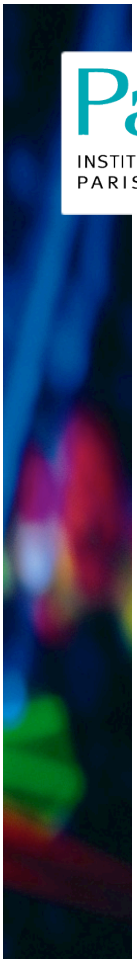
# The Economics of corporate social responsibility : the informational role of third parties

Gabrielle Moineville



ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

MINES ParisTech

École doctorale n° 396 : Economie, Organisation, Société

# Doctorat ParisTech

# T H È S E

**pour obtenir le grade de docteur délivré par**

## l'École nationale supérieure des mines de Paris

### Spécialité "Economie et finance"

*présentée et soutenue publiquement par*

**Gabrielle MOINEVILLE**

le 23 novembre 2012

# L'Economie de la Responsabilité Sociale et Environnementale de l'Entreprise: Le Rôle Informationnel des Tierces Parties

# The Economics of Corporate Social Responsibility: The Informational Role of Third Parties

Directeur de thèse : **Matthieu GLACHANT**
Co-encadrement de la thèse : **Pierre FLECKINGER**

**Jury**

| | |
|---|---|
| **Mme Yolande HIRIART**, Professeur d'Economie, Université de Franche-Comté | Rapporteur |
| **M. Frans DE VRIES**, Reader in Economics, University of Stirling | Rapporteur |
| **Mme Mireille CHIROLEU-ASSOULINE**, Professeur d'Economie, Paris School of Economics, U. Paris 1 | Examinateur |
| **M. Philippe MAHENC**, Professeur d'Economie, Université de Montpellier | Examinateur |
| **M. Pierre FLECKINGER**, Professeur Assistant d'Economie, Paris School of Economics, U. Paris 1 | Examinateur |
| **M. Matthieu GLACHANT**, Professeur d'Economie, MINES ParisTech | Examinateur |

T H È S E

**MINES ParisTech**
**Cerna**
60 Boulevard Saint Michel 75006 Paris

# Acknowledgments

Je veux d'abord remercier Matthieu Glachant. Rien de tout cela ne serait arrivé sans lui. Il m'a insufflé, jour après jour, la motivation et la curiosité nécessaires pour aller au bout du challenge qu'a représenté pour moi cette thèse. Auprès de lui, j'ai appris la rigueur intellectuelle, la persévérance, la capacité d'accepter qu'on s'est trompé et de rebondir. Merci d'avoir cru en moi, même quand je n'y croyais pas moi-même, et de m'avoir offert autant d'opportunités.

Merci à Pierre Fleckinger. Merci pour avoir été particulièrement présent dans les moments les plus importants de ces trois ans. Cette thèse ne serait pas ce qu'elle est sans lui (ne serait pas du tout?). Je le remercie pour sa hauteur de vue, sa finesse d'analyse et sa connaissance époustouflante de la littérature économique, qui m'ont poussée à me dépasser. Merci pour tout ça.

Vous avez été deux directeurs de thèse incroyablement complémentaires. Merci surtout de m'avoir confirmé qu'on peut s'amuser en travaillant très efficacement. Ce fut un grand plaisir de travailler avec vous au quotidien.

Merci à Sesaria Ferreira. Le Cerna ne serait rien sans elle. Elle a été là pour me remettre la tête sur les épaules dans les moments de crises (aussi divers soient-ils), probablement sans même s'en rendre compte la plupart du temps. Mais elle a aussi été là pour partager les joies et sa présence a rendu mon séjour aux Mines incroyablement plus riche et plus humain.

Merci à tous mes fellows thésards pour tous les chouettes moments que nous avons partagés. Par ordre d'apparition, Jan Eilhard, Paula Restrepo, Henry Delcamp, Justus Baron, Arnaud de la Tour, Mattia de Grassi, Michel Berthelemy, Lina Escobar, Marine Lefort, Yasmine Benamrane et Damien Dussaux.

Merci à Vanina Forget, Aurélien Petit et Samer Hobeika pour tous les tips de thésards sur la RSE, mais aussi pour le fun.

Thanks to Tom Lyon for having welcomed me in Michigan University. It was a very rich experience of the American way of living and doing research.

Merci à Patrick d'Humières pour avoir connecté notre recherche au monde réel.

Merci à la Caisse des Dépôts et des Consignations pour l'intérêt qu'elle porte à nos travaux à travers le financement d'un projet de recherche.

Je voudrais aussi remercier mes DJs, grâce à qui j'ai bénéficié d'une super ambiance musicale pendant 3 ans : Gregory van der Heijden, Matthieu Glachant et Florian Bersier.

Dans la même veine, je voudrais remercier mes tech supports, qui m'ont sauvé la vie maintes fois : Pierre Fleckinger, Anaïs Khuong et Florian Bersier.

Il me faut aussi remercier tous mes amis qui m'ont supporté avec patience et avec le sourire (la plupart du temps) à travers ces 3 années : Mathilde, Hélène, Louise, Guillaume, Clémence, Fabien, Stéphane, Tanguy et Nicolas.

Finalement, je voudrais remercier ma famille : Papa, Maman, Marie et Jean. Tout ce que je suis prend sa source en vous.

# Preface

This dissertation deals with the informational mechanisms at stake with regard to Corporate Social Responsibility (CSR), but some of my work is relevant outside this field. More precisely, it contains a literature review and three theoretical articles, organized as four chapters.

The first chapter, *Evaluating Corporate Social Responsibility: a Review of the Literature*, is the first step of a research project financed by the French agency Caisse des Dépôts et des Consignations that aims, on the one hand, to summarize the main findings of the economic literature about Corporate Social Responsibility and, on the other hand and in the light of these findings, propose orientations for welfare improving public policies. I was involved in this project with Pierre Fleckinger and Matthieu Glachant.

The second chapter, *Incentives for Quality in Friendly and Hostile Informational Environments*, is coauthored with Pierre Fleckinger and Matthieu Glachant. An earlier version of this article was presented at the $18^{th}$ Annual Conference of the European Association of Environmental and Resource Economists, Rome, in June 2011.

The third chapter, *The Informational Role of Non-Governmental Organizations to Induce Self-Regulation: Cheering the Leaders or Booing the Laggards?*, is coauthored with Matthieu Glachant and Pierre Fleckinger. This article was presented at the $19^{th}$ Annual Conference of the European Association of Environmental and Resource Economists, Prague, in June 2012. It is an *i3 working paper*. The Interdisciplinary Institute for Innovation (*i3*) brings together all the MINES ParisTech economics, management and sociology research teams (from the CERNA, CGS and CSI) and those of the Department of Economics and Social Sciences at TELECOM ParisTech.

The fourth chapter, *Should Imperfect Labels be Strict or Lenient?*, is coauthored with Pierre Fleckinger and Matthieu Glachant.

# Contents

# Introduction

## Context

Corporate Social Responsibility (CSR) is firm's interpretation of the concept of sustainable development. It entails that a company should not only worry about its profitability, but also take into account its impacts on society and the environment. According to this approach, companies are expected to pay attention to the concerns of all their stakeholders, including employees, shareholders, customers, suppliers and the civil society. It consequently comprehends a wide variety of firms' practices, ranging from waste reduction plan to promoting sustainable fishing, fighting global warming, as well as investing in local communities' projects.

Authors in the economic literature have highlighted two criteria to label a given practice as CSR. A first view is that it consists for firms in voluntarily setting environmental or social policies going beyond the existing mandatory requirements. A second view is that it consists for firms in voluntarily sacrificing a share of their profits to provide a public good, either social or environmental. Many authors consider only the first criteria to be relevant. For example Lyon and Maxwell (2008) "define environmental CSR as environmentally friendly actions not required by law." Or Denicolò (2008) states that CSR amounts to "firms voluntarily and unilaterally over comply(ing) to current regulation". For other authors the crucial criteria is on the contrary the second one. For Ek (2005) CSR is "sacrificing profits in the social interests." Baron (2001) defines two different types of CSR: strategic CSR, driven by profit-maximizing strategies, and CSR, driven by altruistic preferences that entail profit losses. Finally, both dimensions can be reconciled, as in Mc Williams and Siegel (2000) who define CSR as "actions that appear to further some social good beyond the interests of the firm and that which is required by law".

I adopt the first criterion: in this thesis, CSR refers to practices going beyond legal

obligations. I thus use interchangeably the terms self-regulation and CSR in all my chapters. Importantly, throughout this dissertation CSR is considered as a potential trigger for increased profits.

CSR is a trendy concept. Firms growingly claim that they behave responsibly. For instance, Bazillier and Vauday (2009) report that 406 firms out of the 593 of their sample have released at least one sustainable development report. The consulting agency CapitalCom indicates that in 2009, almost half of CAC40 firms were registered at the GRI,[1] 37 of them had declared their support to the Global Compact[2] and all of them had answered the Carbon Disclosure Project[3] questions.

As self-regulatory efforts are arguably costly, the reasons why corporations are willing to self-regulate have been extensively explored in the economic literature. Many works stress the fact that some stakeholders are willing to reward CSR leading firms or, alternatively, to punish laggards. Some consumers may accept to buy their products at a higher price or boycott 'dirty' corporations (Bagnoli and Watts, 2003; Arora and Gangopadhyay, 1995). Employees may work in a more productive way, or they may accept lower wages in environmentally -or socially- responsible firms (Brekke and Nyborg, 2008). Socially responsible investors propose capital to leading firms at a reduced cost (Heinkel et al., 2001). Finally, CSR may also impact the regulatory process, for example by preventing the adoption of new mandatory standards (Fleckinger and Glachant, 2011), influencing the strictness of the mandatory standards (Denicolò, 2008), or reducing the frequency of inspections by authorities in charge of the regulation's enforcement (Maxwell and Decker, 2006). I provide a detailed review of the economic literature on self-regulation in **Chapter**

[1]The Global Reporting Initiative produces one of the world's most prevalent guidelines for sustainability reporting. It was formed by the United States based non-profits Ceres (formerly the Coalition for Environmentally Responsible Economies) and Tellus Institute, with the support of the United Nations Environment Programme.

[2]The United Nation initiative Global Compact is a principle-based framework for businesses, stating ten universally accepted principles in the area of human rights, labour, environment and anti-corruption.

[3]The Carbon Disclosure Project is an independent not-for-profit organization holding the largest database of primary corporate climate change information in the world.

