

Wasteful Labeling

By Philippe Mahenc*

July 9, 2010

Abstract

The role of labeling is to solve the adverse selection problem caused by unsubstantiated claims from firms. The problem however is likely to remain unsolved if the labeling agency is not trustworthy. She can be suspected to divert the fees charged for labeling from their primary purpose of collecting information in order to raise excessive revenue. This paper addresses this issue and shows that labeling may be wasteful due to consumers' scepticism about the trustworthiness of labeling. To award firms green labels, the agency may charge fees that exceed the Ramsey level at which the revenue needed for collecting information is raised with a minimal loss in terms of efficiency.

Keywords: Greenwashing, Labeling, Signaling.

JEL Code: D82, H21, H23, Q28.

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

Wasteful Labeling

1 Introduction

One prominent example of problems due to asymmetric information is the adverse selection caused by unsubstantiated claims that goods are produced with “environmental friendly” technologies, the so-called greenwashing phenomenon. It 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. A growing number of economic studies have recently recognized ecolabeling as a possible solution to this adverse selection problem. Most of the papers insist that information about the environmental type of firms should be supplied by a third party such as a governmental agency or a nongovernmental organization, rather than by firms themselves. Caswell and Mojduszka (1996) claim that signaling quality through labeling requires a reputable certification agent whom consumers can trust¹. In a more formal treatment, McCluskey (2000) shows that third-party monitoring is necessary to ensure truth in labeling. Ben Youssef and Lahmandi-Ayed (2008) argue that the presence of a credible labeling authority is enough to induce firms to provide quality levels that are perfectly identified by consumers. To our knowledge, Mason (2006) is the first analysis in which labeling by a third party sends a “signal” in the

¹See also Kuhn (2005) for a recent survey on labeling.

sense of Spence (1974) to uninformed consumers. A crucial assumption is that consumers are Bayesian, meaning that, after observing the firms' choice of certification, consumers form inferences concerning the severity of the environmental harm to make their consumption decision. Hence, consumers can rationally infer whether firms are brown or green upon seeing the decision to seek certification or not. In Mason (2006), labeling enables green firms to signal their environmental friendliness to consumers in a way that is both profitable (green firms are better off disclosing information rather than concealing it) and credible (brown firms are worse off imitating their green counterpart). As a result, firms have the opportunity to prove that they are trustworthy thanks to labeling by a third party.

The central question here is what if the third-party herself is not trustworthy? That is, the labeling agency is suspected by consumers to manipulate their beliefs and supply unreliable information. This paper addresses this issue and shows that, if a labeling agency is to provide believable information, she must convince consumers that she is trustworthy and is not deceiving them. In the proposed model, labeling is handled by a benevolent agency who has superior information about the harm from pollution generated by firms, which is imperfectly observed by consumers. As in Crespi and Marette (2001), the labeling agency charges firms per-unit fees to finance the fixed cost of collecting information. As they modify consumer price, the fees paid by firms to get a truthful label play the role of signaling how harmful the good is to consumers. This signaling motive for labeling is closely related to

the Bayesian approach initiated by Mason (2006) in which pursuing or not the label is the signal sent by firms to consumers. In contrast, the signal here is sent by the agency via the fee encompassed within the consumer price.

The main result is that labeling may be wasteful due to consumers' scepticism about the truthfulness of labeling. If consumers trust the agency, then the fee charged for labeling ought to raise revenue with a minimal loss in terms of efficiency: it is a pure Ramsey tax dedicated to funding the agency's information. If, however, consumers mistrust the agency, she may charge fees that exceed the Ramsey level to prove her trustworthiness and award firms truthful labels.

Greenwashing encompasses all practices that range from vague claims to misleading advertising about the environmental performance of firms. From an early lab experiment, Boulding and Kirmani (1993) had concluded that consumers do not necessarily perceive warranties as a credible signal of product quality. Regarding environmental quality, 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 Authority². To tackle the problem of greenwashing, two main forms of labeling are usually distinguished by the literature (see Kuhn, 2005): labeling by the firm itself and labeling by a third

²See Lyon and Maxwell (2006) for more on greenwashing, and Mahenc (2008) for reasons why consumers might be misinformed about the environmental performance of products.

party. The focus here is on the credibility of third-party labeling.

The model relies on two specific assumptions regarding consumers. First, consumers have heterogeneous tastes for a polluting good: they want the good but dislike the “bad” bundled to the good with varying degrees. One possible interpretation is that consumers experience heterogeneous personal troubles from purchasing the available variety of the good. These troubles range from simple fears to authenticated risks to personal health or the environment. They may be caused, for instance, by the chemicals consumers can hardly see or detect in consumption goods (such as pesticides, nitrates and heavy metal³) and indoor air (like radon gas, formaldehyde or asbestos⁴), or in genetically modified foods⁵. 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. This interpretation is somehow reminiscent of the chapter 3 of John Stuart Mill’s essay (1863) on utilitarianism, in which Mill defends the

³See Ibanez and Stenger (2000) for the serious consequences for health in the long term all these chemicals can have.

