The Credibility of Certification

By Philippe Mahenc*

September 25, 2009

Abstract

Certification is meant to solve the adverse selection problem raised by unsubstantiated claims from firms about environmental safety. Things are getting complicated by the suspicion of uninformed consumers that the certification agency might manipulate their beliefs. To make her certification credible, a government agency must prove she is not opportunist. This entails a signaling cost: certification fees for green firms must be distorted upward relative to the Ramsey benchmark level. Certification is shown to be credible only if the agency's cost of

*LERNA-INRA, Toulouse School of Economics, Manufacture-bât F, 21 Allée de Brienne 31 000 Toulouse, France. University of Perpignan. Email: mahenc@univ-perp.fr.

collecting information is sufficiently high. Otherwise, fees are pooling brown and green firms and certification discloses no information.

Keywords: Greenwashing, Certification, Signaling.

JEL Code: D82, H21, H23, Q28.

The Credibility of Certification

1 Introduction

Antoine-Augustin Parmentier remained famous for the astute means by which he managed to signal that the potato, a new good unknown to Frenchmen in the XVIIIth century, was edible and tasty. As his claims in favour of the root were persistently ignored, Parmentier thought up the following device with the help of Louis XVI's government. An armed guard worthy of the most precious diamond was set over a potato patch in broad daylight. After dark, the guard was withdrawn to let the people steal potatoes with impunity and grow them in their own garden plots. Thanks to this trick, potatoes were successfully introduced into France in 1787.

This is an early example of mechanism used by an informed and benevolent person to credibly signal the social value of a good to misinformed people. What is instructive in this example is that people can infer the actual value of the good from the observed government's behavior only because it is biased away from what should be efficient from the social standpoint. Clearly, so much fuss about ordinary potatoes is wasteful. Such a waste however is necessary to simultaneously prove that the potato is valuable and Parmentier's claim is credible.

Of central concern of this paper is how the optimal behavior of a benevolent agency in charge of certification is affected by consumers' imperfect knowledge of the environmental safety of a good. It is shown that the credibility of certification always entails some welfare loss, as illustrated by Parmentier's trick. Certification fees are charged by the agency to award the label "brown" or "green" to firms producing the potentially harmful good. Not only are these fees used to raise revenue and finance the agency's cost of collecting information, but they also play the role of a signal for environmental safety. The main result is that, in the unique "reasonable" separating equilibrium, the fee signaling green firms is biased upward relative to the Ramsey fee, that is, the fee aimed at raising revenue with a minimal loss in terms of efficiency. This distortion is necessary to turn the agency's claim into a credible certification and fully reveal information on environmental safety to consumers. The distortion proves that the agency is trustworthy because, by restricting too much consumption, it creates an efficiency loss which would be worthless if the agency were opportunistic.

Opportunism is defined here as the temptation to manipulate consumers' beliefs. Usually, it is attributed to non-competitive firms by the literature of industrial organization concerned with the various means of signaling the quality of experience goods (see the seminal articles by Milgrom and Roberts (1982 and 1986)). Firms opportunism is the source of adverse selection problems and greenwashing practices[1] which lead rational consumers to view claims by firms suspiciously. Thanks to Spence (1974), the signaling literature has long recognized that firms often transmit hidden information about quality through wasteful behaviors such as upward- or downward-biased prices, burned money or any observable expenditure. This paper argues that the provision of credible information by a third-party certification agency may also generate distortions for signaling reason. Consumers' lack of information vindicates the presence of a certification agency in the economy: information is a public good, which provides a rationale for government intervention[2]. Nevertheless, if consumers mistrust firms, there is no

[1] Greenwashing refers to the opportunism of firms that benefit from a rent of information about the harm from the pollution they generate, at the expense of imperfectly informed consumers. Greenwashing encompasses all practices that range from vague claims to misleading advertising about the environmental performance of firms. Cason and Gangadharan (2002) attribute greenwashing to some laxity in the regulation of environmental claims. Some evidence that greenwashing is becoming widespread in the U. S. can be found in the growing number of complaints about green ads received by the Advertising Standard Authority.

[2] Thoughout the world, third-party certification of environmental quality is not always handled by the government. The EU Ecolabel is adopted by the Commission as part of a broader action plan on sustainability. In Japan, the Eco Mark program is operated by

reason why they should give up all suspicion vis-à-vis a government agency and blindly rely on the labels she awards.

Certification agencies generally belong to the public bureaucracy through which government funds are channeled. Due to their monopoly over the provision of public services, some bureaus may be tempted by opportunistic behaviors. This idea was primarily promoted by Niskanen (1971) who argues that bureaucratic control over information gives the bureaucrats discretion to pursue goals other than the maximization of social welfare. A recent instance of bureaucratic opportunism can be found in Viscusi and Hamilton (1999). These authors provide convincing evidence that risk regulators often reason on the basis of people's perception about risks rather than the actual risks. In the terminology of Salanié and Treich (2006), a regulator who maximizes welfare computed with the consumers' beliefs is called "populist". This will be precisely what consumers have in mind here, when they suspect the agency of opportunism. Their suspicion is motivated by the agency's care for the budget size: she might be tempted to manipulate beliefs in order to generate revenue in excess of the amount required by her certification task and divert

---

the Japan Environment Association which cooperates with the Ministry of Environment. The certification mark Green Seal in the US is awarded by an independent non-profit organization.

the extra resources to some other use.

The mere consequence of consumers' suspicion is that the agency ought to prove she is trustworthy while disclosing information through labels, otherwise consumers will ignore them. This requires the agency's strategies to be credible in the sense that the fee specified for one label environmental safety would not be worth imitating by an opportunistic agency if the environmental safety were different. Such a requirement is fairly similar to the traditional incentive-compatibility constraints imposed on firms' behavior in the literature on quality signaling. One essential departure here is that the certification agency plays the role of the signal sender instead of firms. The signal discloses information on environmental safety through the fee encompassed within the consumer price. The firm price definitely cannot be a signal because firms are competitive, as pointed by Spence (1977).

The modeling of certification fees as a signal is one response to recent works on ecolabeling which leaves it to further research to allow for asymmetric information between firms and consumers regarding environmental qualities of firms products (see Amacher *et alii* (2004), one exception being Mason (2006); see also Kuhn (2005) for a recent survey on labeling). A first result is that separation between brown and green firms cannot be achieved by fees when the cost of collecting information is too low. In such a case, the social opportunity cost of the revenue raised from fees is low too. The reason for the failure in revealing information is that the agency incurs the same signaling costs for both types of environmental safety, hence she cannot resist the temptation to charge brown firms the same fee as that intended for green firms. As a result, certification cannot be credible in these circumstances. In contrast, certification is credible provided that collecting information is sufficiently costly. Then, the agency's behavior tends to resemble that of a profit-maximizing monopolist on the market for certification. As the foregone revenue of increased fees is less when firms are green than when they are brown, separation of the two types can be effective with high fees.

A prominent result is that the fee intended for guaranteeing the credibility of the green label never coincides with the Ramsey fee. In fact, the presence of suspicious consumers induces the agency to charge green firms a fee higher than that for brown firms, contrary to what would occur if consumers were trustful. The analysis characterizes separating equilibria in which consumers must pay an extra premium when purchasing the good from green firms. This is the price paid for credibility as illustrated by Parmentier's story. As consumers exert a negative externality on the agency by free riding on her information, the role of the distortion *à la* Parmentier is to force them to internalize this externality in much the same way that the Pigovian tax leads polluters to internalize correctly the costs of pollution.

The paper is organized as follows. Section 2 sets out the structure of the model. Section 3 states the results in the benchmark case where consumers are trustful and the credibility of certification is not an issue. Section 4 investigates the signaling model designed to address this issue and characterizes separating and pooling equilibria satisfying the "intuitive criterion" of Cho and Kreps (1987). Section 5 offers conclusions.

2 The model

Consider a horizontally differentiated market structure similar to a Hotelling model with one good, where consumers differ in their tastes for the good. The good is located at 0, consumer taste is represented by distance x from 0 and x is uniformly distributed along a segment of unit length. Producing the good is risky, either for the environment or health, and the source of consumers' heterogeneity is the good safety. The harm caused by the good will be called pollution for concreteness. Following this interpretation, the transportation cost borne by each consumer resembles his personal harm from pollution, measured by a pure monetary loss. Another interpretation is that taste heterogeneity reflects the degree of social conscience of consumers. If, for instance, the good is fossil energy, consumers may differ in their dislike of the negative impact on global warming, and if it is nuclear energy, they may differ in their dislike of the potential risks imposed on future generations by nuclear repositories.

To handle the problem of adverse selection due to greenwashing, it will be assumed that consumers are not accurately informed about the level of transportation costs. Such an assumption captures the idea that environmental safety is an experience attribute of the good, in the sense that consumers have not perfect knowledge about the harm they experience. It is consistent with the observation made by Karl and Orwat (2000) that the individual costs of ensuring the environmental characteristics of goods are likely to be prohibitive for consumers. The good provides consumers with the same gross surplus of value v.

