).$$

To the firm that does not disclose, the market assigns an expected value of

$$\begin{aligned} E[0,0|1,1] &= \frac{(1-\theta)^2 2(ru + (1-r)d) + 2(1-r)\theta(1-\theta)(d + (ru + (1-r)d))}{1 - \theta r\left(2 - (2-r)\theta\right)} \\ &\quad + \frac{(1-r)^2\theta^2 2d + 2r(1-r)\theta^2(u+d)}{1 - \theta r\left(2 - (2-r)\theta\right)}. \end{aligned} \tag{9}$$

A non-disclosure equilibrium requires $E[0,0|1,1] > E[1,1|1,1]$ and $E[0,0|1,1] > E[1,0|1,1]$. As in section 4.2, the first of these simplifies to

$$r > \frac{1}{2-\theta}. \tag{10}$$

The second requirement, $E[0,0|1,1] > E[1,0|1,1]$ is equivalent to

$$(1-r)\left(r^2\theta^2 + 1\right)(u-d) < (1 - \theta r\left(2 - (2-r)\theta\right))\alpha P. \tag{11}$$

For $u = 1.5$, $d = .5$ and $\alpha P = .3$, the three strategies (full disclosure in dark shading, non-disclosure in light shading, and partial disclosure in medium shading) produce payoffs that are represented in Figure 2. The region in which non-disclosure is a pure-strategy equilibrium is the triangular region where the dark surface is the highest. Note that the exposed non-disclosure region is smaller here than in Figure 1. This is because the market believes the firm will choose to not disclose when the state is $(1,1)$, and this belief reduces the market value of non-disclosure.

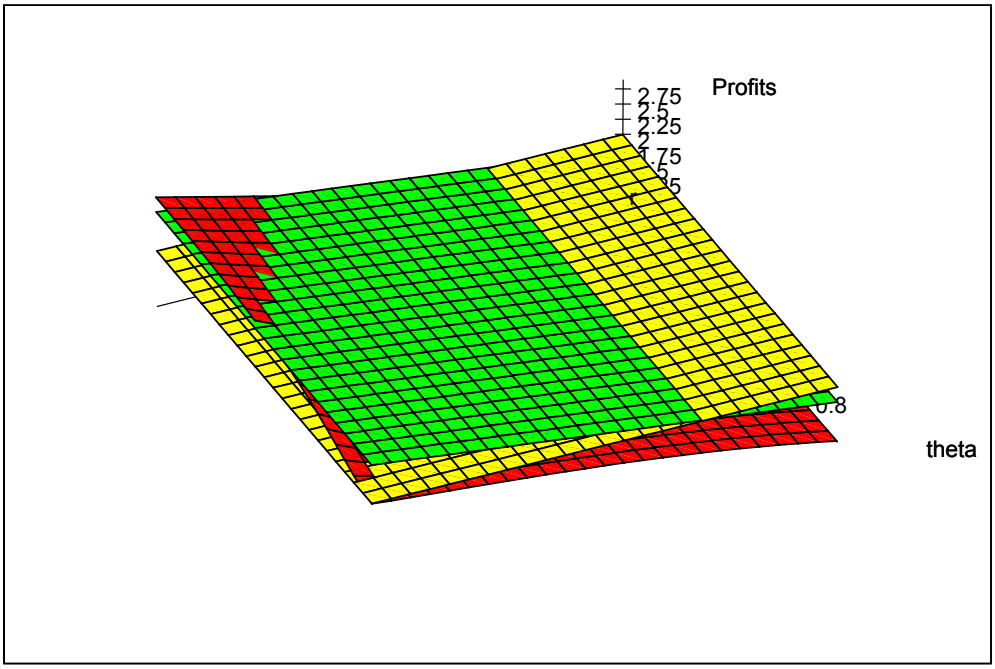

Figure 2: Non-Disclosure Equilibrium

The following proposition summarizes the above analysis regarding the non-disclosure equilibrium.

Proposition 2 *A non-disclosure equilibrium exists when* $r > 1/(2-\theta)$ *and* $(1-r)\left(r^2\theta^2+1\right)(u-d) > (1-\theta r\,(2-(2-r)\theta))\alpha P$.

Intuitively, the non-disclosure equilibrium exists when the probability of a success is high, in which case a firm with a failure gains significantly from hiding it. It is worth noting that $r > 1/(2-\theta)$ implies that the non-disclosure equilibrium can only exist for $r > 1/2$, since $1/(2-\theta) = 1/2$ at $\theta = 0$ and increases with θ.

4.4 Partial-Disclosure Equilibrium

The formal requirements for this type of equilibrium are $E[1,0|1,1] > E[1,1|1,1]$ and $E[1,0|1,1] > E[0,0|1,1]$. The beliefs consistent with a partial disclosure equilibrium are $\mu(1,1|1,1) = \mu(0,0|1,1) = 0$, and $\mu(1,0|1,1) = 1$.

As in the previous sections, the payoff to full disclosure does not depend upon beliefs, and

$$E[1,1|1,1] = u + d.$$

In the partial disclosure equilibrium, the market believes the firm in state $(1,1)$ will disclose $(1,0)$. Hence, the payoff to making this disclosure is different than

it was in the two previous types of equilibrium. Now, there are two situations when firms disclose $(1,0)$—the state is $(1,0)$ and the state is $(1,1)$. Thus, the total probability that a firm discloses $(1,0)$ is

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

Using this information, we can compute the expected payoff to partial disclosure as

$$E[1,0|1,1] = \frac{u(1+r(1-2\theta)) + d(1-r)}{1-r\theta} - \alpha P$$

The non-disclosure payoff is now the same as it was in the full disclosure equilibrium, since the market believes there are three types of firms that opt not to disclose: $(0,0), (0,1)$, and $(0,2)$. Thus, the total probability of non-disclosure in this equilibrium is

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

The expected payoff to non-disclosure is

$$E[0,0|1,1] = \frac{2\left(d(1-r) + ru(1-\theta)\right)}{(1-r\theta)}$$

For $u = 1.5$, $d = .5$ and $\alpha P = .3$, the three strategies (full disclosure in dark shading, non-disclosure in light shading, and partial disclosure in medium shading) produce payoffs that are represented in Figure 3. The partial-disclosure region is the crescent-shaped region where the medium-shaded surface is the highest.