⁴These are three of the four most dangerous substances responsible for indoor air pollution, according to the US Environmental Protection Agency (see Lomborg 2001, p. 183)

⁵The assumption of taste heterogeneity due to the safety of the good is consistent with the evidence that there is a marked difference in the consumer perception of genetically modified foods in the EU and the US. According to Lomborg (2001), “In the EU, 59 percent of the consumers view GM foods as risky and consistent majorities reject their usefulness and find them morally unacceptable and to be discouraged. In the US, 60 percent are positive about food technology...” (see p. 343).

possibility of humans having some social conscience that enables us to care about what happens to others and perceive others' interests as their own⁶. Another philosophical motivation for households' environmental awareness can be derived from the concept of "categorical imperative" proposed by Kant (1785). Individual concern for the environment might be one of the moral obligations generated by the ultimate commandment of reason called the "categorical imperative". This would make environmental concern an unconditional and self-imposed requirement that must be obeyed in all situations and circumstances.

The second specific assumption here is that consumers have not perfect knowledge about the harm they experience. As in Mason (2006), consumers cannot perfectly observe whether firms are brown or green. 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 labeling agency is responsible for disclosing information about the environmental type of firms. She has the analytical skills and technical background to acquire full information at some fixed cost, and recovers this cost by charging competitive firms a per-unit fee in exchange for either the brown or the green label. The agency however can spend the revenue from fees

⁶Two sentences from the chapter three of *Utilitarianism* (1863) illustrate this:

"There is at least a temporary feeling that the interests of others are their own interests" and "The deeply rooted conception which every individual even now has of himself as a social being, tends to make him feel it one of his natural wants that there should be harmony between his feelings and aims and those of his fellow creatures".

on objectives other than information disclosure. This makes the budget size valuable to her and fosters her opportunism. A novel problem of adverse selection may then arise, which has been so far neglected by the literature on labeling⁷: knowing that the fee determines the level of the consumer price, the agency might be tempted to manipulate consumers' beliefs on the severity of the harm from pollution through her fee choice, in order to raise revenues in excess of the spending on information. Aware of this, consumers will place little confidence in the labels from the agency. Such a scepticism in turn requires the agency's strategies to be both beneficial and credible: beneficial in the sense that information disclosure must be valuable to the agency, and credible in the sense that the fees specified for one environmental harm would not be worth imitating if the environmental harm were different.

The agency's choice of fee constrained by the budget requirement is formalized as a signaling game in which labeling is truthful as long as fees separate the environmental types of firms in equilibrium. A prominent result is that the fees charged for the green label do not necessarily coincide with that designed to raise revenue with a minimal loss in terms of efficiency. The analysis characterizes separating equilibria robust to the intuitive criterion in which the agency awards firms the green label with fees that exceed the Ramsey level.

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 the

⁷See Amacher et *alii* (2004), Crampes and Ibanez (1996), Kuhn (1999) and Rege (2000).

agency is trustworthy. Section 4 investigates the signaling model designed to address this issue and characterizes separating equilibria satisfying the “intuitive criterion” of Cho and Kreps (1987). Section 5 offers conclusions.

2 The model

Consider a horizontally differentiated market structure similar in spirit to a Hotelling model where consumers’ tastes for a good are uniformly distributed along a segment of unit length. One novelty here is that taste heterogeneity is due to negative externalities generated by the good and transportation costs resemble the harm generated by these externalities. Moreover, externalities are experience or credence attributes of the good in the sense that consumers have not perfect knowledge about the harm they experience. Externalities work through a single aggregate, called pollution for concreteness, which lowers their willingness to pay for the good. The consumer taste is represented by distance x from the good. 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 ($i = g$) that are produced using respectively a dirty and an environmentally friendly technology. 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 . All consumers agree to rank the brown variety and its green counterpart in the same way with

respect to environmental friendliness, hence $\varepsilon_b > \varepsilon_g$. 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 not restricted a priori but 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 . Hence, the market area X solves equation:

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

and the demand for variety i is given by:

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

The social harm (or environmental damage) caused by pollution is given by the sum of individual harms, that is, $\int_0^{X(p_i, \varepsilon_i)} \varepsilon_i x dx = \varepsilon_i X(p_i, \varepsilon_i)^2 / 2$.

The good is supplied by competitive firms. The technology (production plus pollution abatement) required to produce variety i will be represented by the firms' cost function $c(q, \varepsilon_i) = c(\varepsilon_i)q$, where q is the output. Unlike brown firms, green firms use advanced abatement technologies and clean up wastes. One usually expects such efforts to reduce the environmental damage at the expense of significant private costs. Hence, a natural assump-

tion is $c(\varepsilon_b) < c(\varepsilon_g)$, that is, the marginal production costs of green firms are higher than those of brown firms⁸. 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.⁹

A labeling agency is responsible for providing information about the environmental friendliness of the good, hence the actual harm from pollution. Once she has collected full information on ε_i , she can award firms ecolabels that certify their environmental type. 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_i from the observed ε_i , thereby sharing the same information as firms about their production costs, hence their environmental type. Informing consumers about the pollution harm is not the only environmental service to be supplied by the agency. She also earmarks funds for various purposes such as clean-up programs or transfers to special-interest groups. 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

⁸The 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 Wilcoxon (1990) for the U. S. economy.