**1.**

While self-regulating may increase firms' profits, it is a credence attribute for stakeholders. Neither search nor experience can guarantee them that a firm is really self-regulating. Consequently, firms can be tempted to advertise about a so-called responsible behavior in order to ripe these benefits from socially conscious stakeholders, without really bearing the cost of implementing additional practices. This is called greenwashing. Because of that incentive to cheat, firms' credibility in self-reporting their CSR performance is an issue.

This creates an opportunity for specialized actors that have sufficient resources to observe self-regulatory activities of individual firms and to convey this information to the stakeholders. These specialized actors can be extra-financial rating agencies, Non-Governmental Organizations (NGOs) or labeling institutions.

## Research questions and structure of the document

The starting point of this thesis is that real-world auditors are usually specialized. Certain organizations send mostly good news to the general public: they confirm that a firm – or a product – is socially or environmentally responsible. An illustration is the Marine Stewardship Council that rewards sustainable fishing by certifying about 12 percent of the world catch intended to human consumption. Other auditors send bad news: they get specialized in denouncing firms that perform badly. Greenpeace is the emblematic example of an organization that fiercely conducts campaigns against firms that they do not deem to behave responsibly.[4] Another example is the NSW (New South Wale) food authority in Australia and its "Name & Shame" program that aims at listing publicly firms that have failed to follow the NSW quality standards.[5] Lie in between neutral auditors

[4]Think of the campaign against Shell organized in 1995 to prevent the disposal of the Brent Spar oil storage buoy in deep water.

[5]http://www.foodauthority.nsw.gov.au/aboutus/offences/

that send both good and bad news, as for example WWF which alternates between cooperative and offensive behavior toward firms depending on their compliance with their demands. Or extra-financial rating agencies, such as Vigeo or KLD, that disclose firm's scores whatever their performance. In the following, we refer to the auditors which cheer CSR leaders as good cops whereas bad cops are auditors which boo the laggards. Note that this heterogeneity may reflect selective monitoring - the auditor focuses on the observation of positive activities - or selective disclosure - the auditor chooses to communicate only some of its monitoring results.

My Ph.D. thesis focuses on two main questions: how do the different auditors' informational behaviors impact firms' incentives to invest in self regulation? When and why do auditors choose one informational behavior or another?

To answer these questions, we have developed a simple theoretical model that explores the provision of quality by a firm under friendly informational environments, in which quality is more often disclosed when it is high than when it is low, and hostile environments, in which the converse holds (**Chapter 2**). This chapter is the core contribution of this thesis. We have then applied this base model to two types of third party: NGOs and labels. First, we have endogenized the production of good and bad news in order to understand when and why NGOs choose between being good cops or bad cops (**Chapter 3**). Second, we have used the base model to analyze what is the more efficient level of label's strictness to spur social welfare and investment in high quality (**Chapter 4**). In this chapter, the interpretation of friendly and hostile is the following: a lenient label generates a hostile environment because it entails a clearer identification of worst-in-class than best-in-class firms, while a strict label better identifies best-in-class firms, thus generating a friendly environment.

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

# Quality provision under hostile or friendly informational environments

The issue of providing a credence attribute under asymmetric information is known as a lemon problem and, since the publication of Akerlof 's paper in 1970, has been extensively studied in the economic literature. For this reason, the model and results presented in **Chapter 2** are relevant outside the field of self-regulation and are consequently embedded in a more general theoretical framework.

This paper is obviously far from being the first to explore the impact of informational environments on quality provision. The distinctive feature of our analysis is to build a model that highlights the key differences between hostile and friendly environments. In previous works, these differences, albeit crucial as we point out, are either left implicit as in works on statistical discrimination (Coate and Loury , 1993) and collective reputation (Tirole, 1996); or the focus is on friendly environments as in the literature on quality disclosure which looks at the incentives for firms to voluntarily disclose quality and for certifiers to provide unbiased certification about quality (for a recent survey, see Dranove and Jin, 2010).

We develop a model that accommodates all those types of environments at the same time and we seek to identify their impacts on quality provision. It depicts a continuum of agents who are willing to produce and sell a good whose quality can either be low or high. Agents choose to produce low quality at no cost, while choosing to produce high quality entails a cost varying across agents. This decision is imperfectly observable by the potential buyer. More precisely, once the agent has selected its level of quality, some monitoring occurs, which discloses quality with a probability $g$ if quality is high ("good news") and probability $b$ if quality is low ("bad news"), as described in Figure 1. After monitoring, the buyer updates its beliefs about quality and decides to purchase or not the good.

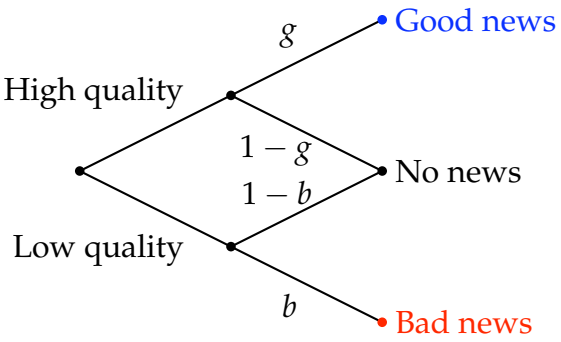



Figure 1: Information structure.

The model describes how agents' quality investments interact with these belief revisions. We find clear-cut differences between hostile and friendly environments. Hostile environments give rise to a bandwagon effect among agents, which leads beliefs to be self-fulfilling in some cases, so that there may be multiple equilibria. The intuition is the following: suppose that the buyer is ex ante very optimistic, supplying high quality is an equilibrium because hostile monitoring would easily reveal the agents supplying low quality. But when the buyer 's prior belief is pessimistic, incentives to increase quality are limited because it is relatively difficult to ascertain high quality, and there is few hope for the seller to prove wrong a pessimistic belief.

In contrast, friendly environments create a form of free riding across agents, which always induces a unique equilibrium. Comparative statics results are also contrasted. In particular, we show that moving from a hostile to a friendly environment increases the average equilibrium quality when the resources allocated to monitoring -reflected in the sum $g + b$ in our model- is low. The reverse is true for a higher $g + b$. The idea is that little information induces little reward to high quality firms and thus little incentive to produce high quality. Knowing that, the buyer is pessimistic about a good's quality on which he did not receive any feedback. It is more effective to go against the buyer 's belief by increasing the probability to get the full reward when producing high quality.

This paper allows to understand finely how a given informational environment shapes

firms' incentives to provide high quality, but one of its limit is that auditors are not explicitely modeled and thus can not be analyzed. We adapted the model to endogenize auditors' choice of informational behavior (good or bad cop) and applied it to the case of NGOs aiming to spur firms' CSR investment.

## The role of Non-Governmental Organizations to induce self-regulation

As highlighted precedently, NGOs and their impacts on firms' behavior is an important issue in the economic literature about self-regulation. Nevertheless, to our knowledge, no one has explicitely model why and when they choose one informational behavior or another. This is what we attempt to do in this chapter.

The economic literature on self-regulation has already explored the role of NGOs in triggering CSR. For instance, "Good Cop/Bad Cop" is the title of a recent book edited by (Lyon, 2010) which contrasts the heterogeneity of NGO strategies towards business. A few theoretical papers specifically deal with bad cops. For instance, (Baron, 2001) and (Lyon and Maxwell, 2011) investigate the impact of NGOs which are able to penalize firms that do not make sufficient environmental or social efforts. Others deal with good cops, and in particular with the role of NGOs in product labeling (ecolabels, fair trade, etc.). For example, (Bottega and De Freitas, 2009) compare two certifiers: a NGO and a for-profit organization. None of these papers deal with the different types of NGOs in a unified framework nor develop a theoretical analysis of what drives such strategies, as I do in this chapter.

To do so we build up on the model introduced in the previous chapter, but now we endogenize the choice of communication behavior of NGOs. There is a continuum of firms that may individually decide to self-regulate or not, a stakeholder who is willing to pay for self-regulation and a NGO whose aim is to maximize the average level of self-

regulation under a budget constraint and that provides the stakeholder with imperfect information on firms real actions. We later extend the analysis to a setting with multiple NGOs.

Information disclosure by the NGO can be hostile or friendly: the NGO tells the firm's action to the stakeholder with a probability $g$ if it self regulates, and with a different probability $b$ if it carries on with business-as-usual. The parameters $g$ and $b$ are endogenously selected by the NGO which can decide to be neutral ($g = b$), a good cop ($g > b$) or a bad cop ($g < b$). We use the model to investigate how the NGO disclosure probabilities $(g, b)$ shape firms' behavior and then to characterize the probabilities chosen in equilibrium by the NGO.

We show that the NGO decides to specialize: If its budget is limited, it chooses to be a good cop ($g > 0$ and $b = 0$). When more resources are available, it opts for being a bad cop ($b > 0$ and $g = 0$). This result confirms those obtained in **Chapter 2**: it is more effective to go against the stakeholder's expectations. Indeed, a high budget expands disclosure probabilities, implying high firms' incentives to self-regulate, which make the stakeholder hold optimistic beliefs on the expected level of self-regulation. And conversely when resources are low.

We develop several extensions of the model. First, we assume multiple NGOs which non cooperatively select their technology. A coordination problem arises for the two possible equilibria -NGOs being all bad cops and NGOs being all good cops- sometimes coexist. We also introduce the option for firms to cooperate with NGOs. More specifically, we distinguish individual partnership, whereby the NGO specifically publicizes the partnering firm's behavior, and collective partnership whereby industry support leads to an increase of the aggregate disclosure probability.

Although the prime focus of the analysis is positive, it is possible to derive a general policy lesson. When the amount of resources available to NGOs is limited and/or

the number of NGOs is low, everything is going smoothly and the role of a welfare-maximizing regulator is limited: There exists a single equilibrium which is socially optimal. As all NGOs are good cops, firms are willing to engage in collective or individual partnerships in order to increase the quantity of information, which in turn increases self-regulation and thus social welfare. Things get more complicated when the quantity of information is such that the social outcome is the bad cop equilibrium. In this case, the coordination problem can lead to the survival of the good cop equilibrium. It is reinforced by the fact that firms which do not self-regulate do not form partnerships anymore and self-regulating firms are less prone to enter in partnership for the benefits get smaller. To sum up, a public intervention promoting bad cops becomes increasingly useful as NGOs financial resources increase.