Potentially, there are two varieties of the good on the market: either a brown variety ($i = b$) or a green one ($i = g$), which are produced using respectively a dirty technology associated with a low level of environmental safety, and a clean technology providing a high level of environmental safety. Consumers are assumed to have the same aversion to pollution, which is modelled as the transportation cost ε_i per unit of distance to variety i. Hence, all consumers agree to rank the brown variety and its green counterpart in the same way with respect to environmental safety, i. e., $\varepsilon_b > \varepsilon_g$. Consumer x's personal harm from the pollution generated by variety i is measured by the linear transportation cost $\varepsilon_i x$ to variety i.

The market area is determined by market prices and consumer willingness to pay. Consumers purchase at most one unit of the good and get zero surplus if they do not buy. A consumer located at x derives a surplus $v - p_i - \varepsilon_i x$ from purchasing variety i at price p_i. Under complete information, the market area X solves equation:

$$v - p_i - \varepsilon_i X = 0, \tag{1}$$

and the demand for variety i is given by:

$$X\left(p_i, \varepsilon_i\right) \equiv \frac{v - p_i}{\varepsilon_i}. \tag{2}$$

Hence, we have both that consumers suffer from environmental degradation and it does affect the way they behave. This sounds realistic even though it departs from much of the literature on externalities that traditionally assumes consumers' behavior to be independent of the environment. As argued by Sandmo (2000), "... air and water pollution influences working conditions, peoples' choice of residential area, their pattern of leisure activities... ", which gives rise to the environmental feedback on demand captured by the Hotelling approach used here. Moreover, it is worth mentioning that, under complete information, the environmental externality is fully internalized by consumers' behaviors. Thus, besides asymmetric information, there is no rationale for government intervention. This modelling allows to focus on the provision of a single public good: information.

The social welfare function is the sum of the consumer surplus and the global harm (the sum of transportation costs) caused by pollution. Under complete information, this welfare function is given by:

$$\int_0^{X(p_i,\varepsilon_i)} (v - p_i - \varepsilon_i x)\, dx = \frac{(v - p_i)^2}{2\varepsilon_i}. \tag{3}$$

The good is produced by competitive firms at a marginal cost $c(\varepsilon_i)$ that depends upon the environmental safety, but not upon the quantity produced. The technology (production plus pollution abatement) required to produce the clean variety is more costly than that used for the brown variety, and so $c(\varepsilon_b) < c(\varepsilon_g)^3$. Moreover, to eliminate corner solutions, v will be taken in the parameter configuration such that $c(\varepsilon_g) < v < \min\{c(\varepsilon_g) + \varepsilon_g, c(\varepsilon_b) + \varepsilon_b\}$. These inequalities guarantee, first, that producing either variety is socially efficient; and second, that the market is never fully covered.[4]

Among various activities, a certification agency is responsible for providing information about the actual harm from pollution and the environmental safety of the good. Once the agency has collected full information on ε_i, she can award firms labels that certify environmental safety. Learning the true harm from pollution is assumed to be prohibitively costly to consumers, whereas the agency can secure full information about ε_i at a fixed cost I. Moreover, the agency is able to infer $c(\varepsilon_i)$ from the observed ε_i, thereby sharing the same information as firms about their production costs, hence environmental safety. To finance the cost I, the agency charges a specific (per-unit) fee t_i on variety i, which is a common practice among public agencies (see Mason (2006) and Crespi and Marette (2001)). Certification directly

[3]The statement that there is a trade-off between environmental improvements and firms efficiency is consistent with the conclusions of Palmer, Oates and Portney (1995) or Jorgenson and Wilcoxen (1990) for the U. S. economy.

[4]Indeed, $v - c(\varepsilon_b) > v - c(\varepsilon_g) > 0$ implies $X(c(\varepsilon_i), \varepsilon_i) > 0, i = b, g$. Furthermore, $v < c(\varepsilon_g) + \varepsilon_g$ and $v < c(\varepsilon_b) + \varepsilon_b$ imply $X(c(\varepsilon_g), \varepsilon_g) < 1$ and $X(c(\varepsilon_b), \varepsilon_b) < 1$, respectively.

follows from the fee choice, that is, the label "brown" is associated with t_b and the label "green" with t_g. Hence, the label and the fee charged for must be consistent, or, to put it differently, the fee turns the costless signal of a green claim into the costly signal of a label.

Because the market is competitive, firms in equilibrium will supply variety i subject to a zero profit constraint $p_i = c(\varepsilon_i) + t_i$. It will be assumed that consumers observe this equilibrium price as a whole and have no way of isolating the producer price $c(\varepsilon_i)$. Although consumers cannot directly observe neither ε_i nor c_i, they have formed beliefs about ε_i before making their purchase decision. Consumers perceive the good to be green with the prior probability $\mu_0 \equiv prob(\varepsilon_i = \varepsilon_g)$, and brown with the prior probability $1 - \mu_0 \equiv prob(\varepsilon_i = \varepsilon_b)$, $\mu_0 \in (0, 1)$. Having observed the consumer price $p_i = c(\varepsilon_i) + t_i$, consumers will revise their prior estimate of environmental safety in an attempt to infer ε_i. Consumers' posterior beliefs will be denoted by $\mu(t_i) : \mathbb{R}^+ \rightarrow [0, 1]$ giving the probability weight consumers attache to the possibility that the good is green after observing $p_i = c(\varepsilon_i) + t_i$. If $\mu \equiv \mu(t_i)$, then $\varepsilon^e(\mu) \equiv \mu\varepsilon_g + (1 - \mu)\varepsilon_b$ is the perception that consumers have of the environmental safety after observing the consumer price.

Suppose that consumers get an accurate perception of the environmental safety from the observed p_i. Anticipating this, the agency aims to maximize social welfare subject to the constraint that the revenue $R_i(t_i) \equiv t_i X(c(\varepsilon_i) + t_i, \varepsilon_i)$ raised from fees covers all her expenditures, that is, not only the cost I of collecting information, but also any additional revenue requirement such as clean-up programs, transfers to special-interest groups and other "perquisites". The part of funds diverted from collecting information about ε_i is normalized to zero without loss of generality. It suffices to keep in mind that the budget size is valuable to the agency even in the case where $I = 0$.

Let us introduce the following notations[5]:

- $V(\varepsilon_i) \equiv v - c(\varepsilon_i)$ is the consumer surplus at equilibrium price in the absence of tax, and so $R_i(t_i) = t_i \frac{V(\varepsilon_i) - t_i}{\varepsilon_i} - I$.
- $\underline{t}_i$ and $\bar{t}_i$ are the lowest and highest fee for which the agency breaks even when consumers perceptions are accurate, i. e., $\underline{t}_i$ and $\bar{t}_i$ solve

$$R_i(t_i) = 0. \tag{4}$$

[5] Here and throughout, subscripts denote partial derivatives and primes denote derivatives with respect to a single variable

To ensure the existence of $\bar{t}_i$ whatever the environmental safety, we will restrict the parameters of the model to satisfy the following assumption

$$I \leq \frac{V\left(\varepsilon_g\right)^2}{4\varepsilon_b} \tag{5}$$

Note that $R_i'\left(t_i\right) = \frac{V(\varepsilon_i)-2t_i}{\varepsilon_i} > 0$ for all $t_i < \frac{V(\varepsilon_i)}{2}$, which rules out any Laffer effect for fees lower than $\frac{V(\varepsilon_i)}{2}$, a common assumption in the literature. As will appear in the remainder of the analysis, there is no need to make such a restriction here.

- For a given price p, $\eta \equiv -X_p\left(p, \varepsilon_i\right) p / X\left(p, \varepsilon_i\right) = \frac{p}{v-p}$ will denote the price elasticity of demand.

- For a given fee t, $\eta^e \equiv \left[X_p\left(c\left(\varepsilon_i\right)+t, \varepsilon_i\right) c'\left(\varepsilon\right) + X_\varepsilon\left(c\left(\varepsilon_i\right)+t, \varepsilon_i\right)\right] \varepsilon_i / X\left(c\left(\varepsilon_i\right)+t, \varepsilon_i\right)$ will denote the pollution elasticity of demand at the equilibrium price. It measures the overall effect of changes in environmental safety on demand. The sign of η^e depends on two opposite effects: a price effect (more harmful goods are sold at lower prices, which encourages their demand relatively to less harmful goods) and a green effect (demand is lower for more harmful goods due to consumers' aversion to pollution). Straightforward calculations yield that $\eta^e = \frac{V'(\varepsilon_i)-X(c(\varepsilon_i)+t,\varepsilon_i)}{X(c(\varepsilon_i)+t,\varepsilon_i)}$. When

$V'(\varepsilon_i) - X(c(\varepsilon_i) + t, \varepsilon_i) < 0$, the marginal valuation of environmental safety by the marginal consumer is negative and so is η^e. In this case, the green effect dominates the price effect and an increase in the harm, hence a decrease in environmental safety, reduces the market area.