⁹Indeed, $v - c(\varepsilon_b) > v - c(\varepsilon_g) > 0$ 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.

size is valuable to the agency even in the case where $I = 0$.

Consumers mistrust the environmental claims made by firms due to widespread practices of greenwashing. Moreover, there is no evidence that consumers consider the agency to be more reliable than firms. In particular, they may suspect the agency to manipulate their beliefs on the severity of the pollution harm in order to raise revenues in excess of what is strictly required for securing information. Two cases will be distinguished depending on whether consumers find the agency trustworthy or not. The agency will be said to be trustworthy when she is fully benevolent and breaks even regarding her labeling activity. The non-trustworthy agency differs in that she is likely to earmark funds for goals other than information disclosure, which makes consumers sceptical about the credibility of labeling.

In practice, ecolabeling programs entail a per-unit licensing fee (see Mason, 2006, or Crespi and Marette, 2001). This amounts to consider here that the agency charges a specific (per-unit) fee t_i on variety i to finance cost I . As there is no substitute for the available variety, the fee paid by firms turns to be an entry fee into the market, which makes labeling here compulsory¹⁰. Certification directly follows from the fee choice, that is, the label “brown” is

¹⁰As pointed by one referee, this somewhat restricts the analysis since labeling is usually viewed as voluntary (see Ben Youssef and Lahmandi-Ayed, 2008, for a definition of ecolabeling). Nevertheless, it turns out that mandatory labeling is required in some instances such as for genetically modified foods in Australia and New Zealand, water-using products in Australia since 2006 or households refrigerators by 2015 in South Africa. In the U. S., the Nutrition Labeling and Education Act of 1990 involves a mandatory disclosure of information in addition to voluntary nutrient content claims and health claims (see Caswell and Mojduszka, 1996)

associated with t_b and the label “green” with t_g . 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 $\mu_0 \equiv \text{prob}(\varepsilon_i = \varepsilon_g)$, and brown with the prior probability $1 - \mu_0 \equiv \text{prob}(\varepsilon_i = \varepsilon_b)$, $\mu_0 \in (0, 1)$.

Under perfect competition, the price paid by consumers in equilibrium is given by $p_i^e = c(\varepsilon_i) + t_i$. This level is observed by consumers as a whole, in particular, it is assumed that consumers have no way of isolating the part of p_i^e corresponding to $c(\varepsilon_i)$. Hence, p_i^e is a potential signal from which consumers can try to infer the actual harm from pollution. Note that the producer price is not a strategic variable here, and so it cannot directly play the role of a signal on the firms environmental performance as in Mahenc (2008). Consumers’ posterior beliefs will be denoted by $\mu(t_i) : \mathbb{R}^+ \rightarrow [0, 1]$ giving the probability weight the consumer attaches to the possibility that the good is green after observing the consumer price $c(\varepsilon_i) + t_i$. 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 pollution harm after observing $c(\varepsilon_i) + t_i$.

Let us introduce further notations:

- $V(\varepsilon_i) \equiv v - c(\varepsilon_i)$ is the gross consumer surplus at the equilibrium price.
- $D^i(t_i, \mu) \equiv \frac{V(\varepsilon_i) - t_i}{\varepsilon^e(\mu)}$ is the demand resulting from the inference process at equilibrium.

- $\eta \equiv -X_p(p, \varepsilon_i) p / X(p, \varepsilon_i) = \frac{p}{v-p}$ will denote the price elasticity of demand (here and throughout, subscripts denote partial derivatives and primes denote derivatives with respect to a single variable).
- $\eta^e \equiv [X_p(c(\varepsilon_i) + t, \varepsilon_i) c'(\varepsilon) + X_\varepsilon(c(\varepsilon_i) + t_i, \varepsilon_i)] \varepsilon_i / X(c(\varepsilon_i) + t_i, \varepsilon_i)$ will denote the pollution elasticity of demand at equilibrium. It measures the overall effect of harm increases 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 $\eta^e = \frac{V'(\varepsilon_i) - X(c(\varepsilon_i) + t_i, \varepsilon_i)}{X(c(\varepsilon_i) + t_i, \varepsilon_i)}$. When $V'(\varepsilon_i) - X(c(\varepsilon_i) + t_i, \varepsilon_i) < 0$, the marginal valuation of pollution harm 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 reduces the market area.