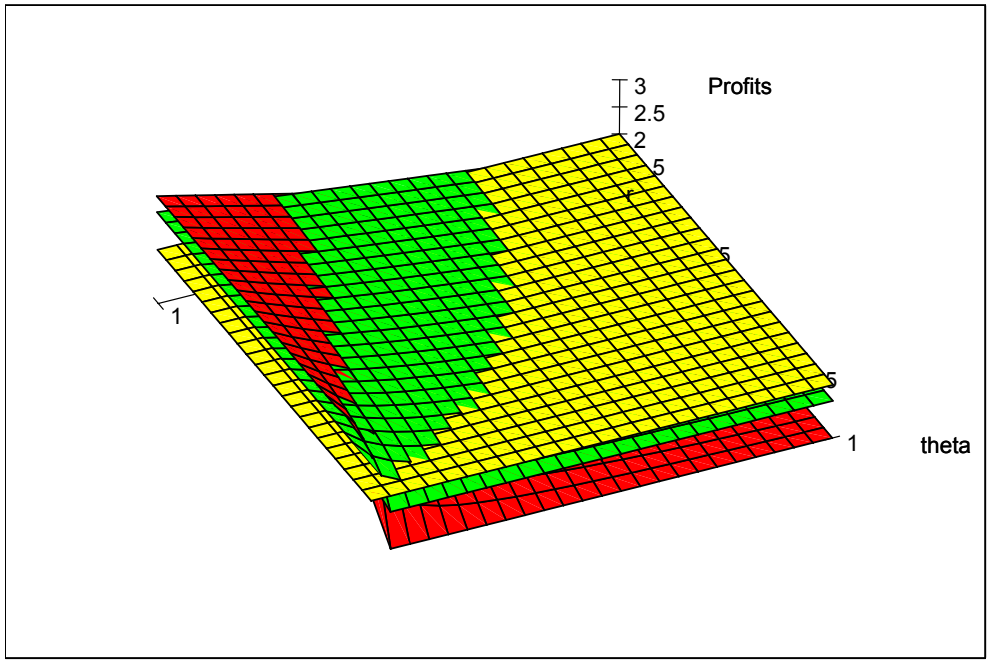

Figure 3: Partial Disclosure Equilibrium

Some algebraic manipulation shows that $E[1,0|1,1] > E[0,0|1,1]$ if

$$r < r_{PD/ND} \equiv \frac{u-d-\alpha P}{u-d-\theta\alpha P}, \tag{12}$$

where the somewhat cumbersome notation $r_{PD/ND}$ indicates the boundary between partial disclosure and non-disclosure.

Similarly, $E[1,0|1,1] > E[1,1|1,1]$ reduces to

$$r > r_{PD/FD} \equiv \frac{\alpha P}{(u-d)(1-\theta)+\theta\alpha P}, \tag{13}$$

where the notation $r_{PD/FD}$ indicates the boundary between partial disclosure and full disclosure. A partial disclosure equilibrium exists for $r \in [r_{PD/FD}, r_{PD/ND}]$. It is easy to see that as αP goes to zero, $r_{PD/FD}$ goes to zero and $r_{PD/ND}$ goes to one. Thus, as the expected penalty becomes negligible, partial disclosure is the unique pure-strategy equilibrium for all values of r and θ. It is also immediate that there is no partial disclosure equilibrium in pure strategies if $r_{PD/FD} > r_{PD/ND}$. The next proposition characterizes conditions for the existence of a partial disclosure equilibrium.

Proposition 3 *For every θ, a partial disclosure equilibrium exists for some set of values of r if and only if $\alpha P < (u-d)/2$. If $\alpha P = 0$, partial disclosure is the unique equilibrium for all values of r and θ.*

Proof. *Imposing the requirement that* $r_{PD/FD} > r_{PD/ND}$ *and simplifying yields* $(u-d)(d-u+2P\alpha)(\theta-1) < 0$. *Since* $u-d>0$ *and* $\theta \leq 1$, *this is equivalent to* $\alpha P < (u-d)/2$. ■

Intuitively, a partial-disclosure equilibrium can only exist when the expected penalty is not too high. As we will show in more detail in the following section, if the penalty is made large enough, it will deter any type of firm from engaging in partial disclosure. Furthermore, it is worth noting that the types of firms most likely to engage in partial disclosure are not those with particularly high or low values of r, but rather those with an intermediate likelihood of positive outcomes. The intuition for this observation is straightforward. Firms with low values of r fully disclose: they gain a lot from trumpeting a success, and lose little by withholding information about a failure (since they are already expected to fail); thus, there is little value in risking public backlash by refusing to disclose. At the other extreme, firms with high values of r do not disclose anything: they gain little by disclosing information about successes (since they are already expected to succeed), and lose a lot by disclosing a failure; thus, there is little value in risking public backlash by disclosing a success. For firms with moderate values 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, they are willing to risk public backlash by disclosing only partially.

It is also interesting to characterize the set of r for which partial disclosure is an equilibrium as θ increases.

Proposition 4 *Let* $R_{PD}(\theta, \alpha P) = r_{PD/FD} - r_{PD/ND}$ *be the set of values of* r *that form a partial disclosure equilibrium for some* θ *and* αP. *Then* $R_{PD}(\theta, \alpha P)$ *is decreasing in* θ.

Proof. Some calculation shows that

$$R_{PD}(\theta, \alpha P) = \frac{(1-\theta)(u-d-2P\alpha)(u-d)}{((u-d)(1-\theta)+P\theta\alpha)(u-d-P\theta\alpha)}.$$

Differentiating with respect to θ yields

$$\frac{dR}{d\theta} = \frac{-\alpha\theta(u-d)P(u-d-2P\alpha)(u-d-P\alpha)(2-\theta)}{(u-d-P\theta\alpha)^2((u-d)(\theta-1)-P\theta\alpha)^2}.$$

The denominator is positive. Assuming $\alpha P < (u-d)/2$, which is the condition for the existence of a partial disclosure equilibrium, we must have $(u-d-2P\alpha) > 0$ and $(u-d-P\alpha) > 0$, so $dR/d\theta < 0$. ■

As shown in Proposition (4), the band of r values that constitute a partial disclosure equilibrium is larger when θ is small. The reason is that small values of θ mean that it is likely the firm is uninformed about the performance of its

projects, and hence the market does not draw strongly negative inferences if the firm fails to report two outcomes.

It is interesting to think about these issues in the context of Walmart's recent conversion to a more promotional stance regarding its social contributions.[22] For years, Walmart kept a low profile on social issues, but as it has come under attack for its low pay and lack of benefits. the company has begun to promote its good activities more prominently. In terms of our model, this represents a shift from non-disclosure to partial disclosure. Such a shift is consistent with the notion that Walmart has experienced a reduction in r, that is, in the probability that its actions are viewed as socially responsible. Indeed, the empirical literature in accounting suggests that firms are more likely to engage in partial disclosure after some sort of public incident that produces damage to the company's reputation.[23] In our context, this can be interpreted as a reduction in the market's estimated probability that the firm produces socially positive outcomes.

