



**HAL**  
open science

# Information communication about environmental quality by markets and NGOs

Alexandre Volle

► **To cite this version:**

Alexandre Volle. Information communication about environmental quality by markets and NGOs. Economics and Finance. Université Montpellier, 2019. English. NNT: 2019MONTD028. tel-02890603

**HAL Id: tel-02890603**

**<https://theses.hal.science/tel-02890603>**

Submitted on 6 Jul 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Sciences Économiques

École doctorale d'économie et de gestion

Unité de recherche Center for Environmental Economics - Montpellier

## Information communication about environmental quality by markets and NGOs

Présentée par Alexandre Volle

Le 04/12/2019

Sous la direction de Philippe MAHENC

Devant le jury composé de

David MARTIMORT, Professeur des Universités, Paris School of Economics

Mireille CHIROLEU-ASSOULINE, Professeur des Universités, Paris School of Economics

Patrick GONZÁLEZ, Full Professor, Université Laval

Philippe MAHENC, Professeur des Universités, Université de Montpellier

Rapporteur

Rapporteur

Examineur

Directeur de thèse



UNIVERSITÉ  
DE MONTPELLIER

# Acknowledgment

First, and most of all, I would like to thank Professor Philippe Mahenc, for his capacity to transmit his knowledge, assistance, guidance, and patience throughout the process of writing this thesis. Moreover, I have to emphasize that I could not expect a better Ph.D. advisor knowing his benevolence.

I directly extend my gratitude to Professor Patrick Gonzalez to welcome me in University Laval. Moreover, I learned a lot thanks to his capacity for well defining economic problems in a general context. In this spirit, I would like to thank Mireille Chiroleu-Assouline without who I would not enter into a Ph.D. project, and her kindness to involve me in a very exciting research project.

I thank my dear colleague Bilal Elrafhi, for the insightful discussions we had in our small office for 3 years. I also thank Ayad Assoil, Antoine Deymier, Romain Fayat, Agnalys Michaud, Boumedienne Souiki, and Maëté Stefan, who all participated in creating a supportive atmosphere to accomplish a Ph.D.

If I have the capacity to accomplish this work, it is obviously thanks to my parents, Yaël and Alain, and I have no words to be grateful enough. A particular thought to my brother Hadrien, who enlightened me to the academic world. I want to thank all my friends for their kindness and their joie de vivre. At last, I want to thank my beloved Pauline, because, in her smile, I found all the answers I wondered these last 10 years.

# Preface

This dissertation focuses on the path of information about environmental quality and its interaction with regard to firms and Non Governmental Organizations (NGOs), but some of my work is more general and draws results on the credibility itself. More precisely, it contains three theoretical articles organized as three chapters.

The first chapter, Why is Price Useless to Signal Environmental Quality, has its root in a context where the green certification is challenged and questions other plausible tools to transmit information. This article was presented at the 17th Institutional and Organizational Economics Academy in April 2018.

The second chapter, Watching the Watchers: Signaling Social Goodwill with Imperfect Monitoring, is a project cowritten with Professor Philippe Mahenc. This article has its roots in the common project research project GREENGO between the Paris School of Economics with Professor Mireille Chiroleu-Assouline and the University of Montpellier.

The third chapter, Incidence of the Championing and Shaming Strategies of an NGO on the Signaling Strategy of a Firm is a project stemming from my visiting at University Laval to Professor Patrick González. This paper has been presented at Laval University, and at the meeting between the University Montpellier and the University Sherbrooke.

# Contents

|  |           |
|--|-----------|
| <b>Introduction</b>  | <b>1</b>  |
| <b>1 Why is Price Useless to Signal Environmental Quality ?</b>  | <b>12</b> |
| 1 Introduction . . . . .   | 15        |
| 2 Relation to the Existing Literature . . . . .  | 17        |
| 3 Model . . . . .  | 19        |
| 3.1 Perfect information . . . . .  | 19        |
| 3.2 Adverse selection . . . . .  | 21        |
| 3.3 Separating equilibrium . . . . .   | 23        |
| 3.4 Pooling equilibria . . . . .   | 24        |
| 3.5 Intuitive Criterion vs Undefeated Equilibrium . . . . .  | 26        |
| 3.6 Extension: a share of informed consumers . . . . .   | 28        |
| 4 Discussion . . . . .   | 30        |
| 5 Appendix . . . . .   | 32        |
| 5.1 Proof of Proposition 3. . . . .  | 32        |
| 5.2 Proof of Proposition 4. . . . .  | 33        |
| <b>2 Watching the Watchers: The Credibility of Signaling Social Goodwill<br/>with Imperfect Monitoring</b> | <b>35</b> |
| 1 Introduction . . . . .   | 38        |
| 2 Related literature . . . . .   | 43        |

|          |  |            |
|----------|--|------------|
| 3        | The model . . . . .  | 47         |
| 4        | The signaling game with hidden actions . . . . .   | 54         |
| 4.1      | Price signaling with no monitoring: the case $\alpha = 0$ . . . . .  | 70         |
| 4.2      | Price signaling with monitoring: the case $\alpha > 0$ . . . . .   | 73         |
| 5        | Conclusion . . . . .   | 82         |
| 6        | Appendix . . . . .   | 85         |
| 6.1      | Appendix 1: Demand functions . . . . .   | 85         |
| 6.2      | Appendix 2: Proof of Lemma 4 . . . . .   | 87         |
| <b>3</b> | <b>Incidence of the Championing and Shaming Strategies of an NGO on<br/>the Signaling Strategy of a Firm</b> | <b>92</b>  |
| 1        | Introduction . . . . .   | 95         |
| 2        | Illustrations . . . . .  | 99         |
| 3        | The model . . . . .  | 102        |
| 4        | NGO informative behavior . . . . .   | 114        |
| 5        | Conclusion . . . . .   | 117        |
|          | <b>Conclusion</b>  | <b>119</b> |
|          | <b>Bibliography</b>  | <b>123</b> |

# Introduction

The starting point of this thesis is the fact that the credibility of environmental claims is increasingly questionable. The asymmetric information problem between firms and consumers in green markets mostly relies on the green certification efficiency yet to be solved. If green certification starts to be questioned, the “Costs of Dishonesty” ([Akerlof, 1970](#)) -due to the persistence of asymmetric information- may jeopardize environmental markets. The cost of dishonesty does imply not only the amount by which the consumer has been cheated but also the loss incurred from crowding out legitimate green business out of the market. Green goods are endowed with *credence* characteristics, meaning that environmental characteristics remain unobserved even after consumption. In this way, credence goods rely on faith, and therefore, the problem of information is even more acute. To explain on which background the thesis finds its roots, we draw a general picture of the key factors contributing to the credibility of the label in a general context.

## Context

**Reputation.** The first is related to the matter of the reputation of the intermediary and the firm, which is a main determinant on this issue. In a general framework, [Tirole \(1996\)](#) underlines that an individual reputation is determined by collective reputation. This principle has been reused in an environmental issue where [Heyes, Lyon, and Martin \(2018\)](#) explain that the perception of a firm may be affected by the reputation of the industry. If the firm faces an adverse selection, an intermediary that offers

such a reputation mechanism can be of benefit to high-quality firms (Belleflamme and Peitz, 2015). Reputation is built on repeated interactions between agents and can be spread in the market through prices, advertising (Klein and Leffler, 1981; Shapiro, 1983) and word of mouth or referrals: when consumers tell other consumers about their experience (Rob and Fishman, 2005). Nevertheless, Mathis, McAndrews, and Rochet (2009) showed that, in the financial market, some certifying agencies may be tempted to benefit from reputation confidence cycles. Moreover, while the reputation argument may be relevant in some cases, it is reasonable to argue that consumers are not always aware of the reputation of the third party responsible for the certification.

**The level of stringency.** The level of the stringency of the standard is crucial to gain consumers' confidence. According to the Steering Committee of the State-of-Knowledge Assessment of Standards and Certification (2012), most sustainable production standards, which benefit from the best credibility, emerged as a result of a conscious effort by a small group of nonprofit organizations to convene and engage a cross-section of stakeholders within a given sector. Theoretical papers tackling this issue usually assume that the NGO that sets this level has a trade-off between a better environmental quality of the label *per se* and the number of firms adopting the label. Heyes and Maxwell (2004) analyze the impact of the interactions between private and public regulation; the presence of an NGO may either increase or decrease the standard quality set by the producers' associations. Along the same lines, Fischer and Lyon (2014) analyze the influence of the competition on the level of the stringency of the label between an NGO label and a firm label. The effect of competition on the level of stringency is not straightforward when an industry label enters a market where the NGO was operating alone. The NGO has a trade-off between extracting a more considerable environmental effort from its remaining firms and its willingness to regain market share. All these potential effects that would enhance the "race to the bottom" of the level of the standard have an influence on the credibility of the certification.

Moreover, while most of these kinds of studies deal with an assumption of a perfect green certification (see [Bonroy and Constantatos \(2008\)](#) for a complete review on these subjects), [Mason \(2011\)](#) allows the possibility of the certification being noisy. In this paper a noisy test is run on firms, which can be subject to two types of errors, rejecting green producers or accepting brown firms that fuzzy the level of stringency of the label.

**Groucho effect.** This effect takes place when standards are uncertain. If a product that is expected to be of low-quality obtains the ecolabel, then consumers infer that the standard is probably weak if such a product earns it ([Harbaugh, Maxwell, and Roussillon, 2011](#)).<sup>1</sup> Figure 1 underlines the significant increase in the number of ecolabels in recent years, which clearly enhances this uncertainty.<sup>2</sup>

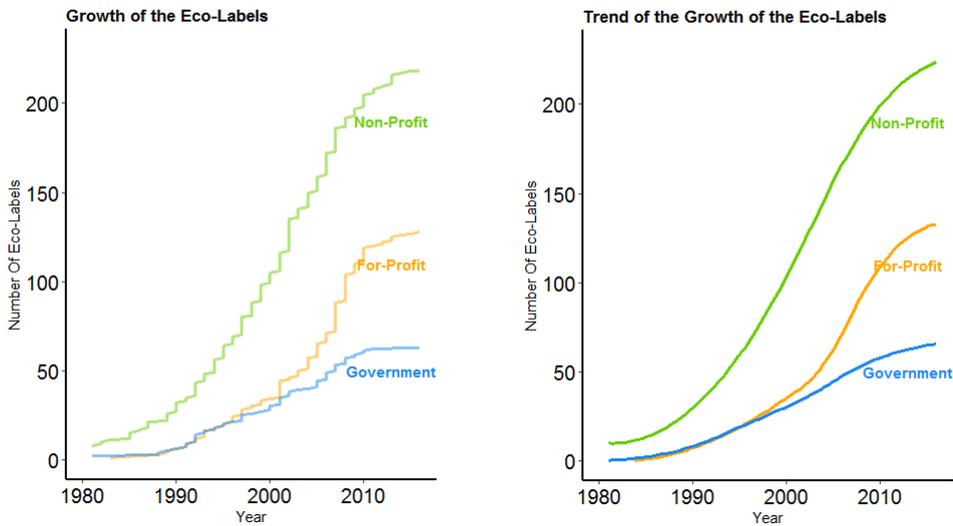


Figure 1: The number of ecolabels has rocketed in the past decades<sup>3</sup>

Moreover, [Civel, Cruz, et al. \(2018\)](#) underline a heterogeneity of perception among consumers toward energy certification. They show that age and previous certification experiences tend to reduce perceived quality and credibility *per se*, of the signal sent

<sup>1</sup>The authors named it after Groucho Marx because he famously joked that he did not want to join a club with standards so low that it would accept him as a member.

<sup>2</sup>All the data have been collected with the help of Python at <http://www.ecolabelindex.com/>

<sup>3</sup>[Heyes and Martin \(2016\)](#) provides an explanation of this label profusion concerning NGOs

by the label. Figure 1 also leads to another pillar of credibility: the issue of the type of stakeholder involved in the certification process.

**The type of third party responsible for the certification.** Generally, it relies on the opportunistic aspect of agents capable of the transmission of green information. As stated in the Steering Committee Report: “Ideally, an ecolabel should be grounded in a diverse multi-stakeholder process that includes consumers”. A main issue with the credibility is associated with the influence of the type of the third party responsible for the certification. Mahenc (2017) shows that, depending on whether the third party maximizes its profit or social welfare, green certification may or may not be misleading. Looking closer at the financial structure of the NGOs ecolabels compared to that of for-profits ecolabels reveals that concerning the certification activity, there does not seem to be many differences.

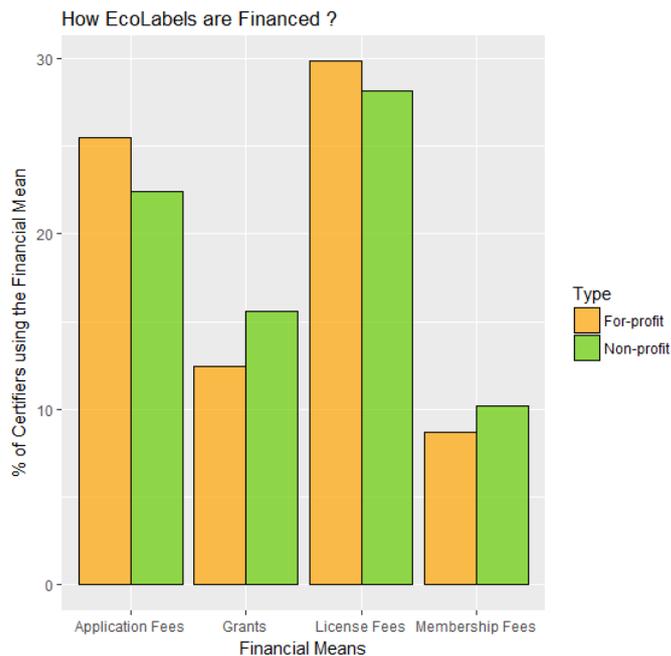


Figure 2: Financial Structure of Ecolabels

Figure 2 underlines that green certification generates funds, and the allocation of these funds is crucial for credibility. Unfortunately, because of the limitation of the

database collected, we cannot know how much these resources weigh in the balance. This graph still raises questions about which stakeholder provides the grants, and about the financial structure of the certification, which seems to depend on the sellers asking for labeling services.

Finally, it is essential to underline that the green demand conditions the certification market. Therefore, if we consider the geographical spatialization of green certifications, we can determine which are the most important markets associated with ecolabeling.

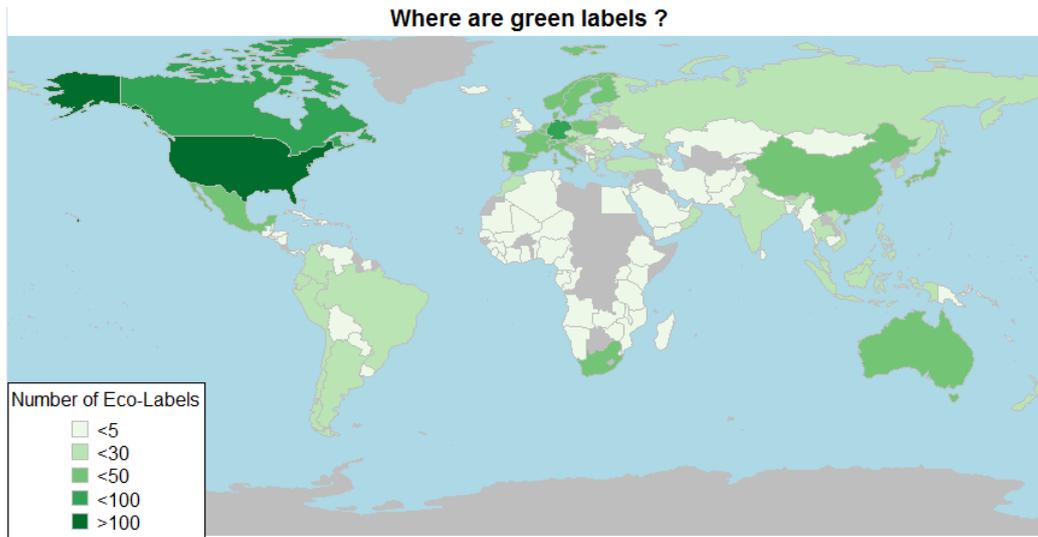


Figure 3: Where are ecolabels

This picture also underlines that some areas of the world have no green certification to provide a guarantee on environmental information.

It leads to a natural question: How is environmental quality information transmitted when certification is not credible or does not exist? What are the viable tools that can bear this role? The idea behind using the price signal in such a context is to consider that the firm has a strategic influence to convince the consumer about its type. Most of the studies skipped this aspect ([Feddersen and Gilligan, 2001](#); [Fleckinger, Glachant, and Moineville, 2017](#)) considering that the firm was undergoing its information environment.

## Structure of the document

This Ph.D. thesis focuses on three main questions: Can the information about environmental quality depend on price? What is the influence of auditors on the signaling strategy of firms? What is the optimal informative behavior of an NGO to credibly promote environmental practices?

To answer these questions, we have built a simple signaling theoretical model that explores the role of the price in signaling the environmental commitment of a firm in a green market (Chapter 1). When certification is not credible, price alone fails to bear the role of the signal, and markets with greenwashing are the most likely to appear. We have then applied to this base model with an activist who provides hard evidence of the signaling model of firms (Chapter 2). The sole presence of the activist reduces the incentive of the low environmental quality producer to cheat consumers. Therefore, it decreases the cost of signaling for firms with high environmental performance. Less acknowledged in the literature, the influence on the initial problem of asymmetry of information directly helps solve the moral hazard on the choice of the technology of firms. Finally, we analyzed the influence of the different informative strategies on the signaling equilibrium path. Then, we endogenized the strategies of the organization that can behave either as a “Good cop ” by championing the virtuous producers, or as a “Bad cop ” by shaming the polluting producers (Chapter 3). Even if both strategies consist of finding hard evidence on the type of the firm, their effect on the signaling behavior of firms is radically different. Moreover, depending on the incentives of the market to reveal the truth or not, the NGO adapts its behavior to maximize the information provided to consumers.

While all papers deal with very similar research topics, each article puts forward distinct research questions and can be read on its own.

## Signaling credence attributes under not Credible Green Certification

Green goods are often endowed with *credence attributes*, which is a characteristic that cannot be verified even after consumption. Providing credence attributes under asymmetric information raises the well-known lemons problem (Akerlof 1970). The results presented in **Chapter 1** would be consistent in a more general framework as long as the two main hypotheses of the analysis hold.

This paper relies on the following idea: if the green certification is not credible, can we rely on the old idea of [Milgrom and Roberts \(1986\)](#) where the prices can be used as signals<sup>4</sup>. While situations in which different firms have market power have been considered ([Daughety and Reinganum, 2008b](#); [Mahenc, 2007](#)) in the literature on the informational role of prices, these situations rarely occur. In many sectors, one firm has a greater market share than the remaining firms ([Mirman and Santugini, 2014](#)). Especially in the green differentiation, it is well known that a direct incentive to commit to green production is to benefit from the market share ([Bonroy and Constantatos, 2008](#)) and thus exiting the very competitive segment.

We develop a standard signaling model that respects general assumptions, and we seek to identify the role of the price in signaling environmental quality. There are two firms, and their types are given independently exogenously. They compete à la Bertrand in a vertically differentiated market. In this spirit, it is more expensive to produce the high-quality (green) good than the low-quality (brown) good, and its marginal cost is normalized to 0. The particularity is that if a firm decides to compete on the brown segment, it earns a null profit because of the idea of the long-term competitive fringe. There is a continuum of heterogeneous Bayesian consumers where some of them are willing to pay more for ecofriendly products. They cannot verify the

---

<sup>4</sup>Completing the reputation explanation, ([Klein and Leffler, 1981](#); [Shapiro, 1983](#)) where past prices signal the current quality

quality produced by the firms but can update their beliefs after observing the prices.

The model describes the structure of signaling equilibria emerging from this framework. There does not exist situations where the green producer can correctly transmit information about its quality. The intuition is the following: because the brown firm would earn a null profit if it reveals the truth, it has strong incentives to mimic any prices of the green segment even if the prices are going high. As a result, the only price that would permit transmitting information correctly is too high such that even the most environmentally motivated consumer would not be ready to pay for it. Depending on the choice of the criteria to reflect different ways of reasoning of consumers, the green market can exist with greenwashing if the a priori of consumers is relatively sufficient, or the market might disappear due to the previously described failure. Facing this extreme outcome, we add the possibility for some consumers to be initially informed ([Bagwell and Riordan, 1991](#)) which may help the firm signal its quality; nevertheless, the failure holds. This paper insists on the importance of the credibility of the certification to transmit information because the price would fail to correctly send information in such a context. As a result, the lemons outcome would be very likely to occur. Nevertheless, one of the limits of the paper is that there remains some hard evidence on the market, even if not entirely credible, and in many cases, green certification still bears its informational role sufficiently. We adapted the model to take into account the third party, which releases hard evidence to consumers about the environmental quality of a production process.

## **The Impact of Non-Governmental Organizations in a Signaling Environment**

As underlined by precedents, even imperfect, NGOs play an essential informative role that impacts firms informative behavior. The informative role of an NGO can be distinguished in two strategies “Good cop/Bad Cop” ([Lyon, 2012](#)). If the NGO behaves as

a “Good Cop”, it works hand-in-hand with industry, for instance, it can design certification. In this case, the economic literature mostly assumes that certification is perfect; therefore, the NGO influences only the differentiation aspect (Amacher, Koskela, and Ollikainen, 2004; Lyon, 2012) as described by Shaked and Sutton (1982). Otherwise, the NGO behaves as a “Bad Cop” and can monitor either firms or certification itself. In this case, the literature analyzes how the sanctions shape the behavior of firms (Baron, 2001; Heyes et al., 2018). Concerning information provision Feddersen and Gilligan (2001) and Fleckinger et al. (2017) showed how different informative environments impact the technology choice of the firm. Nevertheless, in these papers, firms cannot strategically signal their type to consumers. The problem of asymmetric information works in tandem with the moral hazard problem that arises in this situation: the incentives for a firm to engage in a clean production process. It is worth noting that the asymmetric information issue is often viewed as a problem dealing with private characteristics; environmental characteristics turn out to be public characteristics. It remains a void concerning analysis about the interplay between the signal strategy of a firm and the signal provided by the NGO. Moreover, the fact that the type is not chosen by Nature, which is the case in most of the signaling models of similar structure, is not innocuous.

To do so, we build on the model introduced in the previous Chapter, and we open the possibility (1) for an NGO to provide hard evidence in the signaling model and (2) for firms to choose their technology. The information provision given by the NGO is assumed to be exogenous. The perception rule of the consumer is crucial to understand the interplay between the two sources of signals. In this Chapter, the hard evidence released by the NGO works as a Bayesian refinement on the out-of-equilibrium belief. To address the moral hazard issue, we extend the standard signaling model of asymmetric information with hidden knowledge to a model of symmetric information with hidden actions, in which imperfect monitoring by a third party is allowed. The parameter of

detection  $\alpha$  represents the capacity of an NGO to find the correct technology of the firm. We use the model to analyze different cases of equilibrium depending on this level of imperfect efficiency.

We show that the information provided by the NGO solves the previous failure established in the first Chapter. Even with imperfect detection technology, the NGO raises plausible situations where firms can signal their quality again. Therefore, its simple presence permits the market to signal its quality, and an immediate consequence is giving incentives for firms to choose to produce the green good. The reason is that monitoring raises the cost of cheating consumers with misleading prices, and therefore motivates a low-quality firm to reveal the truth about its type. Moreover, due to the NGO, opening up the possibility for the market to solve the asymmetric information problem, it has a direct consequence on the moral hazard issue. In terms of technology choice, it raises a pure asymmetric strategy, i.e., a firm chooses to engage in green production, while the other decides to engage toward brown production and a mixed strategy for both agents. More precisely, the capacity of the NGO to find hard evidence  $\alpha$  also relaxes pressure for the green firm to signal its quality *via* price. If the firm failed by signaling its quality *via* prices due to the fly-by-night behavior of the rival, for instance, the level of information  $\alpha$  would permit the ecofriendly producer to sustain on the green market without signaling by itself. The model highlights the importance of interactions between two different sources of signals.

The end of the analysis leads to a natural extension, which is the endogenization of the strategy of the NGO in such contexts, and the division of the type of the hard evidence discovered in two categories.

## The Optimal Informative Strategy of Non-Governmental Organizations when Firms Can Influence Their Signals

After having analyzed the impact of hard evidence on the informative behavior of firms, we now want to understand how the NGO behaves in this context. The informative strategy is now endogenous, and the technology choice is exogenous again.

We go further than the previous analysis by distinguishing the nature of the hard evidence discovered. The NGO has a probability  $\gamma$ , if it decides to adopt a championing strategy, to discover proof that the firm is truly green and a probability  $\beta$ , if it decides to adopt a shaming strategy, to discover a proof that the firm is definitely brown. Compared to those in the previous analysis, the beliefs in this analysis are only updated according to the Bayes's rule given the two signals received, price and the hard evidence of the NGO.

We aim to understand the optimal informative strategy of an NGO in such context.

Shaming and championing, despite sharing a common nature, that is the discovery of hard evidence, operate on radically different channels shaping the equilibrium strategies. Shaming helps restore the complete credibility and, if sufficiently efficient, permits the perfect information outcome to be reached by deterring the fly-by-night behavior of the low-quality firm. Surprisingly championing, by relaxing the necessity for the high-quality firm to signal, can crowd out market revelation. As soon as the certification is efficient enough, the market can no longer signal the truth other than through certification. The intuition behind this result is that at some point the level of the price that would incite the green producer to reveal its type is no longer sufficient to deter the brown producer from mimicking the high segment price. Concerning the strategy adoption of the NGO, it behaves in accordance with the proportion of firms; typically, we are not likely to observe an NGO practicing championing in a very polluting industry. Conversely, we expect to observe championing in sectors where there are sufficient green production initiatives, such as agriculture.

# Chapter 1

## Why is Price Useless to Signal Environmental Quality ?

## English summary

In a context where the credibility of green certification is questioned, the present paper investigates the role of price as a possible substitute channel of communication as it has been largely developed since [Milgrom and Roberts \(1986\)](#). In this article, the purpose is to examine how a green firm competing against a brown firm can solve the asymmetric information issue. For this purpose, we develop a simple vertically differentiated duopoly model where firms can signal their quality through prices. To represent an increasing pressure of the competition on the low-quality segment, we make the hypothesis that the brown good is sold at the marginal cost. The result is extreme: there do not exist any situations where the green firm can solve the asymmetric information issue. As a result, pooling price equilibria emerge as the most plausible situations as long as the brown firm can mimic the pricing behavior of the green firm. Moreover, depending on the equilibrium refinement, we find that the market can either disappear or sustain in the presence of greenwashing. A green producer is thus constrained to practice uninformative prices, which can lead to the lemons outcome ([Akerlof, 1970](#)).

## Résumé français

Dans un contexte où la crédibilité de la certification verte est remise en question, ce travail analyse le rôle du prix comme un moyen de communication alternatif, comme cela a été développé par [Milgrom and Roberts \(1986\)](#). Dans cet article, le but est d'examiner comment une firme verte en concurrence avec une firme brune peut résoudre le problème d'asymétrie de l'information. Pour cela, nous développons un modèle de duopole verticalement différencié dans lequel les firmes peuvent signaler leur qualité à travers les prix. Afin de représenter une concurrence accrue sur le segment de basse qualité, nous faisons l'hypothèse que le bien brun est tarifé au coût marginal. Le résultat est extrême: il n'existe aucune situation où la firme verte peut résoudre le

problème d'asymétrie de l'information. En conséquence les prix mélangeants émergent comme les situations les plus plausibles tant que la firme brune a la possibilité d'imiter le comportement en prix de la firme verte. De plus, en fonction du critère de raffinement d'équilibre que l'on choisit, nous trouvons que le marché peut, soit disparaître, soit perdurer, avec la présence de "greenwashing". Un producteur vert est donc contraint de pratiquer des prix non informatifs qui peuvent conduire au résultat du "lemon" ([Akerlof, 1970](#)).

# 1 Introduction

Can the price transmit information on the environmental quality of a credence good?

The idea that the price can ensure the correct transmission of information has been largely developed since [Milgrom and Roberts \(1986\)](#). In this literature ([Bagwell and Riordan, 1991](#); [Mahenc, 2007](#); [Daughety and Reinganum, 2008b](#); [Mirman and Santugini, 2014](#)) it is shown that the price ensures the revelation of information. However, authors have to do important specifications to ensure the existence of the market and to raise the possibility of firms revealing their type.

In this article, we examine the consequence of asymmetric information on the price signal in the context of a duopoly, whereby it produces vertically differentiated substitute goods according to an environmental attribute. Firms quality is exogenously determined; it can be green (high) or brown (low), and they operate in a one-shot (three periods) model. The main specification is that the brown market is perfectly competitive.

We find that firms do not use separating price and it implies that price cannot transmit any information about the type of the firm when one of the two segments is perfectly competitive. The selection of plausible equilibrium by some refinement criteria strengthens this result.

While it appears impossible for the green firm to reveal its type, it is still possible that the market exists by forcing the green firm to pool with the brown type. Pooling equilibria, in the case of green markets, may be assimilated to greenwashing as long as these equilibria correspond to a situation where the brown firm makes a misleading claim about the environmental benefits of a product. The recent Volkswagen scandal is a prime example of this where, as stated in [Forbes](#)<sup>1</sup>: “company leaders had the option to attempt to reap the benefits of being green by merely giving the appearance

---

<sup>1</sup>Forbes 15/09/2016: How To Succeed At Sustainability (And Why Greenwashing Doesn't Work)

of an environmentally friendly company on the outside, without making the necessary investments on the inside ”. Moreover as mentioned in the Guardian<sup>2</sup> the means used for greenwashing can take different forms, such as a combination of advertising and higher priced products<sup>3</sup>, as long as it provides information that misleads consumers into thinking they are helping the environment.

Fortunately, the transmission of the correct information on environmental quality can notably rely on certification (Bonroy and Constantatos, 2008) to solve the problem of asymmetric information and to ensure the existence of green markets. There exist natural links between the price and the certification. One of the main argument of the certifiers for label adoption is that consumers are ready to pay a price premium for a green product. Thereby, when a green product is credibly identified, green producers can extract this surplus by charging a higher price. Nevertheless recently there has been an increasing phenomenon of lack of trust toward ecolabels<sup>4</sup> (Jacquet et al., 2010) and green certification (Hamilton and Zilberman, 2006; Mason, 2011). Washington and Ababouch (2011) at the same time made a report wherein they questioned the reality of price premium associated with ecolabels in the fish industry. While there is evidence of this premium for fish retailers, this relation is not established for fish producers (Blomquist, Bartolino, and Waldo, 2015). The report mentioned that there has been no price premium gained from certification for the case of Alaska pollock and pointed to the fact that uncertified Russian pollock was fetching similar prices on the European market. Albeit it is important to mention that in this market the over-riding factor that sets price is still quality, it remains that environmental groups raised objections about the sustainability of the fisheries concerning the Alaska pollock certification. An explanation provided by this analysis is that when the environmental differentiation is not credibly displayed by the certification, an immediate consequence is that green

---

<sup>2</sup>The Guardian 20/08/2016: The troubling evolution of corporate greenwashing

<sup>3</sup>The Guardian 25/05/2014: Five sustainable boondoggles: greenwashing all the way to the bank

<sup>4</sup>Le Monde Diplomatique 01/06/2017: Un label agricole toujours moins exigeant

producers are constrained to practice pooling prices to stay in the market. They cannot signal their environmental quality through prices because (1) it corresponds to credence attributes and (2) some of the standard firms would have incentives to mimic any green prices.