We now turn to the agency's behavior. The agency is assumed to correctly anticipate the equilibrium price of the market. She aims to maximize social welfare subject to the constraint that the revenue $R^i(t_i, \mu) \equiv t_i D^i(t_i, \mu)$ raised from fees covers all her expenditures, I included. Formally, this requires the budget constraint to be non-negative. Clearly, the severity of the budget constraint depends on consumers' beliefs. Optimistic beliefs about the pollution harm, 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 $\bar{t}_i(\mu)$

respectively the lowest and highest fee for which the agency breaks even, i. e., $t_i D^i(t_i, \mu) - I = 0$. To ensure the existence of such fees whatever the value of μ , we will restrict the parameters of the model to satisfy the following assumption

$$I \leq \frac{V(\varepsilon_g)^2}{4\varepsilon_b} \quad (3)$$

Under (3), easy calculations show that $\underline{t}_i(\mu) = \left(V(\varepsilon_i) - \sqrt{V(\varepsilon_i)^2 - 4I\varepsilon^e(\mu)} \right) / 2$ and $\bar{t}_i(\mu) = \left(V(\varepsilon_i) + \sqrt{V(\varepsilon_i)^2 - 4I\varepsilon^e(\mu)} \right) / 2$. It can be checked that, for any fee t ,

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

Hence, whatever the environmental type of firms, we have, first, that optimistic beliefs about the pollution harm loosen the budget constraint for a given harm, and second, that the budget constraint is tighter when the harm is lower for given beliefs. Note that $R_t^i(t, \mu) = \frac{V(\varepsilon_i) - 2t}{\varepsilon^e(\mu)} > 0$ for all $t < \frac{V(\varepsilon_i)}{2}$, 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.

The social welfare function $W^i(t_i, \mu)$ can be written in the following reduced form function:

$$W^i(t_i, \mu) \equiv (V(\varepsilon_i) - t_i) D^i(t_i, \mu) - \int_0^{D^i(t_i, \mu)} \varepsilon_i x dx \quad (4)$$

$$= \frac{(V(\varepsilon_i) - t_i)^2}{\varepsilon^e(\mu)} \left(1 - \frac{\varepsilon_i}{2\varepsilon^e(\mu)}\right) \quad (5)$$

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 $V(\varepsilon_i) - \varepsilon_i X^2/2$. In such a case, there would be no reason to charge a fee since $X^* = X(c_i, \varepsilon_i)$: 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 . Consequently, the derivative $W_t^i(t, \mu)$ is negative for all t such that the demand $D^i(t, \mu)$ is positive. Note that, from the social standpoint, the least favorable belief that consumers can hold is $\mu = 0$ when firms are green, whereas it is $\mu = 1$ when firms are brown.

However, even though she is benevolent, the agency can spend the money collected from fees on various objectives besides that of informing consumers. The agency's discretion about her expenditures motivates her potential opportunism. Hence, the agency might be tempted to take advantage of consumers' imperfect knowledge and divert the fee revenue from its primary purpose of informing. Aware of this, consumers have reasons to be sceptical.

tical about the certification of environmental friendliness provided by the agency. As we shall see, such a scepticism introduces further requirements that guarantee the credibility of labeling.

In the presence of sceptical consumers, the agency's objective function can be written as the following Lagrange function:

$$L^i(t_i, \mu) \equiv W^i(t_i, \mu) + \lambda_i (t_i D^i(t_i, \mu) - I), \quad (6)$$

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. Large values of λ_i will indicate that the agency takes good care about raising revenue. Her behavior then resembles more that of a profit-maximizing monopolist on the market for labeling, rather than the functioning of a competitive market with the free entry constraint $t_i D^i(t_i, \mu) = I$.

3 The benchmark with a trustworthy agency

Of central concern is how the optimal behavior of the agency is affected by consumers' imperfect knowledge of the pollution harm. As a benchmark, we record what would be the optimal behavior of a trustworthy agency under

incomplete information. The sense of trustworthiness here is twofold: the agency is benevolent and she breaks even, that is, all the revenue from the fee is pledged to pay the cost I which strictly corresponds to that of collecting information. The Lagrangian then becomes

$$\frac{(V(\varepsilon_i) - t_i)^2}{2\varepsilon_i} + \lambda_i(t_i X(c(\varepsilon_i) + t_i, \varepsilon_i) - I).$$

Proposition 1: *When the agency is trustworthy, her optimal choice consists of a fee $t(\varepsilon_i)$ and a non-negative Lagrange multiplier $\lambda(\varepsilon_i)$ such that:*

$$\frac{t(\varepsilon_i)}{p_i^e} = \frac{\lambda(\varepsilon_i) - 1}{\lambda(\varepsilon_i)} \frac{1}{\eta}, \quad (7)$$

or, equivalently,

$$t(\varepsilon_i) = \frac{\lambda(\varepsilon_i) - 1}{2\lambda(\varepsilon_i) - 1} V(\varepsilon_i) \text{ with } t(\varepsilon_b) = \underline{t}_b(0), t(\varepsilon_g) = \underline{t}_g(1) \quad (8)$$

$$\text{and } \lambda(\varepsilon_i) = \frac{1}{2} + \frac{V(\varepsilon_i)}{2\sqrt{V(\varepsilon_i)^2 - 4I\varepsilon_i}}. \quad (9)$$

