Strategic Activism in an Uncertain World

van der Made, Allard

Published in:
B E Journal of Theoretical Economics

DOI:
10.1515/bejte-2018-0103

Allard van der Made[1]

Strategic Activism in an Uncertain World

[1] IEEF, Rijksuniversiteit Groningen Faculteit Economie en Bedrijfskunde, Nettelbosje 2, Groningen 9747AE, Netherlands, E-mail: a.van.der.made@rug.nl

Abstract:
We model private politics in the presence of two-sided incomplete information. An interest group (IG) threatens to inform consumers about a firm's damaging technology should the firm not adopt a clean technology. The IG does not know how costly adopting the clean technology is and the firm does not know how much the IG cares about the damages. In equilibrium an IG who cares a lot signals its type to the firm and the firm is more inclined to adopt the clean technology if it receives such a signal. However, impasses can occur: the firm does not adopt the clean technology despite the fact that the IG has signaled that it cares a lot and threatens to inform a large fraction of the consumer population. The IG never informs all consumers: as soon as a certain fraction of the consumers is informed by the IG the firm reduces its price and thereby reveals to the remaining consumers that it is employing a damaging technology. The IG's actions increase consumer well-being, but decrease total welfare unless the cost of adopting the clean technology is likely to be low. Yet, since the IG is inclined to target firms with this property, a regulator might want to delegate information provision to the IG.

Keywords: asymmetric information, impasses, interest groups, private politics, technology adoption

JEL classification: D80, L10, L31

DOI: 10.1515/bejte-2018-0103

1 Introduction

Ethically motivated consumers face a problem when shopping for certain goods. They often cannot observe the production practices of firms whose products they contemplate buying. It could be that seemingly innocuous products have caused grave harm to the environment during the production stage or were manufactured using child labor. This informational problem is difficult to solve: even the act of consumption is not likely to reveal the production practices the manufacturer has employed. Moreover, a firm often cannot credibly convey to consumers that it has abode by certain ethical standards. Government intervention or regulation is the natural candidate for alleviating market failures associated with such *credence goods*. Yet, as Darby and Karni (1973) argue, public attempts to correct these failures can be hampered by the same limitations as private attempts.

Increasingly, interest groups, for instance environmental organizations, take over the role of public institutions in providing consumers with information regarding firms' production practices. Examples include the Rainforest Action Network (RAN) disclosing that the paper and office supplies firm Staples uses trees from old-growth rainforests to produce paper and Greenpeace's Detox Campaign against the use of hazardous chemicals in the clothing industry.[1] Interest groups (IGs) can accomplish two things by taking over this role. In the short run they can steer consumers away from damaging products towards less damaging products. In the long run they can force firms to adopt less damaging technologies by threatening to unveil a firm's harmful practices should it not switch to a cleaner technology. Indeed, Staples agreed to stop using wood from endangered forests and several major clothing brands committed to ban a number of chemicals at all stages of production.[2] This paper deals with the long run effects of this type of strategic activism or *private politics* (Baron 2003).

The impact of strategic activism, which are methods to further a nonmonetary cause without resorting to regular market interactions or electoral processes, has not passed unnoticed among scholars.[3] However, this literature remains rather silent when it comes to some of the informational problems IGs and firms face when dealing with each other. Firms need not know whether an IG views a certain practice as unethical. For instance, does an IG consider letting employees in a developing country work under circumstances that are considered good in that country, but bad in developed countries to be unethical? To give a concrete example: while the oil and gas firm Unocal thought it was not doing much wrong morally or ethically when it started constructing a pipeline in Burma in the mid 1990s and got involved with the junta ruling Burma at that time, activists had a completely different opinion.[4] Even if a firm knows that its production process does not meet an IG's ethical standards, then it need not know how problematic the IG perceives the issue at stake to be. On the other hand,

Allard van der Made is the corresponding author.

an IG cannot easily assess the costs a firm has to incur to adopt a technology that does meet the IG's ethical standards.

We focus on these informational issues that arise when an IG tries to influence the technology choices of a firm.[5] In our framework an IG can request a firm to switch to a cleaner technology after it has learned that the firm is currently using a damaging technology. The IG and the firm do not know each other's type: the IG does not know how costly it is for the firm to adopt the cleaner technology, whereas the firm does not know the *salience* of the issue at stake as perceived by the IG. This salience is either high or low. The IG threatens to initiate a costly campaign aimed at informing consumers about the firm's malpractice should the firm not acquiesce to the IG's request. The IG can enforce its request by trying to convey to the firm that it cares a great deal about the firm's malpractice and would therefore launch a large campaign that would reach a lot of consumers should the firm stick to the damaging technology. It can do so by offering the firm's CEO an ostensibly expensive report regarding the firm's conduct, pay the firm a visit accompanied by experts who require hefty paychecks, state in its annual report that it commits to combat the kind of misconduct the firm is engaging in for years to come, etcetera. In the parlance of information economics the IG can engage in *money burning* to signal to the firm that it is a high type.[6] Such money burning was arguably employed by both RAN and Greenpeace: their campaigns aimed at Staples respectively the clothing industries were preceded by lengthy investigations culminating in big reports.

After the IG has communicated its request to the firm (possibly in conjunction with money burning), the firm decides whether or not to switch to the cleaner technology. If it does not switch, then an IG campaign ensues. The size of the campaign is endogenous: the IG chooses how many consumers it wants to inform. The consumers, who a priori only know that the firm's old technology is damaging with some given probability, are heterogeneous with respect to how ethically motivated they are. Since consumers who are strongly ethically motivated are no longer willing to buy from the firm once they learn about the firm's misconduct, the firm experiences a drop in profits after the IG has initiated a campaign.

We are particularly interested in the occurrence of *impasses*: despite the fact that the IG has conveyed to the firm that it cares a great deal, i. e. is the high type, the firm continues to use the damaging technology. The IG consequently initiates a large and costly campaign which leads to a substantial reduction in profits. Such situations are troublesome from a welfare point of view and occur repeatedly, featuring prominently in case studies of strategic activism.[7] Impasses are less likely to occur as the level of the salience of the high type increases or the cost of informing consumers decreases.[8] This is intuitive: a higher salience level or lower campaign costs lead to a larger campaign should the firm continue to use the damaging technology, implying that the firm has a stronger incentive to avoid the campaign by adopting the clean technology. Since consumers care about any damages the firm might cause, it is forced to charge a relatively low price for its product if consumers think it is likely that the firm is using a damaging technology. This implies that an IG campaign only has a moderate impact on price and hence profits if the a priori probability that the firm's old technology is damaging is large, making an impasse more likely as that probability increases. In other words, the larger the extent to which the potential externality is already reflected in the price, the larger the probability that there is a lasting conflict between the firm and the IG.

We compare the equilibrium of our model with the equilibrium of a benchmark setting in which the IG is not present.[9] Obviously, since in the benchmark the firm never adopts the expensive new technology and it never loses customers who are just informed about the firm's bad practices, the firm is better off in the benchmark setting. As one would expect consumers are on average better off if the IG is present. Yet, consumers who do not care much about the damage the firm is potentially causing prefer the benchmark. The reason is that as long as the firm continues to use the old technology it charges a price that is such that even consumers who are strongly ethically motivated are willing to buy the product. This means that consumers who do not care much about the damage enjoy a surplus when buying the product if the firm sticks to the old technology, which only happens with certainty in the benchmark. Because some consumers as well as the firm are better off in the benchmark, total welfare is often higher in the benchmark. Only if the cost of adopting the new technology is likely to be small does the IG's presence improve welfare. Luckily, firms with this property are the most appealing targets for the IG.