## Optimal strictness of imperfect labels under imperfect competition

After having analyzed NGOs informational behaviors, we focus in this chapter on labels' strictness. A lenient label is akin to a hostile informational environment, where worst-in-class firms are more clearly identified than best-in-class firms, while a strict label stands for a friendly environment, where it is the reverse. There are two important differences with regard to our base model: we introduce competition and consumers beliefs at equilibrium are not fully Bayesian. They hold a prior that a firm produces high quality and update it upon receiving the labels' signals: labeled or unlabeled. Our model entailing competition, unlike the base model, the complexity of the mechanism that a perfectly sophisticated consumer should anticipate at equilibrium turns out to represent quite a challenge. That is why we found more realistic to limit somehow consumers discernment here.

We focus our analysis on the comparison between a strict and a lenient standard. The

stricter the label is, the fewer products will be labeled, but the stronger they will be identified as high quality products and hence rewarded. The more lenient the label is, the more firms will get the label and hence have incentives to invest in higher quality in the first place, but the lesser the reward will be as being labeled won't be perceived as such a great achievement.

We aim to understand why and when one of these two types of label spurs more quality than the other, and what is optimal from a social welfare point of view.

Labels allow firms to differentiate, which has a double effect. On the one hand, it provides incentives for firms to invest and produce high quality which is welfare enhancing since without any information firms that would find profitable to invest under perfect information do not. But on the other hand, differentiation alters competition and firms can charge prices higher than their production cost and make positive profits, which can be welfare damaging. The identification of this issue in the economic literature dates back at least to Shaked and Sutton (1982).

From then on, the interaction between minimum quality standards and the level of competition, alongside its impacts on social welfare, has been an issue well acknowledged in the industrial organization literature (e.g. Ronnen (1991) or Crampes and Hollander (1995)). Our analysis contributes to this literature.

Nevertheless, while these papers assume a perfect certification test, like most of the literature about labels, we model an imperfect one. We deem that this latter approach is more realistic. Indeed, many reasons can lead to mistakes in the attribution of labels. There may be measurement errors, infrequent verification controls, corruption of the labeling institution, etc.

As a result, an important feature of our paper is that we deal both with imperfect competition and imperfect certification.

Our model describes the interaction between a duopoly competing "à la Bertrand"

and consumers who positively but heterogeneously value environmental and social products' attributes (also designated as high quality hereafter). Firms privately decide whether to produce high quality or low quality. They are then submitted to a costless and mandatory labeling process that reveals some information to the consumers about the true level of quality of their products.

More precisely, the monitoring technology is imperfect: it reveals the true level of product quality with a probability $m$, but with a probability $1 - m$ it reveals nothing. The parameter $m$ captures the precision of the certification test. We consider two standard's levels: a lenient label who is granted even to products whose quality level is unknown, and a strict label who is granted only to products that have been confirmed to be high quality products.

Importantly, contrarily to others papers that study standard strictness under imperfect competition, firms have symmetric strategies at equilibrium: there is differentiation ex-post, but not ex-ante. This allows us to focus on informational insights.

We find that the welfare ranking is ambiguous. The strict label dominates the lenient label if consumers' prior belief that the product is of high quality is high and/or if the certification is very noisy. As for the prior, the intuition is the following: under a lenient label, the uncertainty lays on labeled firms, thus a consumer's optimistic state of mind benefits to them. Since producing high quality guarantees to be labeled, it becomes an increasingly more interesting option than not investing as the prior increases. On the contrary, a high prior benefits to unlabeled firms under a strict label. Saving the cost of really producing high quality, which guarantees to get unlabeled, while benefiting from the optimistic consumers' state of mind becomes a more interesting option in these conditions. In conclusion, when consumers hold a high prior, producing high quality is a more interesting option for firms and from a social welfare point of view under a lenient label rather than a strict. And conversely.

# Chapter 1

# Evaluating Corporate Social Responsibility: a Review of the Literature

# English summary

This paper aims to identify the costs and benefits associated with Corporate Social Responsibility (CSR), drawing from the economic literature. More precisely, we first seek to answer the question: is CSR profitable? We review all the mechanisms that have been identified as potential causes for increased profits in the theoretical literature and provide, when possible, existing empirical evidence. Second, we perform a social and environmental evaluation of CSR by focusing on two questions: are firms really implementing the responsible practices that they communicate about? Do firms responsible practices really go beyond mandatory requirements?

We conclude that the more promising mechanism to enhance profits is on the labor market. There also seems to be a little positive effect going through the capital market, but not well understood. On the contrary, improved productivity of environmental inputs appears to be more costly than beneficial. And finally, evidence lack to conclude about the effect on profits of ethical and green consumerism or strategies toward public authorities. As for the socio-environmental impacts of CSR: First, greenwashing appears to be indeed a matter of concern. Even firms acknowledge this reality by implementing strategies to give credibility to their communication and actions (e.g. certification of their sustainable development report by third parties). Second, there only exist a few studies about voluntary programs that show that some of them do spur improved social and environmental performance (e.g. ISO 14001) while others do not (e.g. Climate wise program). Overall, there is a crucial need for more empirical work given the importance of these issues.

# Résumé français

Ce papier a pour objectif d'identifier les coûts et les bénéfices associés à la responsabilité sociale et environnementale de l'entreprise (RSE) en se basant sur la littérature économique. Plus précisément, nous cherchons en premier à répondre à la question: la RSE est-elle profitable? Nous passons en revue tout les mécanismes qui ont été identifiés dans la littérature théorique comme des sources potentielles de profit et fournissons, quand cela est possible, les éléments empiriques existant. Ensuite, nous procédons à une évaluation sociale et environnementale de la RSE en nous concentrant sur deux questions: est-ce que les entreprises mettent réellement en place les pratiques à propos desquelles elles communiquent? Est-ce que les pratiques des entreprises vont vraiment au delà des obligations légales?

Nous concluons que le mécanisme le plus prometteur pour accroître les profits prend place sur le marché du travail. Il semble aussi y avoir un effet positif à travers le marché des capitaux, mais qui n'est pas bien compris. A l'opposé, atteindre une meilleure productivité des facteurs de production environnementaux apparaît plus coûteux que bénéfique. Et finalement, les éléments empiriques manquent pour conclure sur l'effet sur les profits du consumérisme vert et éthique ainsi que des stratégies dirigées vers les autorités publiques. Concernant les impacts socio-environnementaux de la RSE: tout d'abord, il semble que le greenwashing est un vrai sujet de préoccupations. Les entreprises elles-mêmes reconnaissent cette réalité car elles adoptent des stratégies pour donner plus de crédibilité à leur communication et actions (e.g. certification de leur rapport de développement durable par des parties tierces). Ensuite, il existe seulement quelques études à propos de programmes volontaires. Elles montrent que certains d'entre eux entraînent en effet une performance sociale et environnementale supérieure (e.g. ISO 14001) alors que ce n'est pas le cas pour d'autres (e.g. Climate Wise Program). Globalement, il y a un besoin crucial de recherche empirique additionnelle étant donnée l'importance de ces problématiques.

# 1 Introduction

In the last decades, public awareness of environmental and social issues has developed a lot and the concept of sustainable development has become familiar. Corporate Social Responsibility (CSR) is firm's interpretation of this concept and is now widespread. Firms talk about it: An increasing number of them present annual sustainable development reports. Public authorities talk about it: Some countries have set mandatory standards about firms' social and environmental reporting (e.g. France), others are considering it. Non Governmental Organizations talk about it: A growing number of them investigate firms' sustainable programs and/or make precise requests to firms (e.g. WWF). But also, academics talk about it: Ambec and Lanoie (2007) or Kitzmueller and Shimshack (2012) provide reviews of the economic literature about CSR.

There are indeed a lot of interrogations about CSR: What are the rationales behind firms' responsible behavior? Does CSR have an impact on firms' financial performance? Do firms claiming to be socially or environmentally efficient are trustworthy? Do CSR practices bring additional environmental and social results compared to mandatory requirements? Should public authorities intervene to promote CSR and how? etc.

This paper aims to contribute to this debate by identifying the costs and benefits associated with CSR, drawing from the economic literature. More precisely, we first seek to answer the question: is CSR profitable? We review all the mechanisms that have been identified as potential causes for increased profits in the theoretical literature and provide, when possible, existing empirical evidence. Second, we perform a social and environmental evaluation of CSR by focusing on two questions: are firms really implementing the responsible practices that they communicate about? Do firms responsible practices really go beyond mandatory requirements?

The remainder of this paper will be organized as follows. The second section presents the evaluation of CSR profitability while the third proceeds with the socio-environmental

evaluation. The final section concludes.

# 2 Is Corporate Social Responsibility profitable?

By assessing CSR profitability, we take the point of view of the firm and wonder whether the additional costs associated with the adoption of CSR practices are outweighed by additional benefits. This is an important question: if the net impact of CSR on firms' profits is positive, it is economically rational for firms to invest in such programs. In such a case, Friedman (1970)'s argument against CSR, namely that manager's responsibility is to act in the interest of shareholders and therefore prioritize profitability over CSR, would prove invalid.

In the following section, we assess whether CSR has an overall impact on profits. In the next, We list the mechanisms through which CSR could impact positively firms' profits and review theoretical arguments and empirical evidence about their impact on profits for each of them.

## 2.1 General studies

Before investigating mechanism by mechanism, we consider the articles that directly study the link between the level of CSR and corporate financial performance (CFP). The literature on this point is abundant and there exist many meta-analysis (Roman, Hayibor and Agle, 1999; Orlitsky, Schmidt and Rynes, 2003; Ambec and Lanoie, 2007; Wu, 2006). Outstanding by its scope and recent, the meta-analysis offered by Margolis et al. (2009) covers 251 studies over a period of 35 years and highlights the lack of strong consensus in the results: 59% find a non significant relationship, 28% a positive and 2% a negative.[1] Overall, they find that the effect of CSR on financial performance is positive but

[1] An additional 13% did not report sample size, so it was not possible to test for significance.

small (mean $r = 0.11$, median $r = 0.06$ and weighted mean $r = 0.09$) which indicates that SR firms indeed have, even if little, higher profits. Note that close to all studies encompassed in this meta-analysis use data from the United States or the United Kingdom, consequently, these results may not apply other parts of the word. For example, Ambec and Lanoie (2007), whose meta-analysis cover 82 studies, find that the positive impact of CSR on CFP is stronger in the United Kingdom than in the United States,[2] which confirms that results on this question may not be directly transferable from one country to another. There is thus a need for research in other regional areas of the world.