Due to asymmetric information, brown firms have an incentive to exploit consumers' ignorance on environmental safety and the actual harm from pollution with unsubstantiated claims to be offering the green variety. As such claims can freely be copied, consumers will mistrust them. Greenwashing practices are the rationale for the presence of the certification agency on the market, provided that consumers believe she award firms credible labels. However, there is no evidence that consumers should consider the agency's labels to be more credible than firms' claim. Consumers' mistrust about the agency's certification may arise from the multiplicity of tasks she undertakes, which makes the budget size valuable to her. Insofar as the agency needs an amount of revenue in excess of what is just sufficient to secure information, she may be suspected of manipulating consumer beliefs to augment the proceeds from fees. Such would be the case if the agency were, for instance, "paternalist" or "populist" in the sense of Salanié and Treich (2006). Then, she would reason on the basis of the consumer perceived safety rather than the actual safety to maximize social welfare. Consumers' suspicion is a challenge to a trustworthy agency for it imposes the following requirement: to make her label credible, the agency must charge credible fees, meaning that the fees specified for one environmental safety would not be worth imitating by an opportunistic agency if the environmental safety were different. In what follows, two cases will be distinguished depending on whether consumers are trustful or suspicious about the agency's certification. When consumers are trustful, the agency, unlike firms, is not suspected of being opportunistic and consumers spontaneously find her certification to be credible. Suspicious consumers will differ in that they suspect the agency as well as firms of being opportunistic, which makes consumers sceptical about the credibility of certification. In the latter case only will the fee serve as a signal of environmental safety because the agency must prove that she is trustworthy.

3 The benchmark with trustful consumers

As a benchmark, we record what would be the optimal behavior of a benevolent agency under asymmetric information when consumers are trustful.

Trustful consumers do not suspect the agency of opportunism and view the label awarded to firms as substantiated. Were the agency indifferent to the level of the budget, she would have to solve the first-best problem and chooses $X^* = V(\varepsilon_i)/\varepsilon_i$ that maximizes

$$\int_0^X (V(\varepsilon_i) - \varepsilon_i x)\,dx = V(\varepsilon_i)\,X - \varepsilon_i X^2/2 \tag{6}$$

In such a case, there would be no reason to charge a fee since $X^* = X(c(\varepsilon_i), \varepsilon_i)$: the market would implement by itself the first-best optimal allocation. This boils down to consider that the cost I of collecting information is negligible, hence consumers can freely free-ride on the agency to obtain full information on ε_i.

Nevertheless, the agency operates under the budget constraint requiring that all her expenditures, I included, not exceed the revenue generated by the certification fee:

$$R_i(t_i) - I \geq 0. \tag{7}$$

Substituting $c(\varepsilon_i) + t_i$ to p_i in(3), the agency's objective function under the budget constraint can be written as the following Lagrange function:

$$\frac{\left(V\left(\varepsilon_i\right)-t_i\right)^2}{2\varepsilon_i}+\lambda_i\left(t_i X\left(c\left(\varepsilon_i\right)+t_i,\varepsilon_i\right)-I\right), \tag{8}$$

where the Lagrange multiplier $\lambda_i \geq 0$ represents the social opportunity cost of spending money on consumer information relative to other activities of the agency. Hence, λ_i is a choice variable which measures the discretion of the agency about her expenditures. When λ_i is optimally chosen to be low, the budget requirement is not asking for much, so we are close to the case where there should be no tax. Large values of λ_i will indicate that the agency takes good care about raising revenue. Her behavior then resembles that of a profit-maximizing monopolist on the market for labeling.

Proposition 1: *When consumers are trustful, the agency's optimal choice consists of a fee $t\left(\varepsilon_i\right)$ and a non-negative Lagrange multiplier $\lambda\left(\varepsilon_i\right)$*

such that:

$$\frac{t(\varepsilon_i)}{p_i^e} = \frac{\lambda(\varepsilon_i) - 1}{\lambda(\varepsilon_i)} \frac{1}{\eta}, \tag{9}$$

or, equivalently,

$$t(\varepsilon_i) = \frac{\lambda(\varepsilon_i) - 1}{2\lambda(\varepsilon_i) - 1} V(\varepsilon_i) \;\; \textit{with } t(\varepsilon_b) = \underline{t}_b, t(\varepsilon_g) = \underline{t}_g \tag{10}$$

$$\textit{and } \lambda(\varepsilon_i) = \frac{1}{2} + \frac{V(\varepsilon_i)}{2\sqrt{V(\varepsilon_i)^2 - 4I\varepsilon_i}}. \tag{11}$$

Proof : (see Appendix 1)

Converting the specific fee to an ad valorem rate $\frac{t(\varepsilon_i)}{p_i^e}$ yields formula (9) which states that the ad valorem rate should be proportional to the inverse of the price elasticity of demand. Hence, it is optimal to choose a higher fee for varieties with a low price elasticity than for varieties with a high price elasticity. When consumers are trustful, $t(\varepsilon_i)$ is akin to a pure Ramsey tax in the sense that the fee is designed to raise revenue with a minimal loss in terms of efficiency. Therefore, $t(\varepsilon_i) = \underline{t}_i$ will be called the Ramsey fee in what follows. It is worth using the explicit expression of $\lambda(\varepsilon_i)$ given by (11) to interpret (9). When I is close to zero, $\lambda(\varepsilon_i)$ approaches its lowest value 1 and $t(\varepsilon_i)$ also tends to zero. Hence, we have the aforementioned first-best solution at the limit: the market by itself can implement the socially optimal allocation. In this case, there is no budget requirement regarding consumers' information since the cost of collecting information is negligible and the budget requirement for alternative purposes is normalized to zero. By contrast, in the polar situation where I becomes as large as possible under (5), we have that $\lambda(\varepsilon_i) \rightarrow +\infty$, i. e., the budget requirement becomes the main concern of the agency who then charges a fee close to the inverse of the price elasticity. Interestingly enough, this would also be the optimal choice of a private agency enjoying a monopoly position on the market for labeling. Indeed, from (10), when $\lambda(\varepsilon_i) \rightarrow +\infty$, the agency is better off charging a fee close to $\frac{V(\varepsilon_i)}{2}$: this coincides with the price set by a profit-maximizing monopolist selling the true information on ε_i to consumers. Furthermore, it can be checked that $\lambda(\varepsilon_i) = \frac{1}{1-\eta}$. Since $\lambda(\varepsilon_i)$ must be non-negative, the trustworthy agency always operates in a fee region such that the price elasticity of demand is lower than 1, i. e., the good is essential (such as potatoes). Lastly, it can be pointed out that there is no Laffer effect at the optimal fee since $R_i'(t(\varepsilon_i)) = \frac{V(\varepsilon_i)-2t(\varepsilon_i)}{\varepsilon_i}$ is positive. This is consistent with the evidence against the existence of Laffer effect (see Fullerton (1982)).

We end the analysis with trustful consumers by showing how changes in the magnitude of environmental safety affect the Ramsey fee $t(\varepsilon_i)$.

Lemma 1: *For all* $I \in \left[0, \frac{V(\varepsilon_g)^2}{4\varepsilon_b}\right]$, $t(\varepsilon_b) \geq t(\varepsilon_g)$ *and the derivative* $t'(\varepsilon_i)$ *can be written*

$$t'(\varepsilon_i) = -\frac{1}{1-\eta}\frac{t(\varepsilon_i)}{\varepsilon_i}\eta^e \tag{12}$$

Proof : (see Appendix 2)

As a result, the Ramsey fee $t(\varepsilon_i)$ rises as the harm from pollution is more severe and the environment is less safe. Moreover, equation (12) shows that the pollution elasticity of demand at the equilibrium price η^e is negative. As previously seen, this occurs when the green effect dominates the price effect so that demand is pollution sensitive.

4 Suspicious consumers and the credibility requirement

Let us now consider that consumers are suspicious. They suspect the agency of being opportunistic, thereby manipulating their beliefs in order to increase the revenue raised from the certification fee. Consumers' suspicion is motivated by the agency's discretion about her expenditures. In particular, an opportunistic agency might award brown firms the green label associated with the Ramsey fee $t(\varepsilon_g)$ in order to boost demand and spend the money collected from fees on objectives other than that of collecting information. Even though consumers mistrust the labels awarded to firms, they can try to infer how harmful pollution is, hence the actual environmental safety, from the equilibrium price $p_i = c(\varepsilon_i) + t_i$. The central question then is whether t_i can be a credible signal for ε_i as well as the trustworthiness of the agency.

Since firms are competitive, they are unable to use price as a signal for environmental safety. As argued by Spence (1977), raising price is costless to competitive firms: if consumers' perception of safety rose with price, every firm would raise price to signal higher safety at no cost. Therefore a competitive market price cannot play the same signaling role as if it were set by imperfectly competitive firms (see, for instance, Mahenc (2008) for an analysis of the monopoly price as a signal of the firm environmental performance). By contrast, fees may function as truthful and credible signals of environmental safety under some circumstances. As shown below, these circumstances are such that fees are not only costly to the agency, but also the costs are related to environmental safety.

For a given fee t, let $D^i(t,\mu) \equiv \frac{V(\varepsilon_i)-t}{\varepsilon^e(\mu)}$ denote the demand resulting from consumers' inference process at the equilibrium price. Facing this demand, the agency deals with the budget constraint such that the revenue $R^i(t,\mu) \equiv tD^i(t,\mu)$ must finance all the expenditures, I included. Note that the severity of the budget constraint now depends on consumers' beliefs. Optimistic beliefs about the environmental safety, i. e., μ is close to 1, enlarge the market size, thereby increasing the revenue raised from a given fee. We will denote $\underline{t}_i(\mu)$ and $\overline{t}_i(\mu)$ respectively the lowest and highest fee for which the agency breaks even on an expected value basis, i. e., $R^i(t,\mu) - I = 0$.