In light of the previous paragraph it should not be surprising that the IG often overprovides information. In fact, a regulator maximizing consumer surplus plus profit never opts to inform consumers about the damage the firm is causing once the firm has decided to stick to the damaging technology. However, the threat of informing consumers about the firm's bad practices can be welfare-improving from an *ex ante* point of view. The IG therefore does sometimes alleviate a market failure which would persist if it were absent and the regulator could only resort to information provision after the technology choice has been made. In those instances the regulator would from an *ex ante* point of view want to use this instrument, but decides not to inform consumers *ex post*. So, in those cases the regulator might want to delegate information provision to the IG.[10]

Baron (2001) was among the first to argue that incomplete information crucially affects private politics. In his model the IG need not know the salience of the issue at stake as perceived by *consumers*. The firm knows this salience. In equilibrium an IG's request for change is rejected by firms with low-salience issues.[11] In Baron (2001)'s model the IG's campaign brings about an exogenous downward shift in demand. In our model the shift in demand caused by the IG is endogenous and it depends on the salience of the issue as perceived by the *IG*.

Feddersen and Gilligan (2001) consider a setting with two price-taking firms supplying imperfect substitutes. An IG might investigate one of the firms and convey to consumers whether this firm's production practices (a credence attribute) are damaging or clean. This threat can induce one firm or both firms to adopt a more expensive, but clean technology. In particular, there exist mixed strategy equilibria in which both firms are likely to adopt the clean technology if the products are close substitutes. Because the IG does not incur any costs and firms cannot set prices, Feddersen and Gilligan's IG wields considerable power. By contrast, in our model prices are endogenous and the IG's actions are costly. Because of these differences our IG endogenously chooses a campaign size whereas the message of the IG of Feddersen and Gilligan (2001) simply reaches all consumers.

In Heijnen (2013) both an IG and a monopolistic firm can try to convey information regarding the damage the firm's product causes: the IG by choosing an advertising intensity and the firm by choosing the price of its product. Just like in Feddersen and Gilligan (2001) there is no uncertainty regarding cost functions or the salience of the issue and the IG's campaign reaches all consumers. Heijnen (2013) shows that if the firm's technology is exogenously given and production costs are decreasing in damages, then there are always two separating equilibria: one in which the IG signals the damage level and one in which the firm itself takes on this role. If the firm can choose a cleaner, but more costly technology, then the threat of an IG campaign can incentivize the firm to opt for the cleaner technology.

A number of papers focus on other aspects of IGs affecting the behavior of firms offering products with credence attributes. Bottega and de Freitas (2009) investigate the differences between certification of (environmental) quality by an IG and certification by a for-profit private certifier and the interactions of such certification possibilities with minimum quality standards set by a regulator. Baron (2009) looks at how IG pressure can affect the choice of voluntary quality standards that can be credibly communicated by a credence organization established by firms. Krautheim and Verdier (2016) study the interplay between offshoring to countries with laxer environmental or labor regulation and the emergence of IGs that can monitor firms' technology choices in those countries.

The remainder of this paper is organized as follows. In the next section we introduce the model. In Section 3 we derive equilibria and discuss some comparative statics results. Section 4 contains the welfare analysis. Concluding remarks are offered in Section 5. All proofs can be found in the Appendix.

2 The Model

An interest group (IG) is concerned about the damage that a monopolistic firm producing a single good might be causing. If the firm is employing a damaging technology, then the total damage it is causing is proportional to its total sales q and the IG experiences a disutility of αq. The parameter α is the salience of the issue at stake as perceived by the IG. This salience is low ($\alpha = \alpha_L$) with probability ϕ and it is high ($\alpha = \alpha_H$) with probability $1 - \phi$, where $0 < \alpha_L < \alpha_H$.

It is common knowledge that the firm's technology is damaging with probability $\psi \in (0, \frac{1}{2})$. With probability $1 - \psi$ the firm's technology is not damaging. Both the IG and the firm learn whether the technology is damaging, whereas consumers only know the probability ψ. To combat the damage, if any, the IG can contact the firm and request the latter to change its behavior. The firm can do so, but this is costly. It can adopt a new, clean technology with associated constant marginal cost of production c.[12] The choice to adopt the clean technology is denoted $\delta = 1$. The firm can also ignore the IG's request and continue using the old technology ($\delta = 0$). The costs of the old technology are normalized to zero. If the firm adopts the new technology, then this choice becomes public knowledge right after it has been made. In contrast to the firm the IG does not know c, but it does know that it is drawn from a distribution on $[\psi, 1]$ with cumulative distribution function F and density f. As can be gathered from the analysis in the next section the firm would adopt the clean technology if the associated marginal cost were at most ψ even if the IG were absent. Since we are only interested in situations in which the IG can affect the firm's technology choices we confine attention to densities with support $[\psi, 1]$.

If the technology is not damaging (which happens with probability $1 - \psi$ if the old technology is used and happens with certainty if the firm has adopted the clean technology), then each consumer's willingness-to-pay is 1. However, consumers care about the potential damage associated with the firm's good. If a consumer knows that the firm's technology is damaging, then his willingness-to-pay for the good is $1 - \theta$. The parameter θ

varies among consumers and captures the consumer's disutility associated with consuming a good that causes damage. Note that θ does not measure the extent of the damage itself, but how a consumer feels about buying a good from a firm causing damage.[13] A consumer's θ is a draw from the standard uniform distribution. There is a unit mass of risk-neutral consumers and a consumer buys one unit of the good or refrains from buying.

The firm does not switch to the new technology should the IG refrain from further action. The IG can however initiate a campaign if the firm decides to stick to the dirty technology. During the campaign the IG informs some of the consumers about the damaging aspects of the firm's technology. Informing consumers is costly: if the IG informs a fraction γ of the consumers, then it incurs a cost of $\frac{k}{2}\gamma^2$, where $k > 0$ is common knowledge.[14] The IG does not know the θ of individual consumers and is thus unable to target consumers with a particular value of θ.

Of course, the IG only informs consumers if the firm does cause damage. Importantly, we assume that a consumer is unaware of the possibility that the IG has investigated the firm's production practices as long as he has not received any information regarding the firm from it. One could alternatively assume that the number of firms that the IG could potentially investigate is so large that not receiving any information regarding one specific firm has a negligible effect on a consumer's beliefs regarding that firm's product. Either assumption implies that a consumer does not update his subjective probability $\tilde{\psi}$ that the firm causes damage if he does not receive any information on that subject. So, $\tilde{\psi} = \psi$ for an uninformed consumer.[15]

Since a more concerned IG initiates a larger campaign, the firm would be more inclined to switch to the clean technology if it knew that $\alpha = \alpha_H$. However, the value of α is private information. The firm merely knows the distribution of α. The IG can try to signal being the high type, i. e. that $\alpha = \alpha_H$, by 'burning money' while requesting the firm to adopt the clean technology.[16] To give an example, the IG could present an ostensibly expensive report regarding the firm's conduct to the firm. If the firm does believe that it faces the high type after having received such a report, then it is more inclined to switch to the clean technology thereby preempting the IG's campaign. We denote the amount of money burnt by m. If the IG burns an amount m, then its payoff is reduced by m.

The IG's payoff can thus consist of three parts: disutility stemming from damaging production, cost of informing consumers, and the cost associated with signaling its type. If the firm switches to the clean technology and the IG has not burnt any money or if the old technology does not cause damage, then the IG does not initiate a campaign and its payoff is simply 0. However, if the old technology is damaging and the firm sticks to this technology, then the IG incurs the cost of its informational campaign and suffers from the firm causing damage. Its payoff is consequently $G(\gamma;\alpha) = -\alpha q - \frac{k}{2}\gamma^2$, where the quantity sold $q = q(\gamma)$ depends on γ. If on top of that the IG has burnt an amount m, then its payoff is $G(\gamma;\alpha) - m$. If money burning leads to the firm adopting the clean technology, then the IG's payoff is $-m$.

The firm and the IG are involved in a game with two-sided incomplete information. This game consists of five stages. In the first stage Nature determines whether the old technology is damaging and draws c and α. If the old technology is damaging the IG requests the firm to alter its behavior in the second stage. It can opt to enforce its arguments by burning money. The firm subsequently chooses whether or not to switch to the new technology in the third stage. If the firm decides to stick to the old technology and this technology causes damage, then the IG initiates a campaign in the fourth stage. If the firm adopts the clean technology or the old technology is not damaging, then nothing happens in the fourth stage. In the last stage the firm sets a price p, sells its product, and payoffs are realized.