An important issue arises in all these studies: the definition of the key variables at stake, CSR and CFP. For example, in their meta-analysis, Margolis et al. identify up to nine types of measurement for CSR: Charitable contributions,[3] Corporate policies,[4] Environmental performance,[5] Revealed misdeeds,[6] Transparency,[7] Self-reported social performance,[8] Observers' perception,[9] Third party audits[10] and screened mutual funds.[11]

As for CFP, they single out two broad categories: accounting-based measures of financial returns (e.g., Return on Assets, Return on Equity) versus market-based measures of financial value (e.g., stock returns, market/book value ratio).

The former category reflects the present performance of the firm, which can suffer from accounting manipulation due to pragmatic constraints, and the latter reflect the

[2] All other things equal, the dummy variable for the US is statistically significant and negative. There are seven studies using UK data and one using canadian data, most of them show a statistically significant and very positive effect. This means that US studies indicate more often a negative effect than those studies.

[3] Donations to or establishment of a philanthropic fundation.

[4] Adoption of a given ethical stance, e.g. firms that stopped dealing with South Africa to protest apartheid or banks that offer low income loans.

[5] Mesures of some impact of the firm on the environment, either self reported or objective.

[6] Public announcements that reveals an unsocially responsible behavior, e.g. verdicts in lawsuits.

[7] Release of information by a company in publicly available documents, e.g. such as annual reports.

[8] Firms' assessment of their behavior when asked by a researcher or media outlet.

[9] Intuitive perception of observers such industry insiders, executives at other companies, business school faculty members, or undergraduate business students.

[10] Systematic assessment of data by investigators who evaluate a company along a set of criteria, e.g. KLD or Vigeo ratings.

[11] Best vs Worst-in-class approach as presented above.

stock market agents' vision of future profits, which can be misguided. We deem that market-based measures indicate the attractiveness of the firm on the capital market.

We adopt here the point of view of economists and thus consider only accounting-based measures of the profits to answer the question of the CSP-CFP link: on the subgroup of 109 studies that use accounting-based measures, Margolis et al. still find that it is positive but small (mean r=0.15, median r=0.11 and weighted mean r=0.12), but it is slightly stronger.

## 2.2 Studies on specific profit-enhancing mechanisms

We review the different rationales that authors have put forward to explain firms' adoption of CSR strategies: benefits from consumers, benefits on the labor market, benefits due to a more efficient use of environmental inputs, benefits on the capital market, or with regard to governmental bodies.

### 2.2.1 Benefits from consumers: ethical & green consumerism

**Mechanism**

It is now established that some consumers have a higher willingness to pay for products issued from socially and environmentally friendly processes. This creates an opportunity for firms to innovate and access new markets. As early as 1995, Arora and Gangopadhyay proposed a standard model of vertical differentiation where some firms capture the higher willingness-to-pay for environmental quality of part of the consumers by overcomplying with environmental regulation.

One firm decides to produce responsibly if the impact on its profits outweighs the cost of self-regulation. In order to estimate this, there are two key parameters that need to be taken into account:

- **Potential market share**: It depends on the number of consumers who are actually willing to pay higher prices for a responsible product and the intensity of competition.

- **Margin:** While sale price is expected to increase, production costs may well be up too. The key parameter is thus the difference between price and costs for the product responsible version.

**Empirical evidence**

Of the three parameters (demand size, sale price, production cost) that need to be investigated to assess whether it is interesting to self-regulate and produce a responsible product, only the price has been formally studied in the literature.

There exist numerous studies (we listed 42 of them[12]), and they all evidence a positive price premium for responsible products. When percentages of increase with regard to base prices are provided, they always range in dozens of % (from 10% up to 118%[13]). To better understand these results it is important to consider the method used to assess these price premiums and the nature of the benefit brought by the responsible version of the product.

In economics, there exist two broad types of methods to study agents' preferences, and in this case their willingness to pay for a responsible product:

- **Stated preferences methods**: They consist in asking consumers how much would they pay for a product responsible version, but they do not actually purchase the

[12] Henion (1972), Misra et al. (1991), Estes and Smith (1996), Roosen et al. (1998), van Ravenswaay and Blend (1999), Byrnes et al. (1999), Nimon and Beghin (1999), Marymount University (1999), Batley et al. (2000), Gil et al. (2000), Goett et al. (2000), University of Maryland (2000), Batley et al. (2001), Bennett et al. (2001), Elliot and Freeman (2001), Roe et al. (2001), Loureiro and Hine (2002), Teisl et al. (2002), Auger et al. (2003), De Pelsmacker et al. (2003), Loureiro (2003), Zarnikau (2003), Björnel et al. (2004), Nomura and Akai (2004), Prasad et al. (2004), Ek (2005), Loureiro and Lotade (2005), Menges et al. (2005), Parsons (2005), Hiscox and Smyth (2005), Guilloux (2005), Hamschmidt and Dyllick (2006), Borchers et al. (2007), Kotchen and Moore (2007), Delmas and Grant (2008), Hansla et al. (2008), Howard and Allen (2008), Bollino (2009), Yoo and Kwak (2009), Grösche and Schröder (2011), Litvine and Wüstenhagen (2011), Ward et al. (2011)

[13] Estes and Smith (1996) and Loureiro and Hine (2002).

good. Surveys are addressed by mail, email or phone to a representative sample of consumers, containing questions about their willingness to pay for products with different attributes, and also about their socio-economic characteristics. The information collected is then econometrically treated.

- **Revealed preferences methods**: They consist in observing in real life how much consumers pay to buy the product responsible version. There is a higher variety of methods here. For example, one possibility is for authors to directly observe prices or market shares. But they can also set up real life experiments: e.g. ask a store to start selling in two displays the exact same athletic socks but one being certified "sweatshop free" and observe the impact on consumers purchase behavior.

The products studied in the literature bring different types of benefit to consumers. On the one hand, there are responsible goods that bring **private benefit** to the consumer. It consists mainly of food that is perceived as healthier or with better organoleptic characteristics. Typical examples are organic, pesticide free, non Genetically Modified Organisms agricultural products. But devices such as energy efficient equipment are also of this type. On the other hand, there are responsible goods that bring **public benefit** to the whole society: production processes guarantee the provision of some public good or advantages to the workers but there is not direct impact on the final consumer well-being. Typical examples are fair trade, eco-label, good-working-conditions certification, renewable energy. One would expect higher price premiums for the former category. Figure 2 provides examples of studies and results with different types of methods and benefits.

The literature clearly evidences the existence of positive price premiums for some responsible goods. Nevertheless, this does not mean that green consumerism is a strong driver of CSR investment. Firstly, two limits need to be kept in mind: Most studies use stated preferences, and it has been extensively proven that people overestimate their willingness to pay when they do not have to actually spend any money. The sizes of the

| Authors | Method | Product | Type of benefit | Premium |
|---|---|---|---|---|
| Nimon and Beghin (1999) | Hedonic pricing | Cotton | Organic | Price premium of 33.8% of the apparel price expressed in dollars per unit of fiber content. |
| Loureiro and Hine (2002) | Survey | Potatoes | Organic & local | Potential premium of about 9.37 cents per pound over the initial price of $1 per pound-or about a 10% premium. |
| Teisl et al. (2002) | Evaluation of market share | Canned tuna | Dolphin free | 1% increase in market share of tuna after introduction of the dolphin free label which corresponds to a WTP from 0.005 to 0.0125 USD per month for the label. |
| Prasad (2004) | Real life experiment | Athletic socks | Sweatshop free | One in four consumers was willing to pay up to 40% more for the labeled socks. |
| Loureiro and Lotade (2005) | Contingent Valuation | Coffee | Fair trade | Price premium of about 21.64 cents/lb over the original price. |
| Ward et al (2011) | Survey | Refrigerators | Energy efficient | Consumers are, on average, willing to pay an extra $249.82–$349.30 for an ENERGY STAR labelled refrigerator. |

Figure 1: **Examples of studies evidencing a price premium for ethical or green products.**

premiums are thus probably smaller in reality. In addition, information plays a key role: several of these studies underline that the WTP increases with the amount of information the consumer knows about the product's responsible characteristics. Hence one limit to this business strategy is that informing consumers is arguably costly and may impact negatively the margin.

Secondly, prices are only one side of the margin and even if they go up for green or ethical products, it is also important to assess the increase in cost necessary to produce these upgraded products versions, and the potential market for such products. To our knowledge there exists no specific study in the economic literature about these two aspects of ethical and green consumerism. But for example, think of the organic market. Things are clear on the price side: organic agricultural products prices are clearly higher compared to regular agricultural products. But on the costs side, it is not clear whether they are more or less important: while fewer inputs are needed, which induces a decrease in costs, transition costs from conventional to organic farming are important and organic yields are lower than conventional.[14] Finally on the demand size, it is important to remember that organically cultivated surface represents only 0,85% of global cultivated surface in 2010[15] (from the 161 countries included in the survey), suggesting that, despite years of important growth (e.g. 8% in value in 2010), demand for organic food remains limited. To sum up, organic food may be a real opportunity for green consumerism but it is also a limited one.

In conclusion, while empirical research focuses on evidencing a price premium and offers promising results, there is a strong lack of investigation on costs and demands size. No fine analysis of ethical and green consumerism opportunities can be performed without such additional information.

[14] http://www.sare.org/publications/organic/organic07.htm
[15] The World of Organic Agriculture 2012, http://www.organic-world.net/yearbook-2012.html?&L=0

### 2.2.2 Benefits from consumers: boycotts

#### Mechanism

Consumers may also have an impact on firms' profits by boycotting firms that they deem to behave in an irresponsible way. In reaction, firms may self regulate in order to avoid to be targeted (Lyon and Maxwell, 2008). In such a framework, self-regulation becomes more of a risk insurance strategy.

Boycotts usually happen following the coverage in the media of a negative event concerning one firm and at the instigation of NGOs. It is important to note that not all firms are exposed to this risk and will not suffer from it with the same intensity. Bigger firms and firms belonging to critical industrial sector where NGOs scrutiny is particularly acute (e.g. the mining industry) will probably be more prone to undergo a boycott. Firms whom reputation and brand highly matter to their business (e.g. sportswear industry) are likely to suffer the most from boycott.