Under (5), easy calculations show that $\underline{t}_i(\mu) = \left(V(\varepsilon_i) - \sqrt{V(\varepsilon_i)^2 - 4I\varepsilon^e(\mu)}\right)/2$ and $\overline{t}_i(\mu) = \left(V(\varepsilon_i) + \sqrt{V(\varepsilon_i)^2 - 4I\varepsilon^e(\mu)}\right)/2$. Note that the Ramsey fees can now be written $t(\varepsilon_b) = \underline{t}_b = \underline{t}_b(0)$ and $t(\varepsilon_g) = \underline{t}_g = \underline{t}_g(1)$. Furthermore, it can easily be checked that the two following properties are satisfied:

1. Optimistic beliefs about the environmental safety loosen the budget constraint:

    - $\underline{t}_b(0) > \underline{t}_b(1)$ and $\overline{t}_b(0) < \overline{t}_b(1)$, thus $tD^b(t,0) - I \geq 0 \Rightarrow tD^b(t,1) - I \geq 0$;

    - $\underline{t}_g(1) < \underline{t}_g(0)$ and $\overline{t}_g(0) < \overline{t}_g(1)$, thus $tD^g(t,0) - I \geq 0 \Rightarrow tD^g(t,1) - I \geq 0$.

2. The safer the environment, the tighter the budget constraint:

$\underline{t}_b(\mu) < \underline{t}_g(\mu)$ and $\bar{t}_g(\mu) < \bar{t}_b(\mu)$, thus $tD^g(t,\mu) - I \geq 0 \Rightarrow tD^b(t,\mu) - I \geq 0$.

As in the previous section, the agency is benevolent, hence maximizes social welfare. However, the social welfare function is unique in the case of trustful consumers, while the agency's objective encompasses a whole spectrum of social welfare functions when consumers are suspicious. Indeed, the agency must now take into account perceived rather than actual environmental safety. With suspicious consumers, social welfare has the following reduced forms:

$$W^i(t,\mu) \equiv \int_0^{D^i(t,\mu)} \left(V(\varepsilon_i) - \varepsilon^e(\mu)x\right)dx, \tag{13}$$

$$= \frac{\left(V(\varepsilon_i) - t\right)^2}{2\varepsilon^e(\mu)}. \tag{14}$$

Thus, the spectrum of social welfare functions range from the case where consumers have accurate perceptions, i. e., $\varepsilon^e(\mu) = \varepsilon_i$, to the case where they have false certainty, i. e., $\varepsilon^e(\mu) = \varepsilon_j \neq \varepsilon_i$. The latter case will reflect what would be the objective of the agency were she opportunistic. Her behavior would then be similar to that attributed by Salanié and Treich (2009) to a "populist" regulator who maximizes welfare computed with the consumers' beliefs even though they are wrong. Because consumers are suspicious, the agency ought to preclude imitation by an opportunistic agency to simultaneously prove she is trustworthy and make her certification credible. Therefore, the trustworthy agency must also reason as if she were opportunistic. Note that, from the agency's viewpoint, $\mu = 0$ is the least favorable belief that consumers can hold since $\varepsilon^e(0) = \varepsilon_b$ is the worst perception that consumers have of the environmental safety.

Consumer suspicion imposes two further requirements on the agency's behavior. First, the agency must be willing to reveal information, and second, the agency's strategy must be credible in the sense that the fees specified for one environmental safety could not be imitated if the environmental safety were different. In other terms, the agency ought to satisfy an individual-rationality (IR) constraint and an incentive-compatibility (IC) constraint, which follow from using the perfect Bayesian equilibrium concept. This gives the model a signaling structure which merely differs from the standard Spencian game in that the set of signaling strategies is reduced by the budget requirement.

Restricting attention to pure strategies, a perfect Bayesian equilibrium of this game is a set of strategies $\{(t_i^*)_{i=b,g}\}$ and a probability distribution $\mu^*(t)$ such that strategies must be optimal given consumers' beliefs. Formally, this requires that, for each $i = b, g$,

$$t_i^* \in \arg\max_{t_i} W^i(t_i, \mu^*(t_i)) + \lambda_i \left(t_i D^i\left(t_i, \mu^*(t_i)\right) - I\right). \tag{15}$$

Consumers form posterior beliefs from their prior beliefs by using Bayes' rule:

$$\text{If } t_g^* \neq t_b^*, \text{ then } \mu^*\left(t_g^*\right) = 1 \text{ and } \mu^*\left(t_b^*\right) = 0; \tag{16}$$

$$\text{If } t_g^* = t_b^*, \text{ then } \mu^*\left(t_g^*\right) = \mu^*\left(t_b^*\right) = \mu_0. \tag{17}$$

As the equilibrium concept places no restriction on beliefs for fees off the equilibrium path, we will restrict as usual the consumers' beliefs to satisfy the intuitive criterion (see Cho and Kreps (1987)). An equilibrium in which the level of social welfare is W^i when the environmental safety is i fails to survive the intuitive criterion if there exists a deviation d satisfying the budget constraint with $\mu(d) = 1$, such that:

$$W^g < W^g(d, 1), \tag{18}$$

$$W^b(d, 1) \leq W^b. \tag{19}$$

From now on, the reasoning will both consider any t_b^* to be inside $\left[\underline{t}_b(0), \bar{t}_b(0)\right]$ and any t_g^* to be inside $\left[\underline{t}_g(1), \bar{t}_g(1)\right]$, so that the agency's budget constraint is satisfied in equilibrium.

Consider first the (IR) constraints. Recall that $\underline{t}_i(0)$ is the lowest fee for which the agency breaks even, when the true environmental safety is i and consumers believe that firms are brown with certainty (the least favorable beliefs through the agency's eyes). The (IR) constraints can be written:

$$W^b(t_b^*, 0) \geq W^b(\underline{t}_b(0), 0) \tag{20}$$

$$W^g(t_g^*, 1) \geq W^g(\underline{t}_g(0), 0) \tag{21}$$

These constraints guarantee that the agency is better off using the fees that award firms truthful labels, than reasoning on the basis of consumers' worst perception of environmental safety. Moreover, noting that $\underline{t}_b(0) = \underline{t}_b$ in constraint (20), we obtain the following result that, to signal truthfully brown firms, the agency will choose the Ramsey fee.

Lemma 1: *In any separating equilibrium, the regulator charges* $t_b^* = t(\varepsilon_b) = \underline{t}_b$.

Proof : (see Appendix 3)

We can henceforth write that $t_b^* = \underline{t}_b$. Let us now turn to the (IC) constraints. They prevent the agency from behaving as if she were opportunistic. Such a requirement secures the credibility of certification by imposing that the agency should not defect to the equilibrium fee that awards the wrong label. Neglecting the budget constraints, the (IC) constraints would be written as follows:

$$W^b(t_g^*, 1) \leq W^b(\underline{t}_b, 0) \tag{22}$$

$$W^g(t_g^*, 1) \geq W^g(\underline{t}_b, 0). \tag{23}$$

What (22) says is that, when firms are brown, the agency is worse off imitating the fee specified for green firms, and so should not be tempted to deviate from $\underline{t}_b$ to t_g^* . In this case, such a deviation is conceivable as long as t_g^* is inside $\left[\underline{t}_b(1), \bar{t}_b(1)\right]$, otherwise the deviation would not satisfy the budget requirement consistent with the certainty that firms are green. Similarly, condition (23) precludes a deviation from t_g^* towards $\underline{t}_b$ when firms are green. However, we know that $\underline{t}_b(0) < \underline{t}_g(0)$ (see property 2 above). This implies that a deviation to $\underline{t}_b = \underline{t}_b(0)$ would violate the budget requirement associated with $\mu^*(\underline{t}_b) = 0$, hence is not conceivable. In other terms, the severity of the budget constraint stemming from the certainty that firms are brown, prevents the agency from misleading consumers by awarding brown label to green firms. This exempts the agency from taking (23) into consideration. She is thus left with both constraints (22) and (21), which can be respectively expressed as follows:

$$\frac{\left(V(\varepsilon_b) - t_g^*\right)^2}{2\varepsilon_g} \leq \frac{\left(V(\varepsilon_b) - \underline{t}_b\right)^2}{2\varepsilon_b} \tag{24}$$

$$\frac{\left(V(\varepsilon_g) - t_g^*\right)^2}{2\varepsilon_g} \geq \frac{\left(V(\varepsilon_g) - \underline{t}_g(0)\right)^2}{2\varepsilon_b} \tag{25}$$