We look for perfect Bayesian Nash equilibria (simply called equilibria) of this game.[17] Strategies must thus be best responses given the history of play and given beliefs and beliefs are updated using Bayes' Rule whenever applicable.[18] Of course, an equilibrium must be sequentially rational. Given that the old technology is damaging equilibrium behavior consists of a second stage amount of money burnt m^* by a high type IG, a third stage switching rule $\delta^*(c, m)$, a fourth stage campaign size $\gamma^*(\alpha)$ if $\delta = 0$, and a fifth stage pricing rule $p^*(\delta, \gamma, c)$. If the old technology is not damaging, then a fifth stage pricing rule completely characterizes equilibrium behavior. To simplify the analysis we assume that the IG 'wins' any ties. We focus on separating equilibria which satisfy Cho and Kreps (1987) Intuitive Criterion: out-of-equilibrium beliefs only attach probability to types who could possibly benefit from the deviation.

3 Analysis

We determine the equilibria using backward induction. Consider first the firm's pricing decision. Firstly, if no IG were present and the firm sticks to the old technology, then a consumer buys the good if $1 - \psi\theta - p \geq 0$. So, a type θ-consumer then buys the good if $\theta \leq \min\{\frac{1-p}{\psi}, 1\}$ and the firm's profit is consequently

$$\hat{\pi}(p) = \begin{cases} p & \text{if } p \leq 1-\psi \\ \frac{1-p}{\psi}p & \text{if } p \in (1-\psi, 1] \end{cases}.$$

Since $\psi < \frac{1}{2}$, increasing the price beyond $1-\psi$ reduces profit: the loss in revenues from the customers who care a lot about the potential damage (those with θ close to 1) outweighs the extra revenues from the inframarginal consumers. The firm thus maximizes its profit by charging $p^* = 1-\psi$ which leads to sales of $\bar{q} = 1$. Secondly, if the firm adopts the clean technology, then the firm is best off charging $\bar{p} = 1$, which yields a profit of $\bar{\pi}(c) = 1-c$ and zero consumer surplus. Comparing these two profit expressions reveals that in the absence of the IG the firm switches to the clean technology if $c \leq \psi$. As stated before we confine attention to situations in which c is at least ψ.

Suppose now that the old technology is damaging and that the firm continues to use this technology. In response to the IG informing a fraction γ of the consumers the firm has two options. It can either conceal the fact that its technology is damaging from uninformed consumers by charging p^* or it can reveal that fact to all consumers by charging a lower price. If $p = p^*$, then a type θ-informed consumer buys the good if $1-\theta-p^* \geq 0$, i. e. if $\theta \leq \psi$, whereas each uninformed consumer buys the good. The quantity sold thus equals $1-(1-\psi)\gamma$. Setting $p = p^*$ therefore results in a profit of $\pi(\gamma) = 1-\psi-(1-\psi)^2\gamma$. If the firm charges a lower price $\tilde{p}$, then a type θ-consumer buys the good if $1-\theta-\tilde{p} \geq 0$ and hence the firm enjoys a profit of $(1-\tilde{p})\tilde{p}$. The optimal lower price is $\tilde{p} = \frac{1}{2}$ and this yields a quantity sold of $\tilde{q} = \frac{1}{2}$ and a profit of $\tilde{\pi} = \frac{1}{4}$. Note that $\pi(\gamma) > \tilde{\pi}$ if $\gamma < \bar{\gamma} = \frac{1}{1-\psi} - \frac{1}{4(1-\psi)^2}$.[19] So, quantity sold and profit as a function of γ are

$$q(\gamma) = \begin{cases} 1-(1-\psi)\gamma & \text{if } \gamma < \bar{\gamma} \\ \frac{1}{2} & \text{if } \gamma \geq \bar{\gamma} \end{cases}, \quad \pi(\gamma) = \begin{cases} 1-\psi-(1-\psi)^2\gamma & \text{if } \gamma < \bar{\gamma} \\ \frac{1}{4} & \text{if } \gamma \geq \bar{\gamma} \end{cases}. \tag{1}$$

One sees that informing additional consumers only leads to a drop in sales as long as not too many consumers are informed. If sufficiently many consumers are informed, i. e. $\gamma \geq \bar{\gamma}$, then the firm is best off ensuring that all consumers with $\theta \leq \frac{1}{2}$ do buy its good. It does so by lowering its price to $\frac{1}{2}$. So, informing more than a fraction $\bar{\gamma}$ of the consumers accomplishes the same reduction in sales and hence in aggregate damage as informing a fraction $\bar{\gamma}$ of the consumers. It is consequently never optimal to inform more than a fraction $\bar{\gamma}$ of the consumers. Conditional on $\gamma \leq \bar{\gamma}$ the first order condition of the IG's problem in stage four reads

$$\frac{\partial G(\gamma;\alpha)}{\partial \gamma} = -\alpha q'(\gamma) - k\gamma = \alpha(1-\psi) - k\gamma = 0.$$

It follows that $\gamma^* = \frac{\alpha}{k}(1-\psi)$ provided that this expression does not exceed $\bar{\gamma}$, i. e. $\frac{\alpha}{k} \leq \frac{\bar{\gamma}}{1-\psi}$.[20] To ensure that we do not end up with corner solutions that would clutter the analysis without adding much insight we maintain the following condition throughout the remainder of this paper:

Condition 1.
The IG's high-salience-to-marginal-cost ratio is not too large. Specifically: $\frac{\alpha_H}{k} \leq \frac{\bar{\gamma}}{1-\psi}$.

Given that the IG opts for $\gamma^*(\alpha) = \frac{\alpha}{k}(1-\psi)$ and using (1) one sees that its payoff in case the firm sticks to the dirty technology equals

$$G(\gamma^*(\alpha);\alpha) = -\alpha + \frac{\alpha^2(1-\psi)^2}{2k}.$$

Furthermore, combining (1) with the expression for $\gamma^*(\alpha)$ yields the firm's profit as a function of the salience α:

$$\pi^*(\alpha) = 1-\psi-\frac{\alpha}{k}(1-\psi)^3. \tag{2}$$

The second part of (2), $\frac{\alpha}{k}(1-\psi)^3$, is the IG's impact on profit. As one would expect both this impact and the fraction γ^* of informed consumers increases in the IG's *effective salience* $\frac{\alpha}{k}$. By contrast, an increase in ψ leads to a decrease in both γ^* and the IG's impact on profit. The intuition behind this result is as follows. Since an informed consumer buys the good if $\theta \leq \psi$, an increase in ψ reduces the effect of informing consumers on sales and the IG therefore opts for a lower γ thereby making its impact on profit smaller. So, if the a priori probability ψ that the old technology is damaging increases, then the IG's impact on outcomes becomes smaller.