5 The Impact of Alternative Penalties for Greenwash

The analysis in section 4 established conditions for the existence of different types of pure-strategy disclosure equilibria.[24] These equilibria depend upon different sets of beliefs on the part of participants in the disclosure game, and depend upon the parameters r, θ and αP. In this section, we characterize the number of different types of pure strategy equilibria, and their dependence on r and θ, as the expected penalty for greenwash, αP, increases. We begin by characterizing the "extreme" cases, that is, when $\alpha P = 0$ and when αP is so large as to eliminate partial disclosure as a profitable strategy. We then turn to a detailed comparative static analysis of the set of equilibria as αP increases over this domain.

5.1 The Set of Equilibria as Penalties Increase

It is easy to see from our discussion in previous sections that when $\alpha P = 0$, partial disclosure is the only equilibrium strategy for a firm in state $(1, 1)$ 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.

As expected penalties increase, there comes a point where partial disclosure is no longer an equilibrium strategy because of the expected penalties associated

[22] A speech on the topic by Walmart's CEO can be found at http://www.walmartstores.com/Files/21st%20Century%20Leadership.pdf

[23] See Deegan and Deegan (1996).

[24] The conditions presented in Dasgupta and Maskin (1986a,b) imply that there are mixed-strategy equilibria for parameter values (r, θ) for which no pure strategy equilibrium exists. We discuss mixed strategy equilibria in more detail below.

with using this strategy. Indeed, Proposition 3 shows that the partial disclosure equilibrium disappears for $\alpha P \geq (u-d)/2$.

Even if the penalty is high enough to eliminate partial disclosure as an equilibrium, partial disclosure may stll be attractive as a strategy that may possibly overturn one of the other types of pure-strategy equilibria. Recall from Figures 1 and 2 that even in the full disclosure and non-disclosure equilibria, there are parameter values r and θ for which partial disclosure is profitable. Thus, an important question is whether there is some level of penalty sufficient to prevent firms from engaging in partial disclosure, regardless of what type of equilibrium is being played. If so, then the firm must choose between either full disclosure or no disclosure. This is the subject of the following lemma.

Lemma 5 *For $\alpha P \geq (u-d)$, there exists a pure-strategy equilibrium for all values of r and θ, and it involves either full disclosure or non-disclosure.*

Proof. As shown in Proposition 1, a full disclosure equilibrium exists for all $r \leq \min\{1/(2-\theta), \alpha P/(u-d)\}$. We know that $1/(2-\theta) \in [.5, 1]$ for all θ. Hence for $\alpha P \geq (u-d)$, we know that $\alpha P/(u-d) > 1/(2-\theta)$, and full disclosure is an equilibrium for all $r \leq 1/(2-\theta)$. As shown in Proposition 2, a non-disclosure equilibrium exists for $r > 1/(2-\theta)$ and $(1-r)\left(r^2\theta^2+1\right)(u-d) \geq (1-\theta r\,(2-(2-r)\theta))\alpha P$. For $\alpha P = (u-d)$, the second condition reduces to $(1-r)\left(r^2\theta^2+1\right) \geq (1-\theta r\,(2-(2-r)\theta))$ Numerical calculations on a 200 by 200 grid show that this inequality holds for all $r \in [0,1]$ and $\theta \in [0,1]$. ■

The foregoing Lemma shows that if the expected penalty for corporate greenwash is at least $u-d$, then partial disclosure is never an optimal strategy, regardless of the type of equilibrium being played. This is intuitive. The maximum benefit the firm can possibly obtain from greenwash is $u-d$. This would occur if the firm has a very high value of r, so the market grants the firm expected value of 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 firms from using this strategy.

If $\alpha P \geq (u-d)$, the firm in state $(1,1)$ simply chooses between full disclosure or non-disclosure. As shown in section 4, this decision turns upon whether or not $r < 1/(2-\theta)$, with full disclosure the equilibrium if the inequality holds, and non-disclosure the equilibrium if it does not.

One implication of the lemma is that when monitoring is imperfect, i.e. $\alpha < 1$, the actual penalty that must be imposed to deter greenwash is potentially much greater than the difference in value between a successful project and a failure. For companies engaged in high-value acts of corporate social responsibility, this means that the penalties required to prevent partial disclosure may be so high that NGOs are unlikely to be able to produce them.

Having established results for minimal and maximal penalties, we turn now to the task of characterizing equilibria as the expected penalty ranges across this interval.

Proposition 6 *For $\alpha P = 0$, partial disclosure is the unique pure strategy equilibrium for all (r, θ). For $\alpha P \in (0, (u-d)/2)$, each of the three types of pure-strategy equilibria can be supported for some values of (r, θ), and there are also (r, θ) pairs for which no pure strategy equilibrium exists. For $\alpha P > (u-d)/2$, there exist (r, θ) pairs with $r > 1/(2-\theta)$ for which non-disclosure is a pure-strategy equilibrium and (r, θ) pairs with $r < 1/(2-\theta)$ for which full disclosure is a pure strategy equilibrium; there also exist (r, θ) pairs for which no pure strategy equilibrium exists. For $\alpha P > (u-d)$, non-disclosure is the unique pure-strategy equilibrium for all $r > 1/(2-\theta)$ and full disclosure is the unique pure-strategy equilibrium for $r < 1/(2-\theta)$.*

Proof. We approach the parts of the Proposition in sequence. (1) Inequalities (12) and (13) show that when $\alpha P = 0$, a partial-disclosure equilibrium exists for all $r \in [0, 1]$. (2) Inequality (11) shows that when $\alpha P = 0$, a non-disclosure equilibrium exists only in the limit as $r \to 1$. Finally, inequality (8) shows that when $\alpha P = 0$, a full-disclosure equilbrium exists only in the limit as $r \to 0$. (3) Proposition 2 shows that a partial disclosure equilibrium exists for $\alpha P < (u-d)/2$. Inequality (11) shows that when $\alpha P > 0$, a non-disclosure equilibrium exists for large enough r. Finally, Proposition 1 shows that when $\alpha P > 0$, a full-disclosure equilbrium exists for r small enough. (4) See Lemma 5. ■

