Strategic interactions in environmental economics

Heijnen, Pim

Chapter 3

Informative advertising by an environmental group

3.1 Introduction

In Chapter 2, I explored the implications of an environmental group (EG) that was able to influence consumers' preferences. Of course, this is not the only way in which consumers can be influenced. EGs can also inform consumers. To see why this could be useful, note that consuming a product does not (necessarily) reveal the environmental quality of the good. Therefore, for the firm it will be extremely difficult to convey information about environmental quality to the consumers. Hence, there is scope for a third party to act as a monitor of the firm and disseminate information among consumers. In the case of environmental quality, a candidate would be an environmental group (EG). One channel through which an EG can disseminate information is advertising. In case of low environmental quality an EG would be willing to spend large amounts on advertising. The willingness to pay this amount will be less if quality is high. Note that I use the term advertising loosely. It refers to any kind of information transmission by an EG whether it is advertising on television or seeking the attention of the media through campaigning (and getting on the news).

A quick look at the website of Greenpeace (`www.greenpeace.org.uk`) reveals that a large part of the activities of Greenpeace consists of informing consumers. Recently, they have been attaching stickers on Disney pajamas (in stores in the UK) warning the consumers that the product contains dangerous chemicals. This falls under the broad definition of advertising that I employ. Another example is the 'Viswijzer', an informative brochure aimed at Dutch consumers. This brochure tells consumers which fish are on the brink of extinction and of which the consumption should be avoided. The brochure is advertised on television by the World Wildlife Fund. Observe that the product itself does not display this kind of information and, hence, we have something that is distinct from (eco)labeling.

I model this in the following manner. A monopolistic firm and an EG know the damage per unit of production that is caused by the production of the good. Consumers, however, are unaware of this damage. The firm and the EG simultaneously decide on, respectively, the price of the good and the amount of money to be spent on an advertising campaign. Consumers can observe both the price and the amount of money and they form expectations about the damage of the good. Using these expectations, the consumers then decide to buy the good or not at the announced price.

Two variants are examined. In the first variant, the firm does not choose the damage per unit of production. Instead this is chosen at random by Nature prior to the game. I interpret this as the short run in which the firm has already chosen damage and did not anticipate the reaction of the EG or the consumers. I compare a benchmark equilibrium, in which the EG is prohibited to advertise (and no information about the product is transmitted to consumers), to an equilibrium, in which the EG does advertise and transmits information to the consumers.

I show that it is indeed possible for an EG to signal environmental quality through advertising. I show that there are two features of the advertising campaign. First, the EG will always elicit dirty firms. In Diermeier and Baron (Forthcoming)'s model of negotiation between an EG and a firm, a similar phenomenon occurs. In their model, however, by assumption the EG demands that the firm reduces pollution and will hurt the firm if it does not comply: note that the goal of the EG could equally well be attained by praising the firm if it complies. I do not impose that the EG must target dirty firms, it follows from the model. Second, the EG is hurt by their own campaign. This possibility is also noticed by Laffont and Tirole (1991) in their model of interest groups influencing the government. On the other hand, on average, in my model the firm and the consumer do benefit from the information transmission. The results for this first variant cast doubt on the short-run effectiveness of advertising by EGs.

In the second variant the firm does choose damage, which is interpreted as the long run. In this variant, I show that the EG can costlessly 'force' the firm to chose a clean product. Compared to the benchmark of no EG, this turns out to be a Pareto-improvement. The presence of the EG seems to be beneficial in the long run.

The informative advertising model presented here is related to the auditing model presented in Feddersen and Gilligan (2001), which studies a duopoly situation in which firms are price-takers. In their model, the EG can investigate a firm and discover how damaging the production process is. This information can then be transmitted to consumers. The crucial difference is that Feddersen and Gilligan suppose that information needs to be gathered whereas I suppose that information needs to be disseminated. Both approaches emphasize the difficulty the EG faces when it tries to convince the public. Feddersen and Gilligan, in particular, study costless investigation. Consequently, their model relies on cheap talk to transmit information. The fact that cheap talk is possible, is an artifact of their model. Crawford and Sobel (1982) investigate the conditions under which an informative equilibrium arises in general cheap-talk games. Crawford and Sobel reach the conclusion that while such an equilibrium can exist, it is only possible if the disagreement between the receiver and the sender is not too large. Applied to this situation we see that while the EG and consumer may agree that a lower damage per unit of production is better than a higher damage, the consumers also want to keep price as low as possible. Feddersen and Gilligan eliminate this factor by making the firms compete in price.

Besides Feddersen and Gilligan (2001), there is other related literature. First, the model of Chapter 2 is similar to the model presented here, but in that Chapter advertising is solely persuasive instead of informative; in that Chapter I conclude that the threat of advertising is more effective than actually advertising. Second, the model is also closely related to models of costly lobbying (cf. Grossman and Helpman, 2001). In these models, special interest groups spend money to convince a policy maker about the state of the world. In contrast to this literature, I study direct action by an EG aimed at the consumer. Finally, since the model describes a game of asymmetric information the analysis boils down to an investigation of signaling equilibria, where the EG and the firm try to signal the environmental quality of the good by selecting an appropriate level of, respectively, the cost of the campaign and the price of the good. For related studies in the standard advertising literature, see e.g. Hertzendorf and Overgaard (2001) and Milgrom and Roberts (1986).

This Chapter is organized as follows. Section 3.2 introduces the model. Section 3.3 presents the equilibria of the model. Section 3.4 analyzes social welfare. The effect of a firm choosing its damage level is analyzed in Section 3.5. Section 3.6 concludes.

3.2 The model

The model presented in this Section is similar to the one presented in Chapter 2. There are two differences. First, the firm does not choose the damage per unit of production, but instead this damage is drawn from a known distribution. Second, the distribution of preferences of the consumers is fixed and cannot be influenced by an environmental group. In this model consumers are initially unaware of the damage of the good and might be informed by the firm and/or the EG.

Consider a market in which a single profit-maximizing firm sells one good. This good is characterized by the damage d it causes to the environment per product sold. *Ex ante*, with probability ρ damage is low, $d = d_L$, and with probability $1 - \rho$ damage is high, $d = d_H$. Assume that $0 < d_L < d_H$ and $0 < \rho < 1$. The firm and the EG know the true value of d while the consumers, who care about the damage the product causes, only know the distribution of d. The price of the good is p.

The model focuses on the short run; the production technology is fixed. Irrespective of d, it costs $c > 0$ to produce one unit of the good. Usually, marginal cost is assumed to be decreasing in damage level. By abstracting from this, I focus purely on the communication problem. Consumers are indexed by θ, where θ is uniformly distributed on the interval $[0, \bar{\theta}]$. Unlike in Chapter 2, we can set, without loss of generality, $\bar{\theta} = 1$.[1] Consumer θ has the following utility function (Tirole, 1988, pp. 96–97):

$$U(\theta, p) = \begin{cases} V - \theta\mu - p & \text{if one unit of the good is bought,} \\ 0 & \text{if zero units are bought,} \end{cases} \tag{3.1}$$

where V is a positive constant, θ measures the disutility of environmental damage and μ is the expected damage level associated with the consumption

[1] The reason for this is the following. In the previous chapter, the firm chose damage per unit of production to change demand such that the EG would not enter. The shape of the demand function with respect to d did not guarantee an interior maximum and, hence, the assumption that demand is between zero and one could be binding, and could qualitatively affect the behavior of the firm. Here, this turns out to be immaterial. It is true that the objective of the EG is influenced by $\bar{\theta}$. In particular, it is equal to multiplying γ by $1/\bar{\theta}$, but as we will see in Footnote 3, this does not influence the equilibrium qualitatively either.

of one unit of the good. Note that utility is linear in expected damage. All consumers have the same information, which they process in the same manner. After they have observed the actions of the firm and the EG, the consumers have the *ex post* belief that damage is low with a probability ϕ. Note also that $\phi = \rho$ if they receive no new information. Then $\mu = \phi d_L + (1-\phi)d_H$. Note that the relation between ϕ and μ is strictly decreasing. Hence, there is no loss of generality to consider only μ.