Let us now determine whether money burning occurs in the second stage of the game. In a separating equilibrium the firm infers that the IG is a high type after observing money burning. The firm subsequently adopts the clean technology if $\bar{\pi}(c) \geq \pi^*(\alpha_H)$, which is equivalent to $c \leq \psi + \frac{\alpha_H}{k}(1-\psi)^3$. On the other hand, if the IG does not burn any money, then the firm, knowing that it faces the low type, adopts the clean technology only if $c \leq \psi + \frac{\alpha_L}{k}(1-\psi)^3$. Using the fact that the IG's payoff gross of any money burnt is 0 if the firm adopts the clean technology we see that it is only worthwhile for the high type to burn an amount $m > 0$ if

$$\Pr(c > \psi + \tfrac{\alpha_H}{k}(1-\psi)^3)G(\gamma^*(\alpha_H);\alpha_H) - m \geq \Pr(c > \psi + \tfrac{\alpha_L}{k}(1-\psi)^3)G(\gamma^*(\alpha_H);\alpha_H) \Leftrightarrow$$
$$m \leq \overline{m} = (F(\psi + \tfrac{\alpha_H}{k}(1-\psi)^3) - F(\psi + \tfrac{\alpha_L}{k}(1-\psi)^3))\left(\alpha_H - \frac{\alpha_H^2(1-\psi)^2}{2k}\right).$$

To avoid the low type from mimicking the high type m must be such that:

$$\Pr(c > \psi + \tfrac{\alpha_L}{k}(1-\psi)^3)G(\gamma^*(\alpha_L);\alpha_L) \geq \Pr(c > \psi + \tfrac{\alpha_H}{k}(1-\psi)^3)G(\gamma^*(\alpha_L);\alpha_L) - m \Leftrightarrow$$
$$m \geq \underline{m} = (F(\psi + \tfrac{\alpha_H}{k}(1-\psi)^3) - F(\psi + \tfrac{\alpha_L}{k}(1-\psi)^3))\left(\alpha_L - \frac{\alpha_L^2(1-\psi)^2}{2k}\right). \quad (3)$$

Condition 1 ensures that $\underline{m} < \overline{m}$, which implies that separating equilibria do exist:

Proposition 1.
For each x in the nonempty interval $[\underline{m}, \overline{m}]$ there exists a separating equilibrium. In case the old technology is damaging equilibrium behavior is governed by the following[21]:

- *Second stage money burning rule:* $m^* = x$,
- *Third stage switching rule:*

$$\delta^*(c,m) = \begin{cases} 1 & \text{if } c \leq \psi + \frac{\alpha_L}{k}(1-\psi)^3 \\ 1 & \text{if } \psi + \frac{\alpha_L}{k}(1-\psi)^3 < c \leq \psi + \frac{\alpha_H}{k}(1-\psi)^3 \text{ and } m \geq x \\ 0 & \text{else} \end{cases},$$

- *Fourth stage campaign size:* $\gamma^*(\alpha) = \frac{\alpha}{k}(1-\psi)$,
- *Fifth stage pricing rule:* $p^*(\delta, \gamma, c) = \delta + (1-\psi)(1-\delta)$.

If the old technology is not damaging, then the IG does not do anything, the firm sticks to the old technology, and it charges the price $p^* = 1 - \psi$.

Let us call the separating equilibrium in which the high type burns an amount x *equilibrium* x. This equilibrium is supported by beliefs of the firm which attach probability 1 to the event that the IG is a high type after an amount $m \geq x$ is burnt and probability 0 to that event if less than x is burnt. Clearly, the high IG type is strictly better off in equilibrium x_1 than in equilibrium x_2 if $x_1 < x_2$. With the aid of the Intuitive Criterion of Cho and Kreps (1987) we can therefore discard all but one of the equilibria given in Proposition 1:

Proposition 2.
Equilibrium $\underline{m}$ is the unique separating equilibrium satisfying the Intuitive Criterion.

In equilibrium $\underline{m}$ an *impasse* can occur: the firm does not switch to the new technology despite the fact that the IG has conveyed to be a high type by burning money. The probability that an impasse occurs is

$$\rho = \Pr(\alpha = \alpha_H)\Pr(\delta^*(c, \underline{m}) = 0) = (1-\phi)(1 - F(\psi + \tfrac{\alpha_H}{k}(1-\psi)^3)).$$

Of course, this probability is increasing in the probability $1-\phi$ that the IG is a high type. As the effective salience $\frac{\alpha_H}{k}$ of the high type increases the IG's negative impact on profit should the firm stick to the dirty technology becomes more significant. As a consequence, the firm becomes more inclined to adopt the clean technology as $\frac{\alpha_H}{k}$ increases and hence ρ decreases in $\frac{\alpha_H}{k}$. Lastly, if F is replaced by a distribution which is first order stochastically dominated by F, i. e. F is replaced by a distribution which assigns more probability mass to low values for c, then ρ becomes smaller. This is intuitive: if the probability that c is sufficiently low to make adopting the new technology worth the while increases, then an impasse occurs less often.

Because ψ can only change in conjunction with F, it is not straightforward to assess the impact of a change in ψ on ρ or other equilibrium outcomes. One can argue informally that if the shape of the graph of F does not change much, then an increase in ψ, say from ψ_1 to ψ_2, leads to an increase in ρ. The reason is that such an increase must be accompanied by probability mass moving from the interval $[\psi_1, \psi_2)$ to the right. This increases $1 - F(\psi + \frac{\alpha}{k}(1-\psi)^3)$ for fixed ψ. Since the change in $\psi + \frac{\alpha}{k}(1-\psi)^3$ is smaller than the change in ψ itself, the effect of probability mass moving to the right dominates the effect of the change in the position of $\psi + \frac{\alpha}{k}(1-\psi)^3$ as long as the shape of the graph of F does not change much. The intuition behind this result is as follows. If the probability ψ that the old technology is damaging is large, then even without a campaign consumers have relatively low willingnesses-to-pay and hence the price $p^* = 1 - \psi$ that the firm charges if it does not adopt the clean technology is low. Because of the low price the IG's campaign has only a moderate impact on sales and thus profits, making the firm less inclined to adopt the clean technology as ψ increases.

The amount of money burnt

$$\underline{m} = \Pr\left(c \in [\psi + \tfrac{\alpha_L}{k}(1-\psi)^3, \psi + \tfrac{\alpha_H}{k}(1-\psi)^3)\right)|G(\gamma^*(\alpha_L);\alpha_L)|$$

obviously increases in the salience α_H of the high type. However, a change in the salience α_L of the low type has an ambiguous effect on $\underline{m}$. The reason is that the probability $\Pr(c \in [\psi + \frac{\alpha_L}{k}(1-\psi)^3, \psi + \frac{\alpha_H}{k}(1-\psi)^3)$ that the firm adopts the clean technology if $\alpha = \alpha_H$ but not if $\alpha = \alpha_L$ is decreasing in α_L whereas the low type's payoff difference $|G(\gamma^*(\alpha_L);\alpha_L)|$ is increasing in α_L. Which of these two effects dominates depends amongst others on α_H and on ψ. If α_H is large, then the payoff difference is multiplied by a relatively large probability which does not change much as α_L varies. So, if α_H is large, then the payoff difference effect dominates and $\underline{m}$ is increasing in α_L. On the other hand, if the distribution of c is relatively concentrated, i. e. if ψ is large, then the probability is relatively small and an increase in α_L consequently results in a drop in $\underline{m}$. An increase in k also leads to a probability effect and a (positive) payoff difference effect. However, since the sign of the probability effect now also depends on the shape of the distribution from which c is drawn, whether or not an increase in k leads to an increase in $\underline{m}$ depends on the shape of F.

Note that if the IG does not interfere in the market, then all consumers buy the good and aggregate damages are proportional to $\bar{q} = 1$ if the old technology is damaging. In that case the IG's payoff is - α. So, given that the old technology is damaging the IG's gain in expected payoff stemming from engaging in private politics amounts to

$$\begin{aligned}\Delta(\alpha_L) &= (1 - F(\psi + \tfrac{\alpha_L}{k}(1-\psi)^3))G(\gamma^*(\alpha_L);\alpha_L) - (-\alpha_L) \\ &= \tfrac{\alpha_L^2(1-\psi)^2}{2k} + F(\psi + \tfrac{\alpha_L}{k}(1-\psi)^3)(\alpha_L - \tfrac{\alpha_L^2(1-\psi)^2}{2k})\end{aligned}$$

if $\alpha = \alpha_L$ and it amounts to $\Delta(\alpha_H) - \underline{m}$ if $\alpha = \alpha_H$. We can use these expressions to infer when a firm is likely to become the target of private politics. Suppose that the IG knows that each (monopolistic) firm in the set $\mathcal{F}$ employs a damaging technology and that it knows the cost distribution F_i (and hence ψ_i), the values of α_{Li} and α_{Hi}, and its salience type (low or high) vis-à-vis each firm $i \in \mathcal{F}$. Which firm does the IG target given that it only has the means to target a small subset of $\mathcal{F}$? Analyzing $\Delta(\alpha_L)$ and $\Delta(\alpha_H) - \underline{m}$ reveals the following:

Proposition 3.