The remainder of the paper is organized as follows. Section 2 below relates the article to the existing literature. Section 3 provides the model, extensions, and results. Section 4 discusses the implication of results.

## 2 Relation to the Existing Literature

The transmission of information regarding environmental quality can be characterized as vertical or horizontal, according to whether it triggers unanimous consumer reactions or not ([Bonroy and Constantatos, 2008](#)). Nevertheless, the literature treats environmental quality as an excludable characteristic of the good and mainly assumes that adopting a green production is more costly than a standard process ([Amacher et al., 2004](#); [Hamilton and Zilberman, 2006](#); [Mahenc, 2007](#)). Therefore there is an indisputable necessity to signal the quality or else fall into the lemons outcome ([Akerlof, 1970](#)). In this article, we focus on the price only to provide information. Price signaling literature revealed some important insights on the link between the asymmetric information market and price strategies.

[Milgrom and Roberts \(1986\)](#) mostly identify various conditions to ensure the existence of separating equilibria between the low- and high-quality type. However, they find the existence of pooling equilibria if consumers beliefs are high enough. This result comes from the notion that the distortion of the price required to reveal the information may be less attractive when consumers are very optimistic about the quality. Repeat purchases play a major role in the trade-off between the choice to signal by price and/or by advertising.

[Bagwell and Riordan \(1991\)](#) show that, in two types of a quality model, in equi-

librium, firms choose separating prices. The high-quality firm has to set a price above the optimal price in perfect information to prevent the deviations of low-quality firms. This distortion disappears when there is a sufficient number of informed consumers on the market. In both cases, firms reveal their type and successfully solve the problem of asymmetric information.

[Daughety and Reinganum \(2008b\)](#) provide a complete model in which the low quality is associated with a disutility parameter. They examine the interplay of imperfect competition and incomplete information. There are two sources of incomplete information (adverse selection and firms anticipations); nevertheless, in equilibrium, firms charge the separating prices. Even when the disutility is strong and thus incites low-quality firm to mimic the high segment, the high-quality firm raises its price even further to signal its quality.

[Mirman and Santugini \(2014\)](#) point out the role of the competitive fringe in guaranteeing the existence of an equilibrium in which the separation takes place to inform uninformed buyers. This is because the incentives for a low-quality firm to mimic the price of the high-quality firm are reduced since a higher price triggers more sales on the part of the competitive fringe. They also confirm the well-known result that a large fraction of informed buyers is sufficient for the price, without distortion, to convey information about the quality of the good.

[Mahenc \(2007\)](#) examines a model that involves a monopolist using price to signal quality and shows that a firm has the incentive to disclose its quality as long as consumers do not have beliefs that are too optimistic. He underlines that the distortion caused by the signaling occurs when the firm chooses the fully informative price and concludes that the monopolist charges a higher price in the case of incomplete information.

It is noteworthy to mention that, in all of these studies, the market successfully makes full revelation, and the price plausibly solves the problem of asymmetric infor-

mation.

### 3 Model

#### 3.1 Perfect information

One starts by defining the model of perfect information to use it as a benchmark. The quality of the product (the vertical attribute) can be green or brown. In period one, Nature independently draws a type for each firm from a common distribution, and each firm observes its type. The types of firms are denoted  $\theta_i \in \Theta$ ,  $i = g$  (green quality) and  $i = b$  (brown quality). Environmental quality refers here to the cleanness of production.

We normalize the number of consumers to 1, and we assume that the utility of a brown good to consumers is homogeneous and denoted as follows:

$$v - p_b$$

where  $v$  denotes the reservation price of brown consumers for the good and  $p_b$  its price.  $v > p_b$ ; otherwise, it would not be socially profitable to produce the good. The utility of a green good to consumers is heterogeneous and defined as follows:

$$v + s - p_g$$

where  $s$  can be interpreted as each consumer's reservation price for the environmental service provided by the firm. Moreover,  $s$  is uniformly distributed within  $[0, \alpha e]$ , where  $e$  is the effort provided by the green firm and  $\alpha$  the intrinsic valuation of an environmental service for the consumer with the most sensitivity to environmental causes. The combination yields an upper bound  $\alpha e$  representing the reservation price of the household that values the environmental service the most. To ensure that the production of the green good is socially desirable, one has to assume that  $\alpha > 1$ . When the

effort is high, or the valuation is large, green consumers are ready to spend more. This framework is similar to [Mussa and Rosen \(1978\)](#). We denote by  $\tilde{s}$  the customer that is indifferent between consuming the green good and the brown good. In this case the market is assumed to be fully covered.

$$v - p_b = v + \tilde{s} - p_g$$

$$\tilde{s} = p_g - p_b$$

Both demands  $Q_b(p_b, p_g)$  and  $Q_g(p_b, p_g)$  are defined by  $\tilde{s}$  such that:

$$Q_g(p_g, p_b) = \begin{cases} 1 - \frac{p_g - p_b}{\alpha e} & \text{if } \alpha > \frac{p_g - p_b}{e} \\ 0 & \text{otherwise} \end{cases} \quad (1.1)$$

$$Q_b(p_g, p_b) = \begin{cases} \frac{p_g - p_b}{\alpha e} & \text{if } p_g > p_b \\ 0 & \text{otherwise} \end{cases} \quad (1.2)$$

Concerning the supply side, we assume that firms compete in a simplified Bertrand duopoly framework. We assume that the price of the brown market corresponds to the perfectly competitive price  $p_b = c$ . We assume that they have both a linear marginal cost to produce the good such that  $c(\theta_i, q_i) = cq_i$  and that the green firm in order to produce a cleaner good has a variable abatement cost such that  $c(\theta_g, q_g) = (c + e)q_g$ , where  $e \in ]0, 1]$  is the effort of the firm and it is assumed to be exogenously determined. The effort cannot be null because the preferences of [Mussa and Rosen \(1978\)](#) require that green firm that benefits from green demand is at least endowed with the minimum effort possible to be differentiated from the brown firm. The firm producing green quantity solves the following program:

$$\max_{p_g} \pi(\theta_g, p_g) = (p_g - c - e) \left( 1 - \frac{p_g - c}{\alpha e} \right) \quad (1.3)$$

$$p_g^* = \frac{\alpha e}{2} + \frac{e}{2} + c \quad (1.4)$$

The structure of the analysis corresponds to a price competition through product differentiation. The different effects are closely related to [Shaked and Sutton \(1982\)](#). The level of the effort of abatement of the firm here appears as the level of differentiation in a standard model. One can check the classic result that when  $\lim_{e \rightarrow 0} p_g(e) \rightarrow p_b = c$ . For  $e = 1$  products are considerably differentiated and the price competition between firms is softened. As a result, the price of the green firm increases. The profit equilibrium of the green firm is given by:

$$\pi(\theta_g, p_g^*) = \frac{e(\alpha - 1)^2}{4\alpha} \quad (1.5)$$

### 3.2 Adverse selection

In this subsection, the problem of asymmetric information occurs in the green market. An environmental attribute associated with the cleanness of production is defined as a *credence attributes*. Although consumers do not directly observe the environmental performance of the product, the price of the green good can be taken as a signal of the environmental quality. Let  $\mu(p) = \mu \rightarrow [0, 1]$  denote consumers' posterior belief that the product is truly green when the price is  $p$ . Nevertheless, firms are correctly informed on the type of their rival. When information is not complete, the game among the firms and consumers occurs in (one-shot) three stages.

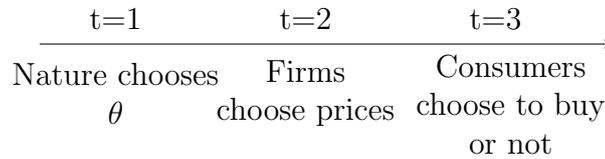


Figure 1: Timing

The utility of green customers is now:

$$v + \mu(p)s_j - p_g$$

$$\tilde{s} = \frac{p_g - c}{\mu(p)}$$

The program of the green firm competing in a duopoly is now given by:

$$\max_{p_g} \pi(\theta_g, p_g) = (p_g - c - e) \left( 1 - \frac{p_g - c}{\alpha e \mu} \right) \quad (2)$$

The equilibrium prices, quantities and profits are now:

$$Q_g^*(p_g) = \begin{cases} \frac{(\alpha\mu-1)}{2\mu} & \text{if } \mu > \frac{1}{\alpha} \\ 0 & \text{otherwise} \end{cases} \quad (2.1)$$

$$Q_b^*(p_g) = \frac{e(\alpha\mu + 1)}{2\mu} \quad (2.2)$$

$$p_g^* = \frac{1}{2}\alpha e \mu + \frac{1}{2}e + c \quad (2.3)$$

$$\pi_g^* = \frac{1}{4} \frac{e(\alpha\mu - 1)^2}{\mu} \quad (2.4)$$

Intuitively the total amount of green quantity (2.1) is an increasing function of the perception of green customers. The condition on  $\mu$  guarantees that the market exists if consumers are not too pessimistic about the environmental attribute of the good. The green firm prefers consumers beliefs about environmental performance to be optimistic. Conversely the total amount of brown quantity (2.2) is a decreasing function of  $\mu$  because of the effect of substitution between the two kind of goods. Equations (2.3) and (2.4) show that the posterior belief of green customers is a parameter that acts in favour of softening competition for the green firm by increasing differentiation. Concerning (2.4) it reaches its maximum when the differentiation is at its highest state and the information is perfect.

For the green firm, the problem is to choose the price to maximize its profit while taking into account consumers' updated beliefs. The particularity here is that even if the brown firm experiences a competitive market, one assumes that it has the possibility to choose between using the informative price, which is equal to the marginal cost, and mimicking the green price. Without loss of generality one supposes that the brown

market still exists even if the brown firm decides to mimic the green price because of the remaining standard firms that do not have the possibility of trying to cheat green consumers. The set of strategies and beliefs that characterize the Perfect Bayesian Equilibrium (PBE) is  $(\phi(\theta_g), \phi(\theta_b), \mu(p))$ , where  $\phi(\theta_g)$  and  $\phi(\theta_b)$  are the strategy of the green and the brown firm, if the two following conditions are satisfied (Mas-Colell, Whinston, Green, et al., 1995).

- *Optimality condition.*

For  $\theta = \theta_b, \theta_g$ ,  $\phi(\theta) = \arg \max_p \pi(p, \theta, \mu)$

- *Bayes's consistency of beliefs.*

If  $\phi(\theta_b) \neq \phi(\theta_g)$ , then  $\mu(\phi(\theta_b)) = 0$  and  $\mu(\phi(\theta_g)) = 1$ . If  $\phi(\theta_b) = \phi(\theta_g)$ , then  $\mu(\phi(\theta_b)) = \mu(\phi(\theta_g)) = \mu_0$ .

The latter condition simply states that consumers' posterior beliefs about the type of the firm is updated according the Bayes's rule along the equilibrium path. As a consequence, when prices are type independent the consumers stick to their prior beliefs. In this analysis we will focus on pure strategies.

### 3.3 Separating equilibrium

In any separating PBE,  $\phi(\bar{\theta}_g) \neq \phi(\bar{\theta}_b)$ ; that is, information is revealed so that  $\mu(p) = 1$  and each type of firm receives a profit equal to its environmental level.

Any separating equilibrium must satisfy a credibility constraint saying that the brown type must resist the temptation to cheat consumers by mimicking the price  $\phi(\bar{\theta}_g)$ . For this constraint, the brown type should prefer  $\phi(\bar{\theta}_b)$  over  $\phi(\bar{\theta}_g)$  that would falsely signal that its product is green. Formally, this requirement is:

$$\pi(\theta_g, \phi(\bar{\theta}_g), 1) \geq 0 \tag{3.1}$$

$$\pi(\theta_b, \phi(\bar{\theta}_b), 0) \geq \pi(\theta_b, \phi(\bar{\theta}_g), 1) \tag{3.2}$$

(3.1) and (3.2) ensure that profits of each type benefit from the best beliefs of consumers and guarantee that informative prices are attractive.

**Proposition 1:** *There exists only one separating equilibrium price such that  $\phi(\theta_g) \neq \phi(\theta_b)$  where  $\phi(\bar{\theta}_g) = \alpha e + c$  and  $\phi(\bar{\theta}_b) = c$  yielding profits of  $\pi(\theta_b, \phi(\bar{\theta}_b), 0) = \pi(\theta_g, \phi(\bar{\theta}_g), 1) = 0$ .*

**Proof.** Straightforward calculation on (2) shows that there is a continuum of possible prices that would satisfy constraint (3.1) with  $\phi(\theta_g) \in [c + e; \alpha e + c]$ . Condition (3.2) gives the incentive constraint of the brown firm. By assumptions, one knows that if the information is revealed then the brown firm is a price taker and thus makes a profit equal to zero. The only price that satisfies condition (3.1) and condition (3.2) is  $\phi(\theta_g) = \alpha e + c$ . At this point the brown firm is indifferent between  $\pi(\theta_b, \phi(\bar{\theta}_b), 0) = 0$  and  $\pi(\theta_b, \phi(\bar{\theta}_g), 1) = 0$ .

This proposition follows the intuition of [Janssen and Roy \(2010\)](#) where they justify their assumption of a positive profit for all firms and contradicts the result of [Mirman and Santugini \(2014\)](#). Note that when the green firm sets this price the demand  $Q_g(p, 1) = 0$ . To explain this result, recall that the brown market is perfectly competitive so that profits are zero for the brown firm. Therefore (3.2) can only hold if the right-hand side is nonpositive. Since production costs are strictly lower for the brown firm than for the green firm, this can only happen when the demand for the green firm is zero.

### 3.4 Pooling equilibria

Because the profit is null in the case of separating equilibrium, pooling equilibria appear more plausible in the model. Since it is the same (uninformative) price charged by the firms, regardless of environmental performance, the consumers' posterior beliefs after observing the price are the same as their prior beliefs denoted  $\mu_0$ . Recall from Section

1 that the nature of these equilibria in this model represents greenwashing because it corresponds to a case where the brown firm by equalizing its price to that of the green firm claims to be environmentally friendly. If such a phenomenon is revealed one cannot be very optimistic about the beliefs of consumers and the sustainability of such a market because of condition (2.1). Note that in that case, both firms share the demand on the green market equally.

In any pooling PBE.,  $\phi(\underline{\theta}_g) = \phi(\underline{\theta}_b)$ ; that is, where the green and the brown firm have incentive to conceal the information.

The incentive constraints of both types are:

$$\pi(\theta_g, \phi(\underline{\theta}_g), \mu_0) > \pi(\theta_g, \phi(\theta_g), 0) \quad (4.1)$$

$$\pi(\theta_b, \phi(\underline{\theta}_b), \mu_0) > \pi(\theta_b, \phi(\theta_b), 0) \quad (4.2)$$

The incentive constraints express the profits of firms at the pooling price given the prior beliefs of consumers supported treating any out-of-equilibrium price as indicating that the firm is definitely a brown type.

**Proposition 2:** *there exists a continuum of pooling equilibria such that  $\phi(\theta_g) = \phi(\theta_b) \in ]c + e; \alpha e \mu + c[$ .*

**Proof.** There is a continuum of possible prices for the green firm which satisfy (4.1). The brown firm has a larger set of possible prices (4.2) because it does not pay an abatement cost. Nevertheless, the set of possible prices is reduced to the set of the green firm because of the definition of PBE.

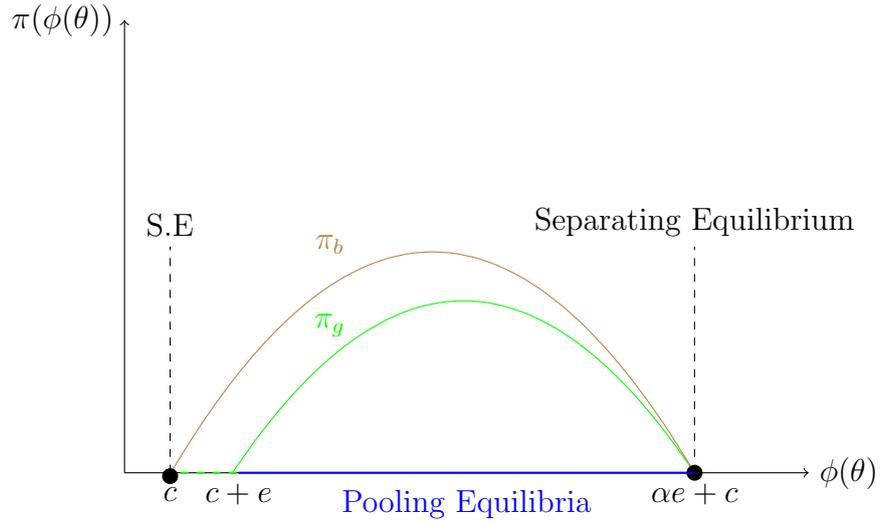


Figure 2: Summary of the Signaling Game

Figure 2 describes a situation between the green and the brown firm. These curves are drawn according to specific posterior beliefs  $\mu$  of green customers. The brown curve is the profit of the brown type and the green curve is the profit of the green type given the chosen price. Dashed green lines on the horizontal axis represent the space where the green firm would not operate given these prices.

### 3.5 Intuitive Criterion vs Undefeated Equilibrium

Facing the multiplicity of equilibrium there is a necessity to use a refinement concept. The idea behind the mobilization of two different refinement criteria is to open the possibility for green consumers to behave in various ways. Price signaling literature usually uses the *intuitive criterion* (Cho and Kreps, 1987) to select a unique separating equilibrium (Milgrom and Roberts, 1986; Bagwell and Riordan, 1991; Daughety and Reinganum, 2008b). Another well-known concept is the *undefeated equilibrium* (Mailath, Okuno-Fujiwara, and Postlewaite, 1993) used by Mahenc (2007) in the case of signaling through prices.

**Proposition 3:** *Intuitive Criterion rules out every pooling equilibria*

**Proof.** *Appendix.*

The most common refinement used in the price literature eliminates every equilibrium associated with possible positive profits for firms and selects the single separating price yielding a null demand. It bolsters the idea that the green firm cannot use the price as a signal. In this case, it is not possible to rely on price to be a substitute to certification for the green credibility as in (Mahenc, 2007).

**Proposition 4:** *The set of undefeated equilibria restricts the set of pooling equilibria to  $[\frac{1}{2}\alpha e\mu_0 + c; \frac{1}{2}\alpha e\mu_0 + c + \frac{1}{2}e]$*

**Proof.** *Appendix.*

Here, an analysis of a pure pooling strategy sequential equilibria where the brown firm mimicks the green firm's price is presented. In other words, these equilibria are characterized by the fact that the brown firm is sending misleading signals to green consumers. This is a situation where the green firm strictly prefers the pooling equilibria in which  $\phi(\theta_g) = \phi(\underline{\theta}_g)$  to the separating equilibrium where  $\phi(\theta_g) = \phi(\overline{\theta}_g)$ . These boundaries characterize Pareto-efficient prices for the brown type and the green type. These optimal prices are changing given the posterior beliefs of customers  $\mu_0$ . As a result, when consumers have sufficient prior beliefs that do not fall into the lemons outcome, the market exists. However, the market experiences greenwashing.

Figure 3 displays the profit associated with both types of firms in the case where prices are informative (thick curves) compared to pooling situations where the equilibrium beliefs correspond to prior beliefs (dashed curves). The results in this section are quite extreme. If agents in the green markets follow the intuitive criterion, the output is that there is no green market at all. Otherwise, using the undefeated equilibrium concept, the market would at least exist.

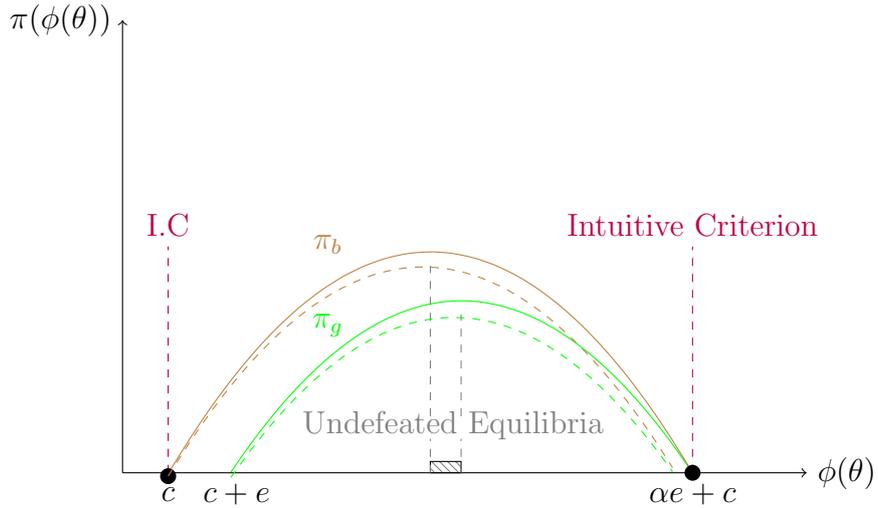


Figure 3: Intuitive Criterion vs Undeclared Equilibrium

### 3.6 Extension: a share of informed consumers

Confronting the extreme output of the previous section, it is important to keep in mind that the price signaling literature also showed that above a critical size of informed consumers, signaling is feasible with full information prices (Mirman and Santugini, 2014; Belleflamme and Peitz, 2015). It is not a new idea in price competition that a share of informed consumers may be extremely efficient as long as it has sufficient impact on the incentive constraints either of the low-quality firm as in Bagwell and Riordan (1991) or of the high-quality firm. Linnemer (2002) extends this framework with the possibility for a firm  $a$  to use advertising. He shows that if the proportion of informed buyers is low, a high-quality firm has to use dissipative advertising in order to lower the necessary price distortion to signal its quality. If this proportion is high, the necessary distortion to signal quality might disappear, which makes advertising even useless.

Now assume that some buyers can ascertain the quality of the environmental good. A fraction  $\beta$  of buyers, for example, the readership of reviews publishing accurate advice on quality, learn the product quality before purchase, while the remaining fraction  $1 - \beta$

of buyers believe that quality is green with probability  $\mu$ . Then, the total demand for the green producer is:

$$Q_g(p_g, \mu, \beta) = \beta Q_g(p_g, 1) + (1 - \beta) Q_g(p_g, \mu)$$

$$Q_g(p_g, \mu, \beta) = \frac{\alpha e \mu + ((1 - \mu) \beta - 1) (p_g - c)}{\alpha e \mu}$$

When  $\beta$  increases, the problem of asymmetric information is reduced; as a result, the green quantity that is sold increases. In this framework, the signaling through price has an influence on the remaining share of uninformed customers, and this quantity is denoted  $Q_g^R(p_g, \mu, \beta)$ :

$$Q_g^R(p_g, \mu, \beta) = (1 - \beta) \left( \frac{\alpha e \mu - p_g + c}{\alpha e \mu} \right)$$

Immediately one can rewrite the credibility constraints for both firms to signal their quality through prices as follows:

$$\pi(\theta_g, \phi(\bar{\theta}_g), 1) \geq (\phi^*(\theta_g) - c - e) \beta Q_g(\phi^*(\theta_g), 1) \quad (5.1)$$

$$0 \geq (\phi(\bar{\theta}_b) - c) Q_g^R(\phi(\theta_b), \mu, \beta) \quad (5.2)$$

**Proposition 5:** *For  $\beta \leq 1$  the share of informed consumers does not raise plausible separating equilibria.*

While it reduces the incentives for the green firm to apply the informative price due to the possibility of setting the optimal price to the share of informed consumers, (5.2) is never satisfied when  $\beta < 1$ . The intuition behind this result is that if the brown firm displays its true quality to consumers, then in a perfect competitive market its profit would be null; therefore, the brown firm always has the temptation to mislead the remaining share of uninformed buyers. The main positive effect of the introduc-

tion of the share of informed customers is that in pooling equilibrium the profit of the green firm is enhanced, which directly supports the possibility for the environmental producer to sustain in the market.

## 4 Discussion

This analysis constitutes a case where assumptions to ensure the existence of separating equilibrium associated with positive payoffs do not hold (Milgrom and Roberts, 1986) or where the disutility associated with the polluting product is very high (Daughety and Reinganum, 2008b). Mentioned but less highlighted than separating equilibria, pooling equilibria remain problematic. The case of green markets selling credence attributes products constitutes an interesting situation as long as the repetitive purchase cannot provide any information. Knowing that some brown firms have incentives to defraud (Hamilton and Zilberman, 2006), one can easily imagine a market where a polluting firm that experiences low profit may be tempted to use a pooling price as long as punishments and regulations are not efficient enough. This study points out the importance of the credibility of the means of providing information such as lobbying (Feddersen and Gilligan, 2001) and certification (Bonroy and Constantatos, 2008; Ibanez and Grolleau, 2008), otherwise, green information could not be revealed. The assumption that the brown side of the market is competitive is more related to the producers of raw materials such as coffee, forests, and fish than retailers that are more likely to have a market share. The output where no separating equilibria exist would be maintained in the case where an industry experiences fixed costs. This particular framework of duopoly where the brown market exists even if the brown firm mimics the green price may be confusing at first, but it is a simplified way to reflect the idea that not every firm operating in the brown market could send misleading signals. The hypothesis on the exogenous choice on quality was to focus on the transmission of information on environmental quality. It would be interesting to investigate the

consequence of such implications on the choice of investments in green technologies.

## 5 Appendix

### 5.1 Proof of Proposition 3.

To simplify notations we note that  $\pi(\theta_i) = \pi_i$ .

An equilibrium  $\phi^*$  is said to violate the Intuitive Criterion if  $\exists \phi' > \phi^*$  with  $\mu(\phi') = 1$  such that:

$$\pi_b(\phi^*, \mu_0) \geq \pi_b(\phi', 1) \quad (5.1)$$

$$\pi_g(\phi^*, \mu_0) < \pi_g(\phi', 1) \quad (5.2)$$

The left hand sides are equilibrium payoffs whereas the right hand sides are the maximum payoff that each type could get by setting a price  $\phi'$ .

- Consider a  $\phi'$  s.t  $\pi_b(\phi^*, \mu_0) = \pi_b(\phi', 1)$  which can be rewritten

$$(\phi^* - c)Q_g(\phi^*, \mu_0) = (\phi' - c)Q_g(\phi', 1) \quad (5.3)$$

One verifies (5.2):

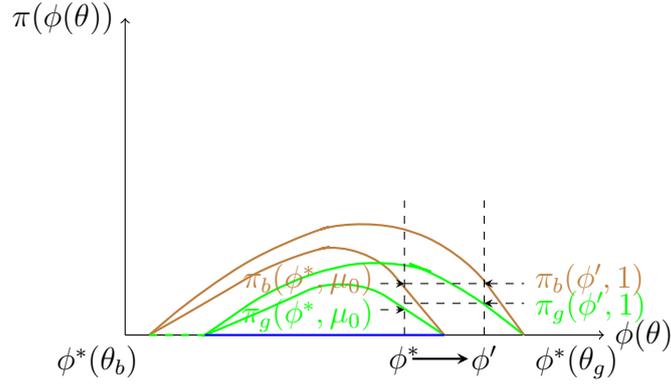
$$\begin{aligned} \pi_g(\phi^*, \mu_0) &= \pi_b(\phi^*, \mu_0) - eQ_g(\phi^*, \mu_0) \\ &= \pi_b(\phi', 1) - eQ_g(\phi^*, \mu_0) \\ \pi_g(\phi', 1) &= \pi_b(\phi', 1) - eQ_g(\phi', 1) \end{aligned}$$

With (5.3)  $Q_g(\phi^*, \mu_0) = \left(\frac{\phi' - c}{\phi^* - c}\right) Q_g(\phi', 1)$

$$\begin{aligned} Q_g(\phi', 1) &< Q_g(\phi^*, \mu_0) \\ Q_g(\phi', 1) &< \left(\frac{\phi' - c}{\phi^* - c}\right) Q_g(\phi', 1) \end{aligned}$$

It implies that  $\phi' > \phi^*$  and verifies the definition.

QED.



Graphic associated to the proof of Proposition 3.

## 5.2 Proof of Proposition 4.

To ensure the existence of an undefeated equilibrium there are three necessary conditions:

$$\frac{\partial \pi(\theta_g, \phi(\theta_g), \mu)}{\partial \mu} > 0 \quad (6.1)$$

$$\frac{\partial \pi(\theta_g, \phi(\theta_g), \mu)}{\partial \phi(\theta_g)} < \frac{\partial \pi(\theta_b, \phi(\theta_b), \mu)}{\partial \phi(\theta_b)} \quad (6.2)$$

$$\frac{\partial^2 \pi(\theta_g, \phi(\theta_g), \mu)}{\partial^2 \phi(\theta_g)} < 0 \quad (6.3)$$

(6.1) States that the profit increases if consumers believes the firm more likely to be of environmental type. (6.2) is the single-crossing property. It says that the slope of the pay-off function of the brown firm with respect to the price is higher than the one of the green firm. This captures the idea that higher messages are more profitable for the brown firm. This is counter-intuitive because, usually, condition (6.2) imposes that it's easier for higher type to send a higher message. Finally condition (6.3) says that increasing the price may result into a loss even if the result was the most favorable possible beliefs on green consumers.

Undefeated equilibrium is closely related to the Pareto efficiency. Thus prices survive

the undefeated equilibrium if there exist an equilibrium  $\rho(\phi(\theta_g), \phi(\theta_b), \mu_0)$  where:

$$\exists \theta \in [b, g], \phi(\theta) \text{ and } \exists \theta \in [b, g], \phi'(\theta)$$

$$\text{with } \pi(\rho, \theta) \geq \pi(\rho', \theta)$$

The price which satisfies the equation above is not the same for both type. Anyway it always crowd-out the separating equilibrium and restrict the set of pooling equilibria to  $\forall i \in b, g \phi(\theta_i) \in [\frac{1}{2}\alpha e \mu_0 + c; \frac{1}{2}\alpha e \mu_0 + c + \frac{1}{2}e]$ .

## Chapter 2

# Watching the Watchers: The Credibility of Signaling Social Goodwill with Imperfect Monitoring

## English summary

NGOs play a key role in the informative process concerning the type of firms. We are interested in how their informational behavior impacts the signaling strategy of firms. Even imperfect, certification still indicates a piece of information to the consumer. The interaction between the price signaling strategy of the firm, and the information released by the NGO yield fruitful results concerning the path of green information. The correct transmission of green information has an important impact on the technological choice of the company. The interest of this model is the revelation of hard evidence in a standard signaling model.

When consumers cannot verify corporate social goodwill, firms may be reluctant to uphold a pledge of social goodwill. We show how imperfect monitoring can mitigate this moral hazard problem.

We augment the standard model of price signaling by allowing consumers to use the results of independent monitoring as a complementary source of information. Monitoring corrects for consumers' arbitrary beliefs. Before sending a price signal to consumers, firms pledge or not to invest in social goodwill.

With no monitoring, firms do not abide by their pledges of social goodwill when they fail to send a credible signal *via* price.