#### Empirical evidence

There exist few studies of the financial impact of boycotts on firms (we denoted 9). All of them use data from the United-States and only a few (3[16] out of 9) consider post 2000 boycotts. Results are not clear-cut. Two studies[17] count the number of successful boycotts, in the sense that firms end up complying with the boycotters' requests, in a sample: they find respectively 26.7% (sample of 90) and 44% (sample of 1092). The others[18] investigate the financial harm undergone by the targeted firms: four find a negative impact, another also finds a negative impact on stocks' value but underlines that it is only short term and that it quickly disappears, and finally a last one[19] evidences a positive impact on

[16] Lenox and Eesley (2009), Koku (2012), Vasi and King (2012).
[17] Friedman (1985), Lenox and Eesley (2009).
[18] Pruitt and Friedman (1986), Pruitt, Wei and White (1988), King and Soule (2007), Bartley and Child (2011), Koku (2012), Vasi and King (2012).
[19] Koku et al. (1997).

firms' financial value. The latter result is explained by the authors as a consequence of the efficiency of firms' counteractions.

Overall this suggests that undergoing a boycott is not completely neutral for firms, but we lack finer understanding of the phenomenon to conclude about the strength of this mechanism with regard to driving CSR investment. Especially, since boycotts are usually organized by NGOs and the Internet has significantly lowered the internal and external communication costs of NGOs (Lyon and Maxwell, 2008), boycotts' extent and impact have probably evolved a lot in the last decade and studies with recent data would probably shed a more accurate light.

### 2.2.3 Benefits on the labor market

**Mechanism**

On the labor market, a SR firm offers the opportunity to work in a professional environment in congruence with one's moral values or that boosts self-esteem, which can be valuable from two points of view:

- With regard to employees: they may accept lower wages, work harder or extra hours for free, be more loyal to the firm and thus reduce employee turnover, etc. (Frank, 1996).
- With regard to potential employees: SR can help to attract morally motived employees who will accept lower wages or will exert higher effort than in a business-as-usual firm (Brekke and Nyborg, 2008).

In addition, Chief Executive Officers (CEOs) are employees who, thanks to their status, can play key roles in firms' CSR strategies. Aside from all the reasons to promote social behavior exposed in this section that can convince a CEO, two others are put forward by the literature. The first one is that CEO may do so out of personal preferences.

Such a stance is heavily condemned by Friedman (1970) who considers that such a manager is deviating shareholders money to illegitimately take upon herself governmental functions while her only responsibility as a manager is to act in the shareholders' interest, namely to maximize profitability. The second one is due to the entrenchment strategy: inefficient CEOs may invest in stakeholders protection in order to gain activists' support and thus secure their position (Cespa and Cestone, 2007).

**Empirical evidence**

This mechanism started to receive more attention only a decade or so ago, so the literature is sparse (we listed 11 studies). Overall, the results confirm the existence of these mechanisms (see Figure 3), but precise quantifications of the financial size of this effect lack. One noticeable exception is a study by Nyborg and Zhang (2011) that shows a negative, substantial and statistically significant association between wage and CSR among Norwegian firms, with decrease in wages ranging from 20% to 42%.[20] Another attempt, by Lanfranchi and Pekovic (2011), evidences in their sample of 7700 French firms that the sub group of "green workers", identified as such because they have joined a sport, cultural or social association existing in their firm, are significantly more likely to work uncompensated extra hours.

Similarly to the green consumerism mechanism, a "market size" factor must be taken into account alongside with the individual effect size. Not only do we need to understand how much firms gain from workers who value SR, but also what type of workers do value SR and how many are they. For example, Albinger and Freeman (2000) empirically find that it only concerns highly qualified employees. They separate the actual or potential job seekers that they study in three groups depending on their high, medium or low chances to find a job. They find that only the high and medium groups are significantly more

[20]The authors use a CSR score for firms that range form 0 to 1. The percentages given are the wage decreased observed for firms with the higher CSR score, for two different specifications of their econometric model.

| Studies | Method | Results |
|---|---|---|
| Nyborg and Zhang (2011) | Econometric estimation | Negative, substantial and statistically significant association between wage and CSR among Norwegian firms. |
| Lanfranchi and Pekovic (2011) | Survey | Some workers are significantly more likely to work uncompensated extra hours. |
| Turban and Greening (1997), Albinger and Freeman (2000), Greening and Turban (2002) | Survey | Firms with a high corporate social performance are more attractive to workers. |
| Ali et al. (2000), Brammer et al. (2007), Rego et al. (2007) | Survey | Corporate social performance is positively correlated with employees' organizational commitment (individual's psychological attachment to the organization), which is known to be positively correlated with job performance and negatively with job turnover. |
| Frank (2003), Montgomery and Ramus (2003) | Survey | Students (Cornell University graduates and MBA's from business schools) are willing to forego financial benefits in order to work for socially responsible firms. |
| Koppel and Regner (2012) | Laboratory experiment | Employees exert more effort the more the firm donates to a charity. |

Figure 2: **Studies evidencing advantages for self-regulating firms on the labor market.**

attracted to self-regulating firms.

In conclusion, there seems to be a real profit opportunity for self regulating firms on the labor market, but its magnitude for a given corporation highly depends on the distribution of its workers' ethical and environmental values. In particular, empirical evidence suggests that this would probably be a stronger opportunity for firms whose workforce encompasses an important proportion of highly skilled workers.

### 2.2.4 Technical benefits: productivity of environmental inputs

**Mechanism**

Improving environmental performance may entail efforts to reduce the use of some environmentally-critical inputs, such as energy and raw materials. This is the case for energy or water saving programs for example.

Such policies can result in an enhanced productivity. Indeed, total factors productivity (TFP) is defined as the following ratio:

$$TFP = \frac{Y}{X_1 + X_2 + \cdots + X_n}$$

With $Y$ being total production and $(X_1, X_2, \cdots, X_n)$ being the factors of production. For the same level of production $Y$, an increase in environmental performance may mean a decrease in the amount needed of some factors of production that are environmentally related, leading to a higher productivity.

Improving environmental performance entails changes in the production process that can imply research and development, new technologies, implementation of new methodologies, etc. All these changes are undoubtedly costly, which means that there is a short term negative impact on profit. The key question is thus to assess whether long term benefits due to an increased productivity exist and if so, if they outweigh the initial and costly investment.

**Empirical evidence**

We draw our evidence from 25 studies of two types: A first type, investigates the link between pollution control expenditures (PCEs) and productivity. Since it is reasonable to assume that higher PCEs entail higher environmental performance, their results are directly applicable here. A second type of studies analyses the impact of regulation stringency, measured by frequency of inspection, date of introduction of new laws, norms stringency, etc., on productivity. Since environmental regulation is bound to induce enhanced environmental performance, we consider the results of these studies to be relevant to our issue even though in this case changes are not due to firm voluntary efforts.

We have three types of results:

- Most studies[21] focus on the direct effect of PCE or regulation on productivity and evidence a negative trend, the declines in TFP measured ranging from 1% to 30%.
- Three papers proceed in two steps: they first find a positive impact of PCE on R&D and then a positive impact of R&D on productivity.[22]
- Finally, three papers that also study direct effect of PCE or regulation on productivity find an overall long term positive effect but short term negative ones.[23]

These last results confirm the intuition that, in the short term, investments and changes in process are needed and costly while in the long term positive effects on productivity due to learning by searching (investment in research & development (R&D)) and/or learning by doing can be expected.

[21] Christainsen and Haveman (1981), Gollop and Roberts (1983), Gray (1987), Barbera and McConnell (1990), Gray and Shadbegian (1983), Conrad and Wastl (1995), Jaffe (1995), Dufour et al. (1998), Gray and Shadbegian (1998), Boyd and McClelland (1999), Gray and Shadbegian (2003), Barla and Perelman (2005), Shadbegian and Gray (2005), Färe et al. (2005), Managi et al. (2006), Lee (2007), Lanoie et al. (2011).

[22] Hamamoto (2006), Chih-hai et al. (2011), Lanoie et al. (2011). Numerous authors have limited their scope to the second step and they all find a positive link between R&D and productivity, which is good news about R&D efficiency.

[23] Berman and Bui (2001), Lanoie et al. (2005), van der Vlist et al. (2007).

Now, as highlighted before, the key issue remains to determine which effect is predominant and under which conditions: even though long term effects may be positive in some cases, they may not always be large enough to compensate for negative short term ones.

It is not easy to draw clear conclusions about the net result from the literature. Indeed, it is difficult to determine the relevant time frame to measure this net effect: how much time is needed for positive effects to overcome negative ones, if it ever happens? What is acceptable from a business point of view? Answering these questions is key to infer the profitability of this mechanism.

Further research, clearly separating short and long term effects, would prove handful. In addition, results could be impacted if the way to measure environmental performance is designed outside the box of environmental regulation. For example, one could use the environmental notations provided by extra-financial rating agencies.

Nevertheless, we can conclude that so far there are little evidence supporting the idea that this is a profitable mechanism.

### 2.2.5 Benefits on the capital market

**Mechanism**

There is an ambiguity about the nature of the effect of CSR on capital market with regard to other mechanisms, that is due to the fact that there exist two types of investors:

- Ethical and green investors that are ready to reward responsible firms because they suit better their personal values. This type of investors' preferences has given birth to socially responsible investment (SRI) that seeks to take into account both extra-financial and financial criteria. An example of SRI is sector approaches where investors exclude all firms from a sector considered harmful (tobacco, alcohol, pornog-

raphy, etc.). In this case, the effect of CSR on capital market is an independent mechanism to increase profits.

- Regular investors may favor self-regulating firms because they observe that they perform financially better than others, thanks to one or several of the mechanisms exposed here. In such a case, the effect of CSR on capital market is merely an amplification of the other mechanisms that increase profits.

Of course, it is possible that the two causes add up. Nevertheless, it is important to keep in mind that if there are not a lot of ethical and green investors, beneficial effects of CSR on the capital market are strictly dependent to the efficiency of the other mechanisms. Now let us study more precisely how CSR can impact profits through the capital market.

Regularly, firm need funds to invest. Outside financing is costly and CSR may mitigate this cost. Namely, firms have two main channels to raise capital:

**Debt**: Firms can borrow money from a bank. When they negotiate a loan, there are three key elements to discuss:

- The interest rate, which represents a direct economic cost paid for bank's services. The lower the interest rate, the lower the debt value.
- The maturity date, or final payment date, which is closely linked to the interest rate.
- The loan covenants that are conditions that the borrower must comply with in order to adhere to the terms in the loan agreement. If the borrower does not act in accordance with the covenants, the loan can be considered in default and the lender has the right to demand payment (usually in full).