The IG becomes more inclined to engage in private politics vis-à-vis some firm as the level of the IG's salience vis-à-vis that firm's malpractice (either α_L or α_H) increases or if that firm's cost distribution F is replaced by a distribution $\hat{F}$ that is first order stochastically dominated by F (keeping ψ fixed).

The observation that the IG becomes more inclined to engage in private politics vis-à-vis some firm as the level of the IG's salience vis-à-vis that firm's malpractice increases is not surprising. More interesting is the observation that the IG is more inclined to target firms for which c is likely to be small. Those firms are relatively likely to cave in to the IG's request and are hence attractive targets.[22]

4 Welfare Implications of Strategic Activism

In order to assess the welfare implications of strategic activism in the presence of multiple sources of uncertainty we compare the *ex ante* expected payoffs of the consumers and the firm in equilibrium $\underline{m}$ with their payoffs in the equilibrium of a benchmark in which no IG is present. In this benchmark the firm never adopts the new technology, consumers know that the IG's technology is damaging with probability ψ, the firm consequently charges $p^* = 1 - \psi$ for its good, and each consumer buys the good. The relevant payoffs of equilibrium $\underline{m}$ and the benchmark are collected in the following lemma:

Lemma 1.

In equilibrium $\underline{m}$ the ex ante expected payoffs are:

– *Ex ante expected consumer surplus:*

$$\mathbb{E}(CS^*) = (1-\psi)\psi + \psi[\phi(1-F(\psi+\frac{\alpha_L}{k}(1-\psi)^3))(\frac{\alpha_L}{2k}(1-\psi)^3 - \frac{1}{2} + \psi)$$
$$+(1-\phi)(1-F(\psi+\frac{\alpha_H}{k}(1-\psi)^3))(\frac{\alpha_H}{2k}(1-\psi)^3 - \frac{1}{2} + \psi)].$$

– *Ex ante expected profit:*

$$\begin{aligned}\mathbb{E}(\pi^*) = (1-\psi)^2 + \psi[\phi F(\psi+\frac{\alpha_L}{k}(1-\psi)^3)(1-\mathbb{E}(c|c\leqslant\psi+\frac{\alpha_L}{k}(1-\psi)^3))\\ +\phi(1-F(\psi+\frac{\alpha_L}{k}(1-\psi)^3))(1-\psi-\frac{\alpha_L}{k}(1-\psi)^3)\\ +(1-\phi)F(\psi+\frac{\alpha_H}{k}(1-\psi)^3)(1-\mathbb{E}(c|c\leqslant\psi+\frac{\alpha_H}{k}(1-\psi)^3))\\ +(1-\phi)(1-F(\psi+\frac{\alpha_H}{k}(1-\psi)^3))(1-\psi-\frac{\alpha_H}{k}(1-\psi)^3)].\end{aligned} \tag{4}$$

In the benchmark the ex ante expected payoffs are:

– *Ex ante expected consumer surplus:* $\mathbb{E}(\widehat{CS}) = \frac{\psi}{2}$.

– *Ex ante expected profit:* $\mathbb{E}(\hat{\pi}) = 1-\psi$.

The firm clearly prefers the benchmark to equilibrium $\underline{m}$: in the latter it is sometimes forced to adopt the new, costly technology and if it continues to use the old, damaging technology, then it does not sell its product to all consumers.

In both settings each consumer buys the good and enjoys some surplus if the old technology is not damaging. Suppose now that the old technology is damaging. In the benchmark the firm never adopts the clean technology and all consumers buy the good. This leads to a positive surplus for consumers who do not care much about the externality ($\theta < \psi$) and a negative surplus for consumers who care a lot ($\theta > \psi$). In equilibrium $\underline{m}$ the firm either adopts the clean technology, in which case each consumer breaks even when buying the good, or some of the consumers who care a lot learn from the IG that they are better off not buying the good and hence refrain from doing so. It turns out that the second, positive effect on consumer surplus dominates the first, negative effect on consumer surplus.

Whether total welfare, which is the sum of ex ante expected consumer surplus plus profit, is higher or lower in equilibrium $\underline{m}$ depends on the distribution of c. Only if c is likely to be quite small is total welfare larger in equilibrium $\underline{m}$. In that case the decrease in profit brought about by the IG's actions is in general outweighed by the increase in consumer surplus. Yet, for a large class of distributions, including the uniform distribution, total welfare is higher in the benchmark.

The next proposition summarizes these payoff comparisons:

Proposition 4.

Compared to the benchmark consumers are better off whereas the firm is worse off in equilibrium $\underline{m}$. *Total welfare in equilibrium* $\underline{m}$ *exceeds that of the benchmark only if F assigns sufficient probability mass to low values of* c.

Proposition 4 reveals that the IG often overprovides information. In fact, a regulator maximizing total welfare never informs consumers about the damaging aspects of the firm's production practices after the firm has made its technology choice.[23] The reason is that given that the firm sticks to the dirty technology informing consumers is always detrimental to total welfare. The mechanism behind this result is as follows. If the regulator would inform some consumers, then informed consumers with a sufficiently high θ would no longer buy the good. Even though such consumers are worse off when buying the good, the total surplus associated with such a transaction, i. e. $1-\theta$, is positive: the firm gains more than the consumer loses.

To try to improve welfare the regulator could also force the firm to disclose information about its production practices.[24] Suppose that the old technology is damaging. Then knowing that its profit is a mere $\tilde{\pi} = \frac{1}{4}$ if $\delta = 0$ and equals $\bar{\pi}(c) = 1-c$ if $\delta = 1$, the firm now adopts the clean technology if $c \leq \frac{3}{4}$. However, total welfare is now[25]

$$\begin{cases} 1-c & \text{if } c \leq \frac{3}{4} \\ \frac{3}{8} & \text{if } c > \frac{3}{4} \end{cases},$$

which is less than $1 - \frac{\psi}{2}$, the total welfare that would be generated if neither the IG nor the regulator would interfere in the market. So, the regulator prefers not to force the firm to disclose the information.

The above observations reveal that disclosing information regarding the firm's production practices appears to be a rather blunt policy tool: in the hands of the regulator it is not useful and in the hands of the IG it only improves welfare if F meets a strong requirement. However, Proposition 3 shows that the IG is especially keen on targeting firms that do meet this requirement. Furthermore, if the regulator is particularly concerned about consumer well-being or the damage the firm is causing, then informing consumers could be optimal.[26] The regulator could take these issues into account by maximizing $\mathbb{E}(\lambda CS + \pi + \epsilon q)$ instead of $\mathbb{E}(CS + \pi)$, where $\lambda > 1$ is the weight attached to consumer well-being and $\varepsilon > 0$ measures the severity of the externality per unit of production. Of course, informing consumers or forcing the firm to disclose information is optimal if either of these parameters is sufficiently large.

Note that, because the IG only cares about the damages, it does use this tool after the firm has decided to stick to its dirty technology. The reason that the regulator is unable to improve welfare by informing consumers is that *ex post* it is never optimal to do so. If the regulator could commit *ex ante* to informing a certain fraction of the consumers, then it would be able to improve welfare in some cases. If the regulator is unable to commit itself to information provision, then it might want to delegate this to the IG.

5 Concluding Remarks

We have developed a simple model of private politics that deals with strategic interactions between a firm and an interest group in an uncertain world. The IG seeks to mitigate the amount of damage the firm is causing, but it does not know how costly it is for the firm to adopt a clean technology. On the other hand, the firm is unaware of how salient the IG perceives the issue at stake to be. If the firm does not meet the IG's request to adopt the clean technology, then the IG initiates a campaign during which it informs some of the consumers about the firm's malpractice. Without the IG's campaign all consumers remain ignorant about the firm's behavior. To avoid such a costly campaign, the IG can signal to the firm that it cares a lot by burning money, e. g. it offers the firm's management an ostensibly expensive report regarding the firm's production practices.