Proof : (see Appendix 1)

Converting the specific fee to an ad valorem rate $\frac{t(\varepsilon_i)}{p_i^e}$ yields formula (7) 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 the agency is trustworthy, $t(\varepsilon_i)$ 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. It is worth using the explicit expression of $\lambda(\varepsilon_i)$ given by (9) to interpret (7). 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 (3), 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 (8), 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 $R_t^i(t(\varepsilon_i), \mu) = \frac{V(\varepsilon_i) - 2t(\varepsilon_i)}{\varepsilon^e(\mu)} > 0$, whatever μ . This is consistent with the evidence against the existence of Laffer effect (see Fullerton (1982)).

We end the analysis of the trustworthy agency by showing how changes in the magnitude of the harm from pollution affect the optimal 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

$$t'(\varepsilon_i) = -\frac{1}{1-\eta} \frac{t(\varepsilon_i)}{\varepsilon_i} \eta^e \quad (10)$$

Proof : (see Appendix 2)

As a result, the optimal fee $t(\varepsilon_i)$ rises as the harm from pollution is more severe. Moreover, equation (10) 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 Untrustworthy agency

Let us now consider that the agency uses fees to spending other than that strictly required by information collection. This amounts to consider that the agency can raise an amount of revenue higher than I with the fee. In addition, the information conveyed by labeling becomes doubtful since the fee may be diverted from its primary purpose of collecting information. However, consumers can use the consumer price as a signal to get information about the actual harm and the environmental type of firms. This inference process 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 actions specified for one

pollution harm could not be imitated if the harm were different. In other terms, the untrustworthy 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 $\mu^*(t_i)$ such that strategies must be optimal given consumers' beliefs. Formally, this requires that, for each $i = b, g$,

$$t_i^* \in \arg \max_{t_i} W^i(t_i, \mu^*(t_i)) + \lambda_i (t_i D^i(t_i, \mu^*(t_i)) - I). \quad (11)$$

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

$$\text{If } t_g^* \neq t_b^*, \text{ then } \mu^*(t_g^*) = 1 \text{ and } \mu^*(t_b^*) = 0; \quad (12)$$

$$\text{If } t_g^* = t_b^*, \text{ then } \mu^*(t_g^*) = \mu^*(t_b^*) = \mu_0. \quad (13)$$

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 firms type is i fails to survive the intuitive criterion if there exists a deviation d satisfying the budget constraint with

$\mu(d) = 1$, such that:

$$W^g < W^g(d, 1), \quad (14)$$

$$W^b(d, 0) \leq W^b. \quad (15)$$

Consider first the (IR) constraints. Recall that $\underline{t}_b(1)$ (resp. $\underline{t}_g(0)$) is the lowest fee for which the agency breaks even, when the true environmental type of firms is b (resp. g) and consumers believe that they are green (resp. brown) with certainty (the least favorable beliefs through the agency's eyes). The (IR) constraints can be written as follows for any t_b^* inside $[\underline{t}_b(0), \bar{t}_b(0)]$ and any t_g^* inside $[\underline{t}_g(1), \bar{t}_g(1)]$:

$$W^b(t_b^*, 0) \geq W^b(\underline{t}_b(1), 1) \quad (16)$$

$$W^g(t_g^*, 1) \geq W^g(\underline{t}_g(0), 0) \quad (17)$$

These constraints guarantee that the agency is willing to disclose information and award firms truthful labels, rather than conceal information.

Let us now turn to the (IC) constraints. They secure the credibility of labeling by imposing that the agency should not defect to the equilibrium fee that awards the wrong label. Neglecting the budget constraints, the (IC)

constraints is written as follows:

$$W^b(t_g^*, 1) \leq W^b(t_b^*, 0) \quad (18)$$

$$W^g(t_g^*, 1) \geq W^g(t_b^*, 0). \quad (19)$$

What (18) says is that the agency should not be tempted to deviate from t_b^* to t_g^* when firms are brown. In this case, such a deviation is conceivable as long as t_g^* is inside $[\underline{t}_b(1), \bar{t}_b(1)]$, otherwise the deviation would not satisfy the budget requirement consistent with the certainty that firms are green. As $\underline{t}_b(1) < \underline{t}_g(1)$, if (16) is satisfied, then (18) is satisfied as well. Condition (19) precludes a deviation from t_g^* towards t_b^* when firms are green. To be consistent with consumers' certainty that firms are brown, such a deviation must satisfy $t_g^* \geq \underline{t}_g(0)$, thereby yielding $W^g(t_g^*, 0) \leq W^g(\underline{t}_g(0), 0)$. However, by (17), this would not generate a higher welfare than that obtained with t_g^* . Thus, if (17) is satisfied, then (19) is satisfied as well. The signaling game has the unusual feature that mimicry is not an issue: from the moment that the agency fulfills the (IR) constraints, she is worse off imitating the fees specified for a pollution harm different from reality. For labeling to be credible, it suffices that the agency be willing to reveal information, that is, conditions (16) and (17) are fulfilled under the budget requirements. These