CSR can impact the interest rate both directly and indirectly. First, the bank assesses an interest rate according to a given grid of criteria that intend to measure the likelihood

that firms won't default on their loan. Banks may consider that SR firms develop a better long term apprehension of their business and may better resist changes (new legislation, financial crisis, etc.) in the framework where they evolve, thus decreasing the firm's exposure to risks. CSR, by decreasing the default's risk, could play in favor of a smaller interest rate. Similarly, CSR may be interpreted as a sign of business stability over time and consequently drive the bank to propose longer maturities,[24] which in turn impacts negatively the interest rate.

As for loan covenants, it is not straightforward to precisely anticipate how CSR can impact them and how this would affect firm's profit. For example, one can hypothesize that having a convincing CSR strategy could be a necessary condition to borrow money from some banks (e.g. the NEF) or some banks' affiliate specialized in responsible financing.

**Increased equity:**

How investors' preferences impact firms' cost of capital differs depending on whether the firm belongs to a stock market or not. In the latter case, firms privately negotiate with their investors. Here, CSR may play a role on two decisions: the investor's decision to invest and, if applicable, the amount that the investor is willing to invest. In the case of firms listed on a stock exchange market, increased equity works through the selling of stocks. The firm can either sell existing stocks that were in its possession (non-dilutive increase in equity) or create and sell new stocks (dilutive increase in equity). CSR can impact this process on two levels. First, stocks' value matters. If it is higher thanks to CSR, firms earn more when selling their existing stocks. Or, since new stocks are sold at a value close to the market value, firms need to issue less stocks to raise the same amount of money, hence mitigating the dilutive effect. Second, a responsible firm can attract ISR

[24] Note that firm's balance between short term and long term debt is a whole area of debate and strategies *per se*, that we do not treat here.

investors, and some of them have long term investment strategies which ensures that they be stable shareholders.

**Empirical evidence**

To our knowledge, no academic study exists about the impact of CSR on the loan conditions (interest rate, maturity and loan covenants) proposed by banks. Research is needed on this point.

With regard to equity, the key questions regarding firms listed on stock exchanges are: do SR firms' have comparatively a higher value on stock exchange? And do firms owned by SRI investors perform better? There exists an important literature on both issues. First, in their meta-analysis, Margolis et al. (2009) compute the effect of CSR on financial performance for the subgroup of 156 studies that use financial market-based measures (e.g., stock returns, market/book value ratio) of financial performance. They find that it is positive but small (mean $r = 0.11$, median $r = 0.06$ and weighted mean $r = 0.09$) which indicates that SR firms are indeed, even if little, advantaged when they raise capital through the selling of stocks. However, this effect is smaller than the one found on the subgroup of studies using accounting based measures of financial performance (e.g., Return on Assets, Return on Equity). This result suggests two possibilities:

1. Greens investors do not weigh a lot on the financial market.

2. Financial actors deem that the positive economic effect of CSR is short termed. Indeed, on a perfect capital market stocks correspond to the financial actors' opinion about the value of the discounted flow of future profits. Thus if they expected CSR to have long term positive impact on profits as important or greater than the present effect, the overall effect of CSR should be greater on the financial market than on present profits.

It is more difficult to assess whether firms owned by SRI investors perform better. It

would necessitate identifying the proportion of stocks owned by SRI investors for each firm and this is not very accessible and tractable data. A good proxy is to look at the performance of SRI funds since it is expected to reflect the economic performance of the companies included. In turn, these companies are included in a SRI fund, thus part of their capital is owned by responsible investors. Margolis et al. (2009) also analyses separately the subgroup of 30 studies that examine the performance of mutual funds that include only companies meeting certain criteria of social performance. They find a positive effect but, again, small (mean $r = 0.032$, median $r = 0.026$). This effect is smaller than the one found on stock markets for responsible firms at large. This is surprising and would suggest that being part of a SRI fund actually decrease financial performance.

Lastly, there exists close to no literature about capital costs for firms that do not belong to a stock market. One noticeable exception is the work of Crifo et al. (2012). They conducted an experiment involving 33 professional Private Equity investors (including both venture capital and buyouts specialists). Investors participated in closed auctions in which they competed to acquire fictive firms based on case studies carefully built to ensure realism and credential context. Investors were provided with accounting and financial information, together with non-financial information. The authors' results highlight that non-financial information matters for equity financing, the main finding being the existence of an asymmetrical impact of corporate non-financial performance. Whereas good environmental and social practices respectively increased firm value by about +5%, bad practices devaluated them by about -10%. Bad governance practices decrease firm value by -15% but there is no significant effect of good governance practices.

In conclusion, it seems that investors are not indifferent to firms social and environmental behaviors and, more precisely, that they value it. Nevertheless, overall they apparently value it little and the impact on firms profits is not striking. A finer understanding of the mechanisms on the capital market and the conditions under which they can be

profit enhancing is needed.

### 2.2.6 Benefits from governmental bodies

**Mechanism**

CSR is sometimes presented as an alternative to regulation. Authors consider that it can impact the legislative process in different ways, depending on the stage on the public policy life cycle (Baron, 2006):

- During the **politicization stage**, the issue acquires a label, opinion leaders begin to discuss the problem in public, the news media become more active in covering the issue, and interest groups begin to mobilize around the issue (Lyon and Maxwell, 2008). Firms can preempt law by self-regulating, and so doing avoid costly regulation. For example, Fleckinger and Glachant (2011) show that in some circumstances CSR can harm the total welfare by preempting regulation.

- During the **legislative stage**, when it has already been decided to legislate but details are still debated, leading and efficient firms can invest a lot in self-regulation to signal to the government that it is not very costly in an attempt to increase the level of the mandatory standard and thus raise their rivals' costs. For example, Denicolò (2008) shows that a firm's voluntary adoption of a clean technology can signal to the regulator that the cost of adoption is low. Consequently, the regulator, in balancing profits, consumer welfare, and environmental externalities, may find it socially desirable to mandate the adoption of the clean technology.

  Similarly but to opposite purposes, laggards and inefficient firms can invest in self-regulation to signal to the government that it is very costly and it is the most they can do, in an attempt to decrease the level of the mandatory standard. For example, Heyes (2005) states that, when firms find themselves in an environment

where regulation is forthcoming and firms have private information about compliance costs, then high-cost firms can use voluntary self-regulatory behavior in the pre-regulatory phase to credibly signal their type and induce a lower mandatory standard.

- Finally, during the **implementation stage**, once new legislation has been adopted and public authorities enforce it, firms can invest in CSR to deflect monitoring. For example, Maxwell and Decker (2006) show that if a firm voluntarily makes an observable investment in pollution control that lowers its marginal cost of abatement, then it is optimal for the regulator to monitor this firm less frequently.

The profitability of this mechanism is more difficult to assess, because here we do not compare a scenario where CSR increases profit compared to a business-as-usual scenario. Rather, we deal with a hypothetical scenario of business as usual that would trigger new or tougher legislation that could negatively impact profits, and here CSR is a way to prevent this scenario from happening. Since this scenario remains hypothetical, there is no baseline available to compare self-regulating firms profits to.

**Empirical evidence**

It is thus very difficult to assess empirically the financial impact on firms of such strategies and consistently, to our knowledge there exists no attempt to do so. Maxwell et al. (2000) provides evidence of the existence of these mechanisms though: they show that North American firms engage in more voluntary pollution abatement in states that pose a greater threat of regulation. There also exists anecdotal evidence: For example the German Association of the Automotive Industry (Verband der Automobilindustrie or VDA) agreed in 1995 to reduce average CO2 emissions from new German passenger cars by 25% between 1990 and 2005 in order to avoid a law.

Further research on this topic should focus on collective and coordinated self-regulation

actions, such as voluntary programs, since in this case CSR is more a collective good, used in the framework of a lobbying strategy.

## 2.3 The key role of information to trigger higher profits

So far, we have described a number of mechanisms that may allow firms to financially benefit from their investment in CSR. The problem is that self-regulation is a credence good: it is very difficult for an individual stakeholder (be it a consumer, an employee, an investor, etc.) to ascertain the veracity of firms' claims. This provides a rationale for the existence of specialized agents that have sufficient resources to observe firms' self-regulatory activities and to convey this information to stakeholders.

These specialized actors can be extra-financial rating agencies, Non-Governmental Organizations, labeling institutions, public authorities, etc. Extra-financial rating agencies, such as KLD or Vigeo, provide neutral information about firms behaviors through the attribution of scores. Access to this information is usually costly.

NGOs usually convey information to public for free but with a bias. Some NGOs are specialized in the transmission of good news: they certify that a firm – or a product – is socially or environmentally responsible. An illustration is the Marine Stewardship Council that rewards sustainable fishing by certifying about 12 percent of the world catch intended to human consumption. Other NGOs mostly send bad news: Greenpeace is the emblematic example of an organization that fiercely conducts campaigns against firms that they do not deem to behave responsibly. Lie in between neutral NGOs which send both good and bad news, as for example WWF which alternates between cooperative and offensive behavior toward firms depending on their compliance with their demands.

Labeling institutions can be private or public. In France, for example, the French public authorities launched the Social Responsibility Label[25] in 2005, and in 2011 the asso-

[25]This label is granted by the Association pour la promotion et le développement du label de la respons-

ciation Qualité-France created the LUCIE Label,[26] intended for organisms respecting the new ISO26000 standard, which provides guidelines for being socially responsible.

Finally, it also happens that public authorities give information about firms responsible or irresponsible behavior. An example is the NSW (New South Wale) food authority in Australia and its "Name & Shame" program aiming at listing publicly firms that have failed to follow the NSW quality standards.[27]

NGOs, defined by Baron (2001) as "private" politics actors, are different from the other actors because they do not only convey information. They also have a proactive behavior toward firms. With the rise of the Internet, these organizations have known a surge in their power to pressure firms thanks to a sharp decrease in internal and external communication costs (Lyon and Maxwell, 2008). A pioneering work is Baron and Diermier (2007), who develop a theory of adversarial NGO campaigns where an NGO makes a demand to a firm, along with the promise of a reward (public endorsement) in case of compliance and a threat of punishment (negative propaganda to encourage boycott) otherwise. Sinclair-Desgagné and Gozlan (2003) show that when an NGO wields a big threat, it can induce green firms to distinguish themselves by issuing a detailed CSR report; whereas if weak, they release only moderately informed CSR reports as other firms do.