In equilibrium a high-salience IG does signal its type to the firm and the firm is more inclined to adopt the clean technology if it receives such a signal. In particular, the firm adopts the clean technology for a larger range of values of its marginal cost associated with the clean technology in case money burning occurs. If the firm decides to stick to the damaging technology and an IG campaign ensues, then the firm faces a trade-off. It can either charge a high price and sell only to a small group of consumers who are either uninformed or do not care much about the damage or it can ensure that sales are relatively high by reducing its price. However, the latter strategy reveals to uninformed consumers that it is employing a damaging technology and hence increases sales only moderately. Because the firm does opt for this reduced price as soon as the IG informs a certain fraction of the consumer population, the IG never informs all consumers.

The fact that both the IG and the firm are imperfectly informed not only causes socially wasteful signaling, but it can also lead to impasses. An impasse occurs if the firm ignores the IG's request despite the fact that the IG has signaled to care a lot. Our welfare analysis reveals that the IG often overprovides information. The reason is that the IG's actions inhibit socially desirable trades. Only if the cost of adopting the clean technology is likely to be small does the positive effect of the IG's actions on consumer surplus outweigh its negative effect on the firm's profit. Luckily, firms with this characteristic are the most appealing targets in the eyes of the IG. Since for a regulator it is never optimal to inform consumers about the damage the firm is causing once the firm has made its technology choice, a regulator might want to delegate information provision to an IG.

Our results provide a number of empirically testable hypotheses. We conjecture that firms are more inclined to cave in to IG demands if the IG has committed more resources (bigger report written by a larger group of experts, longer research period leading up to the start of a campaign) to the issue at stake prior to the actual campaign. However, if a firm observing that an IG is committing substantial resources does not immediately alter its behavior after the IG has formulated its requests, then our model predicts that a very large campaign reaching a lot of consumers ensues. By contrast, if an IG only expends modest resources prior to a campaign, then we expect mainly moderately sized campaigns instead of a bimodal distribution of campaign sizes. The comparative statics regarding firm characteristics indicate that *ceteris paribus* IGs first target firms that can relatively easily improve their behavior. We therefore expect IGs to move on to tougher targets within an industry only after the low-hanging fruit has been harvested. This translates into increases over time of the cost differences between the technologies employed before and after IG interference.

Our analysis is admittedly limited. The current framework precludes competition. It would be interesting to investigate a situation with multiple firms. We conjecture that with multiple firms only a subset of the firms adopt the less damaging technology, for this creates vertical product differentiation between adopters and non-adopters. This differentiation can be beneficial for both adopters and non-adopters. Secondly, we have considered a one-shot game, eschewing any dynamics. Adding dynamics to the model might allow one to quantify the duration of impasses. Lastly, we have assumed that both the firm and the IG know how consumers' well-being is affected by the damage. It could be interesting to introduce a third source of informational asymmetries by relaxing this assumption. We leave these issues for future research.

Acknowledgements

I would like to thank Lambert Schoonbeek, attendees at EARIE 2015 and ASSET 2016, and two anonymous referees for useful comments.

Funding

Financial support by the Netherlands Organization for Scientific Research (NWO) (Funder Id: http://doi.org/10.13039/501100003246, Grant Number: 400-04-135) is gratefully acknowledged.

Appendix

A Appendix

Details regarding Proposition 1

It remains to prove that $\underline{m} < \overline{m}$, i. e. that $\alpha_L - \frac{\alpha_L^2(1-\psi)^2}{2k} < \alpha_H - \frac{\alpha_H^2(1-\psi)^2}{2k}$. Differentiating $\alpha \mapsto \alpha - \frac{\alpha^2(1-\psi)^2}{2k}$ yields $1 - \frac{\alpha(1-\psi)^2}{k}$. Invoking Condition 1 reveals that

$$1 - \frac{\alpha(1-\psi)^2}{k} \geq 1 - \frac{\bar{\gamma}}{1-\psi}(1-\psi)^2 = 1 - (1-\psi) + \frac{1}{4(1-\psi)} > 0.$$

This proves that $\underline{m} < \overline{m}$.

Proof of Proposition 2

Consider equilibrium x for some $x > \underline{m}$ and fix an out-of-equilibrium amount of money burnt $m \in [\underline{m}, x)$. The low type has no incentive to deviate to m, for the resulting payoff is less than its equilibrium payoff of $\Pr(c \leq \psi + \frac{\alpha_L}{k}(1-\psi)^3)G(\gamma^*(\alpha_L);\alpha_L)$ for *any* out-of-equilibrium beliefs the firm might entertain: condition (3) ensures that the low type is worse off after burning the amount m even if the firm is convinced to be dealing with the high type after m has been burnt. The Intuitive Criterion now prescribes that it is not *reasonable* for the firm to believe that the message m could possibly be sent by the low type. So, upon observing m the firm must be convinced to be dealing with the high type and act accordingly. After m has been burnt the firm thus switches to the new technology. The high type consequently has an incentive to deviate to m and equilibrium x therefore fails to meet the Intuitive Criterion. Such a profitable deviation is not possible in equilibrium $\underline{m}$ and this equilibrium consequently does satisfy the Intuitive Criterion. □

Proof of Proposition 3

Differentiating Δ results in

$$\Delta'(\alpha) = \frac{\alpha(1-\psi)^2}{k} + \frac{(1-\psi)^3}{k} f(\psi + \frac{\alpha}{k}(1-\psi)^3)(\alpha - \frac{\alpha^2(1-\psi)^2}{2k}) + F(\psi + \frac{\alpha}{k}(1-\psi)^3)(1 - \frac{\alpha(1-\psi)^2}{k}),$$

which is positive as long as Condition 1 holds. Since $\frac{\mathrm{d}\underline{m}}{\mathrm{d}\alpha_H} = \frac{(1-\psi)^3}{k} f(\psi + \frac{\alpha_H}{k}(1-\psi)^3)(\alpha_L - \frac{\alpha_L^2(1-\psi)^2}{2k})$, we have that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\alpha_H}(\Delta(\alpha_H) - \underline{m}) = &\frac{\alpha_H(1-\psi)^2}{k} + F(\psi + \frac{\alpha_H}{k}(1-\psi)^3)(1 - \frac{\alpha_H(1-\psi)^2}{k}) \\ &+ \frac{(1-\psi)^3}{k} f(\psi + \frac{\alpha_H}{k}(1-\psi)^3)\Big((\alpha_H - \frac{\alpha_H^2(1-\psi)^2}{2k}) - (\alpha_L - \frac{\alpha_L^2(1-\psi)^2}{2k})\Big) > 0.\end{aligned}$$

So, both types become more inclined to engage in private politics vis-à-vis some firm as their salience vis-à-vis that firm's malpractice increases. Clearly, $\Delta(\alpha_L)$ increases if F is replaced by a distribution $\hat{F}$ that is first order stochastically dominated by F. Since

$$\begin{aligned}\Delta(\alpha_H) - \underline{m} =& \tfrac{\alpha_H^2(1-\psi)^2}{2k} + F(\psi + \tfrac{\alpha_L}{k}(1-\psi)^3)(\alpha_L - \tfrac{\alpha_L^2(1-\psi)^2}{2k}) \\ &+ F(\psi + \tfrac{\alpha_H}{k}(1-\psi)^3)\Big((\alpha_H - \tfrac{\alpha_H^2(1-\psi)^2}{2k}) - (\alpha_L - \tfrac{\alpha_L^2(1-\psi)^2}{2k})\Big)\end{aligned}$$

is increasing in $F(\psi + \frac{\alpha_j}{k}(1-\psi)^3)$, $j \in \{L, H\}$, we can conclude that both types become more inclined to engage in private politics if F is replaced by a distribution $\hat{F}$ that is first order stochastically dominated by F. □

Proof of Lemma 1

Let us first consider equilibrium $\underline{m}$:

- Consumer surplus: If the old technology is not damaging, which happens with probability $1-\psi$, then each consumer enjoys a surplus of $1 - p^* = \psi$. If the old technology is damaging, the firm chooses $\delta = 0$, and the IG informs a fraction γ of the consumers, then consumers' surplus equals

$$(1-\gamma)\int_0^1 (1-\theta-p^*)\,\mathrm{d}\theta + \gamma\int_0^{\psi}(1-\theta-p^*)\,\mathrm{d}\theta = \tfrac{\gamma}{2}(1-\psi)^2 - \tfrac{1}{2} + \psi, \tag{5}$$

  where we used the fact that all uninformed consumers buy the good and a type θ-informed consumers buy the good if $\theta \le \psi$. Lastly, if the old technology is damaging and the firm chooses $\delta = 1$, then consumer surplus is zero. Combining these surpluses with the expression for $\gamma^*(\alpha)$ and taking into account the distribution of α and the expression for δ^* yields the desired result.