Only consumers with a sufficiently low θ will buy the good. The indifferent consumer can be found by equating utility to zero:

$$\hat{\theta} = \frac{V-p}{\mu}. \tag{3.2}$$

The demand for the good $q(p, \mu)$ is $\hat{\theta}$. I will assume an interior solution: there are always some consumers that do not buy the good because they care much about the environment, but there are also some consumers that always buy the good. A consequence is that demand is between zero and one.

The profit of the firm is $\Pi(p, \mu) = q(p, \mu)(p - c)$. It can be easily verified that the optimal price is given by:

$$p^* = \frac{1}{2}V + \frac{1}{2}c. \tag{3.3}$$

Note that if consumers could observe the damage of the product, then the price chosen by the firm would be the optimal price, which does not depend on μ.[2] Finally, observe that:

$$q(p^*, \mu) = \frac{(V-c)}{2\mu}, \tag{3.4}$$

$$\Pi(p^*, \mu) = \frac{(V-c)^2}{4\mu}. \tag{3.5}$$

Notice that demand and profit are decreasing in μ. Furthermore, both functions are convex in μ.

There is an environmental group that knows the true per product damage and cares about the total environmental damage caused by the production of the good. The cost of environmental damage to the EG is $q(p, \mu)d$, which is increasing in demand and increasing in d. Simultaneously with the firm's

[2]This lack of dependency is a result of the following two assumptions. First, the marginal cost of producing the good does not depend on d. Second, the lower boundary of the distribution of θ is zero. These two assumptions considerably ease the calculations while they do not affect the results qualitatively.

price decision, the EG can launch an advertising campaign. This advertising consists of a message ("The firm is of type d") and has cost x. It is assumed that when advertising is observed by the consumers and the firm, then it is possible for these observers to infer the cost of the advertising campaign. For instance, if consumers see an ad on prime-time television, they will note that this would have cost the EG a great deal of money. The cost of the advertising campaign is common knowledge.

The EG tries to minimize the sum of total damage and advertising cost: $\gamma q(p, \mu)d + x$. Here γ converts total environmental damage to its monetary value. Note that q possibly depends indirectly on x through the beliefs of the consumers. For example, the consumers might think that a costly advertising campaign signals a dirty product and therefore the consumers update the expected damage level upwards. For convenience I will set γ equal to one. As we will see, this has no effect apart from scaling the cost of the EG in equilibrium by a factor γ.[3] In the social welfare analysis of Section 3.4, I will therefore reintroduce the γ-parameter.

The members of the EG are activists who never buy the product. Furthermore, they are willing to exert effort to figure out how damaging per unit of production the good really is. I assume that the EG has ample funds to launch an advertising campaign. But they want to spend money as efficient as possible for there might be other projects that need money. This idea is captured by making the EG minimize cost.

It is important to notice that the actual message is not important, but the cost of advertising is. It should be noted that it is *not* the objective of the EG to maximize consumers' surplus. The public's skeptical view of the EG stems from the non-alignment of the public interest and the EG's interest. To see that the content of the messages can never be trusted by the consumers if the EG does not spend money on advertising campaigns, take the following situation. The message of the advert is either "$d = d_L$" or "$d = d_H$". For the consumer to trust the message "$d = d_L$", we need that:

$$q(p, d_L)d_L \leq q(p, d_H)d_L. \tag{3.6}$$

In other words, if $d = d_L$, then the cost of the EG should be lower if they tell the truth instead of claiming that "$d = d_H$". However, since q is decreasing

[3] In particular, all equilibrium values of x that are derived in Section 3.3, are multiplied by γ. It is straightforward to check that in all equilibrium conditions γ drops out and γx is indeed an equilibrium value. The cost of the EG in equilibrium is then $\gamma q(p, \mu)d + \gamma x$. Therefore, γ is just a scaling factor.

in μ and $d_L < d_H$, if $d = d_L$, then the EG will claim that the firm is of a type d_H. Hence, the consumer will not trust the EG. To simplify the analysis (and without loss of generality), I assume that the content of the advert is cheap talk, but by showing how much it is willing to spend the EG can try to convince the public that it is telling the truth.

The timing of the game is as follows:

Period 0: Nature draws damage level per unit of production d .

Period 1: The firm and the EG observe d. The EG chooses the intensity of the advertising x. The firm chooses its price p.

Period 2: Consumers observe x and p. They update their beliefs μ (following Bayes' rule whenever possible) and choose to buy the good or not. This results in a demand $q(p, \mu)$.

The equilibrium concept that will be used, is the perfect Bayesian equilibrium (Mas-Colell, Whinston and Green, 1995, p.285). The two requirements of this equilibrium concept are that the firm and the EG choose optimal actions given the beliefs of the consumers, and the consumers' beliefs are consistent with these actions. Note that usually there is one sender of information (cf. Milgrom and Roberts, 1986), whereas here there are two senders (the firm and the EG).

The order in which the game is played, mainly the fact that the firm and the EG move simultaneously, is the result of the following consideration. If the firm is allowed to determine the price after the EG chooses an advertising campaign, then the firm's reaction on x happens before the consumers observe x and update their beliefs. Considering that the advertising campaign is aimed at the consumers this chain of events is highly unlikely. The correct way of modeling this would be to let the consumers update their beliefs twice: first after witnessing x, then after observing p. For sake of simplicity, the firm and the EG are assumed to choose x and p simultaneously.

3.3 Equilibria

The forces that shape the set of equilibria are the following. Intuitively speaking, the EG wants to convince the consumers that damage per unit of production is high while the firm wants to achieve the opposite. If in an equilibrium the consumers believe that damage is low, then the cost of convincing the consumers that damage is high should be high enough to discourage the EG to choose this action. One way of 'punishing' the EG is to assume that the damage expected by the consumers off the equilibrium path (after a deviation of the EG) is sufficiently low. If in an equilibrium the consumers believe that damage is high, then the gains of convincing the consumers that damage is low should be so low that the firm will not choose this action. By assuming that off the equilibrium path (after a deviation of the firm) the expected damage is sufficiently high, the consumers can punish this type of behavior from the firm. The beliefs that sustain an equilibrium therefore assign low expected damage levels to deviations from the EG and high expected damage levels to deviations from the firm.

Two types of equilibria can be distinguished: separating equilibria and pooling equilibria. In the subsequent sections all equilibria are derived.

3.3.1 Separating equilibria

A separating equilibrium is one in which the consumers can distinguish between each state of the world, i.e. $\mu = d_L$ if $d = d_L$ and $\mu = d_H$ if $d = d_H$. Since consumers must be able to differentiate the two situations, it must be that $(x_L, p_L) \neq (x_H, p_H)$, where $x_L = x(d_L)$, $x_H = x(d_H)$, $p_L = p(d_L)$ and $p_H = p(d_H)$. This denotes advertising levels and prices for both states of the world.

In this game there are two senders, the EG and the firm. One of the consequences of this is that there are two-sided separating equilibria (TSE) and one-sided separating equilibria (OSE). The terminology originates in Bagwell and Ramey (1991) and is defined as follows. In a TSE both senders emit different signals in both states of the world: i.e. $x_L \neq x_H$ and $p_L \neq p_H$. In an OSE only one sender emits a different signal in both states of the world. There are two distinct OSE: one where the firm signals ($p_L \neq p_H$ and $x_L = x_H$) and one where the EG signals ($x_L \neq x_H$ and $p_L = p_H$).

One of the peculiarities of a TSE is that it is impossible for either the firm or the EG alone to make the consumers believe that they are in another state of the world, e.g. pretend that $d = d_L$ when it is in fact d_H. Even if, for instance, the firm decides to choose p_L when the state of the world is d_H, the EG would still choose x_H. The consumers would not automatically infer that $d = d_L$. Since there are two senders in this game, this would require a coordinated deviation where the EG chooses x_L instead of x_H and the firm chooses p_L instead of p_H. Hence, it is clear that a deviation from the equilibrium strategy by one player will always result in an out-of-equilibrium action. I will start with deriving the TSE.