With monitoring, there exist equilibria in which a firm invests in social goodwill and succeeds in signaling its choice *via* price. We conclude that independent monitoring, although imperfect, helps a firm fulfill its pledge of social goodwill by restoring the credibility of price signaling.

## Résumé français

Les ONGs jouent un rôle clé dans le processus informatif concernant le type des firmes. Nous nous intéressons sur la façon dont leur comportement informatif im-

pacte la stratégie de signal des firmes. Même imparfaite, la certification indique tout de même une part de l'information aux consommateurs. L'étude de l'interaction entre, la stratégie de signal en prix de la firme et, l'information donnée par l'ONG mène à des résultats prometteurs concernant la diffusion de l'information sur la qualité environnementale. La bonne transmission de l'information verte a un impact important sur le choix technologique de la firme. L'intérêt de ce modèle réside dans la révélation d'information "hard " dans un modèle de signal standard.

Lorsque les consommateurs ne peuvent pas vérifier l'engagement social et/ou environnemental, les firmes peuvent être défiantes à opérer des changements dans leur processus de production pour prendre en compte ces problématiques. Nous montrons comment la surveillance imparfaite peut atténuer ce problème d'aléa moral. De ce fait, nous complexifions le modèle standard de signal en prix en permettant aux consommateurs d'utiliser les résultats d'enquête sous diverses formes comme complément d'information. La surveillance corrige alors les croyances arbitraires des consommateurs. Avant d'envoyer un signal aux consommateurs, les firmes décident ou non de s'engager à résoudre des problèmes environnementaux.

Sans le contrôle de l'ONG, les firmes ne s'engagent pas dans la résolution de ces problèmes car il est impossible pour elles d'envoyer un signal crédible *via* le prix.

Avec le contrôle de l'ONG, il existe des équilibres dans lesquels les firmes investissent pour atténuer les problèmes environnementaux et arrivent à signaler leur choix *via* le prix. Nous concluons que la surveillance indépendante, bien qu'imparfaite, permet à la firme de s'investir de manière plausible dans ces causes en restaurant la crédibilité des signaux qu'elle envoie au consommateur.

# 1 Introduction

As surprising as it may sound, firms often claim their pledges for social objectives, whether environmental or ethical. [Besley and Ghatak \(2007\)](#) find evidence of a broader social role for private enterprise in the development of corporate social responsibility. [Baron \(2010\)](#) recognizes that an increasing number of firms engage in the mitigation of externalities, the redistribution of wealth or the provision of public goods; the researcher categorizes all these pro-social activities under the generic term of “self-regulation”. A firm may find it worthwhile to champion a social cause, as long as doing so increases the product value to consumers and the consumer surplus that the firm can extract in its self-interest. However, self-regulation can also be motivated by concerns that range from the deterrence of public regulation, as in [Lyon, Maxwell, et al. \(2004\)](#), to moral concerns, as in [Baron \(2010\)](#).

Firms that spend resources on self-regulation must, in turn, communicate information about their social goodwill; otherwise, they may receive no credit for doing so. Firms usually claim social goodwill by displaying their own social labels or relying on external certification, whether public or private. These claims are aimed at persuading consumers that the firm wishes to maintain a balance between its own interest and any social concerns that a catchword has brought to the fore.

There are many examples of ecolabels claiming compliance with good environmental practices. For instance, car manufacturers display “Clean Diesel” to guarantee sustainable business practices. Similarly, the “Forest Stewardship Council” label is meant to reduce illegal logging and improve forest governance. Regarding ethical concerns, certifiers such as Fairtrade International grant the Fair Trade logo to cooperatives and farmers that meet standards promoting sustainable development and more equitable distribution of wealth in developing countries.

However, social goodwill is a product attribute with specific features that make it

hard to verify: it is intangible and public, in the sense that every individual's enjoyment of social goodwill does not preclude its enjoyment by other individuals. Certification may fail to display credible information about the social conscience of a business. The Volkswagen emissions scandal in 2015 has raised doubts on both compliance with environmental standards and the trustworthiness of clean certification in the automobile industry. Similarly, non-governmental organizations have long warned about illegal logging in China, Peru, and Romania by companies carrying the Forest Stewardship Council label.<sup>1</sup> The same issue has arisen with Fair Trade certification<sup>2</sup> or the Marine Stewardship Council granted to sustainable fisheries by a nongovernmental organization.<sup>3</sup>

This paper addresses the question of watching the watchers formulated long ago by the poet Juvenal in his Satires: "Quis custodiet ipsos custodes?"<sup>4</sup> We argue that independent monitoring has a key role to play in self-regulation in a context without trustworthy certification of firms' social goodwill. Our model shows that, although imperfect, monitoring helps firms abide by their pledges of social goodwill by ensuring the credibility of the signal sent by firms through prices.

The issue of misleading certification is central to a strand of theoretical literature. [Feddersen and Gilligan \(2001\)](#) question the honesty of third-party certification in a model where a certifier biased toward environmental protection is responsible for sending misleading messages to consumers. [Hamilton and Zilberman \(2006\)](#) show that markets for environmentally-friendly products suffer from fraudulent labeling. [Baksi and Bose \(2007\)](#) demonstrate that some firms make false claims or display spurious labels. [Mahenc \(2017\)](#) shows that third-party certification turns out to be misleading

---

<sup>1</sup>See 20/02/2018 Greenwashed Timber: How Sustainable Forest Certification Has Failed.

<sup>2</sup>Weitzman, Hal (9 September 2006). "The bitter cost of Fair Trade coffee": <https://www.ft.com/content/d191adbc-3f4d-11db-a37c-0000779e2340>

<sup>3</sup>Jacquet et al. (2010) raises concern about potential conflicts of interest involved in this certification.

<sup>4</sup>Satire VI, lines 347–348.

when the certifier is driven more by profit than by social welfare.

Since the work of [Darby and Karni \(1973\)](#), products with a desirable private attribute, whose utility is difficult if not impossible for consumers to ascertain, have been termed “credence” products in the economic literature. In the present paper, we extend this terminology to the public attribute of social goodwill. Like every credence attribute, asymmetric information about social goodwill provides firms with incentives for fly-by-night behaviors that mislead consumers. This may raise a twofold problem, compounded by misleading certification: a problem of adverse selection—products tied in with social goodwill are not attracted to the market because this attribute is hidden from consumers (see [Akerlof, 1970](#))—and a problem of moral hazard—firms may break with their pledges of social goodwill if they are neither observable nor verifiable by consumers. —.

The adverse selection issue has been widely investigated for private attributes of experience goods ([Nelson, 1970](#)) in the framework of price signaling models ([Milgrom and Roberts, 1986](#); [Bagwell and Riordan, 1991](#); [Daughety and Reinganum, 2008b](#); [Janssen and Roy, 2010](#)). Building on pioneering work of [Spence \(1973\)](#), these models assume that Nature selects the firms’ types, which summarize the private attributes of a product such as quality standards resulting from past investments. If repeat purchases and reputation are not an issue, a firm may want to signal a high-quality product with prices higher than those justified by market power under full information. This is due to the credibility requirement inherent to separating equilibria: the high-quality firm must deter its low-quality counterpart from fooling consumers by mimicking. In some circumstances, the least costly means of preventing the fly-by-night strategy is for the high-quality firm to include the forgone profit from cheating in the product price, thereby incurring further costs for informational purpose. Prohibitive signaling costs may explain why market prices conceal rather than reveal information about quality, and also why firms do not find it profitable to put high-quality products on the market.

The aforementioned examples of fly-by-night behaviors in the automobile and logging industries show that the adverse selection problem also occurs for products tied in with a public attribute such as social goodwill. If consumers cannot rely on trustworthy certification, then products labeled differently depending on firms' claims of social goodwill are sold at the same "pooling" prices, thus making them susceptible to the lemons problem of [Akerlof \(1970\)](#). That pooling prices prevail over separating prices in markets for credence products is a theoretical possibility. As demonstrated by [Mahenc \(2017\)](#), this occurs when the certifier is driven more by profit than by social welfare. Recent experimental field and lab studies provide evidence of fraud in markets for credence goods and services that are mainly private in nature ([Kerschbamer and Sutter, 2017](#)). The impossibility of credible signaling is, in turn, a disincentive for firms to tie social goodwill in with products, even though consumers would be willing to pay a premium for it. There is a moral hazard problem behind the adverse selection problem. Anticipating that signaling cannot be achieved, firms do not find it worthwhile to spend resources on social goodwill.

To address the moral hazard issue, we extend the standard signaling model of asymmetric information with hidden knowledge to a model of symmetric information with hidden actions, in which imperfect monitoring by a third-party is allowed. Hidden actions are the technology decisions made by two firms whether to stick to business as usual or to switch to production with an added value of social goodwill that is unverifiable to consumers. In other terms, we consider that firms select their type before they send a price signal to consumers. The choice of type determines the degree of differentiation between two products available for consumption and hence the intensity of subsequent competition in price between firms. Bayesian consumers try to infer from the prices set by firms whether or not social goodwill is tied in with the product they are purchasing.

First, we examine a reduced version of the model, in which certification is not

trustworthy and price signaling is the sole source of information for consumers. In this scenario, a firm that would have invested in social goodwill fails to prevent fly-by-night behavior in the form of misleading prices with the result that there exists no separating equilibrium in price. Faced with the inability to signal social goodwill, firms renege on any pledge to provide the credence attributes. Thus, an additional and complementary source of information is needed for consumers concerned with social goodwill.

We extend the model by allowing for imperfect monitoring activities performed by an independent third-party auditor. Consumers make purchase decisions using both the auditor's report and firms' price signaling to gather information. We assume that monitoring improves consumers' perception of social goodwill by correcting for the arbitrary beliefs held after observing a deviation from the price equilibrium.

Our main result is that monitoring, albeit imperfect, allows the existence of price separating equilibria. The reasons for this are that monitoring, on the one hand, raises the cost of cheating consumers with misleading prices, and on the other hand, motivates firms to reveal the truth about their type. As monitoring makes credible the price signals firms send about their type, tying social goodwill in with the product becomes worthwhile. As a result, there exists a reasonable probability that a firm switches to production with a pledge of social goodwill that can be fulfilled. In a mixed strategy equilibrium, this probability coincides with the probability that a rival firm mimics the price sent as a signal of social goodwill. We conclude that the auditor's monitoring helps solve the two problems firms encounter: the adverse selection problem raised by signaling costs and the moral hazard problem of pledging social goodwill.

The rest of this paper is organized as follows. Section 2 discusses a brief review of the literature. Section 3 introduces the model of symmetric information with firms' hidden actions in which consumers have two sources of information: independent monitoring and firms' price signaling. Section 4 characterizes the possible Perfect Bayesian Equilibria of the whole game called the "signaling game with hidden actions". In Sec-

tion 5, we analyze a baseline model with no monitoring and we establish the conditions under which there exists no separating equilibrium in the signaling subgame. In Section 6, we analyze the full scenario with monitoring and we show the existence of equilibria that involve separation in the signaling subgame. Our concluding remarks appear in Section 7.

## 2 Related literature

Our models builds on the signal theory initiated by [Spence \(1973\)](#) and further developed in a strand of literature on asymmetric information devoted to the analysis of markets in which firms use price as the only means of signaling product quality. Usually, these markets are imperfectly competitive because firms must have enough control over prices to influence the amount of information disclosure. Signaling through prices can occur in monopolistic markets (as in [Bagwell and Riordan, 1991](#)), as well as in oligopolistic markets (see [Daughety and Reinganum, 2008b](#); [Janssen and Roy, 2010](#)).

In these models, firms, like the worker in Spence’s job-market model, are privately informed about their type. Generally, type refers to an economic ability— firms’ product quality or the worker’s productivity —resulting from past activities that remain outside the scope of the model. Hence, several researchers have taken type to be exogenously given. For simplicity, they have often distinguished between a “good” and a “bad” type that differs in their incentives to reveal the truth about their type. A standard feature of these models, known as the single-crossing property, guarantees that the good type is more willing than the bad type to use price as a costly signal. Firms earn a higher profit margin with lower quality, regardless of whether the market is monopolistic or oligopolistic. Therefore, firms of higher quality are less afraid of losing consumers by raising price to signal their type. This property is necessary for the existence of separating equilibria.

We take this signaling approach one step further by allowing firms to choose their

type. As with product quality, there is vertical differentiation between the two product variants obtained with and without social goodwill in the present setting. The choice to tie social goodwill in with the product differentiates the good type from the bad one. Another difference from models of price signaling quality is that our model combines two sources of information: price (endogenous) and imperfect monitoring (exogenous).

One important insight in the signaling literature is that the bad and good types experience two different trade-offs related to cheating and truth telling, respectively. A credibility constraint reflects the trade-off faced by the bad type. A bad firm may be tempted to use the fly-by-night strategy of setting the same price as the good type to trick uninformed consumers into mistaking the bad type for its good counterpart. However, by doing so, the bad type sacrifices the profit obtained by revealing the truth in the separating equilibrium. Usually, the credibility constraint is binding in the least costly separating equilibrium outcome on which the literature has focused a great deal of attention. Credibility sometimes requires that the good type incurs a positive signaling cost by distorting the price upward relative to the full information case. This signaling distortion may be a burden for both a monopoly ([Bagwell and Riordan, 1991](#)) and an oligopoly, albeit to a lesser extent ([Daughety and Reinganum, 2008b](#); [Janssen and Roy, 2010](#)). [Mahenc \(2008\)](#) obtains the same result when the product quality is environmental in nature: an upward distortion in the monopoly price is needed to signal cleaner products, unless cleaner products are cheaper to produce.

In the present setting, price distortion may restrict sales to such an extent in the absence of monitoring that the good type refrains from deterring fly-by-night strategies. It turns out that monitoring is needed for prices to be informative. We also show that monitoring reduces the size of the signaling distortion; this reduction increases as the accuracy of monitoring improves.

There is also a trade-off for the good type firm when revealing the truth. Although the signaling price yields more profit from uninformed consumers, the good type also

foregoes the profit earned in the best worst-outcome in which consumers mistake it for the bad type. This trade-off is formalized by the individual-rationality constraint of the good type. In the present setting, the profit accruing to the good type in the best worst-outcome depends on the probability that the bad type is cheating. We show that the good type succeeds in signaling social goodwill provided that the bad type is not too likely to cheat consumers.

The price signaling literature leaves open the moral hazard issue raised by the failure of signaling; firms may refrain from making the good-type-specific investment if prices provide no evidence for such an investment. In the present setting, we further examine the decision of whether to incur the sunk costs needed to be the good type, assuming that this decision is unobservable. For this, we consider that the probability of cheating at the time of signaling is exactly the probability of choosing the good type one step backward, in a mixed strategy over the set of types. In backward induction, the existence of a mixed strategy equilibrium requires that the probability of choosing the good type allows the firm first to overcome the opportunity cost of signaling this type (the individual-rationality constraint) and second to prevent the bad type from cheating (the credibility constraint). We find that there exists such a mixed strategy equilibrium, as well as two pure strategy equilibria in which either firm invests in social goodwill.

Dynamic models of quality premia ([Klein and Leffler, 1981](#); [Shapiro, 1983](#)) highlight the role played by repeat purchases and reputation in mitigating the moral hazard problem. In these models, firms choose both quality and price every period of an infinite horizon, while consumers cannot observe quality directly. However, the price and quality provided by firms in the past serve as signals of current quality. If quality has been observed to be high in previous periods, then consumers infer that firms are more likely to provide high quality currently. Moreover, consumers punish undue rents for low-quality products by ceasing to purchase them. Hence, firms have an

incentive to develop a reputation for high quality. For this mechanism to work in equilibrium, high quality must command a rent, i. e., a quality premium associated with a high price, such that the cost of losing repeat purchases exceeds the cost savings of cheating consumers. Our analysis deals with credence goods rather than experience goods. Therefore, we abstract from the reputation mechanism to focus on the role of imperfect monitoring in inducing firms to honor promises of high quality. Another difference from [Klein and Leffler \(1981\)](#) is that, in our analysis, competition between firms is imperfect, while it is perfect in their analysis.

The research [Daley and Green \(2014\)](#) is closely related to the present paper. These authors investigate the consequences for costly signaling of using grade as another instrument for information transmission about the sender's type in the canonical model of [Spence \(1973\)](#). In their model, the informed party is a worker who relies both on his education level and his performance on a test (a grade) to signal his ability to potential employers. The authors demonstrate that the presence of sufficiently informative grades dismisses separating equilibria as being less plausible than pooling equilibria. As usual, the worker's choice of education level serves as a signal that allows employers to update beliefs. These beliefs, in turn, influence the amount of information conveyed by the grades, which are used by employers as a redundant signal about the worker's ability. A high-ability worker finds signaling *via* education less costly than does a low-ability worker. However, above a certain level of education, the information conveyed through grades is more beneficial to the low-ability worker. This deters the high-ability worker from investing further in education to fully reveal information, and thus pooling prevails over separating. In the context of [Daley and Green \(2014\)](#), it is reasonable to assume that the use of education as a signal affects the information conveyed through grades because the caliber of grades obtained by a worker is likely to depend on the quality of that worker's past performance as a student. In contrast, there is no reason to assume that the two instruments available for information transmission are interlinked

in the context of our paper. Rather, we assume that the information conveyed by the auditor’s monitoring is not affected by the price signals sent by firms to formalize the idea that the two sources of information are independent from each other. Unlike grades in [Daley and Green \(2014\)](#), monitoring in our setting is used by the informed party to correct for arbitrary perceptions in case of deviation from the equilibrium path. These perceptions improve when the auditor’s report is more accurate because information released in such a way is complementary rather than redundant. Monitoring can be seen as a refinement of Bayesian equilibria in that it reduces the leeway in specifying off-the-equilibrium path beliefs. Our existence result of separating equilibria due to the presence of monitoring starkly contrasts with the prevalence of pooling equilibria due to grades in [Daley and Green \(2014\)](#). This is because independent monitoring strengthens the incentive of the good type to deter cheating with a costly signal, while grades weaken this incentive in the context of education.

### 3 The model

The protagonists in the economy are two firms, consumers and an independent third-party auditor. Both firms are price-takers for a conventional product. This product is a bundle of observable characteristics that can be properly verified by inspection and therefore are perfectly known to consumers. The product value can be enhanced by the social goodwill a firm chooses to tie in with the product. This is a new attribute that is public in nature and reflects the firm’s concern for the social and environmental externalities of its economic activity. The new attribute has the following features:

- (i) it is vertical in the sense that, everything else being equal, all consumers agree that the product is more valuable with than without the attribute;
- (ii) it entails additional costs for the firm that chooses to invest in all the factors needed to provide the attribute, including labor, managers and capital;

- (iii) consumers cannot directly observe whether the product has or has not the attribute, either prior to or subsequent to consumption;
- (iv) a specific label is intended to guarantee that the product has the new attribute, but this certification may not be credible.

Although the spectrum of externalities associated with the public attribute encompasses all kinds of social concerns, political, ethical and environmental, we consider that the new attribute is environmental for the sake of illustration. Hence, we refer to the product with the new attribute as being the “green” type, and to the conventional product as being the “brown” type.

The market for the brown product is in a long-run competitive equilibrium, representing business as usual. A firm can command a price premium over its product by switching to green production. However, the decision of whether to switch is unobservable, and firms use prices only to signal their type. The brown rival, in turn, may counteract both product differentiation and information disclosure with the fly-by-night strategy of tricking consumers into believing it is also selling a green product. Our goal is to concentrate on this strategic interaction. The absence of strategic behaviors within the brown industry avoids the complexity of simultaneous signaling recognized by [Mailath \(1988\)](#). We assume that the brown technology has constant return to scale, i. e., firms have access to perfectly elastic supplies of all the factors needed for brown production.<sup>5</sup>

---

<sup>5</sup>For instance, we could use the model of monopolistic competition developed by [Salop \(1979\)](#) to formalize business as usual in our setting. Under this framework, the market for the brown product is represented by a circle of unit length on which consumers are uniformly distributed. Each point on the circle corresponds to one consumer’s most preferred variant of the brown product. A large number of firms are symmetrically located around the circle. Each firm produces a single variant of the brown product with an identical linear technology. There is free entry into the brown market so that firms continue to enter until profits are driven to zero. If  $t$  is the transport cost per unit of distance from an ideal location,  $c$  is the constant unit product cost and  $f$  is the fixed cost of entry, the market price for the brown product in the symmetric equilibrium is  $p_b = c + \sqrt{tf}$ . These additional variables would make calculations more cumbersome and risk clouding the issue.

Switching from brown to green production requires investment in real and financial assets, which involves some fixed setup cost  $F > 0$ <sup>6</sup> and an additional marginal cost  $c(e) = e$ , where  $e$  is an indicator of the extent to which the product value is increased by the new attribute.

If either firm chooses to stick to the brown type of product, it will be indexed by  $t = b$ , or else by  $t = g$  if it invests in the green type. The firm  $i = 1, 2$  of type  $t \in T = \{b, g\}$  sells its product at price  $p_{it}$ .

The total number of consumers is normalized to unity. Each consumer has exogenous wealth of  $w$  and purchases at most one unit of either type of the product. Consumers have heterogeneous preferences that differ according to a taste parameter  $x$  for the new attribute, which is assumed to be continuously and uniformly distributed over the interval  $[0, l]$ . Thus, the indirect utility of a consumer with taste  $x$  for the attribute valued  $e_t$ , who purchases one unit of the type- $t$  product at price  $p_{it}$ , is given by

$$V_t(p_{it}, e, x) = w + xe_t - p_{it}, t \in T, i = 1, 2 \quad (1)$$

where  $e_g = e$  and  $e_b = 0$ .

*Demands.*—Market demands are determined from a critical value of the taste distribution. The preference level of the consumer who is indifferent between purchasing the brown product and its green substitute is found by setting

$$w + xe_g - p_{ig} = w - p_{jb} \quad (2)$$

and solving for  $x$ . Doing so yields

$$X = \min\left\{\frac{p_{ig} - p_{jb}}{e_g}, 1\right\}. \quad (3)$$

All consumers with values of  $x$  that satisfy  $x \geq X$  purchase the green product and

---

<sup>6</sup> $F$  includes unsalvageable expenditures in developing the new attribute and specific entrepreneurial skills, sunk investments in specialized machinery for green production, fuel-saving equipment or long-term rental contracts that cannot be resold, as well as the fee paid to have the product certified.

the remaining consumers purchase the brown product. This implies linear demand functions for both products. Generally, the demand for firm  $i$  of type  $g$ , when it charges price  $p_{ig}$  and the rival firm  $j$  charges  $p_{jg}$  will be denoted by  $D_{ig}(p_{ig}, p_{jg})$ . This function can take two different forms depending on whether firm  $j$  is brown or green (see Appendix 1).

The profit earned by firm  $i$  in the green market is

$$\pi_{ig}(p_{ig}, p_{jb}) = (p_{ig} - e) D_{ig}(p_{ig}, p_{jg}). \quad (4)$$

One possibility in the green market equilibrium is that the cost of providing the attribute is so high that consumers have zero demand for the green product, even when it is sold at marginal cost. Let  $p_b^c$  denote the market price in the long-run competitive equilibrium. We substitute  $p_{ig}$  and  $p_{jg}$  for  $e$  and  $p_b^c$ , respectively, into (45) given in Appendix 1. As a result, the following inequality ensures that demand for the green product is strictly positive when it is sold at marginal cost

$$(l - 1)e + p_b^c > 0. \quad (5)$$

Roughly, this assumption says that consumer heterogeneity is sufficient for the existence of a green market.

Till now, we have presented the full information framework in which the new attribute is observable. We now turn to the issue of information transmission over the sale period in the green market. We assume that green certification is not trustworthy due to “greenwash” or “launder” trafficking in illegal products. Consumers are unsure whether a product carrying the seal “green” truly has the new attribute. However, consumers are likely to form perceptions about the type of the product they are purchasing, either by observing prices, or by reading reports and test results about the value of the green product, released by the third-party auditor.

*The third-party auditor.*—The auditor has the technology to monitor whether the

product in the green market really has the new attribute. The monitoring is achieved by observing differences between the firm's claims and the truth about its type. For example, the auditor measures all polluting emissions generated by the product and compares them with the standard required to be green: if the polluting emissions are observed to be lower than the standard, then the product passes the test. The auditor's monitoring is not perfect: it succeeds in learning the true type of a firm with the probability  $\alpha$ ,  $0 \leq \alpha \leq 1$ ; thus,  $\alpha$  is a parameter that indicates the accuracy of the monitoring technology.

The auditor's observation takes place over the sale period simultaneously with firms' pricing. Moreover, firms and the auditor are independent players, so they do not influence each other regarding their information transmission. At the end of the monitoring process, the auditor releases a report that supplements the information consumers infer upon observing firms' prices. Consumers make their purchase decision using the two sources of information: the auditor's report and firms' price signaling.

*The moral hazard issue.*—Bayesian consumers form perceptions of a firm's type based on observed prices. This is a standard signaling issue in the spirit of [Spence \(1973\)](#), that has been widely investigated using incomplete-information models. In these models, Nature selects the types of the signal sender according to some exogenous probability distribution. The type usually refers to an economic ability resulting from past investment made by the signal sender. For instance, the good type in signaling models of quality represents firms eager to truthfully signal quality because they have previously chosen to provide improved quality, which is both more costly to produce and more valued by consumers. What is hidden in signaling games is information about the firm's type, which implicitly assumes that the firm's exogenous choice of becoming the good type was not observed by the other players of the game.

Rather, in the present framework, we investigate firms' quality choice when this decision is unobservable to the other players, i. e., consumers, the firm's rival and the

auditor. For this, we allow firms rather than Nature to select their type and we tackle the problem of moral hazard raised by firms' hidden actions.

The main questions are: Is it worthwhile for a firm to switch to green production if the switch is not observable? What if the expected costs of signaling the switch *via* prices are prohibitive? Does the auditor's monitoring help the industry to convert toward green production under these circumstances?

To answer these questions, we transform the signaling model of asymmetric information with hidden knowledge into a model of symmetric information with hidden actions. Unobserved actions are the technological decisions of whether to switch to green production, and the observed actions are both the prices set by firms and the report released by the auditor. A standard assumption in signaling games is that there exists a fraction  $\sigma$  of good-type firms versus a fraction  $1 - \sigma$  of bad-type firms, and this exogenous distribution is common knowledge in the economy. We reinterpret this probability distribution over the set of types as a mixed strategy over the set of pure strategies for either type of firm in our model. This mixed strategy captures public uncertainty about what a firm does regarding its type. Hence, the distribution of types must emerge as the (Nash) equilibrium randomization of a firm's decision to switch to green production or to stick to brown production.

*The timing.*—The whole game is a three-stage game that proceeds as follows:

- (i) In the first stage, the two firms simultaneously choose their types. If a firm decides to switch to green production, it pays the setup cost  $F$ .
- (ii) In the second stage, firms simultaneously post their prices and the auditor releases its report. Consumers observe these actions.
- (iii) In the third stage, consumers update their beliefs about the firms' types upon seeing prices, supplement this information with that conveyed by the auditor's

report and, finally, decide from which firm to buy. Consumers use Bayes’s rule whenever possible to form posterior beliefs from the observed prices.

The game begins with the technology decisions made by firms regarding their type: each firm  $i = 1, 2$ , selects a type  $t$  from the set  $T = \{b, g\}$  and may randomize over these pure strategies. Firms are committed to their technology decisions until the end of the game: they will not change their types in the next stages because the appropriate production facilities have a second-hand value lower than their initial value. The technology decision is private information to each firm: its type is unknown to every other protagonist—the rival firm, consumers and the auditor—This implies that the setup cost  $F$  specific to the green production is not observable either. A mixed strategy for firm  $i$ ,  $\sigma_i : T \rightarrow [0, 1]$ , assigns probability  $\sigma_i(t)$  that it chooses the type  $t$ , where  $\sum_{t \in T} \sigma_i(t) = 1$ . This probability distribution summarizes all the information publicly available to all but the perfectly-informed firm  $i$  at the end of the first stage.

In stage 2, the setup cost  $F$  is sunk, if ever. The probability distributions  $\sigma_i(t)$ ,  $i = 1, 2$ , are the only statistics for past play; hence they become the prior beliefs of the uninformed protagonists at the beginning of stage 2. With these beliefs in mind, firms charge prices  $p_{it} \geq 0$ ,  $i = 1, 2$ . The sequential order between the decisions regarding technology and price captures the notion that a price can in practice be varied at will, unlike the commitment to a type.

It may happen that one type of product is unavailable on the market if both firms decide to supply the other type. In that event, Bertrand competition in the single product market fully reveals information about the firms’ type. Asymmetric information is a trickier issue in the event that two differentiated products are available on the market. The subgame starting at every “informational node” involving two distinct types works like a standard signaling game, in which the probability distributions  $\sigma_i(t)$ ,  $i = 1, 2$  are common knowledge. In other words,  $\sigma_i(t)$  is the probability assigned

by everyone but firm  $i$  to the event that firm  $i$  has chosen type  $t$ , given the presence of two differentiated products on the market. In particular, firm  $i$ 's rival will use this probability distribution to predict how firm  $i$  price discriminates between the two potential types in a separating equilibrium. Firms' unobserved randomization over types requires that subsequent signaling *via* prices be credible: in equilibrium, the green-type firm must deter its brown-type counterpart from cheating about its type.

In stage 3, consumers draw inferences about the firm's types and cross-check this information with that released by the auditor's report. Finally, consumers make their purchase decisions to maximize their expected payoffs, given their posterior beliefs. The payoff to a consumer is her expected net surplus if she buys, and zero otherwise. The payoff of each firm is its expected profits.

We can focus on firms' strategic interplay in the green market due to the assumption of zero profits (perfect competition) in the brown market. We require that strategies in the green market form a Perfect Bayesian Equilibrium (PBE); that is, strategies must yield a Bayesian equilibrium not only for the whole game, but also for every subgame, including that starting after any possible choice of a type made by firms.

## 4 The signaling game with hidden actions

We solve the game by backward induction. For this, we first investigate the signaling subgame starting at every informational node where the market is segmented between two differentiated products and the setup cost specific to green production is sunk. When firms get to move at the second stage of the game, every protagonist other than firm  $i$  believes that it has chosen the type  $t$  with probability  $\sigma_i(t)$ .

The signaling issue in stage 2 raises the familiar problem of multiplicity of equilibria. To simplify, our analysis of price signaling focuses on pure-strategy separating equilibria in which the brown and the green types choose different prices. Separating prices in equilibrium ensure that green certification is credible by truthfully disclosing

information about the actual types. Following Mahenc (2017), the credibility of green certification requires separating prices in the green market, and conversely, pooling prices (in which the firm’s price is independent of its true type) undermine the reliability of certification. Were prices to pool in the green market, information would be concealed by market prices, making information disclosed by the green label inconsistent.

In stage 2, firms’ pure strategies are vectors of four prices  $\{(p_{it})_{i=1,2,t \in T}\}$ . Given that firms are completely symmetric at the beginning of the whole game, we assume that firms employ the same pricing rule in equilibrium, and so,  $p_{1t} = p_{2t} = p_t$  for each  $t \in T$ . Every protagonist other than firm  $i$  makes the prediction that it will employ the pricing rule  $\rho_i$  phrased as follows: “firm  $i$  will charge price  $p_b$  with probability  $\sigma_i(b)$  and price  $p_g$  with probability  $\sigma_i(g)$ ”, where  $\sigma_i(b) = 1 - \sigma_i(g)$ . We can simplify the notation  $\sigma_i(g)$  and write  $\sigma_i$  instead.