- Profit: If the old technology is not damaging, then the firm chooses $\delta = 0$ and its profit equals $1-\psi$. If the old technology is damaging and the firm chooses $\delta = 0$, then the profit is given in (2). Lastly, if the old technology is damaging and the firm chooses $\delta = 0$, then the profit is 1 - c. Taking into account the equilibrium switching rule δ^*, the probability with which the old technology is damaging, and the distributions of α and c results in the expression for $\mathbb{E}(\pi^*)$.

  We next consider the benchmark:

- Consumer surplus: The expression follows from the facts that each consumer enjoys a surplus of $1 - p^* = \psi$ if the old technology is not damaging and that consumer surplus equals $\int_0^1(1-\theta-p^*)\,\mathrm{d}\theta = \psi - \frac{1}{2}$ if the old technology is damaging.

- Profit: Obvious. □

Proof of Proposition 4

Note that $\mathbb{E}(\widehat{CS}) = \frac{\psi}{2} = (1-\psi)\psi + \psi(-\frac{1}{2} + \psi)$. The only difference between $\mathbb{E}(\widehat{CS})$ and $\mathbb{E}(CS^*)$ therefore resides in the 'ψ-parts' of these expected values. Since $\psi < \frac{1}{2}$, we have that

$$\begin{aligned}&(1 - F(\psi + \tfrac{\alpha}{k}(1-\psi)^3))(\tfrac{\alpha}{2k}(1-\psi)^3 - \tfrac{1}{2} + \psi) \\ >&(1 - F(\psi + \tfrac{\alpha}{k}(1-\psi)^3))(-\tfrac{1}{2} + \psi) > -\tfrac{1}{2} + \psi\end{aligned}$$

for $\alpha \in \{\alpha_L, \alpha_H\}$. Hence, $\mathbb{E}(CS^*) > (1-\psi)\psi + \psi(\phi(-\frac{1}{2} + \psi) + (1-\phi)(-\frac{1}{2} + \psi)) = \mathbb{E}(\widehat{CS})$. To prove that $\mathbb{E}(\pi^*) < \mathbb{E}(\hat{\pi})$ it suffices to show that each of the four terms of (4) inside the square brackets is less than $1-\psi$. The terms featuring $1 - \psi - \frac{\alpha}{k}(1-\psi)^3$, $\alpha \in \{\alpha_L, \alpha_H\}$, clearly are. Since $\Pr(c \le \psi) = 0$, it must be that $\mathbb{E}(c|c \le \psi + \frac{\alpha}{k}(1-\psi)^3) > \psi$, $\alpha \in \{\alpha_L, \alpha_H\}$, implying that the other two terms are also less than $1-\psi$. Let $M(\psi + z, \psi) = F(\psi + z)(1 - \mathbb{E}(c|c \le \psi + z)) + (1 - F(\psi + z))\frac{1-z}{2}$. Note that

$$\mathbb{E}(W^*) = \mathbb{E}(CS^* + \pi^*) = 1 - \psi + \psi\Big(\phi M(\psi + \tfrac{\alpha_L}{k}(1-\psi)^3, \psi) + (1-\phi)M(\psi + \tfrac{\alpha_H}{k}(1-\psi)^3, \psi)\Big).$$

The welfare difference between equilibrium $\underline{m}$ and the equilibrium of the benchmark can thus be written as

$$\mathbb{E}(W^*) - \mathbb{E}(\hat{W}) = -\tfrac{\psi}{2} + \psi\Big(\phi M(\psi + \tfrac{\alpha_L}{k}(1-\psi)^3, \psi) + (1-\phi)M(\psi + \tfrac{\alpha_H}{k}(1-\psi)^3, \psi)\Big).$$

To prove that $\mathbb{E}(W^*) < \mathbb{E}(\hat{W})$ ($\mathbb{E}(W^*) > \mathbb{E}(\hat{W})$) it suffices to show that $M(\psi + z, \psi) < \frac{1}{2}$ ($M(\psi + z, \psi) > \frac{1}{2}$) for all $z \in (0, \bar{\gamma}(1-\psi)^2) = (0, \frac{3}{4} - \psi)$, where we used the fact that $\frac{\alpha_H}{k}(1-\psi)^3 \le \bar{\gamma}(1-\psi)^2$. Since

$$\mathbb{E}(c|c \le \psi + z) = \tfrac{1}{F(\psi+z)}\int_{\psi}^{\psi+z} cf(c)\,\mathrm{d}c = \psi + z - \tfrac{1}{F(\psi+z)}\int_{\psi}^{\psi+z} F(c)\,\mathrm{d}c,$$

we have that

$$M(\psi+z,\psi) = F(\psi+z) - (\psi+z)F(\psi+z) + \int_{\psi}^{\psi+z} F(c)\,\mathrm{d}c + (1-F(\psi+z))\tfrac{1-z}{2}$$
$$= \tfrac{1-z}{2} + \tfrac{1-z-2\psi}{2}F(\psi+z) + \int_{\psi}^{\psi+z} F(c)\,\mathrm{d}c. \tag{6}$$

If F is the uniform distribution, then this expression becomes

$$M(\psi+z,\psi) = \tfrac{1-z}{2} + \tfrac{(1-z-2\psi)z}{2(1-\psi)} + \int_{\psi}^{\psi+z} \tfrac{c-\psi}{1-\psi}\,\mathrm{d}c = \tfrac{1}{2} - \tfrac{\psi z}{2(1-\psi)} < \tfrac{1}{2}.$$

We conclude that welfare is lower in equilibrium $\underline{m}$ if c is uniformly distributed. Let us next construct a distribution such that welfare is higher in equilibrium $\underline{m}$. We start with a distribution that does not have a density function: suppose that $\Pr(c=\psi)=1$, i. e. $F(c)=1$ for all $c \in [\psi, 1]$. Then:

$$M(\psi+z,\psi) = \tfrac{1-z}{2} + \tfrac{1-z-2\psi}{2} + z = \tfrac{1}{2} + \tfrac{1-2\psi}{2} > \tfrac{1}{2}.$$

So, if $c=\psi$ with certainty, then welfare would be strictly higher in equilibrium $\underline{m}$. By Urysohn's Lemma there exists a distribution $\hat{F}$ that does have a density function $\hat{f}$ such that welfare is higher in equilibrium $\underline{m}$ if c is drawn from the distribution with cumulative distribution function $\hat{F}$. Note that this distribution assigns a lot of probability mass to low values of c. □

Regulator Informing Consumers

Suppose the old technology is damaging and the firm does not adopt the clean technology. Using (1) and (5) one sees that if the regulator informs a fraction $\gamma < \bar{\gamma}$ of the consumers, then total welfare amounts to

$$\left(\tfrac{\gamma}{2}(1-\psi)^2 - \tfrac{1}{2} + \psi\right) + \left(1-\psi-\gamma(1-\psi)^2\right) = \tfrac{1}{2} - \tfrac{\gamma}{2}(1-\psi)^2,$$

which is decreasing in γ. If $\gamma \geq \bar{\gamma}$, then the firm charges $\tilde{p} = \frac{1}{2}$ and only consumers with $\theta \leq 1-\tilde{p} = \frac{1}{2}$ buy the good. Total welfare now equals

$$\int_0^{\frac{1}{2}} (1-\theta-\tilde{p})\,\mathrm{d}\theta + \tilde{\pi} = \tfrac{3}{8} < \tfrac{1}{2}.$$

So, the regulator opts for $\gamma = 0$.