Two-sided separating equilibria

Observe that beliefs must be such that $\mu(x_L, p_L) = d_L$ and $\mu(x_H, p_H) = d_H$. In equilibrium the firm selects p_H by using:

$$p_H = \arg\max_p \frac{(V-p)(p-c)}{\mu(x_H, p)} \tag{3.7}$$

if $d = d_H$, and p_L by using

$$p_L = \arg\max_p \frac{(V-p)(p-c)}{\mu(x_L, p)} \tag{3.8}$$

if $d = d_L$. Note that the interaction between the firm and the EG is solely through expected damage per unit of production $\mu(x, p)$. Since out-of-equilibrium beliefs are not restricted, the aim is to find all p_H and p_L with $p_H \neq p_L$ that, for some $\mu(x_H, p)$ and $\mu(x_L, p)$, satisfy (3.7) and (3.8) with respectively $p \neq p_H$ and $p \neq p_L$.

Suppose the firm deviates from the equilibrium strategy. Given the out-of-equilibrium beliefs $\mu(x_H, p)$ and $\mu(x_L, p)$, this deviation should not be optimal. If $d = d_H$, then for the firm we must have:

$$\frac{(V-p_H)(p_H-c)}{d_H} \geq \frac{(V-p)(p-c)}{\mu(x_H, p)} \text{ for all } p \neq p_H, \tag{3.9}$$

which implies that for all $p \neq p_H$ we must have:

$$\mu(x_H, p) \geq \frac{(V-p)(p-c)}{(V-p_H)(p_H-c)} d_H. \tag{3.10}$$

By choosing $\mu(x_H, p)$ large enough, it is not profitable for the firm to deviate to a price p. Observe that in equilibrium the consumers already believe that the product is as dirty as it can be, i.e. $\mu(x_H, p_H) = d_H$. Note that the RHS of (3.10) exceeds d_H (which should not be possible) for some p if:

$$(V-p)(p-c) > (V-p_H)(p_H-c). \tag{3.11}$$

As long as $p_H \neq p^*$, see (3.3), such a price exists. Hence, we must have $p_H = p^*$. In other words, in a TSE there is no distortion for the high-damage firm in the sense that if there were no asymmetric information and the consumers could observe the type of the firm, then the high-damage firm would select the same price.

If $d = d_L$, then for the firm we must have:

$$\frac{(V-p_L)(p_L-c)}{d_L} \geq \frac{(V-p)(p-c)}{\mu(x_L,p)} \text{ for all } p \neq p_L. \tag{3.12}$$

Hence, for all $p \neq p_L$ we obtain:

$$\mu(x_L,p) \geq \frac{(V-p)(p-c)}{(V-p_L)(p_L-c)} d_L. \tag{3.13}$$

The firm must be punished for deviating and this is done by increasing the expected damage per unit of production and thereby decreasing the gains of deviation. As long as the RHS of (3.13) is smaller than or equal to d_H, one can always find $\mu(x_L,p)$ that satisfy (3.13). Note that the RHS of (3.13) reaches a maximum at $p = p^*$. In equilibrium p_L must be such that:

$$\frac{(V-p^*)(p^*-c)}{(V-p_L)(p_L-c)} d_L \leq d_H. \tag{3.14}$$

One can check that the set of feasible prices p_L is of the following form $p_L \in [\underline{p}, \bar{p}] \backslash \{p^*\}$, where $\underline{p}$ and $\bar{p}$ are the two solutions to:

$$\frac{(V-p^*)(p^*-c)}{(V-p_L)(p_L-c)} d_L = d_H. \tag{3.15}$$

If there would be no asymmetric information, then the profit for the firm if $d = d_L$ would be highest when $p_L = p^*$. However, in the TSE we must have $p_L \neq p_H = p^*$, and so this optimum is not reached in this equilibrium. Intuitively speaking, you would expect that the clean product is sold at a higher price than the dirty product, i.e. $p_L > p^*$. While this can indeed be true, note that, if $d = d_L$, then the firm can also signal its type by choosing a lower price for its clean product. What matters is that the firm must sacrifice some of its profit to be able to signal that it indeed sells a low-damage good. It turns out that one way of doing this is by selling the good at a low price, i.e. $p_L < p^*$.

I now turn to the behavior of the EG. In equilibrium the EG determines x_L from:

$$x_L = \arg\min_x \frac{V-p_L}{\mu(x,p_L)} d_L + x \tag{3.16}$$

if $d = d_L$, and x_H from

$$x_H = \arg\min_x \frac{V-p_H}{\mu(x,p_H)} d_H + x \tag{3.17}$$

if $d = d_H$. The aim is to find all x_L and x_H with $x_L \neq x_H$ that, for some $\mu(x, p_L)$ and $\mu(x, p_H)$, satisfy (3.16) and (3.17) with respectively $x \neq x_L$ and $x \neq x_H$.

Suppose $d = d_L$ and the EG deviates. What out-of-equilibrium beliefs are necessary to sustain the equilibrium amount of advertising x_L? We must have:

$$\frac{V - p_L}{\mu(x, p_L)} d_L + x \geq \frac{V - p_L}{d_L} d_L + x_L \text{ for all } x \neq x_L. \tag{3.18}$$

For $x > x_L$, this inequality always holds if $\mu(x, p_L) = d_L$ (i.e. $\mu(x, p_L)$ remains low). This restriction implies that extra money spent on advertising is useless since consumers do not increase their expected damage levels. For $x < x_L$, a necessary condition is $\mu(x, p_L) < d_L$, which is impossible. To make this deviation unattractive, the consumers must believe that the product is cleaner than they know is possible. Hence, there can be no $x < x_L$, and I conclude that $x_L = 0$. If the product is clean, then the EG does not advertise.

Next, suppose that $d = d_H$. Here the out-of-equilibrium beliefs must satisfy:

$$\frac{V - p^*}{\mu(x, p^*)} d_H + x \geq \frac{V - p^*}{d_H} d_H + x_H \text{ for all } x \neq x_H. \tag{3.19}$$

If $x > x_H$, then this inequality holds if $\mu(x, p^*) = d_H$. Again the interpretation is that if the extra cost does not increase the expected damage level, then for the EG such a deviation is not worthwhile. Let $x < x_H$ and rewrite (3.19) as:

$$\mu(x, p^*) \leq \frac{V - p^*}{(V - p^*) + (x_H - x)} d_H. \tag{3.20}$$

Note that the RHS is smaller than d_H. However, without further restrictions, it might also be smaller than d_L, which should be impossible. The RHS is minimal at $x = 0$. If we take $0 < x_H \leq (V - p^*)\frac{d_H - d_L}{d_L} \equiv R$, and set $\mu(x, p^*) = d_L$ for $x < x_H$ (the worst belief for the EG), then inequality (3.20) is true. Thus, out-of-equilibrium beliefs that support the equilibrium under consideration are the following: $\mu(x, p^*) = d_L$ if $x < x_H$ and $\mu(x, p^*) = d_H$ if $x > x_H$. So, the EG is punished severely if it advertises less than x_H, and the beliefs ensure that it is never optimal to advertise more than x_H.

Summarizing, in a TSE the following must hold: $x_L = 0$, $0 < x_H \leq R$, $p_L \in [\underline{p}, \bar{p}] \backslash \{p^*\}$ and $p_H = p^*$. However, so far, the role of $\mu(x_L, p_H)$ and $\mu(x_H, p_L)$ in a TSE has been ignored. Note, e.g., that if consumers observe $x = x_L$ and $p = p_H$, then either the true state of the world is d_H and the EG is deviating from the equilibrium, or the true state of the world is d_L

and the firm is deviating from the equilibrium. Hence, in a TSE $\mu(x_L, p_H)$ must be such that it deters both the EG from choosing x_L if the state of the world is d_H and the firm from choosing p_H if the state of the world is d_L. This is an extra restriction on possible TSE that I have not yet considered. A similar observation can be made with respect to $\mu(x_H, p_L)$. Since $x_L = 0$ and $p_H = p^*$, the extra restrictions only apply to x_H and p_L (obviously they also apply to $\mu(x_L, p_H)$ and $\mu(x_H, p_L)$). Before I continue the discussion, it is convenient to introduce the following notation: $\mu_{LH} \equiv \mu(x_L, p_H)$ and $\mu_{HL} \equiv \mu(x_H, p_L)$.