In stage 3, consumer perception of the firms’ types builds on information from two sources: firms’ prices and the auditor’s report. We assume that the information released by firms through price setting does not influence that released by the auditor’s monitoring. Therefore, we treat the two events {the firm is green conditional on observing prices, the firm is green conditional on reading the auditor’s report} as being both independent and mutually non-exclusive.<sup>7</sup>

To formalize consumer perception of a firm’s type based on price, we define a posterior belief function  $\mu(t, p) : T \times R^+ \rightarrow [0, 1]$ , that specifies the probability assigned to either firm of being type  $t$  in response to a price  $p$  observed in the market for the green product. This belief function is the same function for both firms. Along the equilibrium

---

<sup>7</sup>Let  $A$  be the event that the firm is green after observing prices, and  $B$  the event that the firm is green based on the auditor’s monitoring. The probabilities of  $A$  and  $B$  are  $\mu = \Pr(A)$  and  $\alpha = \Pr(B)$ , respectively. If  $A$  and  $B$  are independent events, then the joint probability of both occurring is  $\Pr(A \text{ and } B) = \Pr(A)\Pr(B)$ . If  $A$  and  $B$  are not mutually exclusive events, then the probability of either occurring is  $\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \Pr(A \text{ and } B)$ . Thus, if  $A$  and  $B$  are both independent and mutually non-exclusive events, the probability that the firm is green after observing either prices or monitoring is  $\Pr(A \text{ or } B) = \mu + \alpha - \mu\alpha$ .

path, consumers use Bayes's rule to update beliefs from the prior distributions  $\sigma_i, i = 1, 2$ , available at the beginning of the signaling game. When consumers observe a price off the equilibrium path, they update beliefs with an arbitrary rule instead of Bayes's rule.

We write  $e(\mu) = \sum_{t \in T} \mu(t, p) e_t$  for the expected value consumers infer from price  $p$ . Consumers supplement their information about a firm's type with the auditor's report. Assuming that firms and the auditor have no influence on each other in providing information, we describe the overall formation of consumer perception by the equation

$$\tilde{e}_t(\mu) = \alpha e_t + (1 - \alpha) e(\mu). \quad (6)$$

Equation (6) gives consumers' expected valuation for a product of type  $t$  sold in the green market, given consumer final beliefs. If the auditor's observation provides no information, then consumer perception relies only on inferences from prices. If the auditor's monitoring is perfectly accurate, then consumers learn the true type. Note that equation (6) applies both on and off the equilibrium path, i. e., regardless of whether consumers use Bayes's rule or an arbitrary rule to update beliefs after observing prices. When Bayes's rule applies in equilibrium to beliefs formed from pure-strategy prices  $(p_b, p_g)$ , posterior beliefs are  $\mu(g, p_t) = 1 - \mu(b, p_t) = 1$  for  $t \in T$ , giving consumers correct perceptions whatever the type, since

$$\tilde{e}_t(\mu(t, p_t)) = \alpha e_t + (1 - \alpha) e \mu(t, p_t) = \begin{cases} e & \text{if } t = g, \\ 0 & \text{if } t = b. \end{cases} \quad (7)$$

Hence, consumers align their readings of the auditor's report with updated beliefs when prices reveal the truth. Otherwise, consumers observe a deviation from price equilibrium. Then, given an arbitrary belief  $\mu$  based on a probability-0 price, consumer perception is

$$\tilde{e}_t(\mu) = \alpha e_t + (1 - \alpha) \mu e = \begin{cases} [\alpha + (1 - \alpha) \mu] e & \text{if } t = g, \\ (1 - \alpha) \mu e & \text{if } t = b. \end{cases} \quad (8)$$

The motivation for this is that consumers use the information provided by monitoring to correct for arbitrary perceptions when something “surprising” occurs. In the extreme case where beliefs based on prices are pessimistic ( $\mu = 0$ ), even if the actual type is green, monitoring reduces the leeway in specifying off-the-equilibrium path beliefs. Moreover, the perception consumers have about the type, i. e.,  $\tilde{e}_g(0) = \alpha e$ , improves with monitoring accuracy. In some sense, monitoring helps consumers refine the multiplicity of equilibria supported by unrestricted beliefs based on unexpected prices.<sup>8</sup>

Finally, consumers make their purchase decisions to maximize their expected payoffs, in accordance with (6).

From now on, we will simplify the notation  $\mu(g, p)$  and write  $\mu(p)$  instead, so that  $\mu(b, p) = 1 - \mu(p)$ . The market split between products depends on consumer conjectures about the firms’ types. Consider the following: (i) firm  $i$  of type  $t$  is perceived by consumers to be green with probability  $\mu$ ; (ii) the rival firm  $j$  of type  $t'$  is perceived to be green with probability  $\sigma$ ; and (iii) the expected valuation of the product of firm  $i$  exceeds the expected valuation of the product of firm  $j$ ; that is,  $\tilde{e}_t(\mu) > \tilde{e}_{t'}(\sigma)$ . Then, the critical value of the taste distribution that determines market demands becomes

$$\tilde{X}(\mu, \sigma) = \min\left\{\frac{p_{it} - p_{jt'}}{\tilde{e}_t(\mu) - \tilde{e}_{t'}(\sigma)}, 1\right\}. \quad (9)$$

---

<sup>8</sup>The way monitoring restricts beliefs in our setting clearly differs from how beliefs update based on grades in the paper of Daley and Green (2014). These authors assume that how beliefs update after observing a signal and grades follows a two-stage process. In the first stage, receivers observe a signal through the investment in education, and they use Bayes’s rule to update beliefs, as consumers do after observing prices in the present paper. This first updating results in “interim beliefs”, based on the history  $h^1$  of the game in the first stage. In the continuation game starting at the second stage, receivers observe grades and again update beliefs from their interim beliefs *via* Bayes’s rule. This requirement is stronger than simply using Bayes’s rule in the usual fashion since it applies to updating from the first stage to the second stage, whether or not  $h^1$  has probability 0, and whether or not the signal sent in the first stage has probability 0. Hence, the likelihood of grades implicitly depends on the observation of prices *via* the second Bayesian updating. Applying Bayes’s rule twice captures a form of redundancy in the information transmission, which is not desirable in the context of our paper. The sender influences the decision to give a grade by sending a costly signal in Daley and Green (2014). Rather, we assume that price signaling does not interfere with monitoring.

We denote by  $D_{it}(p_{it}, p_{jt'}, \mu, \sigma)$  the demand for the product sold by firm  $i$  in the green market resulting from the market split at (9).<sup>9</sup> Clearly, this demand depends on whether consumers have an expected valuation for firm  $i$ 's product that is higher or lower than that for firm  $j$ 's product.

A firm's profit can be written as a function of its true type, its perceived type and its price, given the perceived type and the price of its rival. We denote the profit for firm  $i$  of type  $t$ , when it charges price  $p_{it}$ , its perceived type is green with probability  $\mu$  and the rival firm  $j$  of type  $t'$  is perceived to be green with probability  $\sigma$  and charges  $p_{jt'}$ , by

$$\pi_{it}(p_{it}, p_{jt'}, \mu, \sigma) = (p_{it} - e_t)D_{it}(p_{it}, p_{jt'}, \mu, \sigma), \text{ for } t, t' \in T, i = 1, 2. \quad (10)$$

In the price signaling subgame, a pure-strategy separating PBE consists of a set of price strategies and beliefs  $\{(p_t^*)_{t \in T}, \mu(p_t^*)\}$  such that  $p_b^* \neq p_g^*$ ,  $\mu(p_b^*) = 0$  and  $\mu(p_g^*) = 1$ ; that is, consumer perception of the firms' types is correct after observing separating prices. The separating pricing rule used by firm  $i$  in equilibrium is denoted by  $\rho_i^*$ ; it predicts: "firm  $i$  will charge the equilibrium price  $p_b^*$  with probability  $1 - \sigma_i$  and the equilibrium price  $p_g^*$  with probability  $\sigma_i$ ". Furthermore, we know that  $p_b^*$  is given by  $p_b^c$  in the long-run equilibrium.

Generally, we define  $E[\pi_{it}(p, \mu) / \rho_j^*]$  as the expected profits for firm  $i$  of the actual type  $t$ , perceived to be green with probability  $\mu$  after observing the price  $p$ , where  $\sigma_j$  is firm  $i$ 's prediction about firm  $j$ 's separating pricing rule. In particular,  $E[\pi_{ig}(p_g^*, 1) / \rho_j^*]$  is firm  $i$ 's expected profits resulting from a pure-strategy PBE in the price signaling subgame. In the event that firm  $j$  charges  $p_b^*$ , firm  $i$ 's product sold at price  $p_g^*$  in the green market is more valuable to consumers than the rival product. Alternatively, if firm  $j$  charges the same price  $p_g^*$  as does firm  $i$ , then consumers perceive the rival products as the the same. In both events, firm  $i$ 's demand is calculated from

---

<sup>9</sup>The explicit form of the demand function is given in Appendix 1.

the market split (9) by substituting  $\mu = 1$  and  $\sigma = 0$  (resp. 1) in the former (resp. latter) event. So, we can write

$$E [\pi_{ig}(p_g^*, 1) / \rho_j^*] = (1 - \sigma_j) \pi_{ig}(p_g^*, p_b^*, 1, 0) + \sigma_j \pi_{ig}(p_g^*, p_g^*, 1, 1). \quad (11)$$

So far, we have restricted our attention to the signaling subgame starting with prior beliefs  $\sigma_j$  that reflect the common prediction about the price selection made by firm  $j$ ' for its product. This information is derived from that publicly available at the end of the first stage of the game, in which each firm chooses a mixed strategy from the set of probability distributions over  $T$ . The randomization over types summarizes public uncertainty about what each firm does at the first stage. As there is nothing that can alter beliefs between the end of stage 1 and the beginning of stage 2, we assume that the information summarized by the mixed strategies over types are the same as the information used to make subsequent predictions about firms' pricing. Firm  $i$ 's mixed strategy is the belief that its rival will play the pure strategies  $\{b, g\}$  with the probabilities  $(1 - \sigma_j, \sigma_j)$ . Whatever firm  $i$  might think about the rival's choice in the first stage, firm  $i$ 's expected profits from sticking to brown production are zero since the market for the brown product is in a long-run competitive equilibrium. Hence,  $E [\pi_{ib}(p_b^*, 0) / \rho_j^*] = 0$ . If firm  $i$  now chooses to switch to green production; it pays the setup cost  $F$  and earns expected profits

$$E [\pi_{ig}(p_g^*, 1) / \rho_j^*] - F. \quad (12)$$

Firm  $i$ 's expected profits from playing a mixed strategy in stage 1 are the weighted sum of the expected profits for each of the pure strategies  $\{b, g\}$ , where the weights are the probabilities  $(1 - \sigma_i, \sigma_i)$

$$\Pi_i(\sigma_i, \sigma_j) = (1 - \sigma_i) E [\pi_{ib}(p_b^*, 0) / \rho_j^*] + \sigma_i [E [\pi_{ig}(p_g^*, 1) / \rho_j^*] - F] \quad (13)$$

$$= \sigma_i [(1 - \sigma_j) \pi_{ig}(p_g^*, p_b^*, 1, 0) + \sigma_j \pi_{ig}(p_g^*, p_g^*, 1, 1) - F]. \quad (14)$$

Our goal is now to characterize firm  $i$ 's mixed strategies given by the probabilities  $(1 - \sigma_i^*, \sigma_i^*)$  of choosing a type from  $T$ , and separating prices  $(p_b^*, p_g^*)$ , as a subgame PBE of the whole game. In any subgame PBE, no firm must want to change its decision about its type given the decision made by the other firm about its type. This requirement must be met when firms anticipate that prices will subsequently reveal their true types. The mixed-strategy profile  $(\sigma_i^*, \sigma_j^*)$ ,  $i \neq j$  is a Nash equilibrium in which firm  $i$  chooses to switch to green production if and only if  $\sigma_i^*$  is a best response to firm  $j$ 's equilibrium mixed strategy  $(1 - \sigma_j^*, \sigma_j^*)$  under the individual-rationality constraint of non-negative profits from green production; that is, for every  $\sigma_i \in [0, 1]$ ,

$$\Pi_i(\sigma_i^*, \sigma_j^*) \geq \max\{0, \Pi_i(\sigma_i, \sigma_j^*)\}. \quad (15)$$

Indeed, if  $\Pi_i(\sigma_i, \sigma_j^*) < 0$ , then firm  $i$  chooses the brown type and earns zero profit, which is strictly better than the loss in profit entailed by paying the setup cost  $F$  to become green. From (14), inequality  $\Pi_i(\sigma_i, \sigma_j^*) \geq 0$  determines an upper bound for firm  $j$ 's equilibrium probability of choosing the green type, which must satisfy

$$F \leq E[\pi_{ig}(p_g^*, 1) / \rho_j^*]. \quad (16)$$

Once the setup cost of switching to green production is sunk, firms compete for business. Either firm selling its product on the green market maximizes its expected profit with respect to price, given the auditor's report and given the consumer beliefs function. Beliefs are formed from equilibrium prices and the auditor's report, using Bayes's rule for prices with positive probability, and an arbitrary rule otherwise. This is formalized in the following definition. The set of mixed strategies in type, pure strategies in price and beliefs formed from prices  $\{(\sigma_i^*)_{i=1,2}, (p_t^*)_{t \in T}, \mu(p_t^*)\}$  is a PBE if the following four conditions are satisfied:

1. The market for the brown product is in a long-run competitive equilibrium

$$p_b^* = p_b^c. \quad (17)$$

2. Consumers form posterior beliefs from prior beliefs  $\sigma_i, i = 1, 2$  using Bayes's rule

$$p_g^* \neq p_b^* \text{ and } \mu(p_g^*) = 1. \quad (18)$$

3. Prices in the green market are optimal given the separating pricing rule  $\rho_j^*$ , consumer beliefs and the auditor's report

$$p_g^* = \arg \max_p E [\pi_{ig}(p, 1) / \rho_j^*], i \neq j. \quad (19)$$

4. The mixed strategy profile  $(\sigma_1^*, \sigma_2^*)$  is a Nash equilibrium under the constraint of non-negative profits from green production

$$\sigma_i^* = \arg \max_{\sigma} \Pi_i(\sigma, \sigma_j^*) \text{ subject to (16)}, i = 1, 2. \quad (20)$$

As previously mentioned, separation can occur in the price signaling subgame only if one firm is green and the other is brown. Henceforth, we consider, without loss of generality, that firm 1 has chosen to be green and firm 2 has chosen to be brown. With this convention, (6) says that consumer expected valuation for firm 2's product is  $\tilde{e}_b(\mu(p_b^*)) = 0$  in equilibrium. This yields the expected profit  $E[\pi_{2b}(p_b^*, 0) / \rho_1^*] = 0$ , given the prediction that firm 1 employs the pricing rule  $\rho_1^*$ .

However, it might be profitable for firm 2 to mimic its green-type rival, given the prediction that firm 1 charges the price  $p_b^*$  with probability  $1 - \sigma_1$ . In that case, consumers will definitely perceive firm 1 as being brown. Then, firm 2 has the leeway to set the price  $p_g^*$  upon which consumer perception of its product is  $\tilde{e}_b(\mu(p_g^*)) = (1 - \alpha)e$ , from (6). By claiming that its product is green, firm 2 can differentiate it from the brown product. Obviously, this claim is worthless if firm 1 also charges

the price  $p_g^*$ , which occurs with probability  $\sigma_1$ . Indeed, consumers will then correctly value the product sold by firm 1, inferring that  $\tilde{e}_g(\mu(p_g^*)) = e$ , which strictly exceeds  $\tilde{e}_b(\mu(p_g^*)) = (1 - \alpha)e$  for  $\alpha > 0$ . It turns out that the auditor's monitoring in the green market allows some differentiation between the two products in favor of firm 1, despite firm 2's fly-by-night strategy. In contrast, with no monitoring ( $\alpha = 0$ ) the two products look the same in consumers' eyes.

Assume that consumers infer  $\mu(p) = 1$  upon seeing a price  $p$  charged by firm 2. Using (9), we denote firm 2's expected demand by

$$E[D_{2b}(p, 1) / \rho_1^*] = (1 - \sigma_1) D_{2b}(p, p_b^*, 1, 0) + \sigma_1 D_{2b}(p, p_g^*, 1, 1). \quad (21)$$

The functional forms of (21) depend on whether  $\alpha$  is positive or zero, as shown by (46) in Appendix 1. Setting  $p_g^*$  allows firm 2 to expect the following profits

$$E[\pi_{2b}(p_g^*, 1) / \rho_1^*] = p_g^* E[D_{2b}(p, 1) / \rho_1^*]. \quad (22)$$

To prevent firm 2 from imitating firm 1, the separating price  $p_g^*$  must satisfy the following credibility constraint:

$$E[\pi_{2b}(p_b^*, 0) / \rho_1^*] \geq E[\pi_{2b}(p_g^*, 1) / \rho_1^*]. \quad (23)$$

The left-hand side of (23) is the profit earned by firm 2 in a separating equilibrium, which is zero. The right-hand side is the profit firm 2 can expect if it mimics the green type, thereby tricking consumers into buying at a positive price. Thus, (23) means that the brown firm has nothing to lose from misleading consumers. Everything else being equal, this profit decreases in  $\alpha$  because  $\frac{\partial E[\pi_{2b}(p_g^*, 1) / \rho_1^*]}{\partial \alpha} = \frac{p_g^*(p_g^* - p_b^*)(\sigma_1 - 1)}{e l (1 - \alpha)^2} < 0$ , but it is also increasing in  $l$ . Therefore, the more accurate the auditor's report, the lower is the temptation to cheat for the brown type. However, a larger heterogeneity of consumer preferences for the new attribute strengthens firm 2's incentive to deviate

from equilibrium.

For all  $\alpha \geq 0$ , inequality (23) determines a lower bound for the equilibrium price,  $\bar{p}_b$ , which must satisfy

$$p_g^* \geq \bar{p}_b, \quad (24)$$

where

$$\bar{p}_b = \begin{cases} l(1 - \alpha)e + p_b^c & \text{if } \alpha > 0, \\ le + 2p_b^c \frac{1 - \sigma_1}{2 - \sigma_1} & \text{if } \alpha = 0. \end{cases} \quad (25)$$

**Lemma 1:** *In any separating equilibrium,  $p_b^* = p_b^c$  and  $p_g^* \geq \bar{p}_b$ .*

Hence,  $\bar{p}_b$  defines the threshold below which it would be misleading to set  $p_g^*$  for the green product. Condition (24) guarantees that the equilibrium price in the green market is high enough to deter mimicry by firm 2, which then reverts to  $p_b^*$ .

Let us now turn to firm 1's expected profits under different consumer perceptions of its type. Given consumer beliefs  $\mu = \mu(p)$  and the pricing rule  $\rho_2^*$ , we denote firm 1's expected demand by

$$E[D_{1g}(p, \mu) / \rho_2^*] = (1 - \sigma_2) D_{1g}(p, p_b^*, \mu, 0) + \sigma_2 D_{1g}(p, p_g^*, \mu, 1). \quad (26)$$

Again, firm 1's demand is derived from the market split (9) by substituting  $\sigma = 0$  or 1, depending on whether firm 2 signals its true type or uses the fly-by-night strategy. Appendix 1 provides in-depth analysis of firm 1's expected demand with and without monitoring. When firm 1 charges  $p_g^*$ , consumer perception of its type is correct, that is,  $\tilde{e}_g(\mu(p_g^*)) = e$ . Given that firm 2 uses the separating pricing rule  $\rho_2^*$ , firm 1 predicts that firm 2 charges the price  $p_b^*$  for the brown product with probability  $1 - \sigma_2$ . Then, differentiation between the brown and the green products generates a demand for the green product equal to  $D_{1g}(p_g^*, p_b^*, 1, 0) = \frac{le + p_b^* - p_g^*}{le}$ . If, in contrast, firm 2 chooses the price  $p_g^*$ , which occurs with probability  $\sigma_2$ , consumers correctly infer that firm 1's product is more valuable than firm 2's product sold at the same price, provided that  $\alpha >$

0, since  $\tilde{e}_b(\mu(p_g^*)) = (1-\alpha)e < e$ . The auditor's monitoring makes all consumers switch toward firm 1's product and the demand for the green product is  $D_{1g}(p_g^*, p_g^*, 1, 1) = \frac{le-p_g^*}{le}$ . However, when firm 2 uses the fly-by-night strategy, differentiation between the products vanishes in the absence of monitoring ( $\alpha = 0$ ).

Substituting  $\mu = 1$  in (26), we can write firm 1's expected profit when it charges  $p$  and consumer perception is  $\tilde{e}_g(\mu(p)) = e$  as follows

$$E[\pi_{1g}(p, 1)/\rho_2^*] = (p - e)E[D_{1g}(p, 1)/\rho_2^*], \quad (27)$$

where  $E[D_{1g}(p, 1)/\rho_2^*]$  is given by (50) in Appendix 1. Maximizing these expressions with respect to  $p$ , we compute firm 1's best response to the pricing rule  $\rho_2^*$  to obtain

$$p_{1g}(1) = \begin{cases} \frac{e(1+l)+p_b^c(1-\sigma_2)}{2} & \text{if } \alpha > 0, \\ \frac{e(1+l)}{2} + p_b^c \frac{1-\sigma_2}{2-\sigma_2} & \text{if } \alpha = 0. \end{cases} \quad (28)$$

The maximized profit can be written as

$$E[\pi_{1g}(p_{1g}(1), 1)/\rho_2^*] = \begin{cases} \frac{[e(l-1)+p_b^c(1-\sigma_2)]^2}{4le} & \text{if } \alpha > 0, \\ \frac{[e(l-1)(2-\sigma_2)+2p_b^c(1-\sigma_2)]^2}{8le(2-\sigma_2)} & \text{if } \alpha = 0. \end{cases} \quad (29)$$

We see from (28) that the optimal price  $p_{1g}(1)$  may be too low to satisfy the credibility constraint (24). This occurs when  $p_{1g}(1) < \bar{p}_b$ , in which case signaling the green type becomes costly, or even impossible, due to the loss of consumers switching to the brown substitute.

As previously mentioned, several candidates for  $p_g^*$  may satisfy (24). In order to avoid this, we focus on least costly separating equilibria, which are robust to the Cho-Kreps intuitive criterion (Cho and Kreps, 1987). If  $p_{1g}(1) < \bar{p}_b$  in the present setting, the least costly separation can be achieved by setting the price  $\bar{p}_b$  for the green product. The minimum signaling cost is measured by the price differential

$$\bar{p}_b - p_{1g}(1) = \begin{cases} \frac{\epsilon}{2}(l-1-2l\alpha+p_b^c(1+\sigma_2)) & \text{if } \alpha > 0, \\ \frac{\epsilon}{2}(l-1)+p_b^c \frac{2-\sigma_2(3-\sigma_1)}{(2-\sigma_1)(2-\sigma_2)} & \text{if } \alpha = 0. \end{cases} \quad (30)$$

The upward distortion in price includes the forgone profit from fly-by-night strategies, thereby forestalling misleading prices. It is unsurprising that the signaling cost decreases as the auditor's monitoring accuracy increases because the cheating profit  $E[\pi_{2b}(p_g^*, 1)/\rho_1^*]$  decreases in  $\alpha$ , as previously shown. One can check that  $\bar{p}_b > p_{1g}(1)$  for all  $\alpha < \bar{\alpha} = \min\{1, \frac{l-1}{2l} + p_b^c \frac{1+\sigma_2}{2el}\}$ .

**Lemma 2:** *In any separating equilibrium, the least costly way of signaling the green type is to set*

$$p_g^* = \begin{cases} p_{1g}(1) & \text{if } \alpha \geq \bar{\alpha}, \\ \bar{p}_b & \text{otherwise.} \end{cases} \quad (31)$$

If the auditor's monitoring is not sufficiently accurate, signaling the green type *via* prices becomes costly due to the brown firm's mimicry. With sufficient accuracy, monitoring saves the cost of signaling the green type. Note, however, that the minimum signaling cost increases in  $l$  when  $\alpha < \bar{\alpha}$ : it may be more difficult to signal the green type in economies with very heterogeneous preferences for the new attribute because the temptation to cheat is stronger for the brown type.

When  $p_g^* = \bar{p}_b$ , firm 1's expected profits are

$$E[\pi_{1g}(\bar{p}_b, 1)/\rho_2^*] = \begin{cases} \frac{[p_b^c + e(l(1-\alpha)-1)](el\alpha - p_b^c\sigma_2)}{le} & \text{if } \frac{p_b^c\sigma_2}{el} < \alpha < \bar{\alpha}, \\ \max\{0, p_b^c \frac{[2p_b^c(2-\sigma_1) + e(l-1)(2-\sigma_2)](\sigma_2 - \sigma_1)}{le(2-\sigma_2)^2}\} & \text{otherwise.} \end{cases} \quad (32)$$

Straightforward calculations give  $\frac{\partial E[\pi_{1g}(\bar{p}_b, 1)/\rho_2^*]}{\partial \alpha} = e(l-1) - 2el\alpha + p_b^c\sigma_2$ , which is positive for all  $\alpha < \bar{\alpha}$ . Hence, more accurate monitoring strengthens the green firm's incentive to signal its type with an upward distortion in price.

We now examine firm 1's expected profits under the worst belief that consumers might form from a price,  $\mu = 0$ , that yields consumer perception  $\tilde{e}_g(0) = \alpha e$ . Consider that firm 1 sets the price  $p$  for the green product. Given the separating pricing rule  $\rho_2^*$ , firm 1 predicts that firm 2 charges the price  $p_b^*$  for the brown product with probability

$1 - \sigma_2$ . In this event, differentiation between the brown and the green products gives rise to a demand for the green product equal to  $D_{1g}(p, p_b^*, 0, 0) = \frac{\alpha e + p_b^* - p}{\alpha e}$ , provided that  $\alpha > 0$ . Clearly, this demand is nil in the absence of monitoring since there is no differentiation between the products ( $\alpha = 0$ ).

Alternatively, firm 1 predicts that firm 2 chooses the price  $p_g^*$  for the brown product with probability  $\sigma_2$ . Consumers draw inferences from observing two prices in the green market, i. e.,  $p$  and  $p_g^*$ . It follows that consumer valuations are, respectively,  $\alpha e$  for firm 1's product, and  $\tilde{e}_b(1) = (1 - \alpha)e$  for firm 2's product. Indeed, consumers infer from  $p_g^*$  that firm 2's product is green with probability 1 and the auditor's monitoring supports this belief with probability  $1 - \alpha$ . Then, in consumers' eyes, whether firm 1's product is more or less valuable than firm 2's product depends on whether  $\alpha$  is higher or lower than  $\frac{1}{2}$ , respectively. The market split (9) calculated for  $\mu = 0$  and  $\sigma = 1$  results in demand  $D_{1g}(p, p_g^*, 0, 1)$  for firm 1's product. The functional form of this demand given by (52) in Appendix 1 shows that better accuracy in monitoring within the range  $[0, \frac{1}{2}]$  decreases product differentiation, while it increases product differentiation as soon as  $\alpha$  exceeds the threshold  $\frac{1}{2}$ . As previously mentioned, monitoring somewhat corrects consumer misperceptions of firm 1's product off the equilibrium path. Again, the picture is different with no monitoring, because the two products are no longer differentiated.

In summary, the expected demand of firm 1 when falsely perceived to be brown by consumers is  $E[D_{1g}(p, 0) / \rho_2^*]$  given by (54) in Appendix 1. From this, firm 1's expected profits are

$$E[\pi_{1g}(p, 0) / \rho_2^*] = (p - e)E[D_{1g}(p, 0) / \rho_2^*]. \quad (33)$$

Let  $p_{1g}(0, \sigma_2)$  denote firm 1's best response to the pricing rule  $\rho_2^*$ . If firm 1 deviates from the price equilibrium to set  $p_{1g}(0, \sigma_2)$ , monitoring will correct con-

sumer misperception based on this price. The reduced-form function  $\tilde{\Pi}_{1g}(\sigma_2) = E[\pi_{1g}(p_{1g}(0, \sigma_2), 0) / \rho_2^*]$  represents the spectrum of best worst-outcomes for the green type: this is the best that the green firm can do when it believes that its rival will use the fly-by-night strategy to trick consumers into buying at a positive price by setting  $p_g^*$  with probability  $\sigma_2$ . Everything else being equal, the profit  $\tilde{\Pi}_{1g}(\sigma_2)$  grows with firm 2's probability of cheating, thereby weakening firm 1's incentive to signal its type. Obviously, this incentive also depends on the auditor's monitoring accuracy.

Figure 1 maps out  $\tilde{\Pi}_{1g}(\sigma_2)$  as a function (in grey) of  $\alpha$ , given  $l = 3$  and  $p_b^c = 0$ , in two limit cases:  $\sigma_2 = 0$  and  $\sigma_2 = 1$ ; that is, firm 1 predicts with confidence ( $\sigma_2 = 0$ ) that its brown rival will truthfully signal its type with  $p_b^*$  and firm 1 definitely expects its rival to cheat consumers ( $\sigma_2 = 1$ ). The light grey area between both curves depicts the whole spectrum of best worst-outcomes for the green type. Figure 1 also depicts  $E[\pi_{1g}(p_g^*, 1) / \rho_2^*]$  as a function (in black) of  $\alpha$ , given the same parameter configuration.