The foregoing proposition shows that for penalties of low magnitude, that is, for $\alpha P \in (0, (u-d)/2)$, each of the three types of pure-strategy equilibria exist for at least some value of (r, θ). However, it is also true that pure-strategy equilibria do not exist for all (r, θ) pairs. This is illustrated in Figure 4 for the case where $u = 1.5, d = .5$, and $\alpha P = .3$. This figure effectively projects the equilbrium regions from each of Figures 1-2 onto a single (r, θ) plane The non-disclosure region lies above the uppermost (pink) curve, the partial-disclosure region lies between the next two (light green and dark green, if you are reading the electronic version of the paper) curves, and the full-disclosure region lies below the lowest (brown) line. However there are no pure-strategy equilibria for (r, θ) pairs between these regions (i.e., between the dark green and pink curves, and between the light green and brown curves). In the first of these regions, labeled "Mix$_{NP}$", firms employ a mixed strategy that involves mixing between non-disclosure and partial disclosure.. In the second, labeled "Mix$_{FP}$," firms mix between full disclosure and partial disclosure.[25]

[Figure 4 about here]

It is worth emphasizing that the types of firms most likely to engage in partial disclosure are not those with particularly high or low values of r, but rather those with an intermediate likelihood of positive outcomes. The intuition for this observation is straightforward. Firms with low values of r fully disclose: they gain a lot from trumpeting a success, and lose little by withholding information about a failure (since they are already expected to fail); thus, there

[25] The existence of mixed-strategy equilibria in these regions is established by applying the analysis of Dasgupta and Maskin (1986a,b).

is little value in risking public backlash by refusing to disclose. At the other extreme, firms with high values of r do not disclose anything: they gain little by disclosing information about successes (since they are already expected to succeed), and lose a lot by disclosing a failure; thus, there is little value in risking public backlash by disclosing a success. For firms with moderate values 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, they are willing to risk public backlash by disclosing only partially.

Similarly, for penalties of medium size, i.e. $\alpha P \in ((u-d)/2, u-d)$, two types of pure-strategy equilibria exist, but it is also true that pure-strategy equilibria do not exist for all (r, θ) pairs. This is illustrated in Figure 5 for the case where $u = 1.5, d = .5$, and $\alpha P = .5$. Again, the non-disclosure region lies above the uppermost (pink) curve, and the full-disclosure region lies below the lowest (brown) line. Note that there is no partial disclosure region, because the penalty is large enough to eliminate it as an equilibrium. From a graphical perspective, the two former (green) curves bounding the partial disclosure region have collapsed into what is now shown as a single (red) 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 non-disclosure 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.

[Figure 5 about here]

Finally, for $\alpha P > u - d$, partial disclosure is eliminated even as a part of mixed strategies. As a result, non-disclosure is the unique pure-strategy equilibrium for all $r > 1/(2-\theta)$ and full disclosure is the unique pure-strategy equilibrium for $r < 1/(2-\theta)$. This is illustrated in Figure 6.

[Figure 6 about here]

5.2 Implications for NGO Strategic Targeting of Firms

From both a positive perspective, and from the perspective of designing NGO strategies, it is critically important to understand how increasing the penalty for greenwash affects the equilibrium for different values of r and θ. Consider a shift from a low penalty to a high one, that is, from Figure 4 to Figure 6. Points in Figure 4 that lie between the upper curve and the dashed curve in the middle of the partial disclosure region shift from involving the use of partial disclosure (either in a pure or mixed strategy) to a pure strategy of non-disclosure. In contrast, points that lie between the dashed curve in the partial disclosure region and the straight line at the top of the full disclosure region in Figure 4 shift from strategies that use partial disclosure (either in a

pure or mixed strategy) to a pure strategy of full disclosure. Activists hope that by punishing corporate greenwash they will induce firms to become more transparent, and (in the language of our paper) to engage in full disclosure. Our analysis reveals that whether this hope is realized depends critically upon the values of r and θ for a particular firm. Firms with low levels of r ("irresponsible firms") may well make this behavioral shift in response to large penalties for greenwash. However, firms with relatively high values of r ("responsible firms") and relatively low values of θ are likely to shift to non-disclosure instead. This is not only contrary to the desires of activists, but—since some information is better than none for investors—it makes society as a whole worse off.

The types of firms for which such perverse outcomes are most likely are "responsible" or "clean" firms with limited knowledge of the environmental and social impacts of their actions. This lack of knowledge could be because the firm is engaged in projects with long time horizons, or because the firm lacks strong internal monitoring and management systems for dealing with environmental and social issues. Interestingly, this implies that the firms that are the most appropriate targets for activist pressure are those that are likely to be well informed—which does seem consistent with casual observation of the types of firms singled out for punishment as greenwashers. In an environmental context, our model suggests that firms in "irresponsible" industries that are well informed are the best targets for activist pressure. Companies such as BP, Shell, Monsanto, and Enron have been targeted by activists in the past, and appear to fit our model well. Still, it is important for activists to keep in mind that punishing corporate greenwash is a double-edged sword and should be wielded carefully.[26]

6 Environmental Management Systems and NGO Auditing

As noted above, NGO auditing of greenwash is not guaranteed to be socially valuable. It is particularly likely to backfire, and lead to non-disclosure, for poorly informed firms in clean industries, that is, when r is large and/or θ is small. The policy is more likely to be successful for firms where r is small and/or θ is large. This observation suggests that there is a complementarity between the NGO's auditing activities and the presence of environmental management systems (EMS) within the audited firms, which would be interpreted in our model as increasing θ. In our model, however, firms have no incentive to adopt an EMS, since the firm's market value in the interim period is lower when it adopts an EMS, as is shown in the following proposition.

Proposition 7 *In the interim period, the firm's value in the partial disclosure equilibrium is decreasing in θ. It's value in the full disclosure or non-disclosure equilibria is unaffected by θ.*

[26] For an interesting empirical analysis of the targeting behavior of NGOs, see Easley and Lennox (2006).

Proof. Differentiating equation (3) with respect to θ yields $dV_{PD}/d\theta = (u - d)(N - s)(dq/d\theta)$. All terms in this expression are positive with the possible exception of $dq/d\theta$. Recalling that $q = (r - \theta r)/(1 - \theta r)$, and differentiating this

expression yields $dq/d\theta = -r(1 - r)/(1 - \theta r)^2 < 0$. Thus, $dV_{PD}/d\theta < 0$. It is straightforward to see that $V_{FD} = u + d$ and $V_{ND} = N(ru + (1 - r)d)$, neither of which is a function of θ. ■

The intuition for the proposition is as follows. In the partial disclosure equilibrium, the manager withholds unfavorable information to increase its market value. This strategy works because for each withheld piece of information, the market valuation of the firm reflects only the possibility, not the certainty, of a failure. However, as the likelihood increases that the manager knows the environmental outcomes of the firm's activities, the market increasingly interprets non-disclosure as withheld negative information rather than as true uncertainty on the 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.