Let me start with μ_{LH}. First, suppose that $d = d_H$ and the EG deviates to $x_L = 0$. The firm chooses the equilibrium price $p_H = p^*$. It is optimal for the EG to not deviate if μ_{LH} is such that:

$$(V - p^*) + x_H \leq \frac{(V - p^*)}{\mu_{LH}} d_H, \tag{3.21}$$

or, equivalently:

$$\mu_{LH} \leq \frac{(V - p^*)}{(V - p^*) + x_H} d_H. \tag{3.22}$$

Second, suppose that $d = d_L$ and the firm deviates to $p_H = p^*$. The EG chooses the equilibrium amount of advertising $x_L = 0$. It is optimal for the firm to not deviate if μ_{LH} is such that:

$$\frac{(V - p_L)(p_L - c)}{d_L} \geq \frac{(V - p^*)(p^* - c)}{\mu_{LH}}, \tag{3.23}$$

or, equivalently:

$$\mu_{LH} \geq \frac{(V - p^*)(p^* - c)}{(V - p_L)(p_L - c)} d_L. \tag{3.24}$$

We now have two restrictions on μ_{LH}, which do not exclude each other if the RHS of (3.22) is larger than the RHS of (3.24). Consequently, x_H and p_L in a TSE must be such that:

$$\frac{(V - p^*)}{(V - p^*) + x_H} d_H \geq \frac{(V - p^*)(p^* - c)}{(V - p_L)(p_L - c)} d_L, \tag{3.25}$$

or after some manipulations:

$$x_H \leq \frac{(V - p^*)[1 - f(p_L)]}{f(p_L)}, \tag{3.26}$$

where

$$f(p) \equiv \frac{(V - p^*)(p^* - c)}{(V - p)(p - c)} \times \frac{d_L}{d_H}. \tag{3.27}$$

Observe that (3.26) gives an upper bound for x_H for each $p_L \in [\underline{p}, \bar{p}] \backslash p^*$. Note that $f(\underline{p}) = f(\bar{p}) = 1$ (cf. 3.15) and $f(p^*) = d_L/d_H$. Furthermore, the function $f(\cdot)$ has a unique minimum at p^*. Hence, the RHS of (3.26) has a unique maximum at p^*.

Let me proceed with μ_{HL}. First, suppose $d = d_L$ and that the EG deviates to x_H. The firm chooses the equilibrium strategy p_L. It is optimal for the EG to not deviate if μ_{HL} is such that:

$$(V - p_L) \leq \frac{(V - p_L)}{\mu_{HL}} d_L + x_H, \tag{3.28}$$

or:

$$\mu_{HL} \leq \frac{(V - p_L)}{(V - p_L) - x_H} d_L. \tag{3.29}$$

Second, suppose that $d = d_H$ and the firm deviates to p_L. The EG chooses the equilibrium amount of advertising x_H. It is optimal for the firm to not deviate if μ_{HL} is such that:

$$\frac{(V - p^*)(p^* - c)}{d_H} \geq \frac{(V - p_L)(p_L - c)}{\mu_{HL}}, \tag{3.30}$$

or, equivalently:

$$\mu_{HL} \geq \frac{(V - p_L)(p_L - c)}{(V - p^*)(p^* - c)} d_H. \tag{3.31}$$

We thus obtain two restrictions on μ_{LH}, which do not exclude each other if the RHS of (3.29) is larger than the RHS of (3.31). So, x_H and p_L in a TSE must be such that:

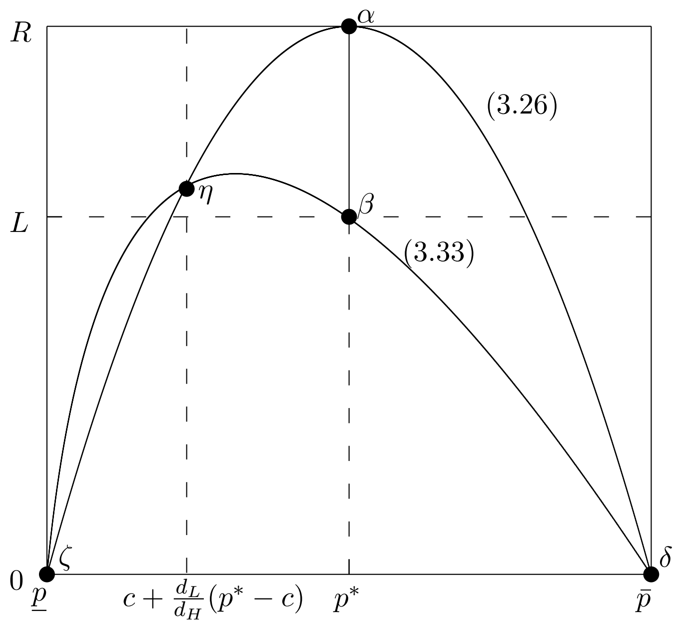
$$\frac{(V - p_L)}{(V - p_L) - x_H} d_L \geq \frac{(V - p_L)(p_L - c)}{(V - p^*)(p^* - c)} d_H, \tag{3.32}$$

which can be rewritten as:

$$x_H \geq (V - p_L)[1 - f(p_L)]. \tag{3.33}$$

Note that (3.33) yields a lower bound on x_H for the different feasible values of p_L. It can be verified that the RHS of (3.33) attains a maximum for some $p \in (\underline{p}, p^*)$. Of course, this lower bound should be consistent with the upper bound derived in (3.26). The following facts show how the upper and lower boundary relate to each other. The RHS of (3.33) is equal to the RHS of (3.26) if $p_L = \underline{p}$, $p_L = \bar{p}$, or $p_L = c + \frac{d_L}{d_H}(p^* - c) \in (\underline{p}, \bar{p})$. One can also check that the RHS of (3.33) is smaller than the RHS of (3.26) if $p_L = p^*$. Using these facts, Figure 3.1, which shows the values of x_H and p_L that are

Figure 3.1: The separating equilibria



NOTE: p_L is on the horizontal axis and x_H is on the vertical axis. Note that $p_H = p^*$ and $x_L = 0$ in all separating equilibria. The area between (3.26) and (3.33) to the right of the point η are the TSE with exception of the line $\alpha\beta$ and the point δ, which are OSE since there either $p_L = p_H$ or $x_H = x_L$. The point ζ is also an OSE.

consistent with a TSE, can be constructed. (In the analysis below I examine OSE which are also shown in Figure 3.1.)

Concluding, there are TSE in which $x_L = 0$, $(V - p_L)[1 - f(p_L)] \leq x_H \leq \frac{(V-p^*)[1-f(p_L)]}{f(p_L)}$, $p_L \in (c + \frac{d_L}{d_H}(p^* - c), \bar{p})\backslash\{p^*\}$ and $p_H = p^*$. These equilibria are supported by beliefs in which, in case of deviations of the firm, perceived damage is high enough and, in case of deviations of the EG, perceived damage is low enough.