Let T_b and T_g denote the sets of fees t_g^* for which, respectively, conditions (22) and (21) are met, and τ_b and τ_g the solutions in t_g^* of the equality version of, respectively, (24) and (25) (see Figure 1 drawn for a parameter configuration such that $\tau_b < \tau_g$). So, $T_b = \left[\tau_b, \bar{t}_b(1)\right]$ and $T_g = \left[\underline{t}_g(1), \min\left\{\tau_g, \bar{t}_g(1)\right\}\right]$ where $\underline{t}_g(1) = \underline{t}_g$ is the Ramsey fee for green firms. One can interpret τ_b as the fee that credibly signals green firms with a minimum of social loss, and τ_g as the fee that truthfully signals green firms with a maximum of social loss. Thus, a necessary condition to disclose credible information and award green firms a substantiated label is that the agency chooses a fee t_g^* inside $T_d \cap T_c$, whenever this set is not empty. Graphically, one can see that the latter condition is fulfilled when $\tau_b < \min\left\{\tau_g, \bar{t}_g(1)\right\}$. Suppose that this inequality is satisfied. Then, to award the green label, the agency is better off charging $t_g^* \geq \tau_b$ which exceeds the Ramsey fee $\underline{t}_g$. In such a case, any t_g^* generates an amount of revenue in excess of what is needed to break even when firms are green, that is, $t_g^* D^g\left(t_g^*, 1\right) - I > 0$. Among all the candidates for a separating equilibrium, the equilibrium fee τ_b is the one that creates the lowest upward distortion relative to the Ramsey fee $\underline{t}_g$. Such a distortion, if ever, is necessary to simultaneously prove that the agency is not opportunistic and reveal that firms are green. Let us call the welfare loss caused by this signaling distortion the "Parmentier distortion". It is the minimum welfare loss due to consumer suspicion about the agency's trustworthiness. When emerging in equilibrium, the Parmentier distortion is the minimum price paid for the credibility of certification.

Proposition 2 establishes necessary conditions for separating equilibrium fees satisfying the intuitive criterion.

Proposition 2: *The least costly way of guaranteeing the credibility of certification is to charge the separating equilibrium fees* $t_b^* = \underline{t}_b$ *and* $t_g^* = \tau_b$ *.*

Proof : (see Appendix 4)

As $\underline{t}_b < \tau_b$ by definition, we have $t_b^* < t_g^*$: the grading of optimal taxes with respect to environmental safety reverses that of the Ramsey fees available with trustful consumers, i. e., $t(\varepsilon_b) \geq t(\varepsilon_g)$. In the presence of suspicious consumers, the agency ought to charge a fee higher for the green than for the brown label. If the agency were opportunistic, she might use $t(\varepsilon_g) = \underline{t}_g$ instead of $t(\varepsilon_b) = \underline{t}_b$ to convince consumers that firms are green while they are in fact brown. Such a misleading strategy would boost consumption in two manners: first, by lowering the fee since $t(\varepsilon_b) \geq t(\varepsilon_g)$, and second, by increasing consumers' perception of environmental safety. To prove her trustworthiness and truthfully reveal information, the agency must distort upward the fee intended for certifying that firms are green, thereby raising more revenue than the amount just sufficient to break even. This will successfully precludes opportunistic behaviors such as charging brown firms the fee dedicated to the green label, provided that reducing the market size is more costly when firms are brown than when they are green. The upward distortion is minimized at τ_b since it is the lowest fee for which the green label cannot be mistaken for the brown one. Thus, it is costly to signal via τ_b that the agency is not opportunistic as well as firms are green, in the sense that it entails a social loss that would not occur if consumers were trustful. This signaling cost is inescapable here. Suspicious consumers rationally accept to pay this cost because it makes the green label credible through their eyes. Indeed, they know that the fee τ_b charged for this label could not be mimicked by an opportunistic agency if firms were brown. By selecting τ_b among the multiple fees that are likely to achieve separation, the intuitive criterion helpfully points out the most efficient way to signal truthfully green firms. This "least-cost separating equilibrium outcome" has received much emphasis in the work of Spence (1974), Riley (1979) and Cho-Kreps (1987), among others. The Parmentier distortion is associated with this outcome. There is some analogy between the logic of this distortion and that followed by a Pigovian tax. Consumers exert a negative externality on the agency by free riding on her information. The extra premium consumers must pay to get trustworthy labels forces them to internalize this externality in much the same way that the Pigovian tax leads polluters to internalize correctly the costs of pollution. Not only must consumers pay for the cost of collecting information, but they also have to pay for the cost of revealing information.

Let us now examine the existence of separating equilibrium fees robust to the intuitive criterion.

For this, we define the functions $\tau_b(I) = \tau_b$ and $\tau_g(I) = \tau_g$ on $\left[0, \frac{V(\varepsilon_g)^2}{4\varepsilon_b}\right]$ (see Appendix 5 for more details). Both functions $\tau_i(I), i = b, g$, are increasing in I, with $\tau_g(I) < \bar{t}_g(1)$ for all admissible I and we have $\lim_{I\to 0} \tau_i(I) = \left(1 - \sqrt{\varepsilon_g/\varepsilon_b}\right) V(\varepsilon_i)$. Hence, if I is sufficiently small, $\tau_b(I)$ clearly exceeds $\tau_g(I)$ so that separation is impossible to achieve. When I is close to zero, the budget requirement is not much demanding and we know, from Proposition 1, that $\lambda(\varepsilon_i)$, approaches its lowest value 1 whatever i. The non-existence of a separating equilibrium in these circumstances suggests that the agency is facing a situation where she does not lose less from increasing the fee when firms are green than when they are brown, otherwise she could achieve separation. This is the sense of the single-crossing property which formally appears by considering the following Lagrange function:

$$\begin{aligned} L^i(t,\mu) &= W^i(t,\mu) + \lambda_i \left(R^i(t,\mu) - I \right) && (26) \\ &= \frac{\left(V(\varepsilon_i) - t\right)^2}{2\varepsilon^e(\mu)} + \lambda_i \left(t \frac{V(\varepsilon_i) - t}{\varepsilon^e(\mu)} - I \right). && (27) \end{aligned}$$

Differentiating (27) with respect to t at the limit when $\lambda_i = 1$ yields:

$$\lim_{\lambda_i \to 1} L^i_t(t,\mu) = -\frac{t}{\varepsilon^e(\mu)}. \tag{28}$$

Clearly, the right-hand side of (28) does not depend on the actual environmental safety. As the welfare cost of charging the fee is the same regardless of whether firms are brown or green, the fee cannot be a signal for environmental safety, thus certification fails to be credible. In particular, charging zero fee whatever the environmental safety cannot be informative for suspicious consumers, while it is the optimal choice of the agency facing trustful consumers in the limiting situation where $I \to 0$. A first conclusion is that certification fails to be credible for too low costs of collecting information. Is it possible that I be sufficiently large for a separating equilibrium to exist in the present framework? The answer is yes, as shown below.

Calculations performed in Appendix 5 show that the functions $\tau_b(I)$ and

$\tau_g(I)$ intersect at some threshold $\widetilde{I}$ which may be lower or higher than $\frac{V(\varepsilon_g)^2}{4\varepsilon_b}$ depending on the parameter values. More precisely, we have the following existence result.

Proposition 3: *A separating equilibrium satisfying the intuitive criterion exists if and only if* $\widetilde{I} \leq I \leq \frac{V(\varepsilon_g)^2}{4\varepsilon_b}$ *and* $\frac{V(\varepsilon_b)}{V(\varepsilon_g)} \leq 1 + \frac{\varepsilon_g/\varepsilon_b}{2\left(1-\sqrt{\varepsilon_g/\varepsilon_b}\right)}$.

Proof : (see Appendix 5)

For all I higher than $\widetilde{I}$, the revenue requirement becomes quite demanding. In this case, there exists a fee specified for the green label too high to be mimicked by an opportunistic agency facing brown firms. To get some intuition for this result, consider the limiting case where I approaches its maximal value so that $\lambda(\varepsilon_b)$ becomes very large. Then, the agency tends to behave as a private monopoly on the market for labeling, in the sense that she only cares for maximizing the revenue $R^i(t, \mu^*(t)) - I = t\frac{V(\varepsilon_i)-t}{\varepsilon^e(\mu^*(t))} - I$. As the brown good is cheaper than the green one at the equilibrium price, we have $V(\varepsilon_g) < V(\varepsilon_b)$ and so the agency has a higher revenue when firms are brown than when they are green, *ceteris paribus*. Consequently, the foregone revenue from a lost consumer is less with green than brown firms. Formally, the single-crossing property previously mentioned in the case where $I \to 0$, is working now in the opposite direction: the agency has less to lose from raising the fee intended for green firms than that for brown firms. As a result, the efficient way of signaling green firms is to associate the green label with an upward-distorted fee. This proves that the agency is not as afraid of contracting the market size as she would be, were she opportunistic. Suspicious consumers can then view the green label as credible.