Admittedly, 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 tends to deter firms from adopting EMSs. Furthermore, our story is broadly consistent with the empirical results of Delmas (2000), who finds that many 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 different rationale for encouraging firms to adopt EMSs. 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 NGO auditing of environmental disclosure and greenwash. With an EMS in place, when a manager discloses nothing about the firm's environmental performance, the market infers that the manager is failing to disclose some negative information, and thus downgrades its rating of the firm's value. The threat that his firm's stock will be devalued makes a manager less willing to adopt a policy of non-disclosure.

In turn, this means that an NGO's threat to punish greenwash is more likely to drive the manager to disclose fully rather than to not disclose at all.

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. In effect, the presence of the EMS brings the market closer to a state of common knowledge, thereby increasing market efficiency. With an EMS in place, the manager is more likely to be well informed about his firm's own environmental impact, but more importantly, *the market knows that the manager is more likely to be well 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.

7 Mandatory Disclosure Requirements

Even when conditions are such that punishing greenwash can actually induce greater disclosure rather than less disclosure, such punishment is never enough to bring about full disclosure of environmental information in all states of the world. The reason is that managers with no successful activities to point to can simply remain silent about their failures without fear of punitive action by the NGO. This observation suggests that it is not greenwash *per se* that is the fundamental problem, it is the failure to fully disclose. In this section, we consider an alternative approach to inducing disclosure of environmental information, namely relying upon legislation that mandates disclosure and penalizes firms that fail to comply. The Public Company Accounting and Reform Act of 2002 (commonly know as Sarbanes-Oxley) was signed into law in July of 2002, and contains a number of provisions that require publicly traded companies to improve the accuracy of their financial disclosures and establish better internal controls for financial reporting. One area where better internal controls will likely be needed is in developing processes to identify, track, quantify and assess the financial impact of potential environmental liabilities.

In addition, the Securities and Exchange Commission (SEC) has promulgated Regulation S-K, which contains several items affecting the disclosure of environmental costs and liabilities. In particular, Item 101 requires companies to disclose material effects of compliance (or non-compliance) with environmental laws, Item 103 requires disclosure of pending, non-routine litigation (with environmental litigation typically being considered non-routine), and Item 303 requires disclosure of business trends or events likely to have a material effect of a company's financial condition. One can easily see how certain environmental "trends or events" such as discovery of environmental contamination (e.g. PCB in fish) might have such a material effect. Of these, Item 303 is perhaps most closely related to our analysis. It is important to note that even this provision leaves substantial room for managerial discretion in determining what is "likely" and what is a "material effect."

Below we revisit the valuation table for the firm, with $F(\widehat{s}, \widehat{f}; s, f)$ the fine levied by the regulator if an audit determines the manager failed to comply with

disclosure regulations.[27] It is unnecessary for the regulator to punish firms that fail to report good news, since market forces will induce firms to report good news without the need for regulation.

2	$V_1(0,2)$		
1	$V_1(0,1) - \alpha F(0,1;s,f)$	$V_1(1,1)$	
0	$V_1(0,0) - \alpha F(0,0;s,f)$	$V_1(1,0) - \alpha F(1,0;s,f)$	$V_1(2,0)$
$\widehat{f}/\widehat{s}$	0	1	2

Table 3: The Firm's Value for Possible Reports $(\widehat{s}, \widehat{f})$ in period 1

There are three states to investigate: $(0,1)$, $(0,2)$, and $(1,1)$. (Firms in states $(1,0)$ and $(2,0)$ have no reason to not disclose, while type $(0,0)$ has no options.) Note that states $(0,1)$ and $(0,2)$ were not part of our analysis in sections 4 and 5, because they do not involve "greenwash" proper, that is, they don't involve any reporting of positive information. We consider the three relevant states in turn.

State $(0,1)$: If the state is $(0,1)$, there are two possible reports: $(0,1)$ or $(0,0)$. If the manager reports $(0,1)$, the market knows the state is either $(0,1)$ or $(0,2)$. (The firm has no incentive to report $(0,1)$ in state $(1,1)$.) Thus,

$$E[0,1|0,1] = [d+(ru+(1-r)d)]\frac{2(1-r)\theta(1-\theta)\mu(0,1|0,1)}{\Psi(0,1)}+2d\frac{(1-r)^2\theta^2\mu(0,1|0,2)}{\Psi(0,1)}.$$

Similarly,

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

State $(0,2)$**:** If the state is $(0,2)$, there are three possible reports: $(0,2)$, $(0,1)$ or $(0,0)$. If the manager fully discloses, the market can confirm this fact and the firm's value is

$$E[0,2|0,2] = 2d.$$

If the manager reports $(0,1)$, the market knows the state is either $(0,1)$ or $(0,2)$. (Again, the firm in state $(1,1)$ has no incentive to report $(0,1)$.) Thus,

$$E[0,1|0,2] = [d+(ru+(1-r)d)]\frac{2(1-r)\theta(1-\theta)\mu(0,1|0,1)}{\Psi(0,1)}+2d\frac{(1-r)^2\theta^2\mu(0,1|0,2)}{\Psi(0,1)}-\alpha F(0,1;0,2).$$

[27] We assume the regulator commits to an audit program in advance. Thus, there is no issue of whether the regulator would really want to follow through on the audit in a truthful reporting equilibrium.

If the firm reports $(0,0)$, the market believes the state could be $(0,0),(0,1)$, $(0,2)$ or $(1,1)$. Noting that $\mu(0,0|0,0)=1$, we can write the firm's expected value as

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

State $(1,1)$**:** If the state is $(1,1)$ and the firm reports $(1,1)$, the market knows the state for certain, and the firm has market value

$$E[1,1|1,1] = u+d.$$

If the firm reports $(1,0)$, then the market believes it is either a $(1,0)$ and revealing truthfully, a $(2,0)$ failing to report a success, or a $(1,1)$ and engaging in greenwash. The firm's expected value in this case is

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

If the firm reports $(0,0)$, then the market will conclude this report might have come from a firm in states $(0,0)$, $(0,1)$, $(0,2)$ or $(1,1)$. The firm receives an expected value of

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

7.1 The Full Disclosure Equilibrium

We are interested in the conditions that will induce the firm to disclose fully in all states of the world. In the full disclosure equilibrium, the market will assesses $\mu(s,f|s,f)=1$ and $\mu(\widehat{s},\widehat{f}|s,f)=0$ for any $\widehat{s}\neq s$ or $\widehat{f}\neq f$. For full disclosure to be incentive compatible, we must have

$$E(s,f|s,f) > E(\widehat{s},\widehat{f}|s,f) \qquad \forall \widehat{s},\widehat{f} \neq s,f.$$