One-sided separating equilibria where the firm informs

Next, OSE will be examined. An OSE is an equilibrium in which only one sender informs. I start by investigating the case in which the firm is the sender and the EG in equilibrium chooses the same advertising expenditure in both states of the world, i.e. $x_L = x_H$. Note that the analysis for finding p_L and p_H in the previous section (until the discussion of μ_{LH} and μ_{HL}) did not depend on the behavior of the EG. Consequently in equilibrium, we still have $p_H = p^*$ and $p_L \in [\underline{p}, \bar{p}] \backslash \{p^*\}$, but this set can be restricted further still. Observe that the behavior of the firm is bound by one additional constraint: since the firm is the only player sending a signal, it can also lie about the state of the world: if $d = d_L$, for instance, then $\mu(x_L, p_H) = \mu(x_H, p_H) = d_H$ whereas $\mu(x_L, p_L) = \mu(x_H, p_L) = d_L$. The firm can mimic its own behavior in the other state of the world as is usual in signaling models with one sender. For the firm this implies that the following extra constraint must be satisfied:

$$\frac{(V - p_L)(p_L - c)}{d_L} \geq \frac{(V - p_H)(p_H - c)}{d_H} \text{ if } d = d_L \tag{3.34}$$

$$\frac{(V - p_H)(p_H - c)}{d_H} \geq \frac{(V - p_L)(p_L - c)}{d_L} \text{ if } d = d_H, \tag{3.35}$$

which immediately leads to:

$$\frac{(V - p_L)(p_L - c)}{d_L} = \frac{(V - p_H)(p_H - c)}{d_H}. \tag{3.36}$$

For given $p_H = p^*$, there are precisely two p_L's that comply with the equation above, namely $\underline{p}$ and $\bar{p}$. Recall that in the TSE a p_L infinitely close to p^* was a possible low-damage price. Here, in the OSE under consideration, we must have that $p_L \in \{\underline{p}, \bar{p}\}$ and $p_H = p^*$. The main difference with the TSE is that the firm has to choose the price p_L far away from the price p_H.

Now consider the behavior of the EG in this equilibrium. Suppose that $x^* = x_L = x_H$. Out-of-equilibrium beliefs must be such that for any $x \neq x^*$:

$$\frac{V - p_L}{d_L} d_L + x^* \leq \frac{V - p_L}{\mu(x, p_L)} d_L + x \text{ if } d = d_L \tag{3.37}$$

$$\frac{V - p_H}{d_H} d_H + x^* \leq \frac{V - p_H}{\mu(x, p_H)} d_H + x \text{ if } d = d_H \tag{3.38}$$

Rewrite (3.37) as:

$$x^* - x \leq (V - p_L) \left(\frac{d_L}{\mu(x, p_L)} - 1 \right). \tag{3.39}$$

Observe that since $d_L \leq \mu(x, p_L)$, the RHS of this inequality is always smaller than or equal to zero. Therefore, $x^* \leq x$. If $x^* > 0$, then there is an $x < x^*$ such that (3.37) cannot hold. We must have $x^* \leq x$ for any $x \neq x^*$. Hence, I set $x^* = 0$.

Take $x^* = 0$. It is easily checked that, e.g., the following out-of-equilibrium beliefs satisfy (3.37) and (3.38): $\mu(x, p_L) = d_L$ and $\mu(x, p_H) = d_H$ for all

$x \neq x^*$. As long as $\mu(x, p_L)$ and $\mu(x, p_H)$ are irresponsive to increasing the cost of advertising, this discourages the EG to advertise.

Concluding, there are OSE in which $x_L = x_H = 0$, $p_L \in \{\underline{p}, \bar{p}\}$ and $p_H = p^*$. In Figure 3.1, these OSE are represented by the points ζ and δ. These equilibria are supported by out-of-equilibrium beliefs in which, in case of deviations of the firm, perceived damage is high enough and, in case of deviations of the EG, perceived damage is low enough. Note that in these OSE where the firm signals if damage is low, the firm has to differentiate its price by a sufficiently large margin instead of a margin as in the TSE.

One-sided separating equilibria where the EG informs

The discussion of OSE is continued by examining OSE in which the EG informs and the firm's strategy is to set the same price in each state of the world.

I start with the decision of the firm. We know that in this kind of equilibrium $p_L = p_H$. In fact, the price the firm will set must be p^* in both states of the world. Suppose on the contrary that $p_L = p_H = \hat{p} \neq p^*$. Take $d = d_H$. Then the deviation to p^* must not be profitable, and we must have:

$$\frac{(V - \hat{p})(\hat{p} - c)}{d_H} \geq \frac{(V - p^*)(p^* - c)}{\mu(x_H, p^*)}. \tag{3.40}$$

Since $\hat{p} \neq p^*$, we know that $(V - \hat{p})(\hat{p} - c) < (V - p^*)(p^* - c)$. But then it must be that $d_H < \mu(x_H, p^*)$ which is impossible. Hence $\hat{p} = p^*$. Take $\mu(x_H, p) = d_H$ for all $p \neq p^*$. These out-of-equilibrium beliefs ensure that it is never profitable for the firm to deviate from the equilibrium since $\frac{(V-p^*)(p^*-c)}{d_H}$ is the minimal profit the firm receives in equilibrium as well as the supremum of profit it could receive out-of-equilibrium. Similarly, $\mu(x_L, p) = d_H$ for all $p \neq p^*$ ensures that it is never profitable for the firm to deviate if $d = d_L$. Of course, these are extreme out-of-equilibrium beliefs and the equilibrium can be supported by more moderate ones.

Note that the analysis for finding x_L and x_H in the section describing the TSE (until the discussion of μ_{LH} and μ_{HL}) did not depend on the behavior of the firm. So, in equilibrium we still have $x_L = 0$ and $0 < x_H \leq R$, but no longer is each combination of x_L and x_H in this set an equilibrium. The reason is that in addition the EG must not be tempted to choose x_H if $d = d_L$

and x_L if $d = d_H$. For the EG these incentive constraints imply:

$$\frac{V-p^*}{d_L}d_L \leq \frac{V-p^*}{d_H}d_L + x_H \quad \text{if } d = d_L \tag{3.41}$$

$$\frac{V-p^*}{d_H}d_H + x_H \leq \frac{V-p^*}{d_L}d_H \text{ if } d = d_H, \tag{3.42}$$

which implies that:

$$x_H \in \left[\left(\frac{V-p^*}{d_L} - \frac{V-p^*}{d_H}\right)d_L, \left(\frac{V-p^*}{d_L} - \frac{V-p^*}{d_H}\right)d_H\right] \equiv [L,R]\,. \tag{3.43}$$

Observe that $R > L > 0$. Thus, in these OSE, the EG always advertises if $d = d_H$. The cost of advertising if the product is dirty must be sufficiently large; this ensures that if the product is clean, then the benefits of campaigning are too small.

Summarizing, in the present OSE the firm no longer needs to inform consumers and chooses its optimal price p^*. The EG now has to advertise with at least an intensity of L if damage is high whereas in the TSE it could advertise less. So, there are OSE in which $x_L = 0$, $x_H \in [L,R]$ and $p_L = p_H = p^*$. In Figure 3.1, these OSE are represented by the line $\alpha\beta$. These equilibria are supported by out-of-equilibrium beliefs in which, in case of deviations of the firm, perceived damage is high enough and, in case of deviations of the EG, perceived damage is low enough.

3.3.2 Pooling equilibria

In a pooling equilibrium neither the firm nor the EG provides the consumer with information. Consequently, in each state of the world they choose the same action.

Let $x_L = x_H = \hat{x}$ and $p_L = p_H = \hat{p}$. Then $\mu(\hat{p},\hat{x}) = \rho d_L + (1-\rho)d_H \equiv \bar{\mu}$. Let us first consider the incentives of the firm. Since profit does not depend on the true value of d directly, in a pooling equilibrium the constraint on out-of-equilibrium beliefs to rationalize the firm's action is:

$$\frac{(V-\hat{p})(\hat{p}-c)}{\bar{\mu}} \geq \frac{(V-p)(p-c)}{\mu(\hat{x},p)} \text{ for all } p \neq \hat{p}, \tag{3.44}$$

which can be rewritten as:

$$\mu(\hat{x},p) \geq \frac{(V-p)(p-c)}{(V-\hat{p})(\hat{p}-c)}\bar{\mu}. \tag{3.45}$$

Note that if the RHS of (3.45) exceeds d_H, then $\hat{p}$ cannot be part of an equilibrium. If $\hat{p} = p^*$, then the RHS is always smaller than or equal to $\bar{\mu} < d_H$. So, there is a pooling equilibrium in which $\hat{p} = p^*$.