Notes

1 See e. g. O'Rourke (2005) for more examples in which IGs try to induce firm behavior by initiating a campaign aimed at informing the public.
2 See O'Rourke (2005) and greenpeace.org/international/publication/17612/destination-zero/ (accessed June 2019), respectively.
3 Examples of strategic activism include organizing or precipitating a consumer boycott, lobbying legislators, establishing certification methods for products that meet the IG's ethical standards, and buying company shares in order to put desired changes in corporate behavior on the agenda during shareholder meetings.
4 Activist engagement (led by the Free Burma Coalition) was precipitated by human rights violations committed by junta troops protecting the pipeline. See Spar and La Mure (2003) for further details.
5 See Lewis (1996) for an account of the various approaches a *regulator* trying to reduce pollution can pursue when costs and benefits of abatement are privately known.
6 The role of money burning in our model is akin to the role played by costly advertising during the introductory phase of an experience good of unknown quality as in Milgrom and Roberts (1986).
7 See for instance the case studies in Spar and La Mure (2003) and O'Rourke (2005).
8 Eesley and Lenox (2006) find using US data that a firm is more likely to acquiesce to a stakeholder's request if that stakeholder's relative power is large. This is in line with our results: lower costs of informing consumers can be interpreted as the IG having larger relative power.
9 The model has multiple separating equilibria, but only one that satisfies Cho and Kreps (1987) Intuitive Criterion. We focus attention on that equilibrium.
10 This is reminiscent of Rogoff (1985)'s observation that it can be optimal to appoint a central banker who has a larger aversion to inflation than the median voter.
11 Note that in our model both the IG and the firm know how consumers respond to information about the firm's production practices.
12 Our results do not change if the clean technology's marginal cost is zero and the firm would instead incur a fixed cost c when adopting the clean technology.
13 The parameter θ only captures an individual consumer's attitude towards buying a damaging good. It does not include any disutility associated with the aggregate damage the firm is potentially causing. Note that since we have a continuum of consumers, an individual consumer's choices has no impact on aggregate damage. The way we incorporate *green preferences* is customary and is borrowed from, amongst others, Moraga-González and Padrón-Fumero (2002) and Heijnen (2013).

14 It is assumed that uninformed consumers do not learn from informed consumers that the technology is damaging. Alternatively, the fraction γ includes all consumers who have learned about the damage indirectly.
15 The price that the firm charges for its product might convey information regarding whether or not the firm causes damage. We do allow consumers to use this information to update $\tilde{\psi}$.
16 Merely requesting the firm to adopt the clean technology comes at no cost.
17 See for instance chapter 8 of Fudenberg and Tirole (1991) for a discussion of this equilibrium concept.
18 Note that since the IG's payoff does not depend directly on the marginal cost c, we can ignore any updating the IG might engage in.
19 The fraction $\bar{\gamma}$ increases from $\frac{3}{4}$ to 1 as ψ increases from 0 to $\frac{1}{2}$.
20 Note that $G(\gamma;\alpha) = -\alpha(1-(1-\psi)\gamma) - \frac{k}{2}\gamma^2$ is strictly concave in γ and hence the relevant second order condition is satisfied.
21 The fact that c and α are not publicly observable has a significant impact on the IG's behavior in the second stage. If the IG knew the value of c, i. e. it knew that $\Pr(c = \bar{c})$ for some $\bar{c} \in [\psi, 1]$, then the high type would only burn money if $\bar{c} \in (\psi + \frac{\alpha_L}{k}(1-\psi)^3, \psi + \frac{\alpha_H}{k}(1-\psi)^3]$. In that case $\overline{m} = \alpha_H - \frac{\alpha_H^2(1-\psi)^2}{2k}$ and $\underline{m} = \alpha_L - \frac{\alpha_L^2(1-\psi)^2}{2k}$. If $\bar{c}$ lies outside of the interval $(\psi + \frac{\alpha_L}{k}(1-\psi)^3, \psi + \frac{\alpha_H}{k}(1-\psi)^3]$, then the firm's behavior does not depend on the IG's type (it either always or never switches to the new technology) and money burning is therefore futile. Of course, if the firm knew the IG's type, then money burning would never occur.
22 IGs might be inclined to focus attention on *soft targets*, i. e. firms that (to some extent) take the external effects they are causing into account when making decisions and are hence more likely to act morally motivated. We do not consider such corporate socially responsible behavior: in our model the firm is a pure profit-maximizer. See e. g. Baron (2009) and Gupta and Innes (2014) for the interplay between corporate social responsibility and private politics.
23 See the Appendix for details.
24 This is equivalent to the regulator committing herself to informing all consumers.
25 If $c > \frac{3}{4}$, then the firm opts for $\delta = 0$, charges $\tilde{p} = \frac{1}{2}$, and consumers' surplus equals $\int_0^{\frac{1}{2}} (1-\theta-\frac{1}{2})\,\mathrm{d}\theta = \frac{1}{8}$.
26 For instance, if the damage jeopardizes consumer health, then the regulator might focus on consumer well-being and largely ignore producer surplus. Indeed, the fact that tobacco companies are often obliged to indicate that smoking causes cancer hints at governments putting an emphasis on consumer health.

References

Baron, D. 2001. "Private Politics, Corporate Social Responsibility, and Integrated Strategy." *Journal of Economics and Management Strategy* 10: 7–45.

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

Baron, D. 2009. "A Positive Theory of Moral Management, Social Pressure, and Corporate Social Performance." *Journal of Economics and Management Strategy* 18: 7–43.

Baron, D. 2011. "Credence Attributes, Voluntary Organizations, and Social Pressure." *Journal of Public Economics* 95 (11–12): 1331–38.

Bottega, L., and J. de Freitas. 2009. "Public, Private and Nonprofit Regulation for Environmental Quality." *Journal of Economics & Management Strategy* 18 (1): 105–23.

Cho, I.-K., and D. Kreps. 1987. "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics* 102: 179–221.

Darby, M., and E. Karni. 1973. "Free Competition and the Optimal Amount of Fraud." *Journal of Law and Economics* 16: 67–88.

Eesley, C., and M. Lenox. 2006. "Firm Responses to Secondary Stakeholder Action." *Strategic Management Journal* 27: 765–81.

Feddersen, T., and T. Gilligan. 2001. "Saints and Markets: Activists and the Supply of Credence Goods." *Journal of Economics and Management Strategy* 10: 149–71.

Fudenberg, D., and J. Tirole. 1991 *Game Theory*. Cambridge, MA: The MIT Press.

Gupta, S., and R. Innes. 2014. "Private Politics and Environmental Management." *Journal of Environmental Economics and Management* 68: 319–39.

Heijnen, P. 2013. "Informative Advertising by Environmental Groups." *Journal of Economics* 108: 249–72.

Krautheim, S., and T. Verdier. 2016. "Offshoring with Endogenous NGO Activism." *Journal of International Economics* 101: 22–41.

Lewis, T. 1996. "Protecting the Environment when Costs and Benefits are Privately Known." *RAND Journal of Economics* 27: 819–47.

Milgrom, P., and J. Roberts. 1986. "Price and Advertising Signals of Product Quality." *Journal of Political Economy* 94: 796–821.

Moraga-González, J., and N. Padrón-Fumero. 2002. "Environmental Policy in a Green Market." *Environmental and Resource Economics* 22: 419–47.

O'Rourke, D. 2005. "Market Movements: Nongovernmental Organization Strategies to Influence Global Production and Consumption." *Journal of Industrial Ecology* 9 (1–2): 115–28.

Rogoff, K. 1985. "The Optimal Degree of Commitment to an Intermediate Monetary Target." *Quarterly Journal of Economics* 100: 1169–89.

Spar, D., and L. La Mure. 2003. "The Power of Activism: Assessing the Impact of NGOs on Global Business." *California Management Review* 45: 78–101.