The following conditions must hold in a full disclosure equilibrium

$$E[0,1|0,1] > E[0,0|0,1] \Rightarrow d+(ru+(1-r)d) > 2(ru+(1-r)d) - \alpha F(0,0;0,1).$$

$$E[0,2|0,2] > E[0,1|0,2] \Rightarrow 2d > d+(ru+(1-r)d) - \alpha F(0,1;0,2).$$

$$E[0,2|0,2] > E[0,0|0,2] \Rightarrow 2d > 2(ru + (1-r)d) - \alpha F(0,0;0,2).$$

$$E[1,1|1,1] > E[1,0|1,1] \Rightarrow u + d > u + (ru + (1-r)d) - \alpha F(1,0;1,1).$$

$$E[1,1|1,1] > E[0,0|1,1] \Rightarrow u + d > 2(ru + (1-r)d) - \alpha F(0,0;1,1).$$

A bit of algebra shows that the fines necessary to induce full disclosure in each state are

$$F(0,0;0,1) > \underline{F}(0,0;0,1) \equiv \frac{r(u-d)}{\alpha}.$$

$$F(0,1;0,2) > \underline{F}(0,1;0,2) \equiv \frac{r(u-d)}{\alpha}.$$

$$F(0,0;0,2) > \underline{F}(0,0;0,2) \equiv \frac{2r\,(u-d)}{\alpha}.$$

$$F(1,0;1,1) > \underline{F}(1,0;1,1) \equiv \frac{r(u-d)}{\alpha}.$$

$$F(0,0;1,1) > \underline{F}(0,0;1,1) \equiv \frac{(u-d)\,(2r-1)}{\alpha}.$$

It is easy to show that $\underline{F}(0,0;0,2) > \underline{F}(1,0;1,1) = \underline{F}(0,0;0,1) = \underline{F}(0,1;0,2) > \underline{F}(0,0;1,1)$. This is intuitively reasonable—the firm has strongest incentives to not disclose when it has two failures. It is straightforward to establish a sufficient condition on penalties that will induce full disclosure, as noted in the following proposition.

Proposition 8 *Full disclosure can be induced through a policy of mandatory disclosure that includes penalties at least as great as* $2r(u-d)/\alpha$ *for failures to disclose.*

The proposition shows that mandatory disclosure requirements, with the requisite level of fines, are more powerful instruments than penalizing greenwash alone. As we found earlier, full disclosure in all states of the world can never be achieved simply by auditing and punishing greenwash. Furthermore, the NGO's ability to deter greenwash depends importantly on the values of parameters such as r and θ. Mandatory disclosure requirements offer the ability to eliminate withholding of information in all states, for any r and θ.

Although mandatory disclosure rules are attractive in principle, in practice they may require the use of fines that are too large to be politically feasible. If so, then there is no guarantee that a mandatory disclosure law will be more effective than auditing by an NGO. We turn to this issue in the following section.

7.2 Limited Regulatory Penalties

In the previous section, we showed that if there are no constraints on penalties for failure to disclose information, then legislative requirements can induce companies to fully disclose their environmental risks. Often, however, government penalties are less than would be required to prevent socially damaging corporate action.[28] Under Sarbanes-Oxley, firms may face fines of up to $5 million, and corporate managers may face up to $1 million in fines. Unfortunately, fines of this magnitude are unlikely to induce truthful disclosure from firms of any substantial size. To get a sense of the magnitudes required, note that Konar and Cohen (1998) find that poor environmental performance significantly reduced the intangible asset value of firms in the S&P 500, with the average intangible liability valued at $360 million. For purposes of calibration, then, suppose the firm has 10 activities with significant negative environmental impacts, and that the figure from Konar and Cohen thus represents the impact of 10 failures. Then, $N(u-d) = \$360$ million. If 10% of firms are likely to be audited, so that $\alpha = .1$, then the fine required to induce full disclosure for the average firm would be $3.6 billion. If this rough calculation is even remotely correct, the $5 million fine that can be levied under Sarbanes-Oxley is nowhere close to enough to discipline the reporting behavior of large firms.

If political constraints limit the fines that can be imposed, then the full disclosure equilibrium will fail to exist. In this case, it is natural to ask whether NGO auditing might complement mandatory disclosure requirements, and thereby restore the full disclosure equilibrium. We explore this question in the remainder of this section.

From section 7.1, we know that it is most difficult to induce full disclosure by a firm in state $(0, 2)$. In addition, we know that NGO punishment of greenwash affects only the incentives of firms in state $(1, 1)$. Suppose the maximum politically feasible fine is $F_{\max}$. Then NGO auditing will be of no additional value if

$$F_{\max} \in \left(\frac{r(u-d)}{\alpha}, \frac{2r(u-d)}{\alpha} \right).$$

In this case, mandatory disclosure rules are sufficient to deter greenwash, but not strong enough to induce a firm with two failures to reveal them. Since greenwash is already deterred by the mandatory disclosure rules, NGO auditing provides no additional effect on behavior.

If $F_{\max} < r(u-d)/\alpha$, then NGO auditing may in principle improve reporting behavior. Consider the case where

$$F_{\max} \in \left(\frac{(u-d)(2r-1)}{\alpha}, \frac{r(u-d)}{\alpha} \right).$$

If mandatory disclosure rules exist, but there is no NGO auditing, then a firm in state $(1, 1)$ will be deterred from disclosing $(0, 0)$, but will report $(1, 0)$. Auditing

[28] For example, many authors have criticized the Occupational Safety and Health Administration (OSHA) for setting fines that are too low to deter corporate safety violations. For details, see Weil (1996).

by the NGO will improve incentives if $\alpha(F_{\max} + P) > r(u - d)$.

7.3 Mandatory Disclosure, NGO Auditing, and EMS

Given the magnitude of the fines needed to induce full disclosure, it is possible that even the combination of government-mandated fines and NGO penalties will fall short of the levels needed to induce full disclosure. If this is the case, then EMSs re-emerge as a complementary tool that may enhance the effectiveness of the other two mechanisms.

If legislatively-mandated fines are very small, e.g., if $F_{\max} < (u-d)(2r-1)/\alpha$, The total penalty that can be imposed on the firm for greenwashing is increased by the amount of the government-imposed fine, but this will not induce full disclosure in all states.

If fines are moderate in size, matters become more complex. If $F_{\max} > (u - d)(2r - 1)/\alpha$, then a firm in state $(1, 1)$ would be deterred from reporting $(0, 0)$ if all other types had incentives to report truthfully. However, if types $(0, 1)$ and/or $(0, 2)$ have incentives to report $(0, 0)$, then $(u-d)(2r-1)/\alpha$ is not a sufficiently large fine to prevent the $(1, 1)$ type firm from opting not to disclose. The firm's adoption of an EMS would improve matters, reducing incentives for non-disclosure.