constraints can be respectively expressed as follows:

$$\frac{(V(\varepsilon_b) - t_b^*)^2}{2\varepsilon_b} \geq \frac{(V(\varepsilon_b) - \underline{t}_b(1))^2}{\varepsilon_g} \left(1 - \frac{\varepsilon_b}{2\varepsilon_g}\right) \quad (20)$$

$$\frac{(V(\varepsilon_g) - t_g^*)^2}{2\varepsilon_g} \geq \frac{(V(\varepsilon_g) - \underline{t}_g(0))^2}{\varepsilon_b} \left(1 - \frac{\varepsilon_g}{2\varepsilon_b}\right) \quad (21)$$

Let T_b (resp. T_g) denote the set of equilibrium fees t_b^* (resp. t_g^*) for which condition (20) (resp. (21)) holds, and τ_b (resp. τ_g) the solution in t_b^* (resp. t_g^*) of the equality version of (20) (resp. (21)). As any t_b^* must also satisfy the budget requirement, T_b is not empty when parameter values are such that

$$\frac{(V(\varepsilon_b) - \underline{t}_b(0))^2}{2\varepsilon_b} \geq \frac{(V(\varepsilon_b) - \underline{t}_b(1))^2}{\varepsilon_g} \left(1 - \frac{\varepsilon_b}{2\varepsilon_g}\right). \quad (22)$$

This existence condition for a separating equilibrium requires that I be sufficiently small, namely lower than $\tilde{I} = \beta V(\varepsilon_b)^2 > 0$ ¹¹.

Proposition 2: *Assume that $I \leq \min \left\{ \tilde{I}, \frac{V(\varepsilon_g)^2}{4\varepsilon_b} \right\}$. Then, separation can be achieved by any pair $(t_b^*, t_g^*) \in [\underline{t}_b(0), \tau_b] \times [\underline{t}_g(1), \tau_g]$.*

Proposition 2 establishes the conditions under which the agency is willing to implement truthful labeling so that separation of the environmental types can be achieved. There is an infinity of possible equilibrium fees that signal the true environmental types, regardless of whether firms are brown

¹¹Cumbersome calculations show that $\beta = \frac{(\sqrt{e(2-e)}-1)\sqrt{e(2-e)}(\varepsilon_g\sqrt{e(2-e)}-\varepsilon_b)}{(\varepsilon_b-\varepsilon_g e(2-e))}$, where $e = \frac{\varepsilon_b}{\varepsilon_g} < 1$.

or green. These equilibrium fees range from the Ramsey level $t(\varepsilon_i)$ characterized in Proposition 1, to fees higher than that level. At $t_i^* = t(\varepsilon_i)$, i. e., $t_b^* = \underline{t}_b(0)$ and $t_g^* = \underline{t}_g(1)$, the agency signals the true harm from pollution while raising revenue with a minimal loss in terms of efficiency. Nevertheless, this is no longer true if the agency sets $t_i^* > t(\varepsilon_i)$. Then, the agency fully reveals information to consumers with a fee that yields an amount of revenue in excess of what is needed to break even, thereby generating a welfare loss. Separation can be achieved with upward-biased fees because consumers do not internalize the changes in revenue they impose on the agency by distorting their consumption. This leaves the agency some freedom to increase her revenue with fees above the Ramsey level, while meeting the twofold requirement of awarding firms true labels and recovering the cost of collecting information. Obviously, such fees create a welfare loss by reducing too much consumption compared to what would occur with fees equal to $t(\varepsilon_i)$, but they also allow the agency to afford the cost of collecting information and finance alternative spending.

Let us now examine the existence of separating equilibrium fees robust to the intuitive criterion. Indeed, some of the separating equilibria identified above do not seem quite reasonable. Suppose for instance that separation is achieved in equilibrium by a pair (t_b^*, t_g^*) such that $t_b^* < t_g^*$, yielding welfare W^b and W^g . Then, according to the logic of the intuitive criterion, the agency may be tempted to deviate from t_g^* to $t_b^* + \epsilon < t_g^*$. Indeed, we simultaneously have $W^g < W^g(t_b^* + \epsilon, 1)$ and $W^b(t_b^* + \epsilon, 0) < W^b$, so that,

consumers should infer upon observing $t_b^* + \epsilon$ that firms are certainly green. Thus, any separating equilibrium in which $t_b^* < t_g^*$ fails to survive the intuitive criterion. Now, if (t_b^*, t_g^*) is such that $t_g^* < t_b^*$, the same reasoning shows that $t_b^* = \underline{t}_b(0)$ is the only fee paid by brown firms to be consistent with the intuitive criterion. Proposition 3 characterizes all the separating equilibrium fees consistent with the intuitive criterion.