Let us now turn to the parameter configuration for which separation cannot occur. Pooling equilibria satisfying the intuitive criterion may nevertheless exist in this case. Let t^* denote the equilibrium fee charged by the agency, regardless of the environmental safety. From (17), the consumers' posterior beliefs after observing the price t^* are the same as their prior beliefs μ_0. The social welfare obtained by charging t^* is $W^i(t^*, \mu_0) = \frac{(V(\varepsilon_i)-t^*)^2}{2\varepsilon^e(\mu_0)}$. Whenever it exists, t^* must meet the budget requirements ensuing from beliefs μ_0, that is, $t^* D^i(t^*, \mu_0) - I \geq 0$ for $i = b, g$. As previously seen, this budget constraint is tighter when firms are green than when they are brown. Thus, we only have to consider inequality $t^* D^g(t^*, \mu) - I \geq 0$. Suppose that the latter holds, then t^* must also satisfy the two following conditions:

$$W^i(t^*, \mu_0) \geq W^i(\underline{t}_i(0), 0), i = b, g, \tag{29}$$

otherwise the agency would have an incentive to defect from t^* and charge $\underline{t}_i(0)$ whatever the consumers' beliefs upon observing this fee. Inequalities (29) can immediately be rewritten

$$\frac{(V(\varepsilon_i) - t^*)^2}{2\varepsilon^e(\mu_0)} \geq \frac{(V(\varepsilon_i) - \underline{t}_i(0))^2}{2\varepsilon_b}, i = b, g. \tag{30}$$

We need to define $\tau_i(\mu_0)$ as the solutions in t^* of the equality versions of (30) so as to characterize the range of pooling equilibrium fees robust to the intuitive criterion. This is established by the following proposition for parameter values such that no separating equilibrium exists.

Proposition 4: *Consider that* $\mu_0 \in (0,1)$. *If*

$$\frac{V(\varepsilon_b)}{V(\varepsilon_g)} \leq 1 + \frac{\varepsilon_g/\varepsilon_b}{2\left(1 - \sqrt{\varepsilon_g/\varepsilon_b}\right)} \text{ and } I \leq \widetilde{I}, \tag{31}$$

or if

$$1 + \frac{\varepsilon_g/\varepsilon_b}{2\left(1 - \sqrt{\varepsilon_g/\varepsilon_b}\right)} < \frac{V(\varepsilon_b)}{V(\varepsilon_g)} \text{ and } 0 \leq I \leq \frac{V(\varepsilon_g)^2}{4\varepsilon_b}, \tag{32}$$

then there exists a whole range of pooling equilibria satisfying the intuitive criterion such that $t^* \in \left[\underline{t}_g(\mu_0), \min\{\tau_b(\mu_0), \tau_g(\mu_0)\}\right]$.

Proof : (see Appendix 6)

For the parameter values stated in Proposition 4, the agency fails to separate green from brown firms because raising the fee above the Ramsey fee has become too costly in terms of welfare. The agency can no longer preclude imitation by an opportunistic agency. The latter facing brown firms can easily behave as if they were green to manipulate consumers' beliefs. In equilibrium, excessive revenue is generated from pooling fees associated with uninformative labels. Certification cannot be credible because it is impossible to use fee as a signal for environmental safety. From the explicit expression of $\tau_i(\mu_0)$ given in the Appendix, $\tau_i(\mu_0)$ is increasing in μ_0. Since, in contrast, $\underline{t}_g(\mu_0)$ is decreasing with μ_0, the range of pooling equilibrium fees gets broader as consumers are initially more optimistic about the environmental safety. Hence, consumers' optimism enlarges the equilibrium possibilities of misleading them.

5 Conclusion

This paper has examined the credibility of certification by a benevolent agency who must prove her trustworthiness to suspicious consumers. When consumers lack information about the environmental safety of a polluting good, firms often try to manipulate consumers' beliefs with unsubstantiated claims to be offering the green variety of the good. In response to this problem of adverse selection, a certification agency can acquire full information at a cost she will recover by charging firms a fee associated with the label "green" or "brown". Consumers however may view these labels as suspiciously as they view claims by firms. Knowing that the agency cares about the size of her revenue, consumers may indeed suspect her to manipulate their beliefs about environmental safety and employ misleading fees in order to generate more revenue than would be collected with truthful labels. In this context, certification by the agency cannot be credible unless she charges fees that she could not duplicate were she opportunistic. This is moreover a necessary requirement to disclose full information about environmental safety to consumers. Unfortunately, it has been shown here that pooling fees associated with uninformative labels cannot always be precluded in equilibrium, in particular when the agency has low costs of collecting information.

The analysis also characterizes the existence of a unique separating equilibrium robust to the intuitive criterion, in which the agency credibly signals green firms with a fee higher than that dedicated to brown firms. This reverses the benchmark result that the Ramsey fee designed for raising revenue with a minimal loss in terms of efficiency should be higher for brown than for green firms. It turns out that the agency must incur sufficiently high costs of collecting information to achieve separation of brown and green firms. The intuition underlying this result is that, when the agency is more greedy for the revenue generated by fees, she is less reluctant to lose welfare when firms are green than when they are brown, and so it is easier for her to prevent an opportunistic agency from mimicking her behavior. As the fee used to credibly signal green firms is distorted upward relative to the Ramsey fee, suspicious consumers must pay an extra premium to get full information about the environmental safety provided by green firms.

The analysis crucially hinges on the assumption that the green good is more costly to produce than the brown one. It follows that consumers are paying more for the brown than the green good sold at the competitive price. This explains why certification fees are more costly to the benevolent agency when firms are brown than when they are green, thereby playing the role of a signal for environmental safety. As a result, a high fee can been shown to be an effective means of signaling green firms provided that the agency has less to lose in terms of revenue from reducing the market size when firms are green than when they are brown. Suspicious consumers then rationally accept the loss in utility necessary for certification to be credible.

From this analysis, we cannot conclude that per-unit fees are a panacea to solve the adverse selection problem due to greenwashing since the market remains uninformed when the sole equilibria robust to the intuitive criterion are pooling. This suggests to explore other communication tools that the certification agency or the market for the good by itself could use to signal the environmental harm caused by firms to consumers. Whatever the means by which the signal is sent, it will be effective only if consumers are ready to accept some direct loss in utility to obtain information.

6 Appendix

6.1 Appendix 1: Proof of proposition 1

The first-order conditions of the agency's constrained optimization problem yield:

$$-\frac{V(\varepsilon_i) - t_i}{\varepsilon_i} + \lambda_i \left(X\left(c(\varepsilon_i) + t_i, \varepsilon_i\right) + t_i X_p\left(c(\varepsilon_i) + t_i, \varepsilon_i\right)\right) = 0, \quad (33)$$

$$\lambda_i \left(t_i X\left(c(\varepsilon_i) + t_i, \varepsilon_i\right) - I\right) = 0, \quad (34)$$

$$\lambda_i \geq 0. \quad (35)$$

Using the expression of demand (2), condition (33) can be rewritten

$$-\left(V(\varepsilon_i) - t_i\right) + \lambda_i \left(V(\varepsilon_i) - 2t_i\right) = 0 \quad (36)$$

This equation yields $t(\varepsilon_i) = \frac{\lambda(\varepsilon_i)-1}{2\lambda(\varepsilon_i)-1} V(\varepsilon_i)$ in (10). Substituting $X(c(\varepsilon_i) + t_i, \varepsilon_i)$ to $\frac{V(\varepsilon_i)-t_i}{\varepsilon_i}$ in the left-hand side of (33), we get (9). When $\lambda_i > 0$, equation (34) admits an upper and lower root in t_i, that is, respectively, $\left(V(\varepsilon_i) + \sqrt{V(\varepsilon_i)^2 - 4I\varepsilon_i}\right)/2$ and $\left(V(\varepsilon_i) - \sqrt{V(\varepsilon_i)^2 - 4I\varepsilon_i}\right)/2$. From (36) and the fact that $\lambda(\varepsilon_i)$ is non-negative, we have that $t(\varepsilon_i)$ must be lower than $V(\varepsilon_i)/2$, thereby implying both $t(\varepsilon_b) = \underline{t}_b$ and $t(\varepsilon_g) = \underline{t}_g$ in (10). The expression of $\lambda(\varepsilon_i)$ given in (11) is obtained from $t(\varepsilon_i) = \frac{\lambda(\varepsilon_i)-1}{2\lambda(\varepsilon_i)-1} V(\varepsilon_i)$ by substituting $\left(V(\varepsilon_i) - \sqrt{V(\varepsilon_i)^2 - 4I\varepsilon_i}\right)/2$ to $t(\varepsilon_i)$.

6.2 Appendix 2: Proof of lemma 1

From proposition 1, we know that $t(\varepsilon_i) = \left(V(\varepsilon_i) - \sqrt{V(\varepsilon_i)^2 - 4I\varepsilon_i}\right)/2$, where $t(\varepsilon_i)$ is the lowest root in t_i of $t_i X(c(\varepsilon_i) + t_i, \varepsilon_i) = I$. Differentiating this budget equation with respect to t_i and I yields

$$
\begin{aligned}
\frac{\partial t(\varepsilon_i)}{\partial I} &= \frac{1}{X(c(\varepsilon_i) + t(\varepsilon_i), \varepsilon_i) + t(\varepsilon_i) X_p} \\
&= \frac{\varepsilon_i}{\sqrt{V(\varepsilon_i)^2 - 4I\varepsilon_i}}.
\end{aligned}
$$

From this expression, we can see that, for all $I \in \left[0, \frac{V(\varepsilon_g)^2}{4\varepsilon_b}\right]$,

$$
\frac{\partial^2 t(\varepsilon_i)}{\partial I \partial \varepsilon_i} = \frac{V(\varepsilon_i)^2 - 2I\varepsilon_i}{\left(V(\varepsilon_i)^2 - 4I\varepsilon_i\right)^{\frac{3}{2}}} > 0.
$$

Thus, $t(\varepsilon_i)$ is an increasing function of I with a higher slope as ε_i rises. Since, at $I = 0$, we have $t(\varepsilon_b) = t(\varepsilon_g)$, we obtain that $t(\varepsilon_b) > t(\varepsilon_g)$ for all $I \in (0, \frac{V(\varepsilon_g)^2}{4\varepsilon_b}]$.