In general, none of the three tools discussed here is likely to be sufficient to induce full disclosure of environmental liabilities by corporations. Combining the three offers promise for improving corporate disclosure, but a full analysis of the interplay between the three is beyond the scope of this paper.

8 Conclusions

This paper has presented what is to our knowledge the first economic analysis of greenwash. We defined greenwash as the selective disclosure of positive information about a company's environmental performance, without full disclosure of negative information on these dimensions. We then modeled the phenomenon using tools from the literature on financial disclosure. In our model, a non-governmental organization (NGO) can audit corporate environmental reports, and penalize firms caught engaging in greenwash. Our model is relatively simple, yet produces some interesting positive implications. We show that 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. We also show that public outrage over corporate greenwash is more likely to induce a firm to become more open and transparent if the firm operates in an industry that is likely to have socially or environmentally damaging impacts, and if the firm is relatively well informed about its environmental 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. We show that 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 socially responsible firms with a high probability of successful projects, yet who are not fully informed about the social impacts of their actions. In an environmental context one might characterize such firms as "poorly informed firms in clean industries." For firms such as this, activist pressures may backfire and produce exactly the opposite of the intended results.

The likelihood that a firm responds to the threat of NGO auditing by opting for non-disclosure is reduced if the firm has adopted an environmental management system (EMS), and the complementarity between EMSs and NGO 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.

We also studied mandatory disclosure policies, and found that they have the potential to induce managers to fully disclose information about their firms' environmental activities in all states of the world. However, the fines necessary to induce full disclosure may be so large as to be politically unpalatable. If this is the case, then disclosure may require the use of a set of complementary instruments, including mandatory disclosure standards, NGO auditing and implementation of EMSs.

There are a number of areas in which further research would be valuable. One need is for empirical study of greenwash, its effects on corporate valuation, and its interaction with NGO information campaigns. Ramus and Montiel (2005) represents one needed step in this direction, as does Kim and Lyon (2006). Yet 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 non-market strategy.[29] 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

[29] Baron and Diermeier (2005) present a model of strategic activism in which the activist punishes firms for their socially or environmentally harmful practices, rather than for greenwash.

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

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

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

In addition,

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

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

Non-Disclosure Equilibrium

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

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

Partial Disclosure Equilibrium

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

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

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

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

References

[1] Baron, David P and Daniel Diermeier. 2005. "Strategic Activism and Nonmarket Strategy," Research Paper No. 1909, Stanford University.

[2] Dasgupta, Partha and Eric Maskin. 1986a. "The Existence of Equilibrium in Discontinuous Economic Games 1: Theory," *Review of Economic Studies*, 53: 1-26.

[3] Dasgupta, Partha and Eric Maskin. 1986b. "The Existence of Equilibrium in Discontinuous Economic Games 2: Applications," *Review of Economic Studies*, 53: 27-42.

[4] Deegan, Craig and M. Deegan. 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.

[5] Delmas, Magali 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.

[6] Easley, Chuck & Michael Lenox. 2006. "Secondary Stakeholder Actions and the Selection of Firm Targets," Working Paper, Duke University.

[7] Greer, Jed and Kenny Bruno. 1996. *Greenwash: The Reality Behind Corporate Environmentalism.* Penang, Malaysia: Third World Network and New York: the Apex Press.

[8] Johnson, Geoff. 2003. *Don't Be Fooled: The Ten Worst Greenwashers of 2003*, Boston, MA: Earthday Resources for Living Green. Available at http://www.thegreenlife.org/dontbefooled.html.

[9] Kim, Eun Hee and Thomas P. Lyon, 2006, "Why Do Firms Participate in Voluntary Environmental Programs?: The Case of DOE's Greenhouse Gas Reporting Program," Working Paper, Ross School of Business, University of Michigan.

[10] Konar, Shameek and Mark A. Cohen. 1998. "Does the Market Value Environmental Performance?," working paper, Owen Graduate School of Management, Vanderbilt University.

[11] Laufer, William S. 2003. "Social Accountability and Corporate Greenwashing," *Journal of Business Ethics*, 43: 253-261.

[12] Lyon, Thomas P. and John W. Maxwell. 2004a. "Astroturf: Interest Group Lobbying and Corporate Strategy," *Journal of Economics and Management Strategy*, 13: 561-597.

[13] Lyon, Thomas P. and John W. Maxwell. 2004b. *Corporate Environmentalism and Public Policy.* Cambridge: Cambridge University Press.

[14] Margolis, Joshua D. and James P. Walsh. 2003. "Misery Loves Companies: Rethinking Social Initiatives by Business," *Administrative Science Quarterly*, 48: 268-305.

[15] Milgrom, Paul and John Roberts. 1992. *Economics, Organization and Management*. Englewood Cliffs, NJ: Prentice Hall.

[16] Ramus, Catherine and Ivan Montiel. 2005. "When Are Corporate Environmental Policies a Form of Greenwashing?," Business and Society, 44: 377-414

[17] Shin, Hyun Song. 2003. "Disclosures and Asset Returns," *Econometrica*, 71: 105-133.

[18] Weil, David. 1996. "If OSHA is So Bad, Why is Compliance so Good?," *RAND Journal of Economics*, 27: 618-640.