The following table presents the calculation results for firm 1's outcomes.

| Firm 1's outcomes for $l = 3$ and $p_b^c = 0$            | Expressions  |
|--|--|
| Firm 1's best worst-profit when $\sigma_2 = 0$           | $\tilde{\Pi}_{1g}(0) = \frac{e(1-3\alpha)^2}{12\alpha}$ if $\alpha \geq \frac{1}{3}$ , and 0 otherwise   |
| Firm 1's best response to $\rho_2^*$ when $\sigma_2 = 0$ | $p_{1g}(0, 0) = \frac{e(1+3\alpha)}{2}$ if $\alpha \geq \frac{1}{3}$ , and $e$ otherwise   |
| Firm 1's sales volume when $\sigma_2 = 0$                | $E[D_{1g}(p_{1g}(0, 0), 0) / \rho_2^*] = \frac{(3\alpha-1)}{6\alpha}$ if $\alpha \geq \frac{1}{3}$ , and 0 otherwise   |
| Firm 1's best worst-profit when $\sigma_2 = 1$           | $\tilde{\Pi}_{1g}(1) = \begin{cases} \frac{e(1-3\alpha)^2}{3(2\alpha-1)} & \text{if } \alpha \geq \frac{1}{2}, \\ \frac{e}{12(1-2\alpha)} & \text{if } \alpha \in (\bar{\alpha}, \frac{1}{2}), \\ \frac{e(2-3\alpha)^2}{12(1-2\alpha)} & \text{if } \alpha \leq \bar{\alpha} = \frac{1}{3}. \end{cases}$ |
| Firm 1's best response to $\rho_2^*$ when $\sigma_2 = 1$ | $p_{1g}(0, 1) = \begin{cases} 3\alpha e & \text{if } \alpha \geq \frac{1}{2}, \\ \frac{3e}{2} & \text{if } \alpha \in (\bar{\alpha}, \frac{1}{2}), \\ \frac{e(4-3\alpha)}{2} & \text{if } \alpha \leq \bar{\alpha}. \end{cases}$   |
| Firm 1's sales volume when $\sigma_2 = 1$                | $E[D_{1g}(p_{1g}(0, 1), 0) / \rho_2^*] = \begin{cases} \frac{e(3\alpha-1)}{3(2\alpha-1)} & \text{if } \alpha \geq \frac{1}{2}, \\ \frac{1}{6(1-2\alpha)} & \text{if } \alpha \in (\bar{\alpha}, \frac{1}{2}), \\ \frac{(2-3\alpha)}{6(1-2\alpha)} & \text{if } \alpha \leq \bar{\alpha}. \end{cases}$    |
| Firm 1's least-costly signaling profit                   | $E[\pi_{1g}(p_g^*, 1) / \rho_2^*] = \begin{cases} \frac{e}{3} & \text{if } \alpha \geq \bar{\alpha}, \\ e\alpha(2-3\alpha) & \text{otherwise.} \end{cases}$  |
| Firm 1's least-costly signaling price                    | $p_g^* = \begin{cases} p_{1g}(1) = 2e & \text{if } \alpha \geq \bar{\alpha}, \\ \bar{p}_b = 3(1-\alpha)e & \text{otherwise.} \end{cases}$  |
| Firm 1's sales volume with $p_g^*$                       | $E[D_{1g}(p_g^*, 1) / \rho_2^*] = \begin{cases} \frac{1}{3} & \text{if } \alpha \geq \bar{\alpha}, \\ \alpha & \text{otherwise.} \end{cases}$  |

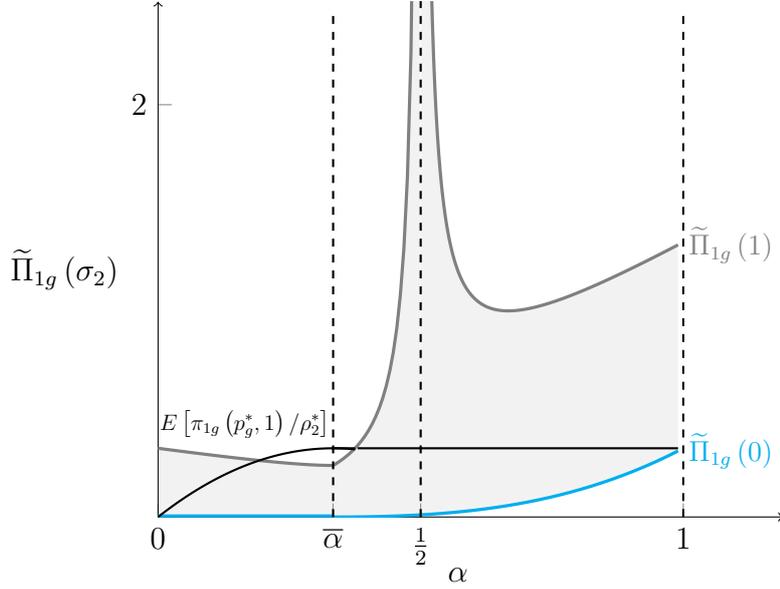


Figure 1: *Spectrum of best worst-profits*

When  $\sigma_2 = 0$ ,  $\tilde{\Pi}_{1g}(0)$  increases with  $\alpha$  when profit margin and demand are positive, which happens when  $\alpha$  exceeds  $\frac{1}{3}$ . This is because better monitoring accuracy increases consumer valuation for firm 1's product.

When  $\sigma_2 = 1$ ,  $\tilde{\Pi}_{1g}(1)$  is the dullest outcome for the green type: its product is falsely perceived as brown by consumers who, furthermore, falsely believe the brown firm to be green upon seeing the price  $p_g^*$ . We must distinguish two cases depending on whether  $\alpha$  is higher or lower than  $\frac{1}{2}$ .

If  $\alpha < \frac{1}{2}$ , monitoring slightly corrects consumer misperceptions, and firm 1 can earn positive profits due to product differentiation. However, after reading the auditor's report, consumers find the truly green product less valuable than the false green one. If, moreover, monitoring is so bad that  $\alpha < \bar{\alpha}$ , the brown firm must distort the price  $p_g^*$  upward to cheat consumers; this distortion relaxes the pressure put on the green firm to deter cheating. Figure 1 shows that  $\tilde{\Pi}_{1g}(1)$  decreases with  $\alpha$  over the interval  $[0, \bar{\alpha}]$  because the fly-by-night strategy of mimicking  $p_g^*$  becomes more aggressive as better

accuracy in monitoring reduces the price distortion. When  $\alpha$  exceeds the threshold  $\bar{\alpha}$ , there is no longer a need to distort prices for signaling. Then, improved accuracy in monitoring increases consumer valuation for the green product and boosts sales volume without affecting the price  $p_{1g}(0, 1) = \frac{3e}{2}$  charged for the green product. Therefore,  $\tilde{\Pi}_{1g}(1)$  increases with  $\alpha$  over the interval  $[\bar{\alpha}, \frac{1}{2}]$ .

If  $\alpha > \frac{1}{2}$ , consumers find the green product more valuable than its false substitute. Figure 1 shows that  $\tilde{\Pi}_{1g}(1)$  is a convex function of  $\alpha$  over the interval  $[\frac{1}{2}, 1]$ , with a minimum at  $\frac{2}{3}$ . Losses and gains of  $\tilde{\Pi}_{1g}(1)$  can be explained by the impact of monitoring accuracy on the green firm's response to the rival fly-by-night strategy. When  $\alpha$  is slightly above  $\frac{1}{2}$ , the two products are close substitutes in consumers' eyes, and the green firm takes the lion's share of the market by attracting consumers with a price  $p_{1g}(0, 1) = 3\alpha e$  far below  $p_g^* = p_{1g}(1)$ . In these circumstances, improved monitoring causes significant losses in sales volume, which are not offset by price increases—note that  $p_{1g}(0, 1)$  increases with  $\alpha$  and exceeds  $p_{1g}(1)$  when  $\alpha$  reaches  $\frac{2}{3}$  —. Finally, the green firm can respond less aggressively to the brown firm's cheating only when monitoring is fairly accurate.

In a nutshell, monitoring plays a dual role when the brown firm is cheating consumers: first, it corrects consumer misperceptions; second, it progressively relaxes pressure on the green firm to deter cheating.

Furthermore, firm 1 may be discouraged from signaling its true type with  $p_g^*$  if it earns more by deviating to the price  $p_{1g}(0, \sigma_2)$ . Therefore, the price  $p_g^*$  should satisfy the following constraint to be a separating equilibrium

$$E[\pi_{1g}(p_g^*, 1) / \rho_2^*] \geq \tilde{\Pi}_{1g}(\sigma_2). \quad (34)$$

This condition guarantees that it is worthwhile for firm 1 to use the signal  $p_g^*$  rather than to be perceived as brown by consumers. The right-hand side of (34) can be

interpreted as the green firm's opportunity cost of signaling its true type. As shown in Figure 1, the emergence of monitoring raises this cost, which makes it harder for firm 1 to reveal the truth about its type. Condition (34) is necessary and sufficient for a least costly separating equilibrium with  $p_g^* \geq \bar{p}_b$ , and supporting beliefs  $\mu^*(p) = 0$  when  $p < p_g^*$ , and  $\mu^*(p) = 1$  when  $p \geq p_g^*$ . Thus:

**Proposition 1:** *There exists a least costly separating equilibrium where  $p_b^* = p_b^c$  and  $p_g^* = \max\{p_{1g}(1), \bar{p}_b\}$  if and only if  $p_g^*$  satisfies (34).*

Figure 1 shows that there is an interval of  $\alpha$  within which  $\tilde{\Pi}_{1g}(1) < E[\pi_{1g}(p_g^*, 1) / \rho_2^*]$ ; thus requirement (34) is met whatever  $\sigma_2$ , in the specific case where  $l = 3$  and  $p_b^c = 0$ . For any  $\alpha$  outside this interval, there exists a whole range of  $\sigma_2$  strictly below 1, for which requirement (34) is met.

We first examine a baseline model in which there is no monitoring by a third-party auditor. Afterward, we extend the model to allow for the possibility of imperfect monitoring.

In the scenario with no monitoring, we find that there exists no separating equilibrium under the following circumstance: signaling costs are so high that the green firm has no incentive to deter the brown firm from cheating consumers. In that case, credible price signaling cannot support the switch to green production in the absence of monitoring. This serves as a benchmark for further comparison with the scenario of monitoring. The full analysis shows that monitoring, albeit imperfect, is likely to ensure the credibility of price signaling, and in turn, makes it worthwhile for a firm to switch to green production.

#### 4.1 Price signaling with no monitoring: the case $\alpha = 0$

As a benchmark, we examine the issue of price signaling in the green market with no monitoring:  $\alpha = 0$ . This presumes that the market is segmented between the brown

product and the product carrying the green seal. As a firm may falsely claim that the brown product is green, the credibility of green certification requires separating prices. These are meant to counteract the brown firm's incentive to cheat consumers. Moreover, we focus on the least costly separating equilibrium:  $p_b^* = p_b^c$  and  $p_g^* = \max\{p_{1g}(1), \bar{p}_b\}$ , from Proposition 1.

In the case where  $\alpha = 0$ , we also know from the credibility condition (23) that the minimum price candidate for deterring mimicry is

$$\bar{p}_b = le + 2p_b^c \frac{1 - \sigma_1}{2 - \sigma_1}. \quad (35)$$

We suppose again that firm 1 is the green type and firm 2 is the brown type. In the separating equilibrium, consumer perception is correct for both types:  $\tilde{e}_b(\mu(p_b^*)) = 0$  and  $\tilde{e}_g(\mu(p_g^*)) = e$  after observing  $p_b^*$  and  $p_g^*$ , set by firm 2 and firm 1, respectively.

Consider firm 1's behavior and suppose that firm 1 charges  $p_g^*$ . Given that firm 2 uses the separating pricing rule  $\rho_2^*$ , firm 1 predicts that firm 2 will charge the prices  $p_b^*$  and  $p_g^*$  with probabilities  $1 - \sigma_2$  and  $\sigma_2$ , respectively. Consumers will draw two different inferences depending on whether firm 2 charges  $p_b^*$  or  $p_g^*$ . Consumers either infer that the two products are imperfect substitutes from observing two distinct prices or consumers infer that the rival products are the same from observing the same price  $p_g^*$  set by both firms, because no product differentiation can be detected in the absence of monitoring.

From (27) and (50) given in Appendix 1, firm 1's expected profit is

$$E[\pi_{1g}(p_g^*, 1) / \rho_2^*] = (p_g^* - e) \frac{(el - p_g^*)(2 - \sigma_2) + 2p_b^c(1 - \sigma_2)}{2le}, \quad (36)$$

which is maximized at

$$p_{1g}(1) = \frac{e(1 + l)}{2} + p_b^c \frac{1 - \sigma_2}{2 - \sigma_2}. \quad (37)$$

Comparing (35) and (37) reveals that firm 1 may be forced to distort upward

$p_g^*$  in order to deter its brown rival from cheating consumers. Therefore, separation is potentially costly for the green type. The following lemma states the parameter conditions under which  $\bar{p}_b$  exceeds  $p_{1g}(1)$ .

**Lemma 3:** *Assume (5),  $\alpha = 0$  and  $p_b^c \frac{\sigma_2(3-\sigma_1)-2}{(2-\sigma_1)(2-\sigma_2)} < \frac{\epsilon}{2}(l-1)$ . In any separating equilibrium,  $p_b^* = p_b^c$  and  $p_g^*$  must be distorted upward relative to  $p_{1g}(1)$ .*

It may happen that the threat of a fly-by-night strategy entails positive signaling costs for the green firm. Price distortion allows the green firm to prove that it is less reluctant than its brown rival to restrict sales volume. The logic is the same as that recognized in most models of price signaling quality in markets of experience goods. In [Bagwell and Riordan \(1991\)](#), a monopolist signals high quality by raising prices up to the level where the loss of sales for the low-quality type is not worth the rent from cheating. Thus, the cost of signaling high quality is determined by the forgone rent from cheating. Compared to the monopoly regime, price competition in oligopolistic markets reduces the degree to which firms distort prices for the purpose of signaling (as shown by [Daughety and Reinganum, 2008b](#); [Janssen and Roy, 2010](#)).

In the present context, a novel insight is that price separation fails if  $p_g^*$  is distorted to the point that firm 1's expected sales volume falls to zero. From (36), we see that  $\bar{p}_g = le + 2p_b^c \frac{1-\sigma_2}{2-\sigma_2}$  is the maximum price for which demand for the green product is positive. Thus,  $\bar{p}_g \leq \bar{p}_b$  when  $p_b^c \frac{\sigma_2-\sigma_1}{(2-\sigma_1)(2-\sigma_2)} \geq 0$ . In that case, credible price signaling is too demanding to be attractive with no monitoring. Following [Mahenc \(2017\)](#), green certification cannot be credible if prices fail to signal the green type. In such circumstances, no firm will pay the setup cost  $F$  to switch to green production. This serves as a baseline model to further investigate the case of imperfect monitoring. It turns out that the signaling failure occurs with no monitoring in the limit case where  $p_b^c = 0$ , and so there is no added value in performing calculations for cases where  $p_b^c > 0$ . In the remainder of the paper, we normalize  $p_b^c$  to zero whenever appropriate

for calculations.

**Corollary 1:** *In the absence of auditor's monitoring, if  $p_b^c = 0$ , then there exists no separating equilibrium in which the green firm can credibly signal its type. In any subgame perfect equilibrium of the three-stage game, no firm will switch to green production with the aim of signaling its true type.*

As in [Akerlof \(1970\)](#), there is an adverse selection problem in the absence of monitoring; namely, there is no incentive for a given firm to provide anything but the brown product with minimum verifiable quality. As a result, the market for the green product is overrun by business as usual.

Corollary 1 also implies that there is a moral hazard problem: a firm does not commit to switching to green production because there is no credible way of signaling the switch. This is closely related to the moral hazard problem initially pointed out by [Klein and Leffler \(1981\)](#) a firm may refrain from producing high-quality products because it would lose all consumers at the minimum price needed to signal high quality. Klein and Leffler show that the threat of losing future business is likely to prevent a firm from renegeing on its promise to enhance product quality.

The analysis of the full scenario below shows that the auditor's monitoring can not only mitigate the adverse selection problem, but also solve the moral hazard problem.

## 4.2 Price signaling with monitoring: the case $\alpha > 0$

We now introduce the auditor to the previous benchmark. The auditor inspects the firm claiming that its product is green in order to detect the presence of the new attribute. The auditor's monitoring reveals the true type of the firm with probability  $\alpha$ . A lack of accuracy in monitoring is perfectly known to consumers. They read the auditor's report while observing the price signals sent by firms. Using both channels of communication, consumers infer information about the firms' types and, finally, make

their purchase decisions.

Unlike the scenario with no monitoring, the auditor's monitoring allows consumers to differentiate products when the brown firm tricks them into buying at price  $p_g^*$ .

From (31), the least costly separating price is distorted upward if the monitoring accuracy falls below the threshold  $\alpha = \bar{\alpha}$ . Assuming that firm 1 is the green type, the expected profit earned in the least costly separating equilibrium depends on the monitoring accuracy. From (29) and (32), we have

$$E[\pi_{1g}(p_g^*, 1) / \rho_2^*] = \begin{cases} \frac{[e^{(l-1)+p_b^c(1-\sigma_2)}]^2}{[p_b^c + e^{(l(1-\alpha)-1)}]^{4le}} & \text{if } \alpha \geq \bar{\alpha}, \\ \frac{[p_b^c + e^{(l(1-\alpha)-1)}]^{4le}}{le} & \text{if } \frac{p_b^c \sigma_2}{el} < \alpha < \bar{\alpha}. \end{cases} \quad (38)$$

Proposition 1 states that firm 1 has no incentive to defect from  $p_g^*$  if and only if the opportunity cost of signaling its true type with  $p_g^*$ ,  $\tilde{\Pi}_{1g}(\sigma_2)$ , does not exceed  $E[\pi_{1g}(p_g^*, 1) / \rho_2^*]$ .

Let us now examine in detail condition (34).<sup>10</sup> The maximized profit  $\tilde{\Pi}_{1g}(\sigma_2)$  depends on the probability  $\sigma_2$  that firm 2 uses the fly-by-night strategy of charging  $p_g^*$ , unlike  $E[\pi_{1g}(p_g^*, 1) / \rho_2^*]$  when we normalize  $p_b^c$  to zero. Setting  $\tilde{\Pi}_{1g}(\sigma_2) = 0$  defines a threshold  $\underline{\sigma}$  for  $\sigma_2$ , above which the margin and demand of firm 1 are positive at the same time. It turns out that  $\underline{\sigma} < 1$  for all  $\alpha \in [0, 1]$ .

This can be seen in Figure 1, where  $\tilde{\Pi}_{1g}(1) > 0$  for all  $\alpha \in [0, 1]$ , given  $l = 3$  and  $p_b^c = 0$ . In this figure, the light grey area between the curves  $\tilde{\Pi}_{1g}(0)$  and  $\tilde{\Pi}_{1g}(1)$  depicts the spectrum of  $\tilde{\Pi}_{1g}(\sigma_2)$  such that  $\sigma_2 \geq \underline{\sigma}$ . If firm 2's probability of cheating falls below  $\underline{\sigma}$ , then firm 1's opportunity cost of signaling its true type is reduced to zero. As a preliminary conclusion, we can state that a least costly separating equilibrium exists for all  $\sigma_2 \leq \underline{\sigma}$ , provided that  $\underline{\sigma} > 0$ .

We now introduce the function  $\tilde{\Delta}_{1g}(\sigma_2) = E[\pi_{1g}(p_g^*, 1) / \rho_2^*] - \tilde{\Pi}_{1g}(\sigma_2)$ , and assume

---

<sup>10</sup>All the calculations and proofs for the full scenario with monitoring can be found in Appendix 2. To reduce the number of cases to review, we have assumed that  $l \geq 2$ . As previously mentioned,  $p_b^c$  is normalized to zero in order to allow a straightforward comparison with the baseline model without monitoring.

that  $\sigma_2 > \underline{\sigma}$ . Under these circumstances, we have  $\tilde{\Pi}_{1g}(\sigma_2) > 0$ . Furthermore, setting  $\tilde{\Delta}_{1g}(\sigma_2) = 0$  defines a critical  $\bar{\sigma}$  for the probability of cheating, below which a least costly separating equilibrium exists, provided that  $\bar{\sigma} > \max\{0, \underline{\sigma}\}$ . If so and inequality  $\underline{\sigma} > 0$  is satisfied as well, then firm 1's opportunity cost of signaling its true type becomes positive at  $\sigma_2 > \underline{\sigma}$  and increases with  $\sigma_2$  until this probability reaches a “no-defect” threshold  $\min\{\bar{\sigma}, 1\}$ . In Figure 1,  $\bar{\sigma}$  is determined according to  $\alpha$  at every point where the black curve  $E[\pi_{1g}(p_g^*, 1)/\rho_2^*]$  crosses the light grey area.

We have previously seen that, in the best worst-situation the green type may face, a green product mistaken for brown is more or less valuable than the brown product mistaken for green, depending on whether  $\alpha$  is higher or lower than  $\frac{1}{2}$ , respectively. For the sake of clarity, we henceforth refer to  $\alpha > \frac{1}{2}$  as a “good” accuracy for monitoring.

Furthermore, price signaling happens to be costless or costly, depending on whether  $\alpha$  exceeds or falls short of  $\bar{\alpha}$ , respectively. Therefore, we must also distinguish between both of these cases. For this, we refer to  $\alpha \in [\bar{\alpha}, \frac{1}{2}]$  and  $\alpha < \bar{\alpha}$ , respectively, as “intermediate” and “bad” accuracy for monitoring.

Calculations yield two limit functions for  $\alpha$ , defined as follows:  $\alpha_0(l) = \frac{5l - \sqrt{l(32-7l)}}{16l}$  and  $\alpha_1(l) = \min\{\bar{\alpha}, \frac{5l + \sqrt{l(32-7l)}}{16l}\}$  are, respectively, the lowest and highest values of  $\alpha$  for which  $\bar{\sigma} \geq 1$ . One can check that these functions are ranked in the following order:  $\alpha_0(l) < \alpha_1(l) < \frac{1}{2}$  for all  $l \geq 2$ . The upper boundary for the probability of cheating that is consistent with the existence of a price separating equilibrium is  $\bar{\sigma}$  or 1, depending on the parameter values of  $l$  and  $\alpha$ . The following lemma lays the groundwork for the parameter values under which there exists a separating equilibrium.

**Lemma 4:** *Assume that  $l \geq 2$  and  $p_b^c = 0$ . For all  $\alpha < \frac{1}{2}$ ,*

$$\max\{0, \underline{\sigma}\} < \bar{\sigma}.$$

*The minimum threshold above which  $\tilde{\Pi}_{1g}(\sigma_2) > 0$  satisfies*

(i)  $0 < \underline{\sigma}$  for all  $\alpha < \frac{1}{l}$ ,

(ii)  $\underline{\sigma} < 1$  whatever  $\alpha$ .

The maximum threshold below which  $\tilde{\Delta}_{1g}(\sigma_2) > 0$  satisfies

(i)  $0 < \bar{\sigma}$  for all  $\alpha \geq \frac{1}{2}$ ,

(ii)  $\bar{\sigma} \geq 1$  for all  $\alpha \in [\alpha_0(l), \alpha_1(l)]$ .

**Proof:** see Appendix 2.

We are now ready to specify the conditions for parameters  $l, \alpha$  and  $\sigma_2$ , under which separation of the brown and the green type can be achieved *via* prices.

**Proposition 2:** *Assume that  $l \geq 2$  and  $p_b^c = 0$ . With imperfect monitoring, a least costly separating equilibrium exists for all  $\alpha \in (0, 1]$  if and only if  $\sigma_2 \leq \sigma^{defect}$ , where*

$$\sigma^{defect} = \begin{cases} \bar{\sigma} & \text{when monitoring accuracy is good,} \\ \min\{\bar{\sigma}, 1\} & \text{when monitoring accuracy is intermediate or bad.} \end{cases} \quad (39)$$

This result contrasts with the non-existence of separating equilibria established in Corollary 1. This shows that imperfect monitoring operates as a mechanism for assuring the credibility of price signaling. In the least costly separating equilibrium, the price of the green product may include an upward distortion (when monitoring accuracy is bad) or not (when monitoring accuracy is good or intermediate), due to the forgone profit from the brown firm's fly-by-night attempt to mislead consumers. Thus, monitoring with good or intermediate accuracy saves the signaling costs for the green firm. Furthermore, when the critical level  $\sigma^{defect}$  falls short of 1, which may occur for every level of monitoring accuracy, the green firm finds it worthwhile to disclose full information about its type, unless the probability of cheating exceeds  $\sigma^{defect}$ .

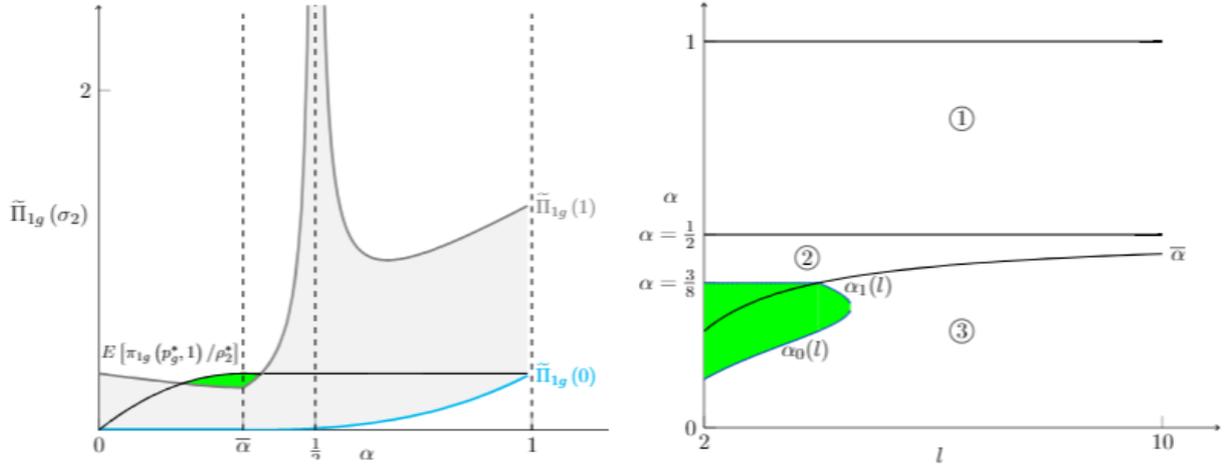


Figure 2: *Existence of Separating Equilibria*

The right-hand picture in Figure 2 divides the  $(l, \alpha)$  space into three regions, in which the least costly separating equilibrium exists.

In Region 1, monitoring accuracy is good. There is no extra cost for signaling the green type, so  $p_g^* = p_{1g}(1)$ . The brown firm cannot trick consumers into buying at this price provided that  $\sigma_2$  remains sufficiently below 1; namely,  $\sigma_2 \leq \bar{\sigma}$ . Therefore, price signaling succeeds unless the brown firm is too likely to cheat. When  $\sigma_2$  exceeds  $\bar{\sigma}$ , the weight on the fly-by-night strategy in the best worst-profit  $\tilde{\Pi}_{1g}(\sigma_2)$  is too heavy for the green firm to use  $p_g^*$  as a signal of its type. Therefore, the green firm prefers to be mistaken for brown for all  $\sigma_2 > \bar{\sigma}$ .

In Region 2, monitoring accuracy is moderate, though still sufficient to save signaling costs for the green firm. Thus, the price signaling the green type is still  $p_g^* = p_{1g}(1)$ . As in Region 1, separation can occur if only if the brown firm is not too likely to cheat:  $\sigma_2 \leq \bar{\sigma}$  when  $\alpha > \max\{\bar{\alpha}, \frac{3}{8}\}$ . In this parameter configuration, the best worst-profit  $\tilde{\Pi}_{1g}(1)$  may be very high when  $\alpha$  is close to  $\frac{1}{2}$ . As previously seen, the reason for this is that monitoring accuracy boosts sales volume for the green firm without affecting the price  $p_g^*$ , if the brown firm ever tries to cheat consumers. The probability of cheating must be sufficiently below 1 to reduce  $\tilde{\Pi}_{1g}(\sigma_2)$  to the point where the green firm

foregoes this profit and instead signals its true type. When  $\alpha$  falls below  $\frac{3}{8}$  (dark grey area), we know that  $\tilde{\Pi}_{1g}(1)$  decreases due to a loss in sales volume for the green firm when faced with the rival's fly-by-night strategy. This provides the green firm with a stronger incentive to signal its type, even if it is sure that the brown firm is cheating consumers. As a result, the least costly separating equilibrium exists regardless of  $\sigma_2$ .

In Region 3, monitoring accuracy is bad and cheating is more attractive to the brown firm than in Region 2. Therefore, an upward-distorted price is needed to signal the green type. As a result, price signaling for the green type is  $p_g^* = \bar{p}_b$ . This is a credible strategy for the green firm because it suffers less than its brown rival from the consequent loss of sales volume, due to the gap in production costs. Moreover, the green firm is better off signaling its true type than being mistaken for brown, even if it is sure that the rival is mimicking  $\bar{p}_b$ , as long as  $\alpha$  lies inside  $[\alpha_0(l), \alpha_1(l)]$  (dark grey area). When  $\alpha$  is outside this area, the fly-by-night strategy becomes less aggressive. Consequently, the best worst-profit  $\tilde{\Pi}_{1g}(1)$  increases to the point that the green firm is worse off with the price  $\bar{p}_b$  than with the lower price  $p_{1g}(0, 1)$  at which it is mistaken for the brown type.  $\sigma_2$  must fall short of  $\bar{\sigma}$  to reduce  $\tilde{\Pi}_{1g}(\sigma_2)$  enough for the green firm to separate with  $\bar{p}_b$ .

Solving the game by backward induction, we now examine how firms choose a type at the first stage of the game. Setting  $p_b^c = 0$  in (13) gives firm 1's expected payoff from playing a mixed strategy as

$$\Pi_1(\sigma_1, \sigma_2) = \begin{cases} \sigma_1 \left[ (1 - \sigma_2) \frac{(\epsilon(l-1))^2}{4\epsilon l} - F \right] & \text{if } p_g^* = p_{1g}(1), \\ \sigma_1 [(1 - \sigma_2)(l(1 - \alpha) - 1)e\alpha - F] & \text{if } p_g^* = \bar{p}_b. \end{cases} \quad (40)$$

Consider now the mixed strategy subgame PBEs of the model satisfying the constraint (16). Define  $\sigma^{IR}$  as the critical  $\sigma_2$  above which firm 1 does not recover the cost

of switching to green production

$$\sigma^{IR} = \begin{cases} 1 - \frac{F4el}{(e(l-1))^2} & \text{if } p_g^* = p_{1g}(1), \\ 1 - \frac{F}{(l(1-\alpha)-1)e\alpha} & \text{if } p_g^* = \bar{p}_b. \end{cases} \quad (41)$$

Inequality  $0 < \sigma^{IR}$  holds for all  $F < \frac{(e(l-1))^2}{4el}$  and all  $\alpha \in (\bar{\alpha} - \frac{\sqrt{(e(l-1))^2 - 4elF}}{2el}, 1]$ . Let us assume this to be the case.

Straightforward calculations yield  $\frac{\partial \sigma^{IR}}{\partial \alpha} = F \frac{l(1-2\alpha)-1}{e\alpha^2(l(1-\alpha)-1)^2} > 0$  for all  $\alpha < \bar{\alpha}$ , such that  $p_g^* = \bar{p}_b$ . Hence,  $\sigma^{IR}$  increases as  $\alpha$  increases. In the case of bad monitoring, cheating problems are less severe with more accurate monitoring.

Using the notation  $\hat{\sigma} = \min\{\sigma^{defect}, \sigma^{IR}\}$ , we can now construct firm 1's best response function. If  $\sigma_2 < \hat{\sigma}$ , then firm 1's unique best response is  $t = g$ , while if  $\sigma_2 \geq \hat{\sigma}$ , then firm 1's unique best response is  $t = b$  because it would incur the loss of  $F$  by changing its decision to switch to green production. In summary, firm 1's best response function is

$$\sigma_1^*(\sigma_2) = \begin{cases} 1 & \text{if } \sigma_2 < \hat{\sigma}, \\ \text{any } \sigma_1 \in [0, 1] & \text{if } \sigma_2 = \hat{\sigma}, \\ 0 & \text{if } \sigma_2 > \hat{\sigma}. \end{cases} \quad (42a)$$

By symmetry, we can find firm 2's best response function. The best response functions of both firms are depicted in Figure 3. The set of mixed-strategy Nash equilibria corresponds to the set of intersections of the best response functions in this figure.