There are also other pooling equilibria. Suppose $\hat{p} \neq p^*$ and the deviation is p^*. For $\hat{p}$ to be part of an equilibrium we must have:

$$\frac{(V - p^*)(p^* - c)}{(V - \hat{p})(\hat{p} - c)}\bar{\mu} \leq d_H, \tag{3.46}$$

i.e. the RHS of (3.45) should not exceed d_H for the (most tempting) deviation p^*. Let I be the interval for $\hat{p}$ implied by (3.46). If $\hat{p} \in I \subset [c, V]$, then it can be part of a pooling equilibrium. Note that $p^* \in I$.

Next, consider the incentives of the EG. For $\hat{x}$ to be an equilibrium, the following two constraints have to be satisfied for all $x \neq \hat{x}$:

$$\frac{V - \hat{p}}{\bar{\mu}}d_L + \hat{x} \leq \frac{V - \hat{p}}{\mu(x, \hat{p})}d_L + x \quad \text{if } d = d_L, \tag{3.47}$$

$$\frac{V - \hat{p}}{\bar{\mu}}d_H + \hat{x} \leq \frac{V - \hat{p}}{\mu(x, \hat{p})}d_H + x \quad \text{if } d = d_H. \tag{3.48}$$

To find the highest $\hat{x}$ that satisfies these constraints, let $\mu(x, \hat{p}) = d_L$ for all $x \neq \hat{x}$. These are the out-of-equilibrium beliefs that make defection the least attractive. The most tempting deviation is, given these beliefs, $x = 0$. Equations (3.47) and (3.48) can now be simplified to:

$$\hat{x} \leq (V - \hat{p})\left[\frac{d_L}{d_L} - \frac{d_L}{\bar{\mu}}\right], \tag{3.49}$$

$$\hat{x} \leq (V - \hat{p})\left[\frac{d_H}{d_L} - \frac{d_H}{\bar{\mu}}\right]. \tag{3.50}$$

Since $d_L < \bar{\mu} < d_H$, it follows that (3.49) is the most binding constraint. In the pooling equilibrium we must have $0 \leq x_L = x_H \leq (V - \hat{p})\left[1 - \frac{d_L}{\bar{\mu}}\right]$.

Observe that pooling equilibria in which the EG advertises ($x_L = x_H > 0$), are not Pareto efficient. Comparing a pooling equilibrium in which the EG advertises to a pooling equilibrium in which the EG does not advertise, we see that the cost of the EG is lowest when the EG does not advertise whereas the profit of the firm is the same in both equilibria. This also reinforces the intuition that non-informative advertising is a waste of money.

Further examining the pooling equilibria in which the EG does not advertise ($x_L = x_H = 0$), it is clear that the out-of-equilibrium beliefs that support these equilibria are strange. In order to see this, consider a pooling equilibrium in which the EG does not advertise. If damage turns out to be high, then for the EG the possible gains of advertising are high compared to the gains of advertising if damage is low. But the out-of-equilibrium beliefs that sustain this equilibrium have to assign a higher probability to low damage than to high damage since the EG has to be punished for a deviation (recall that we take $\mu(x, \hat{p}) = d_L$ for all $x \neq \hat{x}$). Now consider an equilibrium refinement that assigns a higher probability to states of the world in which the possible gains are higher for the sender that deviates.[4] It is clear that this pooling equilibrium would not survive this equilibrium refinement.

3.3.3 Summary of equilibria

The possible equilibria are:

(a) Two-sided separating equilibria in which $x_L = 0$, $(V - p_L)[1 - f(p_L)] < x_H \leq (V - p^*)[1 - f(p_L)]/f(p_L)$, $p_L \in (c + \frac{d_L}{d_H}(p^* - c), \bar{p}) \backslash \{p^*\}$ and $p_H = p^*$,

(b) One-sided separating equilibria in which $x_L = 0$, $x_H = 0$, $p_L \in \{\underline{p}, \bar{p}\}$ and $p_H = p^*$,

(c) One-sided separating equilibria in which $x_L = 0$, $x_H \in [L, R]$, $p_L = p^*$ and $p_H = p^*$,

(d) Pooling equilibria in which $0 \leq x_L = x_H \leq (V - p_L)\left[1 - \frac{d_L}{\bar{\mu}}\right]$ and $p_L = p_H \in I$.

The set of equilibria can be greatly reduced by only considering equilibria which are not Pareto dominated by any other equilibrium. Observe that the profit of the firm in equilibrium does not depend directly on x and that the EG's objective is directly increasing in x. Therefore, in each of the four equilibrium types, I can take x_L and x_H as small as possible.

Another consideration is that the EG's cost is decreasing in price. Since a higher price implies a lower demand for the product, total environmental damage is decreasing in price. Observe that the sets $(c + \frac{d_L}{d_H}(p^* - c), \bar{p}) \backslash \{p^*\}$

[4] The (standard) Intuitive Criterion (Cho and Kreps, 1987) is similar in spirit but is weaker. Suppose a deviation is observed. Furthermore, suppose that under the most favorable belief for the deviator (e.g. either the firm or the EG), it is not optimal to deviate in this manner in a certain state of the world. Then this deviation cannot be rationalized. In the Intuitive Criterion, the receiver assigns zero probability to states of the world for which an observed deviation cannot be rationalized. This criterion is useless here since potentially in both states of the world the EG can profit from a deviation.

and I are symmetric around p^* in the following manner: for every $p < p^*$ that yields a certain profit to the firm, they also contain a $p' > p^*$ that yields the same profit.[5] Since the EG's cost is decreasing in price and the firm is indifferent, it makes sense to consider only $p_L > p^*$. Note that after eliminating the Pareto-inefficient equilibria, the clean product is higher priced than the dirty product.

This leaves us with the following equilibria:[6]

(i) Two-sided separating equilibria in which $x_L = 0$, $x_H = (V - p_L)[1 - f(p_L)]$, $p_L \in (p^*, \bar{p})$ and $p_H = p^*$,

(ii) One-sided separating equilibrium in which $x_L = 0$, $x_H = 0$, $p_L = \bar{p}$ and $p_H = p^*$,

(iii) One-sided separating equilibrium in which $x_L = 0$, $x_H = L$, $p_L = p^*$ and $p_H = p^*$,

(iv) Pooling equilibria in which $x_L = 0$, $x_H = 0$ and $p_L = p_H \in \{p|p \in I \wedge p \geq p^*\}$.

Since I am interested in the effect of advertising by EGs, the OSE in which the EG informs the consumers (i.e. $x_L = 0$, $x_H = L$, $p_L = p^*$ and $p_H = p^*$) is the obvious equilibrium to explore.[7] As a benchmark I will use the case in which it is impossible to advertise (i.e. $x_L = x_H = 0$). Possible benchmark equilibria formally coincide with equilibria derived above in which the EG does not inform the consumers, i.e. the OSE separating equilibrium in which the EG does not inform and the pooling equilibria. To see the full effect of information transmission by the EG I focus on the pooling equilibria. In particular, the pooling equilibrium in which $p_L = p_H = p^*$ will be chosen. There are two reasons for doing this. First, for the firm, the pooling equilibrium with $p_L = p_H = p^*$ clearly dominates the other pooling equilibria in terms of equilibrium payoff.[8] Second, the OSE in which $x_L = 0$, $x_H = 0$, $p_L = \bar{p}$ and $p_H = p^*$ is also dominated by this pooling equilibrium since in the OSE in question the firm's

[5] Since $p^* = \frac{1}{2}V + \frac{1}{2}c$ is in the middle of the interval $[c, V]$ and the profit function is symmetric in the sense that $\Pi(p^* - \epsilon) = \Pi(p^* + \epsilon)$, we have that $p' \in [c, V]$.

[6] The separating equilibria that are not Pareto dominated by another equilibrium coincide with the curve $\beta\delta$ in Figure 3.1.

[7] It may seem arbitrary to ignore the TSE in which the EG also advertises, but these equilibria can be easily dismissed with equilibrium refinements. For instance, Bagwell and Ramey (1991)'s unprejudiced equilibrium refinement rules out any TSE.