To obtain the differential $t'(\varepsilon_i)$, we now differentiate $tX(c(\varepsilon_i)+t,\varepsilon_i)=I$ with respect to t and ε (subscript i is omitted for notational simplicity)

$$t\left(X_p c'(\varepsilon)+X_\varepsilon\right)d\varepsilon+\left(X+tX_p\right)dt=0. \tag{37}$$

Rearranging terms, we get

$$\left(1+tX_p/X\right)t'(\varepsilon)=-t\left(X_p c'(\varepsilon)+X_\varepsilon\right)/X \tag{38}$$

As $1+tX_p/X=1-t(\varepsilon_i)\,\eta/p_i^e$, it can be checked that $1+tX_p/X=\frac{1}{\lambda(\varepsilon_i)}$ by using (9). From the definition of η^e, we obtain

$$t'(\varepsilon_i)=-\lambda(\varepsilon_i)\frac{t(\varepsilon_i)}{\varepsilon_i}\eta^e. \tag{39}$$

Replacing $\lambda(\varepsilon_i)$ by $\frac{1}{1-\eta}$ gives (12).

6.3 Appendix 3: Proof of lemma 1

Suppose for a contradiction that there exists a separating equilibrium in which $t_b^* \neq \underline{t}_b$. As the consumers' expectations are correct at equilibrium, the resulting social welfare is

$$W^b(t_b^*, 0) = \frac{\left(V\left(\varepsilon_b\right) - t_b^*\right)^2}{2\varepsilon_b},$$

which is strictly lower than $W^b(\underline{t}_b, 0) = \left(V\left(\varepsilon_b\right) - \underline{t}_b\right)^2 / 2\varepsilon_b$. Then, the regulator would have an incentive to deviate to $\underline{t}_b$ whatever the consumers' inference μ from observing $\underline{t}_b$ since $\underline{t}_b(\mu) = \left(V\left(\varepsilon_b\right) - \sqrt{V\left(\varepsilon_b\right)^2 - 4I\varepsilon^e(\mu)}\right)/2 \leq \underline{t}_b\left(0\right) = \underline{t}_b = \left(V\left(\varepsilon_b\right) - \sqrt{V\left(\varepsilon_b\right)^2 - 4I\varepsilon_b}\right)/2$. Indeed, for any $\mu \in (0, 1]$, we have $W^b(\underline{t}_b, 0) < W^b(\underline{t}_b, \mu)$. If $t_b^* \neq \underline{t}_b$, then $W^b(t_b^*, 0) < W^b(\underline{t}_b, 0)$ and so $W^b(t_b^*, 0) < W^b(\underline{t}_b, \mu)$.

6.4 Appendix 4: Proof of proposition 2

Assume $\tau_b \leq \tau_g$ so that a pair of separating equilibrium fees $\left(t_b^*, t_g^*\right)$ exists. Lemma 1 states that $t_b^* = \underline{t}_b$ yielding the social welfare level $W^b(\underline{t}_b, 0)$. Suppose that separation is achieved in equilibrium at $t_g^* \neq \tau_b$ yielding a welfare of W^g. Necessarily, t_g^* is higher than $\underline{t}_g$ to meet the budget constraint associated with consumers' certainty that firms are green, and $W^g < W^g(\underline{t}_g, 1)$ since $W^g(t, 1)$ is decreasing in $t > 0$. As $\tau_b > \underline{t}_g$, constraint (22) requires $\tau_b < t_g^*$, hence we simultaneously have $W^g < W^g(\tau_b, 1)$ and $W^b(\tau_b, 1) = W^b(\underline{t}_b, 0)$.

The equilibrium at t_g^* violates the intuitive criterion since the agency would deviate to t_g^*. Thus, any separating equilibrium in which $t_g^* \neq \tau_b$ fails to survive the intuitive criterion.

6.5 Appendix 5: Proof of proposition 3

Using explicit expressions of welfare, equation $W^i(t, 1) = W^i(\underline{t}_i(0), 0)$ can be written

$$\frac{(V(\varepsilon_i) - t)^2}{2\varepsilon_g} = \frac{(V(\varepsilon_i) - \underline{t}_i(0))^2}{2\varepsilon_b}. \tag{40}$$

Solving (40) in t yields

$$\tau_i = \left(1 - \sqrt{\varepsilon_g/\varepsilon_b}\right) V(\varepsilon_i) + \underline{t}_i(0)\sqrt{\varepsilon_g/\varepsilon_b}. \tag{41}$$

As $\underline{t}_i(0) = \left(V(\varepsilon_i) - \sqrt{V(\varepsilon_i)^2 - 4I\varepsilon_b}\right)/2$, we can write explicitly the function $\tau_i(I)$

$$\tau_i(I) = V(\varepsilon_i)\left(1 - \sqrt{\varepsilon_g/\varepsilon_b}/2\right) - \sqrt{\varepsilon_g/\varepsilon_b}\sqrt{V(\varepsilon_i)^2 - 4I\varepsilon_b}/2. \tag{42}$$

Note first that, for all I inside $\left[0, \frac{V(\varepsilon_g)^2}{4\varepsilon_b}\right]$, it can be checked that $\tau_g(I) < \bar{t}_g(1) = \left(V(\varepsilon_g) + \sqrt{V(\varepsilon_g)^2 - 4I\varepsilon_g}\right)/2$. Remark also that $\tau_i(I)$ is an increasing function of I, and second, that $\lim_{I \to 0} \tau_i(I) = \left(1 - \sqrt{\varepsilon_g/\varepsilon_b}\right) V(\varepsilon_i)$. Thus, in the limiting situation where $I \to 0$, we have $\tau_b > \tau_g$. Cumbersome calculations show that inequality $\tau_b \leq \tau_g$ is satisfied for all I such that

$$I \geq \frac{\sqrt{\varepsilon_g/\varepsilon_b} - 1}{\varepsilon_g\left(\sqrt{\varepsilon_g/\varepsilon_b} - 2\right)^2}\left(\left(1 - \sqrt{\varepsilon_g/\varepsilon_b}\right)(V(\varepsilon_b) - V(\varepsilon_g))^2 - (\varepsilon_g/\varepsilon_b)V(\varepsilon_b)V(\varepsilon_g)\right). \tag{43}$$

From the right-hand side of this inequality, we define $\widetilde{I}$ as

$$\widetilde{I} \equiv \frac{e-1}{\varepsilon_g(e-2)^2}\left((1-e)(V(\varepsilon_b) - V(\varepsilon_g))^2 - e^2 V(\varepsilon_b)V(\varepsilon_g)\right). \tag{44}$$

where $e \equiv \sqrt{\varepsilon_g/\varepsilon_b}$. Further calculations show that $\widetilde{I}$ is lower than the maximum value acceptable for I, i. e., $\frac{V(\varepsilon_g)^2}{4\varepsilon_b}$, as long as

$$\frac{V(\varepsilon_b)}{V(\varepsilon_g)} \leq 1 + \frac{e^2}{2(1-e)}. \tag{45}$$

6.6 Appendix 6: Proof of proposition 4

Denoting $\tau_i(\mu_0)$ the solutions in t^* of the equality versions of (30), we obtain the following explicit expression

$$\tau_i(\mu_0) = V(\varepsilon_i)\left(1 - \sqrt{\varepsilon^e(\mu_0)/\varepsilon_b}/2\right) - \sqrt{\varepsilon^e(\mu_0)/\varepsilon_b}\sqrt{V(\varepsilon_i)^2 - 4I\varepsilon_b}/2. \tag{46}$$

Inequality $\tau_b(\mu_0) \leq \tau_g(\mu_0)$ is satisfied for all I such that

$$I \geq \frac{\sqrt{\varepsilon^e(\mu_0)/\varepsilon_b} - 1}{\varepsilon^e(\mu_0)\left(\sqrt{\varepsilon^e(\mu_0)/\varepsilon_b} - 2\right)^2}\left(\left(1 - \sqrt{\varepsilon^e(\mu_0)/\varepsilon_b}\right)(V(\varepsilon_b) - V(\varepsilon_g))^2 - (\varepsilon^e(\mu_0)/\varepsilon_b)\, V(\varepsilon_b)\, V(\varepsilon_g)\right). \tag{47}$$

We denote by $\widetilde{I}(\mu_0)$ the right-hand side of this inequality. Further calculations show that $\widetilde{I}(\mu_0)$ is lower than the maximum value acceptable for I, i. e., $\frac{V(\varepsilon_g)^2}{4\varepsilon_b}$, as long as

$$\frac{V(\varepsilon_b)}{V(\varepsilon_g)} \leq 1 + \frac{\varepsilon^e(\mu_0)/\varepsilon_b}{2\left(1 - \sqrt{\varepsilon^e(\mu_0)/\varepsilon_b}\right)}, \tag{48}$$

where the right-hand side is increasing with μ_0.