It is important to keep in mind that all these actors are essential to the functioning of all the profit mechanisms described above. Without these organizations, stakeholders are not able to reward firms' self-regulation efforts and CSR cannot be a profit opportunity.

abilité sociale: http://www.alrs.asso.fr/

[26]This label is granted by the agency LUCIE, in partnership with the extra-financial rating agency VIGEO and the labeling institution AFNOR Certification: http://www.labellucie.com/

[27]http://www.foodauthority.nsw.gov.au/aboutus/offences/

# 3 What are the environmental and social impacts of Corporate Social Responsibility?

Now that we have assessed CSR from the firm's point of view by evaluating its profitability, we adopt a social welfare point of view. Our aim here is to contribute to the debate about CSR social desirability by considering the two following questions:

**1)** Are firms really implementing the responsible practices that they communicate about?

In order to assess whether CSR programs bring additional socio-environmental results or not, there is one first basic criteria that must be met: there must be real. Indeed, many firms have been criticized for being merely greenwashing, or "the act of misleading consumers regarding the environmental practices of a company or the environmental benefits of a product or service".[28] We begin our evaluation by addressing this issue in section 4.1.

**2)** Do firms responsible practices really go beyond mandatory requirements?

Then, remains the need to check the additionality with regard to what is already legally required. This is far from being easy or trifling. To start with, one can sometimes wonder whether CSR does not merely fall down to increased compliance with existing laws. Indeed, it is easy to find examples of firms that stress in their sustainable development report that they respect the laws, see Figure 1.[29] While no conclusion can be drawn from such anecdotal evidence, it is interesting to note that some firms consider that respecting the laws is an effort worth underlying,[30] but also that it is only one aspect of the

[28] Definition of Greenpeace.

[29] Data retrieved online on August 7, 2012 at
http://www.export.findusgroup.com/filearchive/4/4328/10principles_findusgroup.pdf;
http://www.thecoca-colacompany.com/sustainabilityreport/TCCC_2010_2011_Sustainability_Report_Full.pdf;
http://www.hsbc.com/1/PA_esf-ca-app-content/content/assets/sustainability/120525_sustainability_report_2011.pdf;
http://www.riotinto.com/annualreport2011/performance/sustainable_development/people.html

[30] A counterexample is provided by RWE (German electric power and natural gas public utility) that only mentions in its sustainable development report 2011 that "For us and our stakeholders it goes without

| Firm | Main area of activity | Source | Quotations |
|---|---|---|---|
| Findus | Frozen food | Ten Principles for Responsible Fish Procurement; Second Edition | "1. Legality<br>We take all reasonable precautions to ensure that we do not purchase fish which has been caught, landed or farmed illegally. " |
| The Coca-Cola Company | Non-alcoholic beverages | 2010/2011 Sustainability Report | "Compliance with local work-hours and overtime laws is a fundamental component of our Workplace Rights Policy and Supplier Guiding Principles."<br>"Doing business with integrity means avoiding bribery or corruption in any form. It also means complying with the anti-corruption laws of the countries where we operate." |
| HSBC | Banking | 2011 Sustainability Report | "In 2011, we established a Group Bribery Committee at executive level to ensure compliance with the letter and spirit of all laws covering bribery and corruption and, in particular, the UK Bribery Act which is extraterritorial in nature." |
| Rio Tinto | Mining and resources | Annual report 2011 | "We offer our employees a rounded total rewards package, the principles of which are consistent across the Group, designed to be competitive, in compliance with all applicable laws and regulations, and appropriately balanced in favour of variable pay linked to performance." |

Figure 3: **Examples of firms that mention that they respect the law in their Sustainability reports; *Source:* Firms' institutionnal websites.**

CSR practices they list. This suggests that some other practices may be indeed additional with regard to mandatory requirements.

This is a key aspect of the whole evaluation of CSR. Responsible firms should not be rewarded while others are not if their social and environmental impact is not different. We present the few existing attempts to assess this difference in section 4.2.

saying that we are complying with the relevant environmental regulations and the statutory legislation."

## 3.1 Credibility of firms communication

Firms do not only develop socially desirable activities. They also tell they do so. Almost all corporate websites or annual reports now include a description of the company's CSR policy. For instance, Bazillier and Vauday (2009) report that 406 firms out of the 593 of their sample have released at least one sustainable development report. Indeed, if some stakeholders prefer responsible firms and have thus a higher willingness to pay to support such firms, these firms need to communicate about their CSR strategies in order to signal themselves and ripe these benefits. Nevertheless, as highlighted before, the truthfulness of firms' virtue claims is often impossible or at least difficult to ascertain for an individual stakeholder with limited resources and there exists an incentive for firms to manipulate their communication.

For this reason some stakeholders view corporate communication about CSR with suspicion and denounce greenwashing. Such cases of communication abuse are growingly highlighted, particularly by NGOs. For example, the French branch of the NGO Friends of the Earth organizes annually the Pinocchio Sustainable Development Awards, which intends to "illustrate and denounce the negative impacts of some French companies that behave in total contradiction with the concepts of sustainable development that they boast of extensively". In 2009, EDF "won" the award in the category greenwashing for spending in the same year 8,9 M€[31] for sustainable energies R&D and 10 M€[32] in an advertisement campaign praising their "ambitious" program in sustainable energies R&D.

**Empirical evidence**

It is difficult to track greenwashing on a wide and systematic scale because of limited data on real implementation of CSR practices. Indeed, to prove that a firm is greenwash-

[31] According to their 2008 sustainable development report.
[32] According to computations performed by the magazine TerraEco.

ing, one must first list the firm's CSR claims and then estimate their real implementation to see if there are discrepancies.

On the academic side, one attempt is due to Ramus and Montiel (2005) who use as proxy for real implementation employees' observations about their workplace, drawn from a database containing statements about their workplace from 586 non-management employees of 10 large European environmentally proactive firms. As for firm's CSR claims, they use a second database that gathers the published corporate environmental policy statements of 188 large companies from 20 countries all over the world. They find that commitment to specific environmental policies does not vary greatly amongst industry sectors but implementation does. For example, oil and gas companies, although just as likely to commit to specific policies of sustainable development as the other sectors, were not at all likely to commit to or implement a fossil fuel use reduction policy.

Interesting anecdotal evidence about greenwashing and firms' strategies to be credible can also be drawn from non academic studies:

TerraChoice, an environmental marketing and consulting firm, provides what is, to our knowledge, the only systematic study of firms' greenwashing marketing behaviors. They issued three versions (2007, 2009, 2010) of their report "Sins of Greenwashing". Each time, TerraChoice researchers go to retailers in North America, the United Kingdom, and Australia to record every product whose packages bear environmental claims. There were 4,744 such products in 2010. For each product, the researchers record product details, claim(s) details, any supporting information, and any explanatory detail or offers of additional information. This allowed them to finally define seven sins of Greenwashing, see Figure 4 for their definitions, and count how many times each sin was committed in their sample of environmental claims. They find that 95% of the "greener" products (products that claim to be "green") commit one or more of the seven sins of greenwashing. Even if there is a slight improvement over years, with 1% of products being sin-free

in 2008, less than 2% in 2009, and 4,5% in 2010, greenwashing as defined by TerraChoice seems so far to be the rule rather than the exception.

Examples of firms that mention that they respect the law in their Sustainability reports; Source: Firms' institutionnal websites

Detail of the percentages of each sin committed for each year is provided in Figure 5. TerraChoice indicates that part, but not all, of the sharp decline in the sin of hidden trade-off is due to methodology changes in the way of applying its definition. They highlight that much of the increase in the sin of no proof is related to an important corresponding trend in BPA-free[33] claims on toys and baby products. Finally, they underline that the attempt of shaving off greenwashing by using external certifications to support green claims has given birth to a seventh sin, worshiping of false labels, a sin that is so far increasing along with the use of legitimate labels.

Interestingly, "mature" sectors[34] that have experience in green marketing are less prone to greenwashing than "immature" sectors[35] that more recently started to add green claims on their products. They also have a higher rate of legitimate certification (around 29% vs 14%). There are at least two possible explanations to this phenomenon, which can be complementary. First, there may be reasons why firms would rather avoid extensive use of greenwashing, for example a threat of negative impacts on their reputations, which push them to improve their communication credibility. Second, there may be an adaptation period to this new need for credibility: the time necessary for adequate marketing techniques and appearance of relevant and legitimate certification.

Another source of valuable information are CapitalCom's "barometers" of the CAC40

[33]Concerns about the use of bisphenol A in consumer products have been regularly reported in the news media since 2008, after several governments issued reports questioning its safety, making consumers more aware about this issue.

[34]TerraChoice considers the following sectors to be mature : building, construction products, office products, cleaning chemicals and tissue products.

[35]TerraChoice considers the following sectors to be immature : Toys, baby products and consumer electronics.

| Sin | Example |
|---|---|
| **Sin of the hidden trade-off:** committed by suggesting a product is "green" based on an unreasonably narrow set of attributes without paying attention to other important environmental issues. | Paper, for example, is not necessarily environmentally preferable just because it comes from a sustainable-harvested forest. Other important environmental issues in the paper-making process, including energy, greenhouse gas emissions, and water and air pollution, may be equally or more significant. |
| **Sin of no proof:** committed by an environmental claim that cannot be substantiated by easily accessible supporting information or by a reliable third party certification. | Common examples are tissues products that claim various percentages of post-consumer recycled content without providing any evidence. |
| **Sin of vagueness:** committed by every claim that is so poorly defined or broad that its real meaning is likely to be misunderstood by the consumer. | "All-natural" is an example. Arsenic, uranium, mercury and formaldehyde are all naturally occurring, and poisonous. "All-natural" isn't necessarily "green". |
| **Sin of irrelevance:** committed by making an environmental claim that may be truthful but is unimportant or unhelpful for consumers seeking environmentally preferable products. | "CFC-free" is a common example, since it is a frequent claim despite the fact that CFCs are banned by law. |
| **Sin of the lesser of two evils:** committed by claims that may be true within the product category, but that risk distracting the consumer from the greater environmental impacts of the category as a whole. | Organic cigarettes might be an example of this category, as might be fuel-efficient sport-utility vehicles. |
| **Sin of fibbing:** the least frequent Sin, it is committed by making environmental claims that are simply false. | The most common examples were products falsely claiming to be Energy Star certified or registered. |
| **Sin of worshipping false labels:** committed by a product that, through either words or images, gives the impression of third-party endorsement where no such endorsement actually exists; fake labels, in other words. | |