[8] Observe that in the benchmark the EG cannot advertise and, therefore, I do not include the EG into these considerations.

profit is $\frac{(V-p^*)(p^*-c)}{d_H}$ (cf. 3.36) in both states of the world whereas the firm's profit in the pooling equilibrium with $p_L = p_H = p^*$ is $\frac{(V-p^*)(p^*-c)}{\bar{\mu}}$. Hence, the obvious equilibrium to explore for the benchmark case of no information transmission by the EG is the pooling equilibrium where the price is p^*.

3.3.4 The equilibrium with advertising versus the benchmark

In this section the OSE with advertising will be compared with the benchmark of no advertising. In particular, I will compare the *ex ante* cost of the EG and profit of the firm. Let me start with the equilibrium with advertising. If damage is low, then the total cost of the EG in the equilibrium is:

$$\Delta_L = \frac{V-p^*}{d_L} d_L, \tag{3.51}$$

and the profit of the firm is:

$$\Pi_L = \frac{(V-p^*)(p^*-c)}{d_L}. \tag{3.52}$$

Similarly, if damage is high, then the total cost of the EG in the equilibrium is:

$$\Delta_H = \frac{V-p^*}{d_H} d_H + L, \tag{3.53}$$

and the profit of the firm is:

$$\Pi_H = \frac{(V-p^*)(p^*-c)}{d_H}. \tag{3.54}$$

Thus, for the equilibrium with advertising, the *ex ante* expected total cost and profit are:

$$\Delta_{ads} = \rho \frac{V-p^*}{d_L} d_L + (1-\rho)\frac{V-p^*}{d_H} d_H + (1-\rho)L, \tag{3.55}$$

$$\Pi_{ads} = \rho \frac{(V-p^*)(p^*-c)}{d_L} + (1-\rho)\frac{(V-p^*)(p^*-c)}{d_H}. \tag{3.56}$$

In the benchmark equilibrium, we have:

$$\Delta_{bench} = \frac{V-p^*}{\rho d_L + (1-\rho)d_H}(\rho d_L + (1-\rho)d_H) \tag{3.57}$$

and

$$\Pi_{bench} = \frac{(V-p^*)(p^*-c)}{\rho d_L + (1-\rho)d_H}. \tag{3.58}$$

Remark that $\Delta_{ads} = (V - p^*) + (1 - \rho)L$ and $\Delta_{bench} = (V - p^*)$. Hence $\Delta_{bench} < \Delta_{ads}$, i.e. the cost of the EG increases when it is able to advertise.

The firm's *ex ante* profit is always higher in the equilibrium with advertising compared to the benchmark equilibrium. To see this, note that profit is convex in μ. Applying Jensen's inequality, we obtain:

$$\begin{aligned}\Pi_{bench} &= \frac{(V - p^*)(p^* - c)}{\rho d_L + (1 - \rho) d_H} \\ &< \rho \frac{(V - p^*)(p^* - c)}{d_L} + (1 - \rho) \frac{(V - p^*)(p^* - c)}{d_H} = \Pi_{ads}. \quad (3.59)\end{aligned}$$

Note that the low-damage firm prefers the advertising equilibrium over the benchmark equilibrium since in the former equilibrium the firm sells the good at the same price as in the latter but the consumer values the low-damage good higher. This results in more demand. For the high-damage firm the opposite occurs. However, the gain of the low-damage firm is higher than the loss of the high-damage firm, and on average the firm benefits from advertising. Regarding consumers' surplus, note that in Chapter 2 we derived that in this case consumers' surplus is exactly half of profit. Hence, *ex ante* the consumers also prefer the advertising equilibrium.

Concluding, the EG can transmit information by advertising if damage is high, but it does not benefit the EG. The firm, however, does benefit from the information transmission by the EG. While the EG transmits information, it is its adversary, the firm, that profits from it. The reason for this is that the goal of the EG is not to inform the consumers *per se*. On average and net of advertising cost, the EG is indifferent between having informed or uninformed consumers since $q(p^*, d_L)d_L = q(p^*, d_H)d_H = q(p^*, \mu)\mu$ is constant in this model. Ideally the EG wants to delude the consumers: to make them believe that damage is higher than it actually is. This is precisely what happens in the benchmark case to uninformed consumers if damage is low. However, those same uninformed consumers believe the product is fairly clean when damage is high. This is the point at which the EG starts to advertise if advertising is possible. Consequently, the EG is now paying to inform consumers, but it is, net of advertising cost, indifferent between having informed and uninformed consumers.

3.4 Social Welfare

In this section I will perform an analysis of social welfare. We have already seen in the previous section that both the firm and consumers *ex ante* prefer the advertising equilibrium over the benchmark equilibrium whereas the EG prefers the benchmark equilibrium. In order to determine which effect dominates, social welfare will be examined. I define social welfare as:

$$SW_i = \Pi_i + CS_i - \gamma\Delta_i, \tag{3.60}$$

where $i \in \{\text{ads}, \text{bench}\}$ and CS_i denotes consumers' surplus. Recall the role of the γ-parameter as discussed in Footnote 3: its role is to scale the cost of the EG and hence I will include it here. Since $CS_i = \frac{1}{2}\Pi_i$ we have:

$$SW_i = \frac{3}{2}\Pi_i - \gamma\Delta_i. \tag{3.61}$$

Since $\Pi_{bench} > \Pi_{ads}$ and $\Delta_{bench} > \Delta_{ads}$, it is immediately clear that by taking γ small enough the advertising equilibrium will be preferred and by taking γ large enough the benchmark equilibrium will be preferred. So, I will again set γ equal to 1 and determine the effect of the other parameters, i.e. ρ, d_L, d_H, V and c. Define $D = SW_{ads} - SW_{bench}$ as the difference in social welfare between the advertising equilibrium and the benchmark equilibrium. So if $D > 0$, then advertising raises social welfare. After some tedious, but straightforward algebra, we obtain:

$$D = \frac{V-c}{4} \times \left[\frac{3}{2}\left(\frac{\rho}{d_L} + \frac{1-\rho}{d_H}\right) - 2\left(1 - \frac{d_L}{d_H}\right) - \frac{3}{2(\rho d_L + (1-\rho)d_H)}\right]. \tag{3.62}$$

Note that V and c have no effect on the sign of D. Also, if ρ is either 0 or 1, then $D < 0$. This is expected since in these two cases there is no uncertainty about damage. Let $\hat{\rho}$ be such that $D = 0$. It can be shown that:

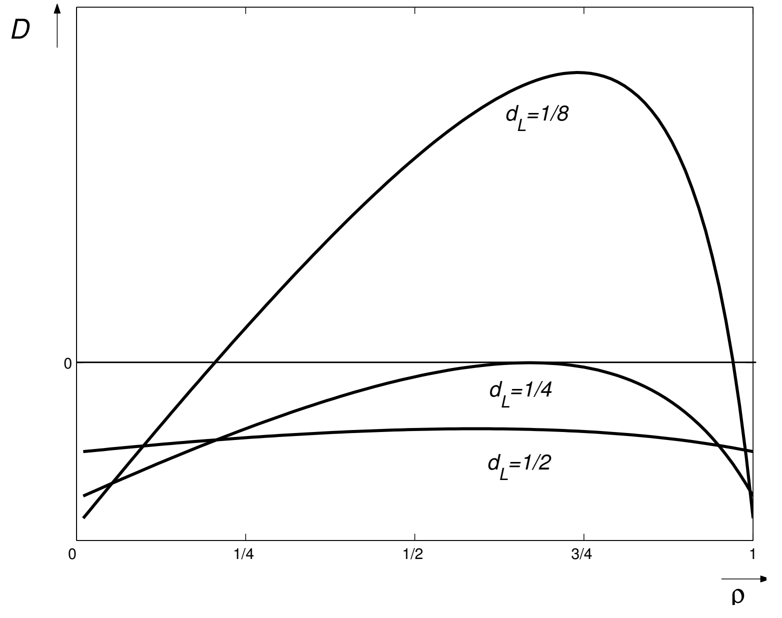
$$\hat{\rho} = \frac{1}{6}(3 + 4d_L) \pm \frac{1}{6}\sqrt{\frac{(d_H - d_L)(3 + 4d_L)^2 - 48d_Ld_H}{(d_H - d_L)^2}}. \tag{3.63}$$