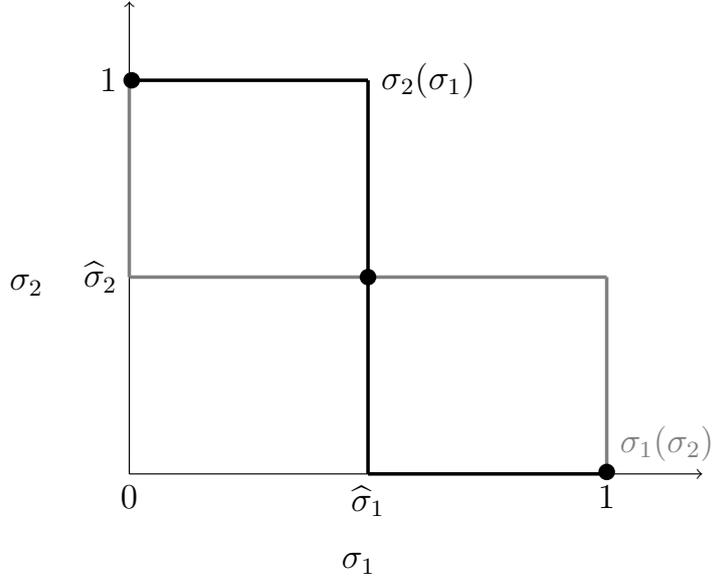


Figure 3: *Best response functions and Nash Equilibria*

**Proposition 3:** *Assume that  $l \geq 2$ ,  $p_b^c = 0$ ,  $F < \frac{(e(l-1))^2}{4el}$  and  $\alpha \in (\bar{\alpha} - \frac{\sqrt{(e(l-1))^2 - 4elF}}{2el}, 1]$ . With imperfect monitoring, there exists three PBEs in which  $(\sigma_1^*, \sigma_2^*) = (0, 1), (1, 0)$  and  $(\hat{\sigma}, \hat{\sigma})$ .*

In the first two equilibria, firms choose asymmetric pure strategies: one firm switches to green production with probability 1, while the other sticks to brown production with probability 1. In these equilibria, all the protagonists of the game, including consumers, are certain that the firms put two differentiated products on the market. The auditor's monitoring mitigates the threat of fly-by-night strategies on the brown firm's side. If consumers observe such a deviation from price equilibrium in the signaling subgame, monitoring partly corrects the misperception, thereby maintaining some degree of differentiation between the products. The cheating profit resulting from Bertrand competition off the equilibrium path decreases sufficiently to make the fly-by-night strategy unattractive. Thanks to the auditor's report, the green firm can rely on price to curtail cheating and hence credibly signal its type. Simultaneously, the brown firm becomes fully convinced that cheating is worthless. If either firm is sure that its rival

switches to green production, then it is better off sticking to brown production. A dilemma remains in the pure strategy equilibria: there is no explanation for how firms know which equilibrium will play out.

There is no such a dilemma in the symmetric mixed-strategy equilibrium. Both firms choose to switch to green production with the same probability  $\hat{\sigma}$ . This probability turns into the probability of cheating in the signaling subgame. Intuitively, cheating is as likely to occur as the switch to green production because cheating boils down to mimicking the green behavior as often as possible. The probability  $\hat{\sigma}$  is sufficiently low in equilibrium to meet the two requirements for curtailing cheating and revealing the truth. Another interpretation of the mixed-strategy equilibrium is that the proportion of firms that choose to switch to green production in the economy must remain reasonably low for price signaling to be credible. If too high a proportion of firms turn into green firms, the likelihood that brown firms will mimic is just as high and this threat may be too strong to afford the cost of revealing the truth.

We have seen that price signaling entails no further cost for the green firm when monitoring accuracy is good or intermediate. In those cases, the problem of cheating is less a matter of concern than that of revealing the truth for the green firm. This firm must forego the profit earned in the best worst-outcome in which consumers mistake it for the brown type. The opportunity cost of signaling the green type decreases as the brown rival becomes less likely to cheat. Therefore, the probability of cheating must offer the green firm a balanced compromise between revealing the truth and being mistaken for the brown type.

In the case where monitoring is bad, the problem of cheating is more severe because price signaling turns to be costly for the green firm. In equilibrium, the probability of cheating must be sufficiently low to make the brown firm indifferent between signaling its true type and using the fly-by-night strategy. Increased accuracy in monitoring mitigates the signaling distortion, thereby reducing the problem of cheating. When

cheating is no longer a problem, the probability of switching to green production coincides with the probability of cheating for which the firm exactly covers the cost of switching.

## 5 Conclusion

When social goodwill enhances product value in consumers' eyes, firms can increase profits by tying social goodwill in with their product. However, social goodwill is a credence attributes of the product that may be difficult to assess and certify. Asymmetric information about social goodwill may raise a twofold problem of adverse selection and moral hazard: firms renege on their pledge of social goodwill and the market is overrun by products carrying spurious labels.

In a simple model of signaling where price alone fails to send a credible signal of social goodwill, we show that independent monitoring, although imperfect, restores the credibility of price signaling, motivates firms to reveal the truth and, finally, helps a firm fulfill its pledge of social goodwill. The argument for this stems from the existence of a mixed-strategy equilibrium. The fly-by-night strategy of mimicking the price signal of social goodwill occurs as often as a firm pledges to tie social goodwill in with its product. Furthermore, a firm can afford the cost of using price as a signal of social goodwill provided that the fly-by-night strategy is not used too often to mislead consumers. Therefore, in a mixed-strategy equilibrium, there may exist a reasonably low probability that a firm will pledge social goodwill. This promise is credible because the firm, aided by monitoring, can find a price aimed at deterring the rival from cheating while making it worthwhile to reveal the truth about its goodwill.

There also exists pure strategy equilibria in which monitoring mitigates the threat of fly-by-night strategies by improving consumer perception based on prices. In these equilibrium outcomes, only one firm pledges social goodwill because, again, it can rely on the signal of social goodwill sent *via* price. Knowing this pledge, the rival firm has no

incentive to make the same pledge and any attempt to cheat consumers is unsuccessful.

Our findings highlight the role played by independent monitoring beyond the informational content of the auditor's report. Monitoring underpins the credibility of price signaling, and hence the honesty of certification. When signaling social goodwill through prices, a firm can rely on monitoring to correct for any arbitrary beliefs consumers might hold after observing a deviation from the equilibrium path. In other words, monitoring acts as a refinement of Bayesian equilibrium. It may happen that signaling social goodwill through price is costly due to the fly-by-night strategy of mimicking prices. In that case, monitoring relaxes the pressure that the firm must withstand to deter cheating. It turns out that the signaling cost decreases with enhanced accuracy in monitoring because the cheating profit is lower. Beyond a certain level of accuracy, monitoring saves the full cost of signaling social goodwill.

In the present world, monitoring is increasingly conducted by non-governmental organizations (NGOs). Among various activities, they operate as watchdogs for public and private certification of social goodwill by inspecting and testing products to verify an industry's compliance with certification standards. Their monitoring is constantly improving with advances in information and communications technology. In various recent cases, NGOs have demonstrated skill and ability in disclosing accurate information on fly-by-night practices in industry.<sup>11</sup> For instance, the International Council on Clean Transportation commissioned a study on emissions discrepancies between European and US models of diesel vehicles that has cast serious doubts on Volkswagen's compliance with environmental standards. Greenpeace and the Environmental Investigation Agency have produced evidence that FSC certification has been granted to logging companies operating with little regard for sustainability or even legality. While some firms see NGOs' monitoring as a threat, others may be willing to invite them to strengthen their commitment to social goodwill. This suggests that firms and

---

<sup>11</sup>Several examples can be found in Delmas and Burbano (2011) and Heyes and Martin (2017).

independent auditors somehow interact to release information to the public, in a way that can be cooperative or not.

In order to investigate this strategic interaction, it would be worthwhile to endogenize the auditor's behavior in the present model.

## 6 Appendix

### 6.1 Appendix 1: Demand functions

**Demand functions under full information** In the case where firm  $i$  of type  $g$  charges price  $p_{ig}$  and the rival firm  $j$  of the same type charges  $p_{jg}$ , products are undifferentiated. Then,  $D_{ig}(p_{ig}, p_{jg}) = D_{ig}^u(p_{ig}, p_{jg})$ , where

$$D_{ig}^u(p_{ig}, p_{jg}) = \begin{cases} \frac{le-p_{ig}}{le} & \text{if } p_{ig} < p_{jg} \\ \frac{1}{2} \frac{le-p_{ig}}{le} & \text{if } p_{ig} = p_{jg} \\ 0 & \text{if } p_{ig} > p_{jg} \end{cases} \quad (43)$$

In the case where firm  $i$  is green and its rival is brown, the rival products are differentiated. Then,  $D_{ig}(p_{ig}, p_{jg}) = D_{ig}^d(p_{ig}, p_{jg})$ , where

$$D_{ig}^d(p_{ig}, p_{jb}) = \begin{cases} 1 & \text{if } p_{ig} \leq p_{jb} \\ \frac{le+p_{jb}-p_{ig}}{le} & \text{if } p_{jb} < p_{ig} < le + p_{jb} \\ 0 & \text{if } p_{ig} \geq le + p_{jb} \end{cases} \quad (44)$$

**Demand functions under asymmetric information** In the case where consumers have higher expected valuation for the product  $i$  than for the product  $j$ , the demand for firm  $i$  resulting from the market split at (9) is given by

$$D_{it}(p_{it}, p_{jt'}, \mu, \sigma) = \begin{cases} 1 & \text{if } p_{it} < p_{jt'} \\ \frac{l\tilde{e}_t(\mu)-p_{it}}{l\tilde{e}_t(\mu)} & \text{if } p_{it} = p_{jt'} \\ \frac{l(\tilde{e}_t(\mu)-\tilde{e}_{t'}(\sigma))+p_{jt'}-p_{it}}{l(\tilde{e}_t(\mu)-\tilde{e}_{t'}(\sigma))} & \text{if } p_{jt'} < p_{it} < l(\tilde{e}_t(\mu) - \tilde{e}_{t'}(\sigma)) + p_{jt'} \\ 0 & \text{if } p_{it} \geq l(\tilde{e}_t(\mu) - \tilde{e}_{t'}(\sigma)) + p_{jt'} \end{cases} \quad (45)$$

In the case where firm 1 has chosen to be green while firm 2 has chosen to be brown, the pricing rule  $\rho_i^*$  predicts that firm  $i$  will charge price  $p_b^*$  with probability  $1 - \sigma_i, i = 1, 2$ .

**Firm 2's expected demand** Given  $\rho_1^*$  and consumer beliefs  $\mu(p) = 1$  after observing the price  $p_g^*$  charged by firm 2, its expected demand (21) takes two different

forms depending on whether monitoring is active or not; namely,

$$E [D_{2b}(p_g^*, 1) / \rho_1^*] = \begin{cases} (1 - \sigma_1) \frac{l(1-\alpha)e + p_b^* - p_g^*}{l(1-\alpha)e} & \text{if } \alpha > 0, \\ (1 - \sigma_1) \frac{le + p_b^* - p_g^*}{le} + \frac{\sigma_1}{2} \frac{le - p_g^*}{le} & \text{if } \alpha = 0. \end{cases} \quad (46)$$

**Firm 1's expected demand when consumer perception is correct** If firm 1 charges  $p_g^*$ , then consumer perception of its type is correct, i. e.,  $\mu(p_g^*) = 1$ . From the pricing rule  $\rho_2^*$ , two events may occur: firm 2 signals its true type with  $p_b^*$  or firm 2 plays the fly-by-night strategy of mimicking  $p_g^*$ . In the first event, firm 1's demand is derived from (9) by substituting  $\mu = 1$  and  $\sigma = 0$ , yielding

$$D_{1g}(p_g^*, p_b^*, 1, 0) = \frac{le + p_b^* - p_g^*}{le}. \quad (47)$$

In the second event, firm 1's demand with monitoring is derived from (9) by substituting  $\mu = 1$  and  $\sigma = 1$ , yielding

$$D_{1g}(p_g^*, p_g^*, 1, 1) = \frac{le - p_g^*}{le}. \quad (48)$$

Demands (47) and (48) are those faced by firm 1 in the presence of monitoring. With no monitoring, firm 1's demand is the same as (47) when firm 2 signals its true type. But when firm 2 plays the fly-by-night strategy, consumers perceive the two rival products as the same, and so firm 1's demand becomes

$$D_{1g}(p_g^*, p_g^*, 1, 1) = \frac{1}{2} \frac{le - p_g^*}{le}. \quad (49)$$

To sum up, when firm 1 charges  $p_g^*$ , firm 1's expected demand is

$$E [D_{1g}(p_g^*, 1) / \rho_2^*] = \begin{cases} (1 - \sigma_2) \frac{le + p_b^* - p_g^*}{le} + \sigma_2 \frac{le - p_g^*}{le} & \text{if } \alpha > 0, \\ (1 - \sigma_2) \frac{le + p_b^* - p_g^*}{le} + \frac{\sigma_2}{2} \frac{le - p_g^*}{le} & \text{if } \alpha = 0. \end{cases} \quad (50)$$

**Firm 1's expected demand when consumer perception is wrong** In the case where consumers hold the worst belief  $\mu = 0$  after observing a price  $p$  set by firm 1, consumer perception of its type is  $\tilde{e}_g(0) = \alpha e$  from (6). Again, it may happen that

firm 2 signals its true type with probability  $1 - \sigma_2$  or plays the fly-by-night strategy with probability  $\sigma_2$ . Then, consumer perception of its type is  $\tilde{e}_b(0) = 0$  in the first event, and  $\tilde{e}_b(1) = (1 - \alpha)e$  in the second event.

If firm 2 charges  $p_b^*$ , then the market split (9) calculated for  $\mu = 0$  and  $\sigma = 0$  gives

$$D_{1g}(p, p_b^*, 0, 0) = \frac{l\alpha e + p_b^* - p}{l\alpha e}, \quad (51)$$

when  $\alpha$  is positive, and 0 otherwise.

If firm 2 chooses price  $p_g^*$  while firm 1 sets  $p$ , then the market split (9) calculated for  $\mu = 0$  and  $\sigma = 1$  yields the following demand for firm 1's product

$$D_{1g}(p, p_g^*, 0, 1) = \begin{cases} \frac{l(2\alpha-1)e-p+p_g^*}{l(2\alpha-1)e} & \text{if } \alpha \geq \frac{1}{2}, \\ \frac{p_g^*-p}{l(1-2\alpha)e} & \text{if } \alpha \in (0, \frac{1}{2}). \end{cases} \quad (52)$$

If  $\alpha = 0$ , then the demand for firm 1's product, given that firm 2 chooses the price  $p_g^*$  with probability  $\sigma_2$ , is

$$D_{1g}(p, p_g^*, 0, 1) = \begin{cases} 1 & \text{if } p \leq p_g^* - e \\ \frac{p_g^*-p}{le} & \text{if } p_g^* - e < p < p_g^* \\ 0 & \text{if } p \geq p_g^* \end{cases} \quad \text{if } \alpha = 0. \quad (53)$$

When consumers falsely perceive firm 1 to be brown, its expected demand is

$$E[D_{1g}(p, 0) / \rho_2^*] = \begin{cases} (1 - \sigma_2) \frac{l\alpha e - p}{l\alpha e} + \sigma_2 \frac{l(2\alpha-1)e-p+p_g^*}{l(2\alpha-1)e} & \text{if } \alpha \geq \frac{1}{2}, \\ (1 - \sigma_2) \frac{l\alpha e - p}{l\alpha e} + \sigma_2 \frac{p_g^*-p}{l(1-2\alpha)e} & \text{if } \alpha \in (0, \frac{1}{2}), \\ (1 - \sigma_2) \times 0 + \sigma_2 \frac{p_g^*-p}{le} & \text{if } \alpha = 0 \text{ and } p_g^* - e < p < p_g^*. \end{cases} \quad (54)$$

## 6.2 Appendix 2: Proof of Lemma 4

In the best worst-situation, the green firm believes that its rival uses the fly-by-night strategy by setting  $p_g^*$  with probability  $\sigma_2$ . The best worst-profit (33) reaches a maximum at  $p_{1g}(0, \sigma_2)$ , where it takes the following values, depending on whether  $\alpha$  is

higher or lower than  $\frac{1}{2}$ ,

$$E[\pi_{1g}(p_{1g}(0, \sigma_2), 0) / \rho_2^*] = \begin{cases} \frac{(p_g^* \alpha \sigma_2 + e(\sigma_2 - 1 + \alpha(2 + l - \sigma_2(l+3)) + 2l\alpha^2(\sigma_2 - 1)))^2}{4le\alpha(2\alpha - 1)(1 - 2\alpha + \sigma_2(3\alpha - 1))} & \text{if } \alpha \leq \frac{1}{2}, \\ \frac{(p_g^* \alpha \sigma_2 - e(1 - \sigma_2 + \alpha(\sigma_2 - l - 2) + 2l\alpha^2))^2}{4el\alpha(1 - 2\alpha)(1 - \sigma_2 + \alpha\sigma_2 - 2\alpha)} & \text{otherwise.} \end{cases} \quad (55)$$

From Lemma 2, the least costly price that signals the green type is  $p_g^* = p_{1g}(1)$  if  $\alpha \geq \bar{\alpha}$ , and  $p_g^* = \bar{p}_b$  otherwise. Substituting  $p_{1g}(1)$  and  $\bar{p}_b$  for  $p_g^*$  in (55) leads to consider three cases.

1. Monitoring accuracy is good:  $\alpha > \frac{1}{2}$

For all  $\sigma_2 > \underline{\sigma}$ , the best worst-profit is

$$\tilde{\Pi}_{1g}(\sigma_2) = \frac{(e(2(1 - 2\alpha)(l\alpha - 1) - \sigma_2(3\alpha + l\alpha - 2)))^2}{16le\alpha(1 - 2\alpha)(1 - \sigma_2 + \alpha\sigma_2 - 2\alpha)}, \quad (56)$$

and the margin profit is

$$p_{1g}(0, \sigma_2) - e = \frac{e(2(1 - 2\alpha)(l\alpha - 1) - \sigma_2(3\alpha + l\alpha - 2))}{4(1 - \sigma_2 + \alpha\sigma_2 - 2\alpha)}. \quad (57)$$

In order that margin and demand be positive at the same time for the green firm falsely perceived to be brown, a necessary condition is  $\sigma_2 > \underline{\sigma} = \frac{2(1 - 2\alpha)(l\alpha - 1)}{3\alpha + l\alpha - 2}$ .

Note that  $1 - \sigma_2 + \alpha\sigma_2 - 2\alpha < 0$  and  $3\alpha + l\alpha - 2 > 0$  when  $\alpha \geq \frac{1}{2}$ , and hence  $\underline{\sigma} > 0$  for all  $\frac{1}{2} < \alpha < \frac{1}{7}$ .

Let us consider

$$\tilde{\Delta}_{1g}(\sigma_2) = 0 \quad (58)$$

This is a quadratic equation in  $\sigma_2$  with at most two real roots,  $\bar{\sigma}^-$  and  $\bar{\sigma}$ , such that  $\bar{\sigma}^- < \bar{\sigma}$  whenever they exist. Furthermore,  $\tilde{\Delta}_{1g}(\sigma_2)$  is a concave function of  $\sigma_2$  because its second derivative with respect to  $\sigma_2$  is negative. Thus,  $\tilde{\Delta}_{1g}(\sigma_2) > 0$  if the discriminant  $D(l)$  of (58) is positive and  $\sigma_2$  lies inside  $[\bar{\sigma}^-, \bar{\sigma}]$ . It turns out that  $D(l) > 0$  for all  $l > \frac{1 - \alpha - \alpha^2}{1 - 3\alpha + 3\alpha^2}$ , which is satisfied for any  $l > 1$  because  $1 - \frac{1 - \alpha - \alpha^2}{1 - 3\alpha + 3\alpha^2} = \frac{2\alpha(2\alpha - 1)}{1 - 3\alpha + 3\alpha^2} > 0$ .

The calculations done by Mathematica produce the following expressions

$$\bar{\sigma} \text{ (resp. } \bar{\sigma}^-) = 2 \frac{2 - (8 + l + l^2) \alpha + (9 + 3l + 4l^2) \alpha^2 - 2(1 + l + 2l^2) \alpha^3 + (\text{resp. } -) \sqrt{(l-1)^2 (l+1) (1-2\alpha)^2 \alpha^2 (l-1 + (1-3l) \alpha + (1+3l) \alpha^2)}}{(3+l)\alpha-2}.$$

From these expressions, further calculations show that  $\bar{\sigma}^- < \underline{\sigma} < \bar{\sigma} < 1$  for all  $\alpha > \frac{1}{2}$ . We can conclude that  $\bar{\sigma}$  is the critical  $\sigma_2$  below which separation is possible.

2. Monitoring accuracy is intermediate:  $\alpha \in [\bar{\alpha}, \frac{1}{2}]$

For all  $\sigma_2 > \underline{\sigma}$ , the best worst-profit is

$$\tilde{\Pi}_{1g}(\sigma_2) = \frac{(e(((l+5)\alpha - 2 - 4l\alpha^2)\sigma_2 + 2(2\alpha - 1)(l\alpha - 1)))^2}{16le\alpha(1-2\alpha)(\sigma_2(3\alpha - 1) + 1 - 2\alpha)}, \quad (59)$$

and the margin profit is

$$p_{1g}(0, \sigma_2) - e = \frac{e((2 - (5+l)\alpha + 4l\alpha^2)\sigma_2 - 2(2\alpha - 1)(l\alpha - 1))}{4(\sigma_2(3\alpha - 1) + 1 - 2\alpha)}. \quad (60)$$

In order that both the margin and demand be positive for the green firm falsely perceived to be brown, a necessary condition is  $\sigma_2 > \underline{\sigma} = \frac{2(2\alpha-1)(l\alpha-1)}{2-(5+l)\alpha+4l\alpha^2}$ . Note that  $2 - (5 + l) \alpha + 4l\alpha^2 > 0$  and so  $\underline{\sigma} > 0$  for all  $\alpha < \min\{\frac{1}{2}, \frac{1}{l}\}$  and  $\underline{\sigma} < 1$ .

Equation (58) is quadratic in  $\sigma_2$  with at most two real roots,  $\bar{\sigma}^-$  and  $\bar{\sigma}$ , such that  $\bar{\sigma}^- < \bar{\sigma}$  whenever they exist. Furthermore,  $\tilde{\Delta}_{1g}(\sigma_2)$  is a concave function of  $\sigma_2$  because its second derivative with respect to  $\sigma_2$  is negative. Thus,  $\tilde{\Delta}_{1g}(\sigma_2) > 0$  if the discriminant  $D(l)$  of (58) is positive and  $\sigma_2$  lies inside  $[\bar{\sigma}^-, \bar{\sigma}]$ .  $D(l)$  is a quadratic and convex function of  $l$ , such that  $D(l) > 0$  when  $l \in (1, \frac{1}{1-2\alpha})$ . The highest root in  $l$  of equation  $D(l) = 0$  is given by  $\frac{\alpha(1-2\alpha) + \sqrt{(8\alpha-3)(1-4\alpha+3\alpha^2)^2}}{-1+7\alpha-15\alpha^2+8\alpha^3}$ : it does exist for any  $\alpha > \frac{3}{8}$  and falls short of 1. Otherwise,  $D(l) > 0$  when  $\alpha \leq \frac{3}{8}$ . Finally,  $D(l) > 0$  when  $\frac{(l-1)}{2l} < \alpha \leq \frac{1}{2}$  and so both roots  $\bar{\sigma}^-$  and  $\bar{\sigma}$  exist.

The calculations done by Mathematica produce the following expressions

$\bar{\sigma}$  (resp.  $\bar{\sigma}^-$ )

$$= 2 \frac{2 - (10 + l + l^2)\alpha + (15 + 5l + 6l^2)\alpha^2 - 6(1 + l + 2l^2)\alpha^3 + 8l^2\alpha^4}{(2 - (5+l)\alpha + 4l\alpha^2)^2} + (\text{resp. } -) \sqrt{(l-1)^2(1-2\alpha)^2\alpha^2(3-11\alpha+9\alpha^2+2l\alpha(1-2\alpha)+l^2(1-7\alpha+15\alpha^2-8\alpha^3))}$$

From these expressions, further calculations show that, for all  $\alpha \in [\bar{\alpha}, \frac{1}{2}]$ ,  $\bar{\sigma}^- < \underline{\sigma} < \bar{\sigma}$ , and, moreover,  $0 < \bar{\sigma} < 1$  when  $\alpha > \max\{\bar{\alpha}, \frac{3}{8}\}$ . We can conclude that  $\min\{\bar{\sigma}, 1\}$  is the critical  $\sigma_2$  below which separation is possible.

### 3. Monitoring accuracy is bad: $\alpha < \bar{\alpha}$

For all  $\sigma_2 > \underline{\sigma}$ , the best worst-profit is

$$\tilde{\Pi}_{1g}(\sigma_2) = \frac{e((1-3\alpha+l\alpha^2)\sigma_2 - (2\alpha-1)(l\alpha-1))^2}{4l\alpha(1-2\alpha)(\sigma_2(3\alpha-1) + 1-2\alpha)}, \quad (61)$$

and the margin profit is

$$p_{1g}(0, \sigma_2) - e = \frac{e((1-3\alpha+l\alpha^2)\sigma_2 - (2\alpha-1)(l\alpha-1))}{2(\sigma_2(3\alpha-1) + 1-2\alpha)}. \quad (62)$$

In order that both the margin and demand be positive for the green firm falsely perceived to be brown, a necessary condition is  $\sigma_2 > \underline{\sigma} = \frac{(2\alpha-1)(l\alpha-1)}{1-3\alpha+l\alpha^2}$ . Note that  $1-3\alpha+l\alpha^2 > 0$  and hence  $\underline{\sigma} > 0$  for all  $\alpha < \min\{\frac{1}{2}, \frac{1}{l}\}$ .

Equation (58) is quadratic in  $\sigma_2$  with at most two real roots,  $\bar{\sigma}^-$  and  $\bar{\sigma}$ , such that  $\bar{\sigma}^- < \bar{\sigma}$  whenever they exist. Furthermore,  $\tilde{\Delta}_{1g}(\sigma_2)$  is a concave function of  $\sigma_2$  because its second derivative with respect to  $\sigma_2$  is negative. Thus,  $\tilde{\Delta}_{1g}(\sigma_2) > 0$  provided that the discriminant  $D(l)$  of (58) is positive. This turns out to be true when  $\alpha < \frac{(l-1)}{2l}$ , or, equivalently,  $l > \frac{1}{1-2\alpha}$ , because, first,  $D(l)$  is a quadratic and convex function of  $l$ , and second,  $\frac{1}{1-2\alpha}$  exceeds  $\frac{1}{1-\alpha}$ , which turns out to be the highest root in  $l$  of equation  $D(l) = 0$ . Thus, both roots  $\bar{\sigma}^-$  and  $\bar{\sigma}$  do exist.

The calculations done by Mathematica produce the following expressions

$\bar{\sigma}$  (resp.  $\bar{\sigma}^-$ ) =

$$\frac{1 - (5 + l)\alpha + (6 + 8l - 2l^2)\alpha^2 + l(11l - 18)\alpha^3 - 4l(5l - 3)\alpha^4 + 12l^2\alpha^5}{(1 - 3\alpha + l\alpha^2)^2} + (\text{resp. } -) 2\sqrt{\alpha^3(l - 2l\alpha)^2 \left( \begin{array}{c} -1 + 6\alpha - 12\alpha^2 + 9\alpha^3 + l(1 - 8\alpha + 24\alpha^2 - 34\alpha^3 + 18\alpha^4) \\ + l^2\alpha(1 - 7\alpha + 19\alpha^2 - 22\alpha^3 + 9\alpha^4) \end{array} \right)}$$

From these expressions, further calculations show that, for all  $\alpha < \bar{\alpha}$ ,  $\bar{\sigma}^- < \underline{\sigma} < \bar{\sigma}$ , and  $0 < \bar{\sigma} \leq 1$  when  $\alpha$  lies outside  $\left[ \frac{5l - \sqrt{l(32 - 7l)}}{16l}, \min \left\{ \bar{\alpha}, \frac{5l + \sqrt{l(32 - 7l)}}{16l} \right\} \right]$ .

Moreover,  $\bar{\sigma}$  reaches a minimum of 0 at  $\alpha = \frac{3l - \sqrt{l(9l - 16)}}{8l} < \frac{1}{4}$  for all  $l \geq 2$ . We

can conclude that  $\min\{\bar{\sigma}, 1\}$  is the critical  $\sigma_2$  below which separation is possible.

## Chapter 3

# Incidence of the Championing and Shaming Strategies of an NGO on the Signaling Strategy of a Firm

## English summary Chapter 3

It is worth mentioning that the hard evidence displayed by NGOs can be of two different natures. In this Chapter, the types of discovered evidence go in pairs with shaming and championing strategies. In this spirit, we endogenize the signaling choice of the NGO, which interacts with the signaling strategy of a firm. We seek to analyze (1) the effect of these two different strategies on the equilibrium of signaling of the firm and (2) the optimal informative behavior of the NGO.

The technology choice here is exogenous, and the consumer is perfectly Bayesian according to the two different signals received. Either a proof discredits the firm, or a proof accredits the firm. We find that the shaming strategy reduces the signaling cost for the high type and can even restore the perfect information outcome. The championing strategy can make incentives for the market to reveal the truth disappear. Concerning the optimal informative behavior of the NGO, if the market reveals information, the NGO is indifferent between adopting a shaming strategy and adopting a championing strategy. If we added a heterogeneous cost would definitely explain the informative strategy adopted by the NGO. When the market conceals information, the strategy depends on the difference between the shaming and championing efficiency and the distribution of the firm in the market.

## Résumé français

Il est important de mentionner que les preuves montrées par les ONGs peuvent être de deux natures différentes. Soit une preuve accréditant la firme, soit une preuve discréditant la firme. Dans ce chapitre, ces types de preuves vont de pair avec les stratégies de “championing” et de “shaming”. Dans cet esprit, nous rendons endogène le choix du signal de l’ONG qui interagit avec la stratégie de signal de la firme. Nous analysons (1) l’effet de ces deux différentes stratégie sur l’équilibre en signal des firmes

et (2) la stratégie d'information optimale de l'ONG.

La technologie ici est exogène, et le consommateur est parfaitement bayésien par rapport aux deux signaux qu'il reçoit. Nous trouvons que la stratégie de "shaming" réduit le coût du signal pour le bon type et peut même rétablir le prix en information complète. La stratégie de "championing" peut faire disparaître les incitations du marché à révéler la vérité. Concernant le comportement informatif de l'ONG, si le marché révèle l'information, celle-ci est indifférente entre adopter une stratégie de "shaming" ou de "championing". Si on prenait en considération le coût hétérogène entre les stratégies, cela expliquerait définitivement le choix de l'ONG. Quand le marché cache l'information, la stratégie dépend de la différence d'efficacité dans la détection du type de la firme, ainsi que de la distribution du type des firmes sur le marché.

# 1 Introduction

Green non-governmental organizations (G.NGOs) play an essential role in shaping our perception. Gathering civil society, they aim to prevent harmful projects and practices. NGOs' rising influence is one of the most significant changes in business over the past four decades (Daubanes, Rochet, et al., 2019). Heyes et al. (2018) present the management of salience of pollution as a key battleground between polluters and the G.NGO with which they interact, something both sides strategically seek to influence. Currently, the NGOs cooperate at an international level which increases their signaling power. Two directions can be taken by the NGO to influence the firm's image. One direction consists of adopting a good cop strategy, such as the World Wide Fund (WWF), by working hand-in-hand with industry to design environmentally friendly products with the use of ecolabels. Alternatively, the NGO can behave like a bad cop, such as Greenpeace, through strikes, investigations, or public revelations (Lyon, 2012).