Proposition 3: *There exists a whole range of separating equilibria satisfying the intuitive criterion such that $t_b^* = \underline{t}_b(0)$ and $t_g^* \in [\underline{t}_g(1), \underline{t}_b(0)]$.*

The untrustworthy agency can truthfully signal the environmental types of firms by behaving as if she were trustworthy. To award labels, she then charges brown and green firms respectively $t(\varepsilon_b) = \underline{t}_b(0)$ and $t(\varepsilon_g) = \underline{t}_g(1)$, thereby raising revenue with a minimal loss in terms of efficiency. However, unlike the trustworthy agency, the untrustworthy agency can also charge green firms a fee t_g^* higher than $\underline{t}_g(1)$ to the extent that t_g^* remains below $\underline{t}_b(0)$. Consumers' scepticism is then substantiated by a whole range of equilibrium fees that simultaneously reveal that firms are green and raise an amount of revenue in excess of what is required to balance the cost of collecting information. Clearly, such fees create a welfare loss compared to the situation prevailing with fees set at the Ramsey level. Nevertheless, the refinement of the intuitive criterion fails to eliminate the separating equilibria in which t_g^* is biased upward with respect to $\underline{t}_g(1)$. To obtain full information through labeling, sceptical consumers must pay for the cost of collecting information, but they are also likely to pay for alternative goals pursued by the agency.

5 Conclusion

This paper has examined the signaling role of labeling in a context where consumers lack information about the environmental type of firms, that is, the actual harm caused by the good they sell. In response to the standard problem of adverse selection due to firms' opportunism, a labeling agency can acquire full information at a cost and charge a fee to award firms green or brown labels. This may raise another problem of adverse selection now due to the agency's opportunism and the consumers' scepticism about the label trustworthiness. To the extent that she cares about the size of her revenue, the agency might be tempted to raise an amount of revenue in excess to what is required solely by the cost of collecting information. It turns out that such a wasteful behavior may emerge in equilibrium with the fees signaling green firms.

The analysis characterizes the separating equilibrium fees associated with truthful labeling and shows that the credibility of labeling is not an issue because mimicry is worthless. Although the incentive-compatibility constraints required by the concept of Bayesian equilibrium are not binding, the individual-rationality constraints give rise to an infinity of separating equilibria. Moreover, there exists a whole range of separating equilibria robust to the intuitive criterion, in which the fees used to truthfully signal green firms do not necessarily coincide with that designed to raise revenue with a minimal loss in terms of efficiency. In other terms, revealing that firms are

green is consistent with fees that are biased upward relative to the Ramsey level. Hence, labeling green firms is likely to generate wasteful expenditures when consumers do not trust the labeling agency.

As previously mentioned, one limit of the present model is that labeling is mandatory since firms cannot enter the market unless they pay the labeling fee. To allow for voluntary labeling, one possibility is to address the signaling issue in a model of vertically differentiated products à la Shaked and Sutton (1982) where labeling serves the dual task of differentiating products and transmitting information on quality.

6 Appendix

6.1 Appendix 1: Proof of proposition 1

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

$$-\frac{V(\varepsilon_i) - t}{\varepsilon_i} + \lambda_i (X(c(\varepsilon_i) + t, \varepsilon_i) + tX_p(c(\varepsilon_i) + t, \varepsilon_i)) = 0, \quad (23)$$

$$\lambda_i (tX(c(\varepsilon_i) + t, \varepsilon_i) - I) = 0, \quad (24)$$

$$\lambda_i \geq 0. \quad (25)$$

Using the expression of demand (2), condition (23) can be rewritten

$$-(V(\varepsilon_i) - t) + \lambda_i (V(\varepsilon_i) - 2t) = 0 \quad (26)$$

This equation yields $t(\varepsilon_i) = \frac{\lambda(\varepsilon_i)-1}{2\lambda(\varepsilon_i)-1}V(\varepsilon_i)$ in (8). Substituting $X(c(\varepsilon_i) + t, \varepsilon_i)$ to $\frac{V(\varepsilon_i)-t}{\varepsilon_i}$ in the left-hand side of (23), we get (7). The equation derived from the budget constraint (24) admits an upper and lower root in t , 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 (26) 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 $t(\varepsilon_b) = \underline{t}_b(0)$ and $t(\varepsilon_g) = \underline{t}_g(1)$ in (8). The expression of $\lambda(\varepsilon_i)$ given in (9) 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 $t(\varepsilon_i) = \left(V(\varepsilon_i) - \sqrt{V(\varepsilon_i)^2 - 4I\varepsilon_i}\right)/2$ provided that $\lambda_i > 0$, where $t(\varepsilon_i)$ is the lowest root of $tX(c(\varepsilon_i) + t, \varepsilon_i) = I$. Differentiating the budget equation $tX(c(\varepsilon_i) + t, \varepsilon_i) = I$ with respect to t 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 $I \in \left[0, \frac{V(\varepsilon_g)^2}{4\varepsilon_b}\right]$,