Thus, $\tau_b(\mu_0) \leq \tau_g(\mu_0)$ when $1 + \frac{\varepsilon_g/\varepsilon_b}{2\left(1-\sqrt{\varepsilon_g/\varepsilon_b}\right)} < \frac{V(\varepsilon_b)}{V(\varepsilon_g)} \leq 1 + \frac{\varepsilon^e(\mu_0)/\varepsilon_b}{2\left(1-\sqrt{\varepsilon^e(\mu_0)/\varepsilon_b}\right)}$, and $\left[\underline{t}_g(\mu_0), \tau_b(\mu_0)\right]$ is the interval of pooling, otherwise it is $\left[\underline{t}_g(\mu_0), \tau_g(\mu_0)\right]$ when $1 + \frac{\varepsilon^e(\mu_0)/\varepsilon_b}{2\left(1-\sqrt{\varepsilon^e(\mu_0)/\varepsilon_b}\right)} \leq 1 + \frac{\varepsilon_g/\varepsilon_b}{2\left(1-\sqrt{\varepsilon_g/\varepsilon_b}\right)}$.

Moreover, as $\tau_g < \tau_b$, all the existing pooling equilibria are robust to the intuitive criterion. To demonstrate this, we define d^i as the highest fee that could tempt the agency to defect from a pooling equilibrium at t^* if firms of the environmental safety i were then believed to certainly be green, that is, d^i solves

$$W^i(t^*, \mu_0) = W^i(d^i, 1), \tag{49}$$

or, equivalently,

$$\frac{\left(V(\varepsilon_i) - t^*\right)^2}{2\varepsilon^e(\mu_0)} = \frac{\left(V(\varepsilon_i) - d^i\right)^2}{2\varepsilon_g}. \tag{50}$$

This yields

$$d^i = \left(1 - \sqrt{\varepsilon_g/\varepsilon_b}\right) V(\varepsilon_i) + t^* \sqrt{\varepsilon_g/\varepsilon_b}. \tag{51}$$

Clearly, we have that $d^g < d^b$. Thus, for all $d \geq d^b$, we simultaneously have $W^b(d, 1) \leq W^d(t^*, \mu_0)$, and $W^g(d, 1) < W^g(t^*, \mu_0)$. Thus, it is not possible to find a defection d from a pooling equilibrium at t^*, that simultaneously fulfills (18) and (19). This shows that the intuitive criterion cannot restrict beliefs for any pooling equilibrium fee.

References

[1] AMACHER, G. S., KOSKELA E. and M. OLLIKAINEN (2004), "Environmental Quality Competition and Eco-Labeling ", *Journal of Environmental Economics and Management* 47, 284-306.

[2] CASON T. N. and L. GANGADHARAN (2002), "Environmental Labelling and Incomplete Consumer Information in Laboratory Experiments", *Journal of Environmental Economics and Management* 43, 113-134.

[3] CHO, I-K. and D. KREPS (1987), "Signaling Games and Stable Equilibria", *Quarterly Journal of Economics* 102, 179-221.

[4] CRESPI, J. M. and S. MARETTE (2001), "How Should Food Safety Certification Be Financed?", *American Journal of Agricultural Economics* 83, 852-861.

[5] FULLERTON, D. (1982), "On the Possibility of an Inverse Relationship between Tax Rates and Government Revenues" 19, 3-23.

[6] KARL H. and C. ORWAT. (1999), "Economic Aspects of Environmental Labelling" in H. Folmer and T. Tietenberg (eds), *Yearbook of Environ-*

mental and Resource Economics 1999/2000, Elgar, Cheltenham, UK, 107-170.

[7] KUHN, M. (2005), *The Greening of Markets: Product Competition, Pollution and Policy Making in a Duopoly*. Cheltenham: Edward Elgar.

[8] MAHENC, P. (2008), "Signaling the Environmental Performance of Polluting Products", *International Journal of Industrial Organization*, Vol. 26, 1, 59-68.

[9] MASON, C. F. (2006), "An Economic Model of Eco-Labeling", *Environmental Modelling and Assessment* 2006 11(2), 131-143.

[10] MILGROM, P. and J. ROBERTS (1982), "Limit pricing and Entry under Incomplete Information: An Equilibrium Analysis", *Econometrica* 50, 443-459.

[11] MILGROM, P. and J. ROBERTS (1986), "Price and Advertising Signals of Product Quality", *Journal of Political Economy* 94, 796-821.

[12] NISKANEN, W. A. Jr. (1971), *Bureaucracy and Representative Government*. Chicago: Aldine-Atherton

[13] RILEY J. (1979), "Informational Equilibrium", *Econometrica* 47, 331-360.

[14] SALANIE, F. and N. TREICH (2009), "Regulation in Happyville", *The Economic Journal*, 119, 665-679.

[15] SANDMO, A. (2000), *The Public Economics of the Environment.* Cambridge: Harvard University Press.

[16] SPENCE, M. (1974), *Market Signaling.* Oxford University Press.

[17] SPENCE, M. (1977), "Consumer Misperceptions, Product Failure and Producer Liability", *Review of Economic Studies* 44, 561-572.

[18] VISCUSI, W. K. and J. T. HAMILTON (1999), "Are Risk Regulators Reliable? Evidence from Hazardous Waste Cleanup Decisions", *American Economic Review* 89, 1010-1027.

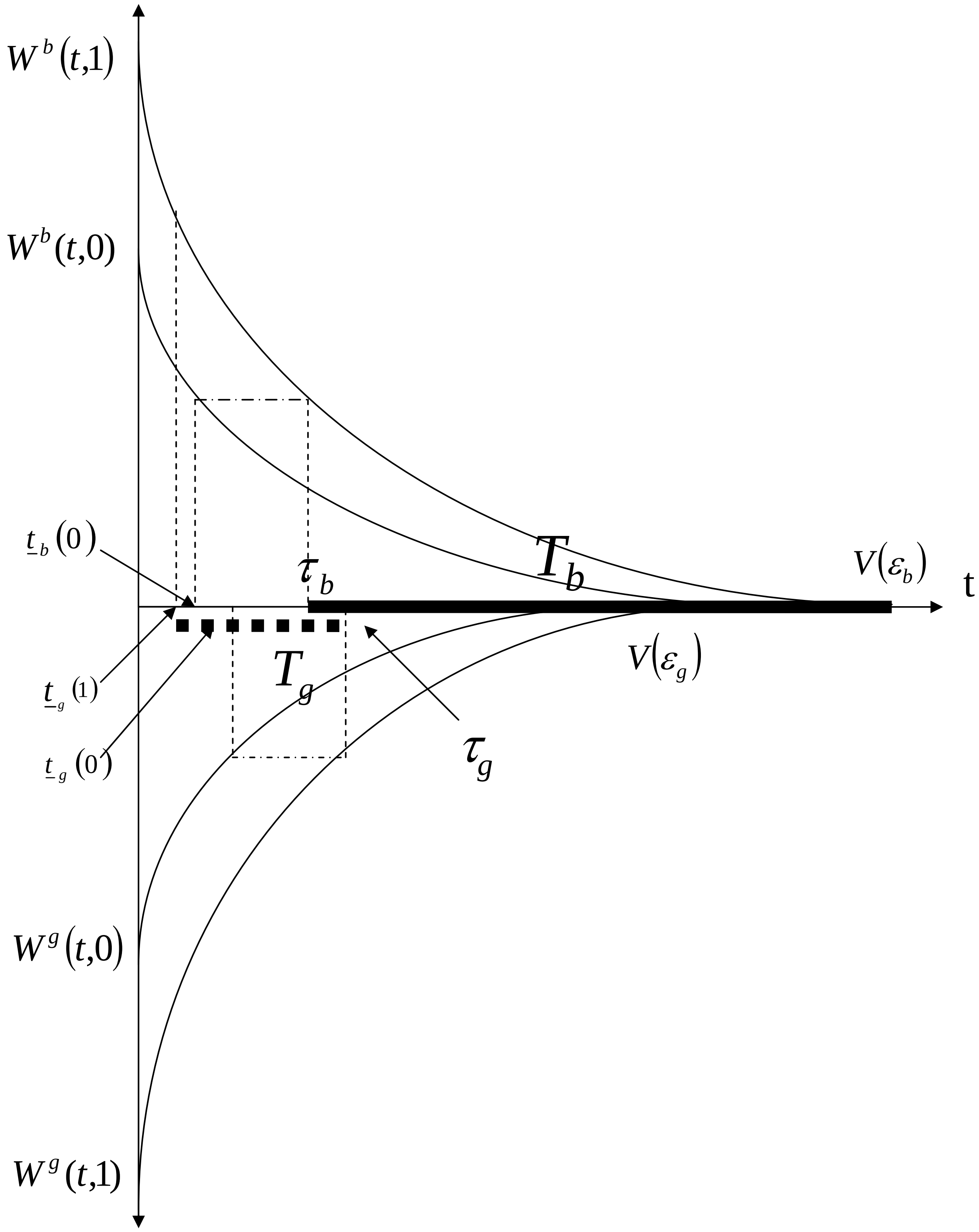

Figure 1
The credibility constraints