Observe that, given d_L and d_H, these two roots are either imaginary (and in that case $D < 0$ for all ρ) or both roots are in the interval $[0, 1]$. Since the roots should be real for there to be a situation in which there exist ρ such that $D > 0$, we need that:

$$(d_H - d_L)(3 + 4d_L)^2 - 48d_Ld_H > 0. \tag{3.64}$$

For instance, if $d_H = 1$, it can be shown that this inequality holds if $d_L < \frac{1}{4}$. Hence, the dirty good has to cause at least four times more damage per unit of

Figure 3.2: The difference in social welfare ($d_H = 1$ and $\gamma = 1$)



production such that, for at least some value of ρ, the advertising equilibrium is better than the benchmark equilibrium in terms of social welfare. Numerical calculations suggest that in general d_L has to be small compared to d_H for the advertising equilibrium to be socially optimal. Figure 3.2 plots D for different values of d_L (and $d_H = 1$). This figure shows two things. First, the maximal D is reached at $\rho > \frac{1}{2}$, i.e. it should be more likely that the product is dirty than clean. Second, it shows that the interval of ρ's for which $D > 0$ can be quite large once the ratio d_L/d_H is small enough.

Concluding, if it is almost certain that the good is clean or dirty, then the benchmark equilibrium is socially optimal. Only if the clean good is very clean relative to the dirty good, can it be possible that advertising is socially optimal. And if the clean good is clean enough for this to happen, then there must also be a higher probability that the good is dirty. This (partial) numerical analysis also shows that the losses of the EG can be substantial and that advertising can decrease social welfare.

3.5 Choosing damage: the long run versus the short run

So far I have assumed that the firm cannot influence the amount of damage per unit of production of the good. In the short run this might be plausible, but in the long run the firm should be able to choose damage per unit of production itself. In the short run, the firm might unexpectedly face consumers that have newly become aware of the fact that the product they consume causes harm to the environment. As an example one could think of a material like asbestos, whose detrimental effects were only discovered after the material had been used for several years. When the firm started the production of the good, its choice of d was based on other considerations than the consumers' attitudes about the environment. In that case it is reasonable to assume that the consumers attach an *a priori* probability that the good causes either a low damage or a high damage. However, in the long run, e.g. when the firm builds a new plant, the environmental awareness of the consumers may be an important issue. Consequently, in a perfect Bayesian Nash equilibrium the consumers must have expectations about the level of damage that is consistent with the decision of the firm concerning the level of damage per unit of production. As we will see, this drastically changes the role of the EG.

So, I allow the firm to choose $d \in \{d_L, d_H\}$ in period 0 instead of Nature. For the remainder of this Section I will assume that if the firm is indifferent between choosing d_L and d_H, then it will choose d_H. While this will seem to drive the result, one should remember that if the marginal cost of producing the clean good would be infinitesimally higher than the marginal cost of producing the dirty good, then the firm would prefer d_H.

In absence of an EG, the firm must choose d_H. The reason for this is almost trivial. Suppose the firm mixes its strategy in period 0 and with probability ρ chooses low damage. Given this choice, the analysis in Section 3.3 shows the possible equilibria of the subsequent game. Observe that in all equilibria in which $x_L = x_H = 0$, the firm's profit does not depend on the actual value of d. So in period 0, the firm is indifferent between d_L and d_H and by assumption will choose d_H.

Observe that in a perfect Bayesian Nash equilibrium, the consumers' belief about which damage level the firm chooses must be consistent with the firm's choice. Hence, if the firm chooses d_H, then the consumers also believe that damage is d_H. Consequently, the profit of the firm in this equilibrium is

$\frac{(V-p^*)(p^*-c)}{d_H}$.

Now suppose that the EG can advertise. I will show that the EG can without cost force the firm to choose d_L. Consider the following strategy by the EG: $x(d_L) = 0$ and $x(d_H) = L$. Furthermore, suppose that the strategy of the firm is $p(d_L) = p(d_H) = p^*$ and choose d_L with probability ρ in period 0. Also the beliefs of the consumers given p^* are:

$$\mu(x, p^*) = d_L \quad \text{if } x < L, \tag{3.65}$$

$$\mu(x, p^*) = d_H \quad \text{if } x \geq L. \tag{3.66}$$

The analysis in Section 3.3 shows that conditional on the firm's choice of ρ this is a OSE where the EG informs. Regarding the firm's decision of ρ in period 0, the following considerations must be taken into account. If the firm chooses d_L, then its profit will be $\frac{(V-p^*)(p^*-c)}{d_L}$. And its profit will be $\frac{(V-p^*)(p^*-c)}{d_H}$ if it chooses d_H. Observe that its profit will always be higher if the firm chooses d_L. Hence, it sets $\rho = 1$ and the product will be clean. Moreover, since $d = d_L$, the EG never actually advertises.

The difference with the short-run analysis is that now the advertising equilibrium is a Pareto improvement over the benchmark equilibrium with no advertising. The firm and the consumers still benefit from advertising by an EG, but since the EG only threatens to advertise, it incurs no losses. Hence the presence of an EG increases social welfare.

3.6 Conclusion

In this Chapter I studied informative advertising by an EG. A monopolist sold a good which causes a certain damage per unit of production to the environment. The consumers care about this damage, but are unable to discern the level of damage. An EG subsequently uses an advertising campaign to disseminate information about product quality.

First, I considered the case in which the firm did not choose the damage level of the good that it produced. In this case, I find that the EG is able to transmit information about the level of damage to the consumer. While there are equilibria in which the firm informs the consumers, in absence of an EG, the pooling equilibrium in which the firm does not inform the consumers dominates equilibria in which the firm does inform the consumers.[9] I conclude

[9] Moreover, equilibria in which the firm informs the consumers only exist if the marginal cost of producing a clean good are lower than or equal to the marginal cost of producing a

that, unlike the EG, the firm cannot inform consumers. However, the ability of the EG to inform consumers is not beneficial for the EG. I showed that if the EG could commit to a silent strategy this would improve its situation. The EG is hurt by its own power. A further result is that the EG will punish the bad firm by advertising only if damage is high and not praise the good firm although this is in principle possible. This result follows here directly from the incentive constraints. The social welfare results are less clear: unlike the EG, the firms and the consumers on average benefit from advertising. However, numerical calculations suggest that often the gains of the firms and the consumers are smaller than the loss of the EG.

Second, I consider the case in which the firm was able to choose damage and, hence, the firm could possibly preempt. If the firm can choose damage before the EG can advertise, then advertising is effective; in equilibrium the EG only threatens to advertise and helps the firm maintain a low level of damage. This equilibrium is comparable to the 'entry deterrence'-equilibrium as discussed in Chapter 2.

A natural interpretation of being able to choose damage (or not) is the long run versus the short run. I have shown that for the EG the long-run cost, unlike the short-run cost, is improved by advertising. Moreover, profit and consumers' surplus are not harmed by advertising. The interpretation would be that pressure by an environmental group, while harmful in the short run, may in the long run push society in the right direction. A policy implication for the short run is that while the EG can transmit information, it is probably not the most efficient way to achieve this goal. Regulation (i.e. required labeling, taxation of dirtier firms) is probably the way to solve this information problem since the government can rely on the legal system to enforce its credibility.

The implications in the short run for the policy of the EG itself seem dour: the EG cannot engage in campaigns aimed at supplying consumers with information and be better off. However, since the cost of campaigning is created by the EG's lack of credibility, a possible solution might be the building of a reputation. If the claims of an EG are occasionally verified by a trustworthy authority, then the EG could possibly inform consumers at substantially lower cost. The results in this Chapter stress the need for reputation building.

---

dirty good. The assumption that dirty products are cheaper to produce than clean goods is a standard one in the literature.