Figure 4: Low-Penalty Case

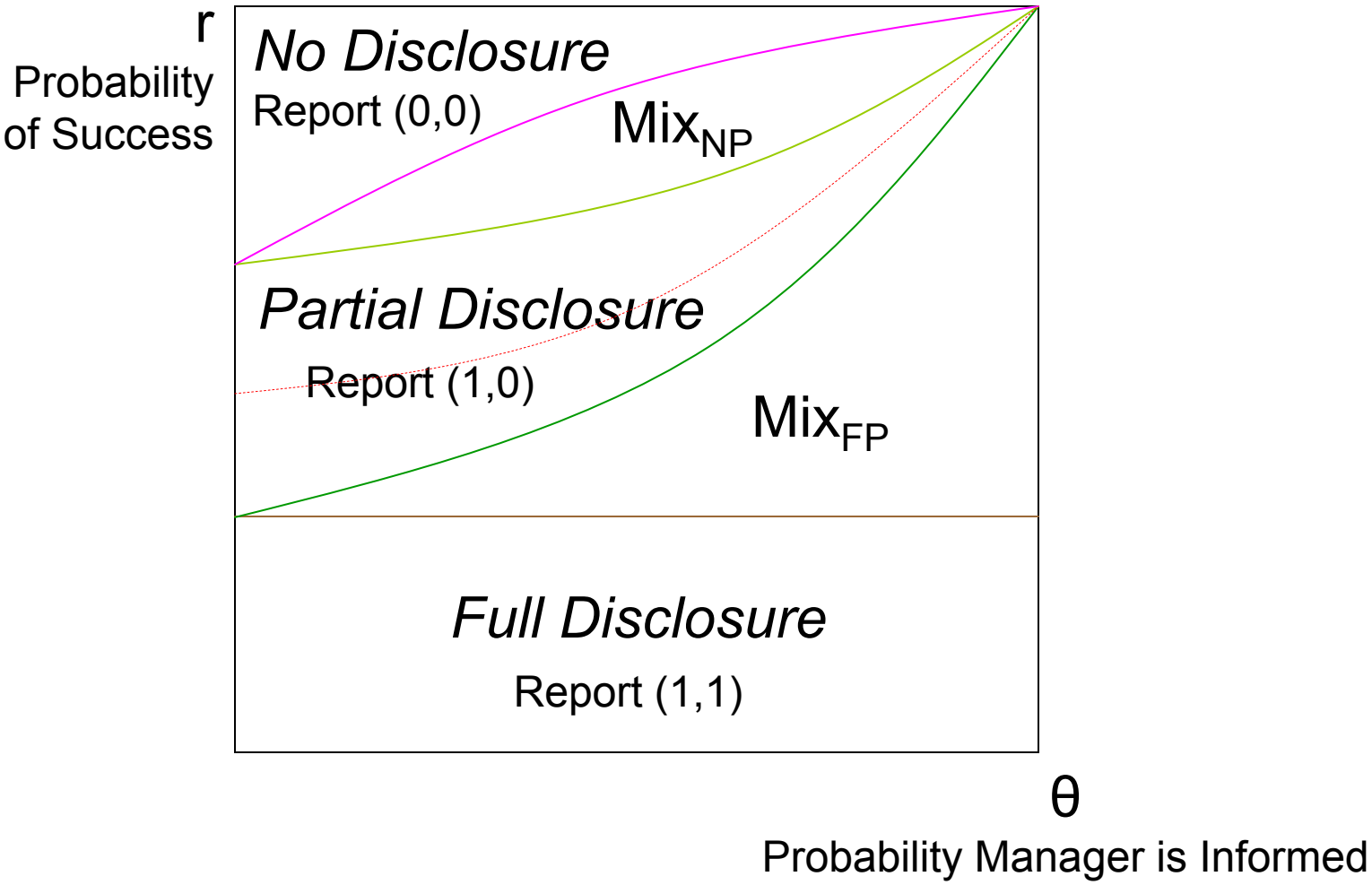

Figure 5: Medium-Penalty Case

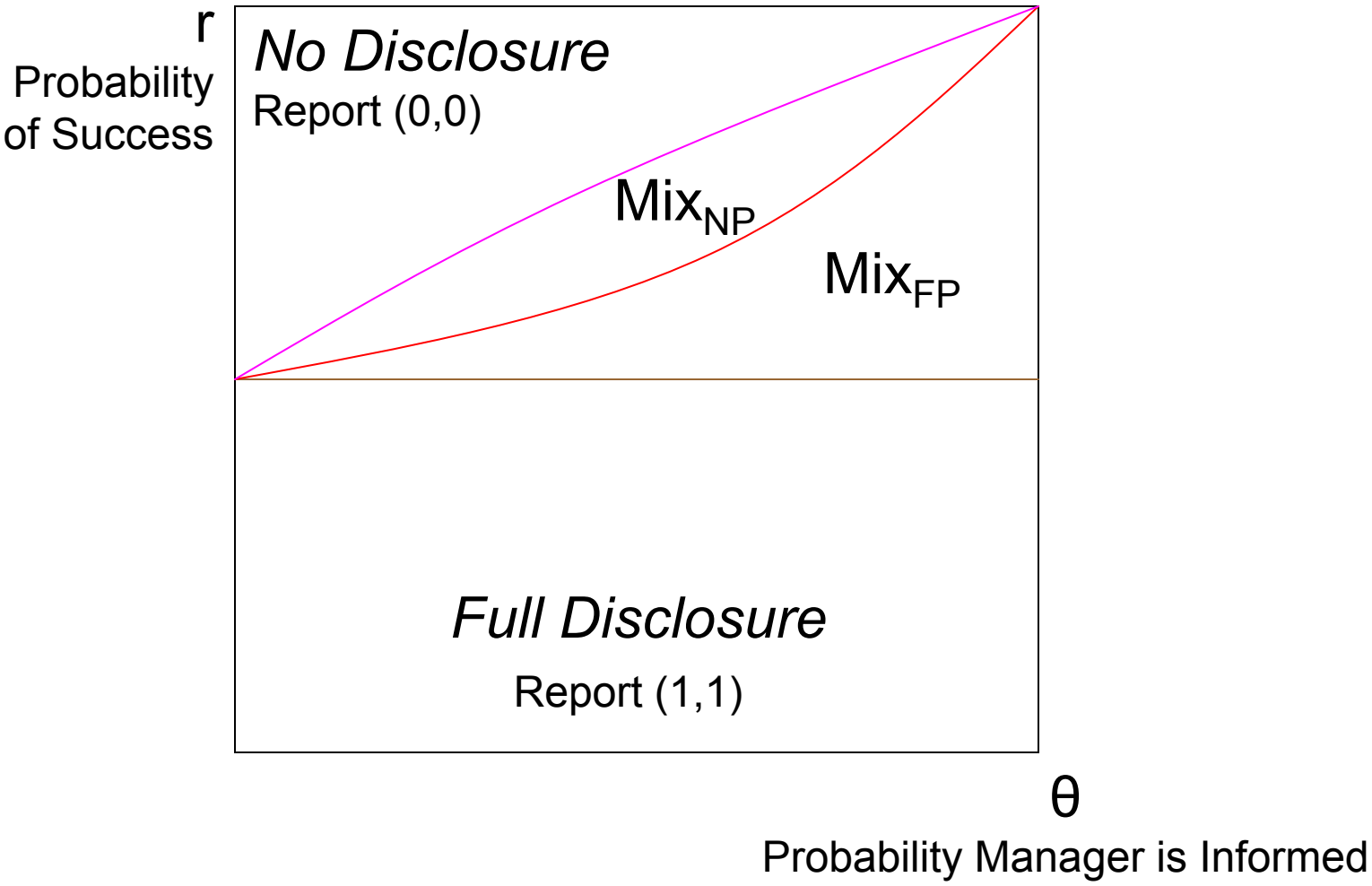

Figure 6: High-Penalty Case