(a) Bad Cop: Greenpeace action



(b) Good Cop: W.W.F label

Figure 1: Shaming and Championing

The exposure of events highlighting errors made in labeling or monitoring challenges the effectiveness of the informational role of the G.NGO. Figure 1 above shows a Greenpeace campaign against tuna fishing against some brands such as Saupiquet, which obtained the Marine Stewardship Council (MSC) label where WWF is a stakeholder. In the case of MSC controversy (Wijen and Chiroleu-Assouline, 2019), Greenpeace and

the Pew Environment Group were increasingly raising doubts about the sustainability promoted by the label. [Jacquet et al. \(2010\)](#) stated that the incentives of the market had led the MSC. certification scheme away from its original goal. Concerning imperfect monitoring, Greenpeace admitted having overestimated the impact of sinking a drilling platform when engaging a campaign against Shell.<sup>1</sup>

Nevertheless, it appears that firms are often much more concerned - and therefore reactive - than governments by image and reputation issues which constitute a high share of their wealth.<sup>2</sup> In reputation models ([Klein and Leffler, 1981](#); [Shapiro, 1983](#)), the price and quality provided by firms in the past are signals of current quality. The incentive to maintain quality comes through positive markups and repeat sales, which are lost once cheating is discovered. These studies are complementary to the signaling explanation we provide in this paper.

In this paper, our theory explains the presence and intervention of G.NGO activists concerning its informational role in green markets. Our framework analyzes the strategy of influencing the perception by providing *hard evidence* in a price signaling model. These pieces of evidence can be of two types: a proof that the firm is environmentally friendly (championing) or a proof that the firm is a polluter (shaming). Championing and shaming operate on radically different channels. For instance, the former supports the profit of the virtuous firm by awarding a label, and the latter sanctions the black sheep if something had been revealed during the monitoring process. Our model directly extends a standard price signaling model by adding a third party that endogenously provides hard evidence to characterize interplay between two information providers, a firm, and a G.NGO

Our main results are as follows. First, the presence of a G.NGO that applies a discrediting strategy directly reduces the cost of signaling for the green producer.

---

<sup>1</sup><https://www.latimes.com/archives/la-xpm-1995-09-06-fi-42780-story.html>

<sup>2</sup>Le Monde 06/05/2019: [https://www.lemonde.fr/economie/article/2019/05/06/les-grands-groupes-cibles-des-ong\\_5458611\\_3234.html](https://www.lemonde.fr/economie/article/2019/05/06/les-grands-groupes-cibles-des-ong_5458611_3234.html)

Moreover, when sufficiently efficient to provide hard evidence, imperfect monitoring is enough to restore the perfect information outcome. Very surprisingly, the imperfect label cannot bring back complete credibility and can even make separating equilibria disappear. When the market reveals full information, the G.NGO is indifferent between championing and shaming. As a result, the informative behavior would definitely be explained by the heterogeneous cost of discovering a specific hard evidence. When the market conceals information, the choice between championing and shaming depends on the proportion of green types versus bad types. For instance, one expects the G.NGO to use shaming in industries where the initiatives to use clean production processes are limited.

*Related literature.* The collective perception of trust toward NGO is fundamental to study information production. Trust is a predictor of whether stakeholders will find the organization credible and will take into account its information. Edelman (2018)<sup>3</sup> dedicated an entire survey to address the issue of trust with regard to organizations. Their report entitled “A Battle of Trust” shows that the trust toward an NGO has many country-specific effects, but is still the most reliable organization for the informed public and general population. In the context of asymmetry of information in green markets, watchdogs of the credibility are sometimes explicitly asked to defend a specific cause. Dewatripont and Tirole (1999) provide rationales for such organizations that consists of seeking and displaying hard evidence. Market incentives or social welfare maximization are not expected to be internalized in such a context. Lizzeri (1999) underlines that market structure strongly influences the intermediary to disclose information. In his paper, different disclosure rules lead to various profits for the intermediary. A monopoly intermediary that certifies a privately informed seller has an incentive to disclose the minimum information needed to induce the trade.

This article aims to explain the informative behavior of the NGO through the sig-

---

<sup>3</sup>Edelman (2018) Trust Barometer.

naling theory. In the previous literature, NGOs help firms to differentiate their product with respect to environmental characteristics ([Bonroy and Constantatos, 2008](#)). [Baron \(2011\)](#) underlines the necessity of the participation of such organizations, which exerts social pressure to market the credence good. [Fischer and Lyon \(2014\)](#) explore the effect of the competition between stakeholders of different natures, NGO and industry, on the level of stringency of a label. [Heyes and Martin \(2016\)](#) identify several other features that explain the level of the standard chosen by an NGO and the overproliferation of different labels provided on the market. [Bottega and De Freitas \(2009\)](#) extend the possibility for the NGO to use not only the certification but also the green advertisement. They analyze the impacts of an NGO on the level of average environmental quality on the market and the optimal environmental regulation by comparing the level of quality provided by the ecolabel chosen by the NGO and the minimum quality standard set by the regulator. [Mason \(2011\)](#) relaxes the hypothesis of a perfect ecolabeling scheme, and uses a signaling game to show multiple effects on welfare. The informational effects reduce the net surplus and he concludes that the introduction of an ecolabel can either increase or decrease welfare. [Mahenc \(2017\)](#) raises the question of the incentives for a label organization to report the truth on the environmental quality of the good when the stakeholder responsible for the certification is overweighting the profit associated with the green industry. From a consumer point of view [Civel et al. \(2018\)](#) provide a complete overview on the question of the assumption of the perfect information transmission of the label. They find in the case of the energy label that the energy performance certificates are not perceived as perfectly informative. Moreover, they state that subjects are more likely to infer the signal into their prior beliefs on energy quality, suggesting that reading can be considered as Bayesian.

Two articles are closely related to the present paper. The first is [Feddersen and Gilligan \(2001\)](#) who investigate the impact of an information-supplying activist on the market for credence goods. The activist, by monitoring, can send a signal about the

type (good or bad) of the product to the representative consumer. The existence of the activist on the market positively affects the quality choices of firms by giving them an incentive to adopt the desirable costly operating practice and the credibility of the good product perceived by the consumer. First, in this paper, the activist minimizes the aggregate social costs of trade: the pollution. The assumption of markets incentives seems credible as long as the activist is taken as an ecolabel provider; nevertheless, the knowledge of the social cost of a market is questionable. Second, in this paper, the firms are not strategic concerning the signal received by customers and omit important literature on this issue ([Milgrom and Roberts, 1986](#); [Bagwell and Riordan, 1991](#); [Mahenc, 2007](#); [Daughety and Reinganum, 2008a](#)). The second is [Fleckinger et al. \(2017\)](#) which focus their research on the effect of the informational environment on the level of the quality chosen by firms. This environment can be friendly when announcing good news or hostile when announcing bad news to consumers. This dissociation permits underlining which strategy is the most efficient to increase incentives for firms to choose the good quality given the market structure.

The rest of the paper is organized as follows. Section II presents illustrations of the theoretical issues tackled in this paper. Section III introduces a price signaling model with NGO influence. Section IV analyzes the informative behavior of a G.NGO. Section V concludes.

## 2 Illustrations

The purpose of this section is twofold. The first illustration aims to underline that firms and NGOs have intricate roles concerning information revelation on the market. It emphasizes that firms are well aware of the signaling capability of NGOs. The second example shows that the issue of asymmetric information is not solved by certification. Therefore, the signaling theory is relevant to forecast the informative behavior of firms and NGOs.

In 2014, the International Council of Clean Transportation conducted a study that revealed excessive emission volume in several models of Volkswagen cars sold in the United States. They alerted the Environmental Protection Agency (EPA), which, accused in September 2015 Volkswagen of installing illegal manipulation devices. In October, Volkswagen sales slipped globally. In November the company admitted that they cheated not only on NOx but also on CO2 emissions.<sup>4</sup> Afterward, this revelation also had consequences on European public policy toward diesel.<sup>5</sup> Moreover, the communication strategy of the firm following these events was deeply oriented toward the reconquest of consumers' confidence.<sup>6</sup> Figure 2 shows the revenue of the I.C.C.T, which monitored products that were supposed to pass environmental norms. The idea behind taking a look at the budget of I.C.C.T. is to materialize the possibility of participation of firms in the financial constraints of an NGO

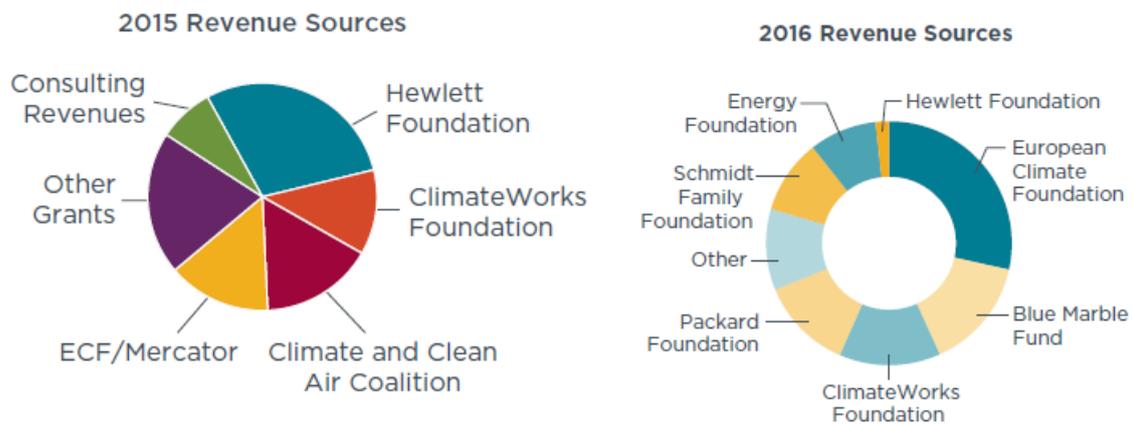


Figure 2: Revenue Sources of the International Council on Clean of Transportation (I.C.C.T.)

To quote a few, the Hewlett Foundation, which represents approximately 40% of the revenue of the NGO in 2015, was established by William R. Hewlett, creator of the

<sup>4</sup><https://www.cleanenergywire.org/factsheets/dieseltgate-timeline-germanys-car-emissions-fraud-scandal>

<sup>5</sup><https://www.theguardian.com/business/2015/nov/27/vw-labour-boss-says-fresh-emissions-crisis-has-hit-orders-for-new-cars>

<sup>6</sup>Advertising Campaign 2018: "We can doubt about everything, but not about our Volkswagen"

multinational Hewlett-Packard, and his wife Lamson Hewlett, (H.P).<sup>7</sup> E.C.F Mercator was created by the family Schmidt-Ruthenbeck, owner of METRO A.G., in 1996. Note that both of these firms continuously engage in green campaigns associated with recycling, water, emissions, etc. The point is to underline how firms took over the informative role of NGOs and not only by subscribing to a label.

The story of the Volkswagen emissions scandal illustrates different aspects of the interactions among consumers, green NGO and firms. First, it shows that an NGO may use hard evidence to reveal that a firm does not comply with environmental standards. Second, it highlights how firms have understood the stakes information production. Here, they participate in the financial constraint of the NGO to facilitate their influence on the transmission of the environmental quality.

To emphasize imperfection of the labeling activity, which is often considered perfect in economic studies, we have a similar story concerning the environmental label Forest Stewardship Council (FSC) launched in 1993. In 2014 Greenpeace released a report that accused FSC-certified logging companies of harvesting “areas that are either slated for legal protection or supposed to be protected as a part of FSC requirements.”<sup>8</sup> Simon Counsell, executive director of Foundation UK and an early proponent of the forest certification idea, argues that the opposite is true. His frustration with FSC led him to co-found the website FSC-Watch.com. He underlines that certifying agencies often display a lack of expertise on visits to logging operations, and this point is crucial to underscore the monitoring role of some green NGOs, and it might also raise skepticism among consumers against ecolabel. It might give pause to the entire wood products industry, which has profited up to now by turning a blind eye to that illegality.

There exist many other examples with a similar structure, and they are becoming ever more numerous. The issue of asymmetric information persists and the signaling

---

<sup>7</sup><https://hewlett.org/about-us/>

<sup>8</sup><https://e360.yale.edu/features/greenwashed-timber-how-sustainable-forest-certification-has-failed>

theory is relevant to understand how the credibility can be restored.

### 3 The model

BigA, a monopoly that grows apples, is either *green* with probability  $\phi$  or *brown* with probability  $1 - \phi$ . A brown BigA grows apples at a constant marginal cost  $\mu_b > 0$ ; a green BigA does so at a higher cost  $\mu_g > \mu_b$  but in an environmental-friendly way (the subscript  $b$  and  $g$  shall be used throughout to distinguish the brown monopoly from the green monopoly). In either case, BigA sets a price  $p$  for a profit  $(p - \mu)q$  where  $q$  is the quantity demanded at that price.

PublicEye is a G.NGO who pursues either a *championing* or a *shaming* strategy. Both strategies involve discovering hard information about BigA's type but they represent very different activities that perform differently contingently on BigA's type. The following probabilities associated with the choice of the strategy are given by Nature. The idea is to represent the technological capacity of PublicEye to discover a hard evidence.<sup>9</sup> With the championing strategy, PublicEye attempts to certify BigA's green pedigree. For instance, she could demonstrate that the technology used by BigA is actually better for the environment than a more commonly used one. With the shaming strategy, PublicEye assumes that BigA produces brown apples despite what the latter may claim. She could try to expose, for instance, an unreported use of pesticides. Although both strategies involve finding BigA's type, they are likely to perform differently, as it is one thing not to find a black swan and quite another to establish that there is no such thing as a black swan.

With the championing strategy, there is a probability  $\gamma$  that PublicEye will succeed in establishing BigA's type. With the shaming strategy, there is a probability  $\beta$  that

---

<sup>9</sup>Note that this capacity is not conditional on the distribution of the firm. For instance, if all firms were brown  $\phi = 0$ , PublicEye would still have a calculable likelihood to find a green hard evidence.

BigA's true brown type will be exposed.<sup>10</sup> Accordingly, PublicEye fails to provide new hard information to the consumer with probability  $1 - \gamma$  if BigA is green and with probability  $1 - \beta$  if it is brown. Allowing for a mixed strategy, we denote by  $s$  the probability that PublicEye chooses the championing strategy. If  $s = 1$ , championing is a pure strategy for PublicEye, while if  $s = 0$ , shaming would be the pure strategy.

Finally, there is a consumer who buys apples from BigA but whose marginal utility of apples depends on the latter type. Leaving the consumer aside, BigA and PublicEye play a Bayesian game (see figure 3 below) where Nature first chooses first BigA's type and whether any inquiry by PublicEye will be successful. Then PublicEye chooses between championing (the lighter edge) and shaming while not being aware of BigA's type, and BigA simultaneously chooses its price (not shown in the figure) without knowing if it is targeted by PublicEye.

Once PublicEye's inquiry is completed, the consumer knows BigA's price and ends up in one of three information states:

1. With probability  $i = s\phi\gamma$ , she thinks that BigA is green (leftmost node).
2. With probability  $j = (1 - s)(1 - \phi)\beta$ , she knows that BigA is brown (rightmost node).
3. With complementary probability  $1 - i - j$ , the consumer believes that BigA is green, if she observed no hard evidence, with probability

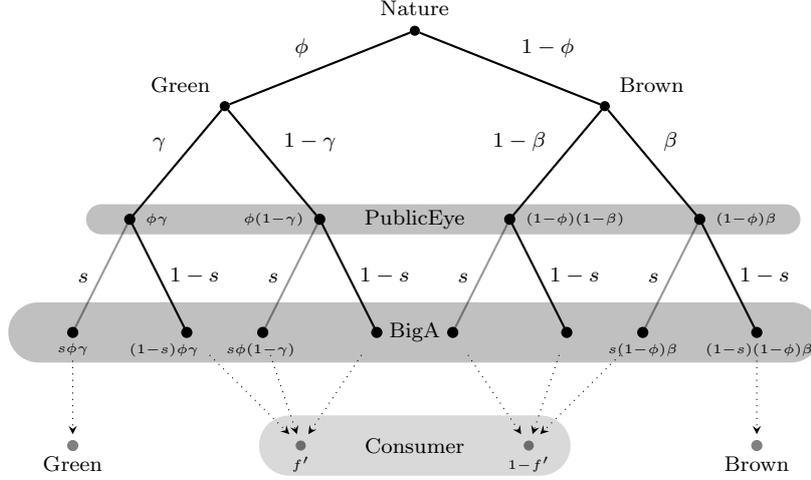
$$f' = \phi \frac{1 - s\gamma}{1 - i - j} \tag{63}$$

prior to factoring in the signal eventually provided by BigA's price. The consumer does not know *a priori* the strategy of the G.NGO.

---

<sup>10</sup>Declaration of Greenpeace UK in 18 May 2011: "We have the evidence to name and shame some of the companies blocking European climate ambitions[...]"

Figure 3: Bayesian Game in (Partial) Extensive Form



The consumer is risk neutral and buys a quantity  $q \in [0, \theta - 1]$  of apples, where  $\theta - 1 > 0$  is a satiety point. She values the consumption of a quantity  $q$  of brown apples (that is, produced by a brown BigA) at  $\ln(1 + q)$  and that of green apples at  $\nu \ln(1 + q)$  where  $\nu \geq 1$ . We assume that  $\mu < 1$  and that

$$\frac{1}{\theta^2} < \frac{\mu}{\nu} \quad (64)$$

for both  $\mu = \mu_g$  and  $\mu = \mu_b$ . These regularity conditions on marginal costs *vs* marginal benefit ensure interior solutions for both BigA types of programs below.

Let  $f$  denote the consumer's belief and  $\nu_f = f\nu + 1 - f$  so that her expected utility of buying and consuming  $q$  apples at price  $p$  is  $\nu_f \ln(1 + q) - pq$ . Her demand becomes

$$d(p, f) = \begin{cases} \theta - 1 & \text{if } p \leq \nu_f/\theta \\ \frac{\nu_f}{p} - 1 & \text{if } \nu_f/\theta < p \leq \nu_f \\ 0 & \text{if } \nu_f < p \end{cases}$$

BigA's profits at price  $p$  amount to  $(p - \mu)d(p, f)$  which is concave in  $p$ . As the price is raised from zero to  $\nu_f/\theta$ , profits first rise. The marginal profit over  $[\nu_f/\theta, \nu_f]$

<sup>10</sup>Note that under (64), the satiety point is socially efficient or not, depending on whether  $\mu < \frac{\nu}{\theta}$  or  $\mu > \frac{\nu}{\theta}$ .

amounts to

$$\frac{\nu_f \mu}{p^2} - 1$$

. It decreases as the price is further increased; however, because of condition (64), it is positive at  $p = \nu_f/\theta$ . By increasing it further, we reach a global maximum at  $\sqrt{\mu\nu_f}$  where it is zero. Profits<sup>11</sup> there amount to

$$\pi(\mu, f) = (\sqrt{\nu_f} - \sqrt{\mu})^2$$

They are strictly increasing in  $\nu_f$  and strictly decreasing in  $\mu$ .

PublicEye's objective is to maximize the utility of the consumer. All Greek letters denote parameters; all the other variables are endogenous. We look for a Bayesian-Nash equilibrium of this game.

## Perfect Information

If either  $\gamma = 1$  or  $\beta = 1$ , then PublicEye has the costless opportunity to turn this game of incomplete information into one of perfect information. It suffices that it plays a pure strategy (in equilibrium), as then, the absence of hard information nevertheless perfectly informs the consumer. For instance, suppose that PublicEye plays the championing strategy ( $s=1$ ). Then, whenever BigA is green, the consumer would learn it, and she would correctly infer<sup>12</sup> that BigA is brown should no hard information come about.

Anticipating a fully informed consumer, BigA sets its price accordingly: if it is

---

<sup>11</sup>The complete form of the profit function of the firm given the price interval is:

$$\pi(p, f) = \begin{cases} (p - \mu)(\theta - 1) & \text{if } p \leq \nu_f/\theta \\ (p - \mu) \left( \frac{\nu_f}{p} - 1 \right) & \text{if } \nu_f/\theta < p \leq \nu_f \\ 0 & \text{if } \nu_f < p \end{cases}$$

<sup>12</sup>With (63), the example yields to the following the Bayes rule:  $\phi \frac{1 - 1}{1 - \phi - (1 - 1)(1 - \phi)\beta} = 0$  meaning that the probability assigned by the consumer to the event that the firm is green given that she received no message is 0.

green, it sets  $p_g^* = \sqrt{\nu\mu_g}$  and realizes profits  $\pi(\mu_g, 1)$  and if it is brown, it sets  $p_b^* = \sqrt{\mu_b}$  and realizes  $\pi(\mu_b, 0)$ ; thus,  $\nu > 1$  and  $\mu_g > \mu_b$  imply that  $p_g^* > p_b^*$ . If we assume in addition that

$$\sqrt{\nu} - 1 > \sqrt{\mu_g} - \sqrt{\mu_b}$$

then a green BigA makes more profits than a brown one under perfect information.

## No Hard Information

If  $\gamma = \beta = 0$ , then PublicEye is powerless and we obtain a familiar signaling game where the consumer may only gather soft information inferred from BigA's chosen price. Endow the consumer with a (conditional) belief function that maps any observed price into a belief  $f$  that BigA is green. In a Perfect Bayesian Equilibrium, such a function should satisfy Bayes's rule given both of BigA's types of equilibrium strategies, and the latter should maximize profits given the beliefs and the other type's pricing strategy.

**Separating Equilibrium.** Consider first a separating equilibrium where a green BigA plays a separating price  $p_g$  and a brown one plays its first best price  $p_b^* \neq p_g$ . Set the beliefs to be one (assurance that BigA is green) if  $p = p_g$  and zero otherwise (BigA is brown). Then, both of BigA's types expect a demand  $d(p, 1)$  if  $p = p_g$  and  $d(p, 0)$  otherwise. Playing its first best price remains optimal for the brown type as long as it would benefit by mimicking the green type's pricing strategy:

$$\pi(\mu_b, 0) \geq (p_g - \mu_b)d(p_g, 1) \tag{IC_b}$$

and signaling one's type is a good strategy for a green BigA if

$$(p_g - \mu_g)d(p_g, 1) \geq \pi(\mu_g, 0) \tag{IC_g}$$

A signaling price must satisfy these two inequalities: we look for an interval of signaling

prices  $[p_0, p_1]$ . Define

$$h(p, \mu) = (p - \mu) d(p, 1) - \pi(\mu, 0)$$

so that (IC<sub>b</sub>) and (IC<sub>g</sub>) simply read  $h(p, \mu_b) \leq 0$  and  $h(p, \mu_b) \geq 0$ . We have already shown that  $(p - \mu) d(p, 1)$  is strictly concave in price and reaches a maximum in  $p = \sqrt{\mu\nu}$  so that  $h$  is strictly concave as well and reaches

$$(\sqrt{\nu} - \sqrt{\mu})^2 - (1 - \sqrt{\mu})^2 > 0$$

at its maximum —notably,  $h(p, \mu_g)$  reaches its maximum in  $p_g^*$ .

In addition,  $h(p, \mu)$  is negative if the price is null or is so large that the demand vanishes. Given  $\mu$ , it follows that it has two roots that define an interval  $[p^-(\mu), p^+(\mu)]$  over which  $h(p, \mu) \geq 0$  and  $h(p, \mu) < 0$  outside of it. If  $p^+$  is strictly increasing, then we may set  $p_0 = p^+(\mu_b)$  and  $p_1 = p^+(\mu_g)$  to obtain an interval of signaling prices. Taking the derivative with respect to  $\mu$  at  $p^+(\mu)$ , we find that

$$\frac{\partial p^+}{\partial \mu}(\mu) = -\frac{\partial h / \partial \mu}{\partial h / \partial p} = \frac{1}{\sqrt{\mu}} - \frac{\nu}{p^+(\mu)} > 0$$

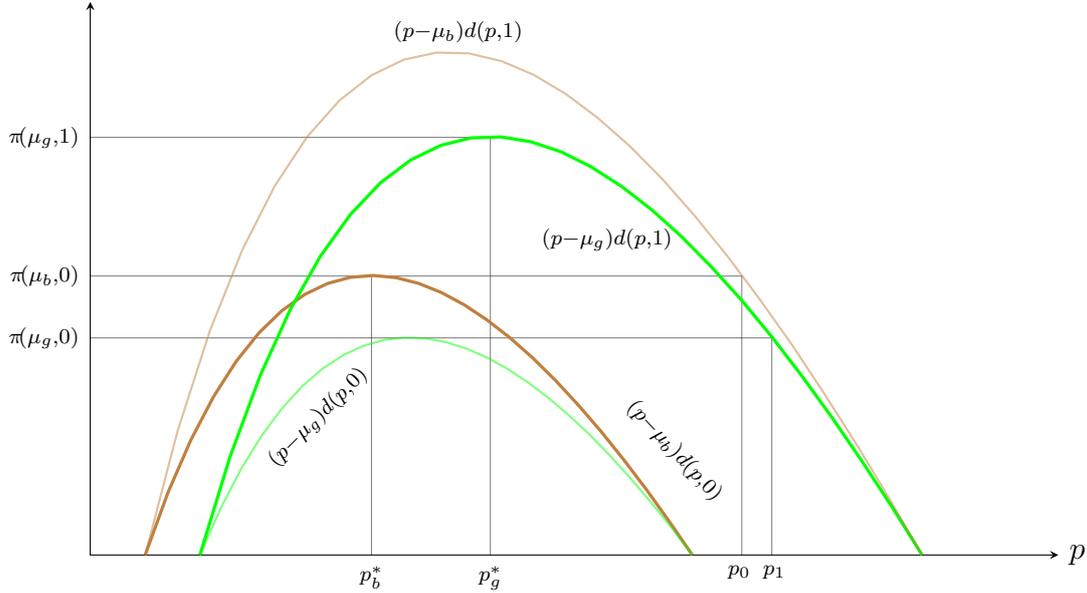
as required.<sup>13</sup>

---

<sup>13</sup>First establish that  $p^+(\mu) = \sqrt{\mu} + \frac{1}{2}(\nu - 1) \left( 1 + \sqrt{1 + 4 \frac{\sqrt{\mu} - \mu}{\nu - 1}} \right)$  where the last square root is greater than one because  $\mu < 1 < \nu$ . Then

$$\begin{aligned} 1 &> \mu \\ 1 + \sqrt{1 + 4 \frac{\sqrt{\mu} - \mu}{\nu - 1}} &> \sqrt{\mu} \times 2 \\ \frac{1}{2}(\nu - 1) \left( 1 + \sqrt{1 + 4 \frac{\sqrt{\mu} - \mu}{\nu - 1}} \right) &> \sqrt{\mu} \times (\nu - 1) \\ p^+(\mu) &> \sqrt{\mu\nu} \\ \frac{1}{\sqrt{\mu}} - \frac{\nu}{p^+(\mu)} &> 0 \end{aligned}$$

Figure 4: Separating equilibria



**Lemma 1.** *There exists a continuum of separating equilibria. The Riley outcome<sup>14</sup> is given by the pair of prices  $p = \bar{p}_0(\mu_b)$  and  $p_0 = p_b^*$ .*

Signaling the quality is costly for the producer of green apples. It has to set a price  $\bar{p}_0$  further than the perfect information price  $p_g^*$  to prevent any profitable deviation for the producer of brown apples.

**Pooling Equilibria.** If both types choose the same price the consumer cannot infer any information from prices. Therefore her *posterior* belief  $f$  will be equal to her *prior* belief  $\phi$  which implies a somewhat average valuation  $\nu_\phi$  for each apple bought. Nevertheless, assuming that any departure by BigA from such a strategy would be interpreted by the consumer as a sure sign that its apples are brown; then, BigA must balance the benefit of a higher valuation  $\nu_\phi > 1$  for its product against the opportunity cost of having to supply the market at that price. Therefore, for such an for such

<sup>14</sup>The least costly separating equilibrium which survives the Intuitive Criterion (Cho and Kreps, 1987)

equilibrium to exist, the following conditions must be met:

$$h^\phi(p, \mu) = (p - \mu)d(p, \phi) - \pi(\mu, 0) \geq 0 \quad \mu \in \{\mu_b, \mu_g\}$$

This condition states that neither the green nor the brown type would deviate from the pooling price strategy. Any price yielding a profit in both states of nature, no lower than the profit the right-hand side, is a candidate for concealing information.

## Hard Information

Recall that the G.NGO can successfully find hard evidence regarding the producer of green apples with probability  $\gamma < 1$  or learn nothing with probability  $1 - \gamma$ . The G.NGO can successfully monitor the producer of the brown apples with probability  $\beta < 1$  or learn nothing with probability  $1 - \beta$ . Moreover, the consumer updates his belief according to the pricing separating rule.

**Influence of championing and shaming when the market reveals the information.** This configuration will impact the incentive constraints on both types of apple producers in a different manner. The main impact is on the incentive of the producer of the brown apples. When the producer of the brown apples attempts to be taken for the green producer, if the G.NGO plays shaming with a positive probability, there is a chance  $\beta$  that the fraud is revealed to consumers. If the fraud is revealed the consumer applies the worst belief and faces a high price. Now, the right hand-side of Condition  $IC_b$  is rewritten as an expected profit on the green market of  $E[\pi(p, s)] = (1 - s)\beta\pi(p, 0) + (1 - (1 - s)\beta)\pi(p, 1)$ . For instance, if the G.NGO plays championing as a pure strategy, i.e.,  $s = 1$ , the brown firm would not face an expected profit anymore, and would have the same incentive constraint ( $IC_b$ ) as in the case of no hard information. If the G.NGO plays shaming for sure, i.e.,  $s^* = 0$ , then it would reveal the fraud if and only if the G.NGO successfully found a hard evidence. In this case, the profit that would be obtained by deviation is written  $\beta\pi(p, 0) + (1 - \beta)\pi(p, 1)$ .