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

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)

$$t(X_p c'(\varepsilon) + X_\varepsilon) d\varepsilon + (X + tX_p) dt = 0. \quad (27)$$

Rearranging terms, we get

$$(1 + tX_p/X) t'(\varepsilon) = -t(X_p c'(\varepsilon) + X_\varepsilon)/X \quad (28)$$

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 (7). From the definition of η^e , we obtain

$$t'(\varepsilon_i) = -\lambda(\varepsilon_i) \frac{t(\varepsilon_i)}{\varepsilon_i} \eta^e. \quad (29)$$

Replacing $\lambda(\varepsilon_i)$ by $\frac{1}{1-\eta}$ gives (10).

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] BEN YOUSSEF, A., and R. LAHMANDI-AYED (2008), “Eco-labelling, Competition and Environment: Endogeneization of Labeling Criteria”, *Environmental and Resources Economics*, 41, 133-154.
- [3] BOULDING, W. and A. KIRMANI (1993), “A Consumer-Side Experimental Examination of Signaling Theory: Do Consumers Perceive Warranties as Signals of Quality?”, *Journal of Consumer Research* 20, 111-123..
- [4] 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.
- [5] CASWELL, J. A. and E. M. MOJDUSZKA (1996), “Using Informational Labeling to Influence the Market for Quality in Food Products”, *American Journal of Agricultural Economics* 78, 1248-1253.
- [6] CHO, I-K. and D. KREPS (1987), “Signaling Games and Stable Equilibria”, *Quarterly Journal of Economics* 102, 179-221.

- [7] CRAMPES, C. and L. IBANEZ (1996), “The Economics of Green Labels”, Toulouse: University of Toulouse, G.R.E.M.A.Q. Papers 96.439.
- [8] CRESPI, J. M. and S. MARETTE (2001), “How Should Food Safety Certification Be Financed?”, *American Journal of Agricultural Economics* 83, 852-861.
- [9] FULLERTON, D. (1982), “On the Possibility of an Inverse Relationship between Tax Rates and Government Revenues” 19, 3-23.
- [10] IBANEZ L. and A. STENGER (2000), “Environmental and Food safety in Agriculture: Are Labels Efficient?”, *Australian Economic Papers*, Vol. 39, pp. 452-464.
- [11] JORGENSON, D. W. and P. J. WILCOXEN (1990), “Environmental Regulation and U.S. Economic Growth”, *Rand Journal of Economics*, Vol. 21, No. 2, pp. 314-340.
- [12] KARL H. and C. ORWAT. (1999), “Economic Aspects of Environmental Labelling” in H. Folmer and T. Tietenberg (eds), *Yearbook of Environmental and Resource Economics 1999/2000*, Elgar, Cheltenham, UK, 107-170.
- [13] KUHN, M. (1999), “Green Lemons - Environmental Labels and Entry into an Environmentally Differentiated Market under Asymmetric Information”, University of Rostock: Thuenen-Series of Applied Economic Theory 20.

- [14] KUHN, M. (2005), *The Greening of Markets: Product Competition, Pollution and Policy Making in a Duopoly*. Cheltenham: Edward Elgar.
- [15] LOMBORG, B. (2001), *The Skeptical Environmentalist*, Cambridge University Press.
- [16] LYON, T. P. and J. W. MAXWELL (2006), "Greenwash: corporate environmental disclosure under threat of audit", Working paper, Ross School of Business, University of Michigan.
- [17] MAHENC, P. (2008), "Signaling the Environmental Performance of Polluting Products", *International Journal of Industrial Organization*, Vol. 26, 1, 59-68.
- [18] MASON, C. F. (2006), "An Economic Model of Eco-Labeling", *Environmental Modelling and Assessment* 2006 11(2), 131-143.
- [19] McCLUSKEY, J. J. (2000), "A Game Theoretic Approach to Organic Foods: An Analysis of Asymmetric Information and Policy", *Agricultural and Resource Economics Review* 29, 1-9.
- [20] PALMER, K., W. E. OATES and P. R. PORTNEY (1995), "Tightening Environmental Standards: The Benefit Cost or the No-Cost Paradigm?", *Journal of Economic Perspectives* Vol. 9, N° 4, 119-132.
- [21] REGE, M. (2000), "Strategic Policy and Environmental Quality: Helping the Domestic Industry to Provide Credible Information", *Environmental and Resource Economics* 15, 279-296.

- [22] RILEY J. (1979), “Informational Equilibrium”, *Econometrica* 47, 331-360.
- [23] SHAKED, A. and J. SUTTON (1982), “Relaxing Price Competition through Product Differentiation”, *Review of Economic Studies*, 49, 3-13.
- [24] SPENCE, A. M. (1974), *Market Signaling*. Cambridge: Harvard University Press.