The right-hand side of Condition  $IC_g$  relaxes the worst beliefs' profits. For the producer of green apple the price  $p^E = \sqrt{s\gamma\mu_g(v-1) + \mu_g}$  optimizes the expected profit  $E[\pi(\mu_g, s)] = s\gamma\pi(p, 1) + (1 - s\gamma)\pi(p, 0)$ . Note that when  $\gamma = 1$  we find  $p_g^*$  defined in the perfect information case. To achieve separation, the producer of the green apples must set a price  $p$  that satisfies the two following conditions:

$$\pi(p, 1) \geq E[\pi(\mu_g, s)] \quad (65)$$

$$\pi(\mu_b, 0) \geq E[\pi(p, s)] \quad (66)$$

**Proposition 1.** *If the technology of monitoring is relatively high (low)  $\beta > \beta^*(s)$  ( $\beta < \beta^*(s)$ )<sup>15</sup>, there exists a unique separating equilibrium where the producer of green apples applies its full information price,  $p = p_g^* = \sqrt{\nu\mu_g}$  ( $p = \bar{p}_b$ ) and the producer of brown apples does also,  $p = p_b^* = \sqrt{\mu_b}$ .*

We denote the separating price that prevents any deviation for the brown producer  $\bar{p}_b$  from (66) and  $\bar{p}_g$  the price ensuring revelation for the green producer from (65). Equalizing  $\bar{p}_b$  with the full information price  $p_g^*$  yields:

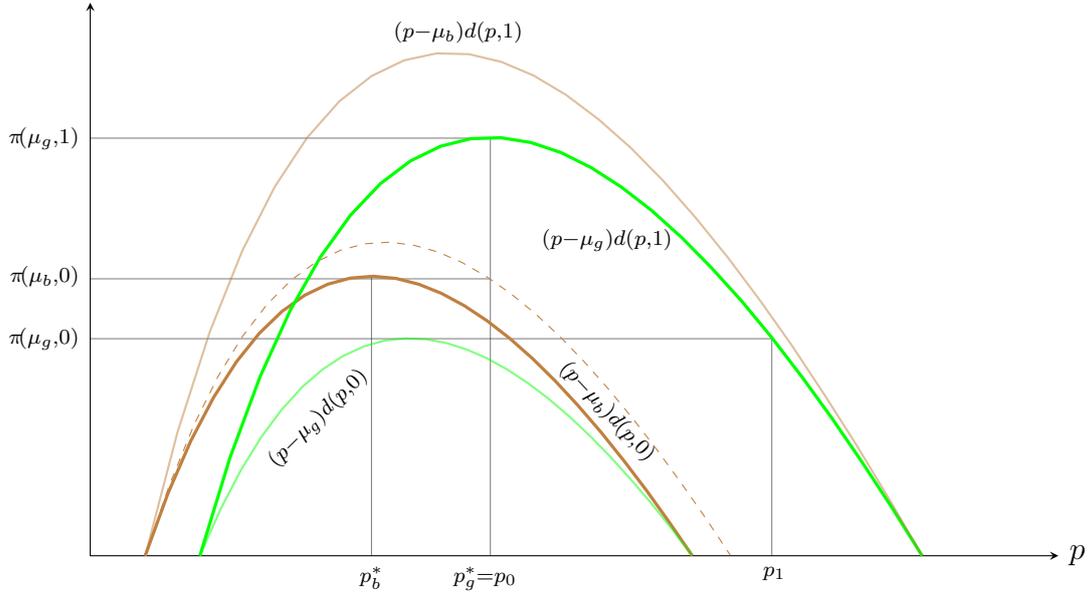
$$\beta^*(s) = \frac{(2\mu_b^{3/2} + (\nu-1)\mu_b)\sqrt{\nu\mu_g} - \nu(\mu_b + \mu_g)\mu_b}{(((-2s+1)\mu_b + (-s+1)\nu-s)\sqrt{\nu\mu_g} + ((-1+s)\nu+s)\mu_b + 2(s-1/2)\mu_g\nu)\mu_b}$$

At this level of monitoring, the producer of the green apples can always apply its full information price  $p_g^*$  and meet the separating conditions which is the strict dominant strategy. The intuition is close to that in (Bagwell and Riordan, 1991). These authors argue that the higher the proportion of informed consumers is, the lower is the cost of signaling for the high-quality firm. If there are enough informed consumers in the market, then it yields the perfect information price outcome. In this paper, the more

<sup>15</sup>To ensure the existence of such a threshold, we need two conditions. We know that  $\beta$  is a probability. We simultaneously need that (1) the consumer sufficiently values the environmental green good  $\nu \geq \frac{(-\mu_b - \mu_g)\sqrt{-8\mu_b^{5/2}\mu_g + (\mu_g^2 + (2\mu_b + 4)\mu_g + \mu_b^2)\mu_b^2 + 4\mu_b^{3/2}\mu_g} - (\mu_g^2 + (2c+2)\mu_g + \mu_b^2)\mu_b}{2\mu_b\mu_g}$ , and that the probability of the G.NGO to play a shaming strategy is high enough such that  $s \leq \frac{(-\mu_b - \nu)\sqrt{\nu\mu_g} + \nu(\mu_b + \mu_g)}{(-2\mu_b - \nu - 1)\sqrt{\nu\mu_g} + (\mu_b + 2\mu_g)\nu + \mu_b}$ .

accurate the PublicEye is, the higher the cost for the low-quality firm to deviate from the truth. As a result, a sufficient technology of detection is enough to reach the perfect information price outcome, as shown in Figure 5 where  $p_g^* = \bar{p}_0$

Figure 5: Influence of shaming on separating equilibria



Now, we turn on the analysis of the effect of the championing strategy on the price signaling strategy of the firm. In Proposition 1, if we use the Intuitive Criterion when  $\bar{p}_g > \bar{p}_b$ , then  $\bar{p}_b$  does not depend of  $\gamma$ . An immediate consequence is contained in the following proposition.

**Proposition 2.** *With imperfect monitoring technology  $\beta$ , the shaming strategy can help restore full credibility. With imperfect labeling technology  $\gamma$ , the championing strategy cannot.*

**Proof:** When  $\bar{p}_b > \bar{p}_g$ , there does not exist a price that satisfies (65) and (66). To keep the core of the reasoning as clear as possible, we set  $\beta = 0$ . The following proof would hold for any  $\beta$  that respects  $\bar{p}_b > \bar{p}_g$ .

Recall that  $E[\pi(\mu_g)] = \max_{p'} s\gamma\pi(p', 1) + (1 - s\gamma)\pi(p', 0)$ . We define  $\underline{\gamma}$  as the sufficient level of labeling satisfying the following equality  $\pi(\underline{\gamma}) = \pi(\mu_b, 0)$ . Since  $\pi(\gamma)$  is strictly increasing in  $\gamma$ , suppose that  $\bar{\gamma} > \underline{\gamma}$  such that  $\pi(\bar{\gamma}) > \pi(\mu_b, 0)$ . In this case (65) is always more stringent than (66); therefore,  $\bar{p}_b > \bar{p}_g$ . Now suppose that there exists a price  $p(\gamma)$  satisfying (65)-(66) simultaneously to confirm the existence of a separating equilibrium. One can rewrite these constraints as follows:

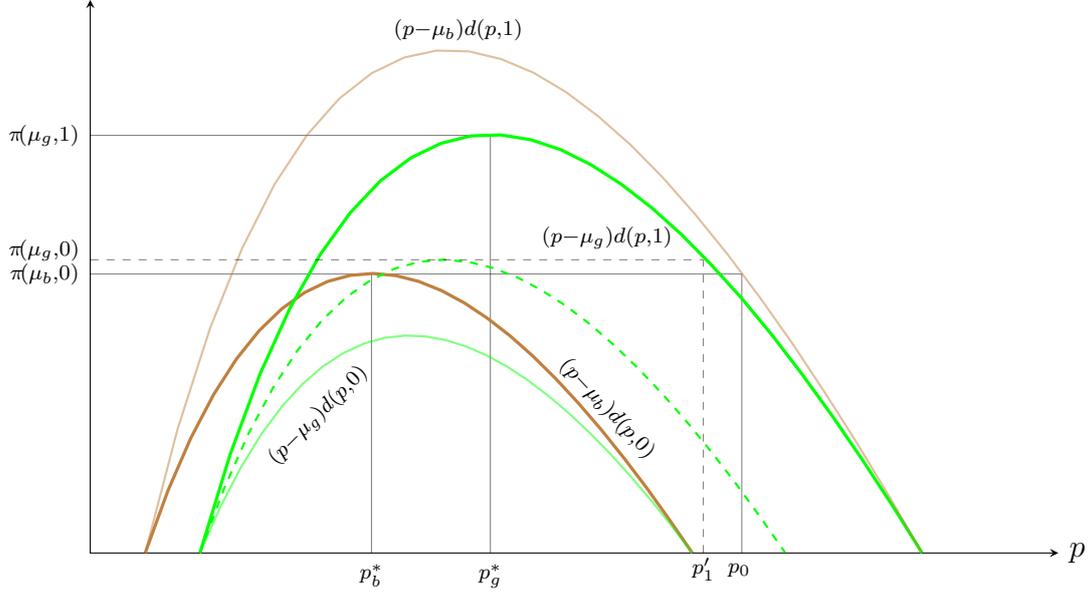
$$(p(\gamma) - \mu_g) d(p, 1) \geq \pi(\bar{\gamma})$$

$$\pi(\mu_b, 0) \geq (p(\gamma) - \mu_b) d(p, 1)$$

Then,  $(p(\gamma) - \mu_g) d(p, 1) \geq \pi(\bar{\gamma}) > \pi(\mu_b, 0) \geq (p(\gamma) - \mu_b) d(p, 1)$ . By transitivity one can simplify it to  $(p(\gamma) - \mu_g) > (p(\gamma) - \mu_b)$  implying that  $\mu_b > \mu_g$ , which is a contradiction. ■

This corollary emphasizes the fact that there does not exist a  $0 < \bar{\gamma} < 1$ , which would reduce the cost for the virtuous producer to signal its quality. It increases the level of confidence *per se*, but it crowds out the incentive for the green type to achieve separation by paying the signaling cost. Another way to understand the result is that it relaxes the pressure of the green producer to signal its quality because the consumer has some hard evidence. Therefore, the resulting price of this effect is not enough to prevent the deviation of the brown type. The main message of the proposition is that when the championing activity is sufficiently efficient, there are no longer revealing prices strategies as shown in Figure 6. Moreover, it is essential to underline that [Mailath \(1988\)](#) states the necessary conditions of a separating equilibrium to exist in a simultaneous signaling game. When we open the possibility of releasing hard evidence in such a game, it might strengthen the necessary condition for existence.

Figure 6: Influence of championing on separating equilibria



**Influence of shaming when the market conceals information.** With the presence of PublicEye the two conditions that have to be fulfilled for the existence of a pooling equilibrium are given by:

$$s\gamma\pi(p, 1) + (1 - s\gamma)\pi(p, \phi^H(\gamma)) > \max_{p'} \gamma\pi(p', 1) + (1 - \gamma)\pi(p', 0) \quad (67)$$

$$s\beta\pi(p, 0) + (1 - s\beta)\pi(p, \phi^H(\beta)) > \pi(\mu_b, 0) \quad (68)$$

For (68), it is clear that when the technology of monitoring augments, the space of pooling equilibrium is reduced for the same arguments as in the case of separating equilibria. For (67) it is less obvious and the reasoning is the following. When  $\phi$  increases it increases the set of possible prices. When  $\gamma$  increases the share of profit associated with the *a priori* of the consumer is decreased. As a result, when  $\gamma$  increases the possible prices satisfying (67) are reduced. In other words, for the green BigA to choose the worst-belief equilibrium profit is increasingly less of a problem as the labeling technology improves. Using an equilibrium refinement would definitely allow

us to conclude that the presence of the G.NGO reduces the situation where firms conceal information.

## 4 NGO informative behavior

The PublicEye is assumed to maximize the utility of the consumer. In this section, we analyze its optimal informative behavior given the level of the technology of monitoring  $\beta$ ;  $\gamma$ , the optimal pricing strategy of the firm; and the equilibrium beliefs of the consumer. We denote  $v(f, p)$  the indirect utility function of the consumer. As shown in Figure 3 the consumer has three possible informational nodes:  $f = 0$ , the apple is brown,  $f = 1$ , the apple is green; and  $f = f'$ , she does not know the type of the apple.

**Informative behavior in separating equilibria.** An equilibrium of the incomplete information simultaneous game is given by  $(p_b^*, p_g, s^*, f)$  where PublicEye plays  $s^*$ , the green type plays  $p_g$ , the brown type plays  $p_b^*$ , and the consumer reacts with belief  $f$ . Then if PublicEye's move is championing, the PublicEye obtains

$$\phi v(1, p_g) + (1 - \phi)v(0, p_b^*) \quad (69)$$

and, if PublicEye's move is shaming, the PublicEye obtains

$$\phi v(1, p_g) + (1 - \phi)v(0, p_b^*) \quad (70)$$

This will be an equilibrium if championing ( $s = 1$ ) is indeed the best response for PublicEye compared to if it never have incentives to deviate from the fixed point to shaming ( $s=0$ ), regardless of the beliefs about  $s$  of other players even if they believe with certainty that the NGO played championing:

$$\phi v(1, p_g(1)) + (1 - \phi)v(0, p_b^*) \geq \phi v(1, p_g(1)) + (1 - \phi)v(0, p_b^*) \quad (71)$$

Ibidem for shaming.<sup>16</sup> When the market fully reveals the information, the outcome of the NGO is limited because the separation in prices crowds out the evidence discovered. It is clear that the payoff would be the same whatever the strategy employed, and therefore, there is an infinity of equilibria with  $s^* \in [0, 1]$ .

**Proposition 3.** *When the firm reveals the truth, PublicEye is indifferent between playing the championing or shaming strategy.*

A direct intuition from this proposition is that if we assumed a heterogeneous cost between the two informative strategies we could easily explain why we observe G.NGO sticking to a specific strategy. However it is important to underline that the separating strategy depends on the presence of PublicEye. PublicEye bolsters the incentive of the green BigA to unravel its type by reducing the signaling cost.

**Informative Behavior in Pooling Equilibria** An uninformed consumer that buys at price  $p$  receives an indirect expected utility of

$$v(f, p) = \nu_f (\ln(\nu_f) - \ln(p))$$

Consider a candidate mixed strategy pooling equilibrium  $(s^*, p)$  where PublicEye plays  $s^*$  and both types of BigA play a pooling price  $p$ . Then if PublicEye's move is championing, the consumer (and thus PublicEye) obtains

$$\phi\gamma v(1, p) + (1 - \phi\gamma)v(f', p) \tag{72}$$

---

<sup>16</sup>As we showed, there exist two regions of separating equilibrium. When  $\beta > \beta^*$ , both firms apply their full information price and the consumer is perfectly informed, when  $\beta < \beta^*$ , BigA of the green type has to pay a signaling cost while BigA of brown type applies its full information price. As a result, we have:

$$v(1, p_g(0)) = \begin{cases} \phi v(1, p_g^*) & \text{if } \beta \geq \beta^* \\ \phi v(0, \bar{p}_g) & \text{if } \beta < \beta^* \end{cases}$$

for shaming to be a best response for PublicEye it needs the following:

$$\phi v(1, p_g(0)) + (1 - \phi)v(0, p_b^*) \geq \phi v(1, p_g(0)) + (1 - \phi)v(0, p_b^*)$$

and, if PublicEye's move is Shaming, the consumer gets:

$$(1 - \phi)\beta v(0, p) + (1 - (1 - \phi)\beta)v(f', p)$$

where  $f'$  is defined above in (63) with  $s = s^*$ . For these payoffs to be part of an equilibrium with  $s^* \in (0, 1)$ , they must be equal, so

$$\frac{v(1, p)}{v(0, p)} = \frac{\frac{1}{\phi\gamma} - 1}{\frac{1}{(1 - \phi)\beta} - 1} \equiv k$$

or

$$p = \nu \frac{1}{1 - k/\nu}$$

Finally, if  $p^*(s)$  is the pooling price that would be adopted by both types should  $s$  be played, then such an equilibrium exists if there is a  $s^* \in [0, 1]$  such that

$$p^*(s^*) = p$$

Notice that if  $p^*$  is monotonic, then a mixed equilibrium is unique when it exists.

Consider a candidate pure strategy equilibrium with pooling where PublicEye plays  $s^* = 1$ . Then the consumer obtains (72) with  $p = p^*(1)$  and

$$f' = \phi \frac{1 - \gamma}{1 - \phi\gamma}$$

This will be an equilibrium if championing ( $s = 1$ ) is indeed a best response for PublicEye that is, if

$$\phi\gamma v(1, p^*(1)) + (1 - \phi\gamma)v(f', p^*(1)) \geq (1 - \phi)\beta v(0, p^*(1)) + (1 - (1 - \phi)\beta)v(f', p^*(1))$$

or more simply

$$\phi\gamma v(1, p^*(1)) \geq (1 - \phi)\beta v(0, p^*(1)) + (\phi\gamma - (1 - \phi)\beta)v(f', p^*(1)) \quad (73)$$

Ibidem for an equilibrium with  $s^* = 0$ .

By analyzing (73), we can obtain several intuitions. We do not expect PublicEye to practice championing in industries where the dynamic of green production is very low. For instance, in a polluting chemical industry, the difficulty of promoting environmental practices will result in a pure strategy of shaming for PublicEye. In a contrasting case, if there is a significant amount of environmental production initiative, as in the case of some agricultural industries, it is likely that PublicEye would practice championing. In a nutshell, there is a correspondence between the proportion of firms adopting environmentally friendly production techniques and the informative behavior of PublicEye. A relevant suggestion would be to analyze the consequence of the difference between the two different discovery technologies  $\gamma$  and  $\beta$ . It seems legitimate to state that it is easier to find one piece of evidence that the firm is definitively brown than to find one piece of evidence that the firm is definitively green. Such a statement would have serious implications on the evolution of the continuum path of the mixed strategy of the NGO.

## 5 Conclusion

The main contribution of the paper has been to provide a rationale for the informative behavior of a G.NGO. After observing that many organizations use a shaming strategy, it is argued that, when imperfect, it has two major impacts on the signaling strategy of a firm. First, shaming reduces the cost of signaling for the high-quality type and can even restore the perfect information output. Second, championing decreases the incentive for the high-quality firm to signal itself, and to a certain extent, separating equilibria might even disappear in the signaling strategy of the firm. In a simultaneous incomplete information game, these results seem robust to different timing assumptions. Depending on whether the market correctly reveals or conceals the information,

the informative behavior strategy of the G.NGO is radically different. However, in situations where the presence of the G.NGO is the most desirable, *i.e.*, when pooling price equilibria occur, the G.NGO tends to match its informative strategy to the environmental technology of the market.

Finally, a natural extension that appears would be to use the dynamic of the reputation over time. If an NGO would repeatedly find hard evidence proving the type of the firm, our intuition is that it might anchor the NGO to a particular informative behavior.

# Conclusion

This thesis contributes to the economic literature concerning the signaling by providing a theoretical analysis of the role of third parties in discovering hard evidence for Bayesian consumers. More precisely, we focus on understanding how different strategies of the NGO shape the signaling behavior of firms. We will not review the results of this conclusion, as they are available in the introduction and in the section In the different chapters of this thesis, we will focus instead on the policy implications and perspectives for future research.

**The necessity of credible certification.** It might seem slightly trivial but it is not only a question of certification efficiency. It is also a matter of the lack of alternatives that could solve the lemons problem. In this case, this thesis showed a dead end. Indeed, these markets are dependent on the prior beliefs of consumers. As a result, if they are pessimistic, we expect the market to disappear even if some firms made a past investment for a greener production, and if they are optimistic, we expect some greenwashing due to the temptation of low-environmental-quality firms to benefit from the informational rent. Governments should not undermine these phenomena and should be more involved in guaranteeing the financial independence of third party responsible for the certification. Moreover, as underlined by the OECD,<sup>17</sup> more impact studies should be conducted to quantify the effect of the green certification, especially concerning apparel and forestry products, it would assuage its efficiency and, at the

---

<sup>17</sup><https://www.oecd.org/env/policy-perspectives-environmental-labelling-and-information-schemes.pdf>

same time its credibility. Moreover, this thesis provides a hint on the signaling behaviors of firms given the degree of imperfection of the label. If the label is slightly imperfect, it removes every incentive for firms to signal by themselves, and in this case, the market relies only on the degree of the precision of the certification. When the label is recognized as highly imperfect then we expect the firm to be more involved in signaling their environmental quality which is line with what we currently observe with the MSC label that has been strongly challenged.<sup>18</sup>

**Information a key factor for an ecological transition.** The Intergovernmental Panel on Climate Change (IPCC) alerts the necessity of shaping behaviors to answer the ecological emergency. The unsolved problem of asymmetric information directly raises the issue of firms engaging in environmental concerns. Although there are efficient and environmentally friendly means of production, we state that all the efforts allocated to technological development are put aside as long as we do not focus on preventing the fly-by-night strategy. This thesis provides hints on how third parties, such as NGOs, labeling or rating agencies, can induce firms to self-regulate through the provision of public information. They provide the market with the opportunity to address the moral hazard to ensure a long-term environmental quality provision.

**A particular attention to the shaming strategy.** Sanctioning the black swan does not result in the same effect as promoting a champion. As a consequence, when the certification fails to perfectly solve the asymmetric information issue, if firms can effectively transmit information, the shaming strategy appears to be the best option for NGOs. There are many “Name and Shame ” campaigns that efficiently corrected opportunistic behavior related to environmental frauds. Some of firms that engaged in green production have strong incentives to participate in the financial constraint of such agents. First, it raises the willingness of the green consumers to pay for a specific

---

<sup>18</sup><https://www.foodmanufacture.co.uk/Article/2019/09/10/Firms-urged-to-use-MSC-to-combat-fraud-in-seafood>

concern, and second giving more information helps the customers avoid being misled. Nevertheless, we must keep an eye on such financial participation in control activities to avoid the known excesses of green certification ([Jacquet et al., 2010](#)).

**Perspectives.** This thesis could be extended in several directions. There are three particular points that seem relevant to investigate in further research.

- First, the financial relationship between firms and NGOs. It has been shown in this thesis that such a relationship exists in both cases, either when a firm asks for a certification or when an NGO runs campaign against polluting behaviors. The NGOs in this context appear to be used as tools promoting the vertical differentiation between standard and green production. It would be relevant to further investigate this track in order to understand which drawbacks can emerge from this financial investigation concerning credibility.
- Second, some generalizations could be further developed concerning the existence of separating equilibria. Signaling models often occur in a context where there exist multiple sources of information. This thesis, by raising some reasonable assumptions, challenged the existence of this nature of equilibria. Again, the separation is crucial to perfectly solve the question of asymmetry of information and there are plenty real-world examples where the signal interacts with hard evidence. As a result, the out-of-equilibrium arbitrary beliefs well known in the signaling model could be further discussed.
- Third, due to the nature of the thesis it is hard to collect relevant data to test the theoretical ideas of the models. The use of some experimental economics could be relevant to further investigate the intuitions provided in this paper. We can easily imagine an experiment combining agents representing the market with and without potential third parties to study their influence on the transmission

of information on environmental quality on credence goods markets.

# Bibliography

- Akerlof, G. A. (1970). The Market for “Lemons”: Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics*, 488–500.
- Amacher, G. S., Koskela, E., and Ollikainen, M. (2004). Environmental Quality Competition and Eco-labeling. *Journal of Environmental Economics and Management*, 47(2), 284–306.
- Bagwell, K., and Riordan, M. H. (1991). High and Declining Prices Signal Product Quality. *American Economic Review*, 224–239.
- Baksi, S., and Bose, P. (2007). Credence Goods, Efficient Labelling Policies, and Regulatory Enforcement. *Environmental and Resource Economics*, 37(2), 411–430.
- Baron, D. P. (2001). Private Politics, Corporate Social Responsibility, and Integrated Strategy. *Journal of Economics and Management Strategy*, 10(1), 7–45.
- Baron, D. P. (2010). Morally Motivated Self-regulation. *American Economic Review*, 100(4), 1299–1329.
- Baron, D. P. (2011). Credence Attributes, Voluntary Organizations, and Social Pressure. *Journal of Public Economics*, 95(11-12), 1331–1338.
- Belleflamme, P., and Peitz, M. (2015). *Industrial Organization: Markets and Strategies*. Cambridge University Press.
- Besley, T., and Ghatak, M. (2007). Retailing Public Goods: the Economics of Corporate Social Responsibility. *Journal of Public Economics*, 91(9), 1645–1663.
- Blomquist, J., Bartolino, V., and Waldo, S. (2015). Price Premiums for Providing Eco-labelled Seafood: Evidence from MSC-certified Cod in Sweden. *Journal of Agricultural Economics*, 66(3), 690–704.
- Bonroy, O., and Constantatos, C. (2008). On the Use of Labels in Credence Goods Markets. *Journal of Regulatory Economics*, 33(3), 237–252.
- Bottega, L., and De Freitas, J. (2009). Public, Private and Nonprofit Regulation for Environmental Quality. *Journal of Economics and Management Strategy*, 18(1), 105–123.

- Cho, I.-K., and Kreps, D. M. (1987). Signaling Games and Stable Equilibria. *Quarterly Journal of Economics*, 102(2), 179–221.
- Civel, E., Cruz, N., et al. (2018). *Green, Yellow or Red Lemons? Artefactual Field Experiment on Houses Energy Labels Perception* (Tech. Rep.).
- Daley, B., and Green, B. (2014). Market Signaling with Grades. *Journal of Economic Theory*, 151, 114–145.
- Darby, M. R., and Karni, E. (1973). Free Competition and the Optimal Amount of Fraud. *Journal of Law and Economics*, 16(1), 67–88.
- Daubanes, J., Rochet, J.-C., et al. (2019). *The rise of ngo activism* (Tech. Rep.).
- Daughety, A. F., and Reinganum, J. F. (2008a). Communicating Quality: a Unified Model of Disclosure and Signalling. *RAND Journal of Economics*, 39(4), 973–989.
- Daughety, A. F., and Reinganum, J. F. (2008b). Imperfect Competition and Quality Signalling. *RAND Journal of Economics*, 39(1), 163–183.
- Dewatripont, M., and Tirole, J. (1999). Advocates. *Journal of Political Economy*, 107(1), 1–39.
- Feddersen, T. J., and Gilligan, T. W. (2001). Saints and Markets: Activists and the Supply of Credence Goods. *Journal of Economics and Management Strategy*, 10(1), 149–171.
- Fischer, C., and Lyon, T. P. (2014). Competing Environmental Labels. *Journal of Economics and Management Strategy*, 23(3), 692–716.
- Fleckinger, P., Glachant, M., and Moineville, G. (2017). Incentives for Quality in Friendly and Hostile Informational Environments. *American Economic Journal. Microeconomics*, 9(1), 242.
- Hamilton, S. F., and Zilberman, D. (2006). Green Markets, Eco-certification, and Equilibrium Fraud. *Journal of Environmental Economics and Management*, 52(3), 627–644.
- Harbaugh, R., Maxwell, J. W., and Roussillon, B. (2011). Label Confusion: the Groucho Effect of Uncertain Standards. *Management Science*, 57(9), 1512–1527.
- Heyes, A., Lyon, T. P., and Martin, S. (2018). Salience games: Private politics when public attention is limited. *Journal of Environmental Economics and Management*, 88, 396–410.
- Heyes, A., and Martin, S. (2016). Social labeling by competing NGOs: A model with multiple issues and entry. *Management Science*, 63(6), 1800–1813.
- Heyes, A., and Maxwell, J. W. (2004). Private vs. Public Regulation: Political Economy of the International Environment. *Journal of Environmental Economics and*

- Management*, 48(2), 978–996.
- Ibanez, L., and Grolleau, G. (2008). Can Ecolabeling Schemes Preserve the Environment? *Environmental and Resource Economics*, 40(2), 233–249.
- Jacquet, J., Pauly, D., Ainley, D., Holt, S., Dayton, P., and Jackson, J. (2010). Seafood Stewardship in Crisis. *Nature*, 467(7311), 28–29.
- Janssen, M. C., and Roy, S. (2010). Signaling Quality Through Prices in an Oligopoly. *Games and Economic Behavior*, 68(1), 192–207.
- Kerschbamer, R., and Sutter, M. (2017). The Economics of Credence Goods—a Survey of Recent Lab and Field Experiments. *CEPR Economic Studies*, 63(1), 1–23.
- Klein, B., and Leffler, K. B. (1981). The Role of Market Forces in Assuring Contractual Performance. *Journal of Political Economy*, 89(4), 615–641.
- Linnemer, L. (2002). Price and advertising as signals of quality when some consumers are informed. *International Journal of Industrial Organization*, 20(7), 931–947.
- Lizzeri, A. (1999). Information Revelation and Certification Intermediaries. *RAND Journal of Economics*, 214–231.
- Lyon, T. (2012). *Good Cop/Bad Cop: Environmental NGOs and Their Strategies Toward business*. UK: Routledge.
- Lyon, T., Maxwell, J. W., et al. (2004). *Corporate Environmentalism and Public Policy*. UK: Cambridge University Press.
- Mahenc, P. (2007). Are Green Products Over-priced? *Environmental and Resource Economics*, 38(4), 461–473.
- Mahenc, P. (2008). Signaling the Environmental Performance of Polluting Products to Green Consumers. *International Journal of Industrial Organization*, 26(1), 59–68.
- Mahenc, P. (2017). Honest Versus Misleading Certification. *Journal of Economics and Management Strategy*, 26(2), 454–483.
- Mailath, G. J. (1988). An Abstract Two-period Game with Simultaneous Signaling—Existence of Separating Equilibria. *Journal of Economic Theory*, 46(2), 373–394.
- Mailath, G. J., Okuno-Fujiwara, M., and Postlewaite, A. (1993). Belief-based Refinements in Signalling Games. *Journal of Economic Theory*, 60(2), 241–276.
- Mas-Colell, A., Whinston, M. D., Green, J. R., et al. (1995). *Microeconomic Theory* (Vol. 1). New York: Oxford University Press.
- Mason, C. F. (2011). Eco-labeling and Market Equilibria with Noisy Certification Tests. *Environmental and Resource Economics*, 48(4), 537–560.

- Mathis, J., McAndrews, J., and Rochet, J.-C. (2009). Rating the Raters: are Reputation Concerns Powerful Enough to Discipline Rating Agencies? *Journal of Monetary Economics*, 56(5), 657–674.
- Milgrom, P., and Roberts, J. (1986). Price and Advertising Signals of Product Quality. *Journal of Political Economy*, 94(4), 796–821.
- Mirman, L. J., and Santugini, M. (2014). The Informational Role of Prices. *Scandinavian Journal of Economics*.
- Mussa, M., and Rosen, S. (1978). Monopoly and Product Quality. *Journal of Economic Theory*, 18(2), 301–317.
- Rob, R., and Fishman, A. (2005). Is Bigger Better? Customer Base Expansion Through Word-of-mouth Reputation. *Journal of Political Economy*, 113(5), 1146–1162.
- Salop, S. C. (1979). Monopolistic Competition with Outside Goods. *Bell Journal of Economics*, 10(1), 141–156.
- Shaked, A., and Sutton, J. (1982). Relaxing Price Competition Through Product Differentiation. *Review of Economic Studies*, 3–13.
- Shapiro, C. (1983). Premiums for high quality products as returns to reputations. *The quarterly journal of economics*, 98(4), 659–679.
- Spence, M. (1973). Job Market Signaling. *Quarterly Journal of Economics*, 87(3), 355–374.
- Steering Committee of the State-of-Knowledge Assessment of Standards and Certification. (2012). Toward sustainability: The roles and limitations of certification. *Washington, DC: RESOLVE, Inc.*
- Tirole, J. (1996). A theory of collective reputations (with applications to the persistence of corruption and to firm quality). *Review of Economic Studies*, 63(1), 1–22.
- Washington, S., and Ababouch, L. (2011). *Private Standards and Certification in Fisheries and Aquaculture* (Tech. Rep.).
- Wijen, F., and Chiroleu-Assouline, M. (2019). Controversy over voluntary environmental standards: A socioeconomic analysis of the Marine Stewardship Council. *Organization and Environment*, 32(2), 98–124.