Figure 4: **The seven sins of Greenwashing;** ***Source:*** **TerraChoice.**

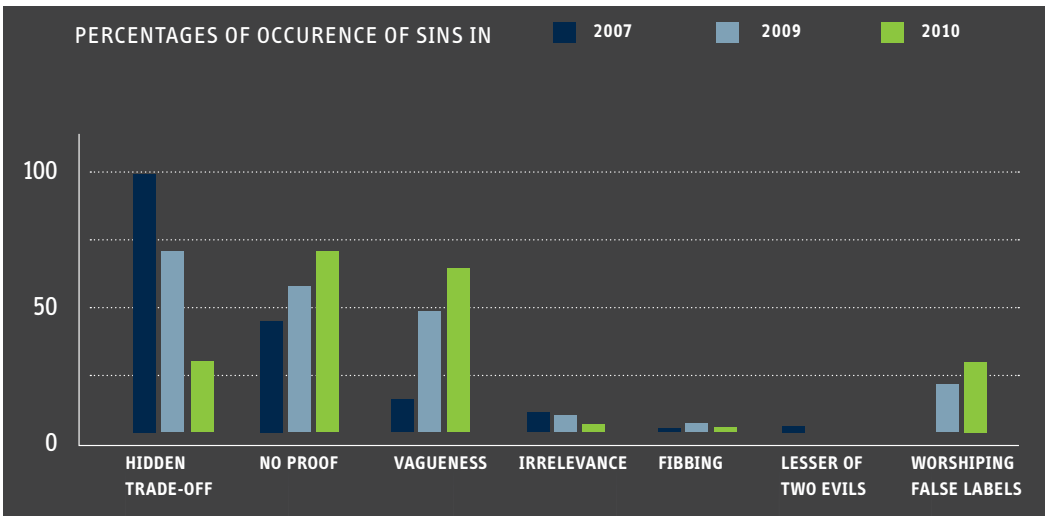



Figure 5: **Percentages of occurence of TerraChoice's sins of greenwashing from 2007 to 2010; *Source:* TerraChoice.**

CSR strategies that provide interesting material to understand firms' strategies to increase their credibility. CapitalCom, a French consulting agency in communication, issued its first analysis of the CSR communication and extra-financial practices of the CAC40 firms in the early 2009 and has since regularly released "barometers" of the CAC40 CSR strategies. In each document they highlight what they deem to have been the most important trends in the last time period in terms of CSR practices but also in CSR reporting habits. Their studies rely on all the data publicly available (annual sustainable report, websites, etc).

From their reports, we can identify four approaches that firms have developed to tackle doubts about their credibility, and that are often combined:

**1) Follow internationally recognized guidelines**

CapitalCom indicates that in 2009, almost half of CAC40 firms are registered at the GRI.[36] In addition, 37 of the firms have declared their support to the Global Compact[37]

[36]The Global Reporting Initiative produces one of the world's most prevalent guidelines for sustainability reporting. It was formed by the United States based non-profits Ceres (formerly the Coalition for Environmentally Responsible Economies) and Tellus Institute, with the support of the United Nations Environment Programme.

[37]The United Nation initiative Global Compact is a principle-based framework for businesses, stating ten

and all of them have answered the Carbon Disclosure Project[38] questions.

These high figures may be interpreted in two ways. Multinationals may consider that international institutions are granted such a moral credit that supporting their standards is compulsory for any multinational. But also, since neither the GRI, nor the Global Compact encompasses mandatory audits, firms can adhere to them without really changing anything in their behavior and they may consider them to be low cost credibility material.

**2) Indicate the results achieved**

In 2009, CapitalCom found that 18 of the CAC40 firms have performed an assessment of their GHGs emissions. This particular result is probably linked to the importance of global warming in the public debate.

In 2011, 12 of CAC40 firms communicated about CSR objectives (from 1 to 3 depending on the firm) that they had failed to meet. Most of them were environmental objectives. If communicating about unmet objectives spreads, those for which no result, good or bad, is provided may end up being considered as pure greenwashing.

**3) Acquire external certifications / integrate SRI indexes**

In 2009, CapitalCom found that CAC40 firms belong to 2,5 SRI indexes in average. The most widespread is ASPI Eurozone[39] (32 firms), followed by Ethibel[40] (26 firms), FTSE4Good[41](24 firms) and DJSI[42] (21 firms). In addition, 21 firms have asked statutory auditors to certify their environmental indicators.

**4) Team up with third parties**

In 2010, CapitalCom found that 25 of the CAC40 firms communicate about subscrip-

universally accepted principles in the area of human rights, labour, environment and anti-corruption.

[38]The Carbon Disclosure Project is an independent not-for-profit organization holding the largest database of primary corporate climate change information in the world.

[39]Advanced Sustainable Performance Indices. It gathers the 120 firms with best Vigeo extrafinancial notation.

[40]This index mixes a best-in-class approach and ethical exclusion criteria.

[41]According to the FTSE website "The FTSE4Good Index Series has been designed to objectively measure the performance of companies that meet globally recognised corporate responsibility standards."

[42]Dow Jones Sustainability Indexes. It includes the 10% of Dow Jones index firms with the best ESG performance.

tion to or partnerships with organizations and major institutes such as the Global Compact, the Comité 21,[43] the ORSE,[44] etc.

While firms team up with third parties for a lot of different types CSR projects, they all impact the firm reputation in the same positive manner: if third parties have accepted to cooperate with these firms on sustainable development issues, it is because they grant them some credibility in this matter. By doing so, they confer their moral guarantee to firms CSR' strategies. Consequently, the better is the reputation on this issue of the third party you team up with, the higher its impact on your image.

In conclusion, even though systematic evidence lack, it seems that greenwashing is indeed a matter of concern. Even firms acknowledge this reality by implementing strategies to give credibility to their communication and actions. As highlighted previously, credible communication is a key issue for CSR to have a positive impact on firms' profits and social welfare, thus further research is crucial to assess the extent of greenwashing and to identify the efficient strategies to temper this phenomenon.

## 3.2 Additionality

We have considered greenwashing and the need to verify whether firms really do what they claim they are doing. The other relevant reality check consists in confirming whether SR firms' social and environmental achievements are really additional with regard to what is already mandatory.

This question raises the issue of self-regulation socially desirability. To investigate this issue, consider a social welfare function that is the sum of firms' profits $\Pi$, individuals utility $U$, total environmental goods $E$ (i.e. the economic value of the environment) and

[43] It is a French network that was created after the Rio Summit in 1994, with the goal to implement sustainable development in the French regions. It includes firms, local authorities, NGOs, etc.

[44] Observatoire de la Responsabilité Sociétale des Entreprises, it is a French network designed to study and promote socially responsible investment. It includes firms, trade unions, NGOs, etc.

total social goods $S$ (i.e. the economic value of the social system).

$$SW = \Pi + U + E + S$$

CSR is socially desirable if the social welfare with CSR, $SW_{CSR}$, is higher than the social welfare when firms carry on with business-as-usual, $SW_{BAU}$.

- If CSR is not additional, the value of $E$ and $S$ remain the same while profits decrease due to the costs associated to these new practices, $c_{CSR}$. It is not socially desirable: $SW_{CSR} - SW_{BAU} = -c_{CSR} < 0$
- If CSR is additional, the value of total environmental and social goods is higher $(E + \delta E, S + \delta S)$: we have $SW_{CSR} - SW_{BAU} = \delta E + \delta S - c_{CSR}$ whose sign decides the social desirability of CSR.[45]

Very importantly, if CSR does provide higher social or environmental benefits, it can also become an alternative to legislation. Indeed, legislation to achieve the same improvements $(E + \delta E, S + \delta S)$ entails a double cost: legislation cost,[46] $c_L$, and the cost endured by firms to reach the new standards, which is assumed to be the same in both cases, $c_{CSR}$. Hence the difference between social welfare under legislation, $SW_L$, and social welfare with CSR only is: $SW_{CSR} - SW_L = c_L > 0$. If CSR is additional and allows to reach a state $(E + \delta E, S + \delta S)$, it is a more socially desirable option than direct legislation making these improvements mandatory. In this perspective, it is utterly important for public debate to

[45] Note:
- Individuals' utility may be a positive function of $E$ and $S$ because individuals are conscious of the improvements and value them. In such a case, when $E$ and $S$ increase, so does $U$, and we have $SW_{CSR} - SW_{BAU} = \delta E + \delta S + \delta U - c_{CSR}$
- Even if individuals may reward firms for being responsible, it does not impact the overall social welfare: what is taken from the individuals is given to firms, it is merely a transfer.

[46] Direct legislation costs include expertise costs to assess the right level of law stringency, the costs of lobbying, the costs of controlling compliance, etc.

evidence the contribution of CSR: if it is additional, further regulation may not be needed, while if it is not, there may be a rationale for more stringent standards.

**Empirical evidence**

Surprisingly, despite its uttermost importance for the political debate, there exists to our knowledge no study about CSR additionality with regard to law requirements. Nevertheless, there exist studies that compare CSR performance of BAU and SR firms and bring valuable evidence. Indeed, if such studies prove that there exist no difference between SR and BAU firms social and environmental performance, it means at least that CSR practices are not additional with regard to law requirements.

To do so, authors must solve two problems. First, they need a criterion to distinguish BAU from SR firms: All authors use the participation to a voluntary program (e.g. Climate wise program,[47] 33/50 Program[48]) or the adoption of an Environmental Management System (e.g. ISO14001[49]). Second, as for greenwashing, it is difficult to assess the real social and environmental performance of firms due to limited existing data: Authors widely use pollutant emissions as a proxy, while a few others use the amount of waste/level of recycling.

A first consequence of these choices is that results must be analyzed at program-level. Indeed, every voluntary program or EMS may be differentially effective in spurring additional CSR performance. We have listed the following:

- **ISO14001**: it is the most widely documented and seems to be an effective EMS. We

[47]Officially established by US Environmental Protection Agency in 1993, Climate Wise was a performance-based voluntary program focusing on the nonutility industrial sector to encourage the reduction of CO2 and other GHGs via adoption of energy efficiency, renewable energy, and pollution prevention technologies.

[48]The 33/50 Program was launched by the EPA in 1991 to induce firms to voluntarily reduce their emissions of 17 high priority toxic chemicals.

[49]ISO 14001 sets out the criteria for an environmental management system and can be certified to. It does not state requirements for environmental performance, but maps out a framework that a company or organization can follow to set up an effective environmental management system.

denoted 8 studies, out of which 7[50] (resp. 1[51]) evidence a positive impact (resp. no impact) on environmental performance (pollutant emissions or production of waste).

- **33/50 Program**: We denoted 4 studies with a majority of positive results. 3[52] of them (resp. 1[53]) evidence a positive impact (resp. no impact) on environmental performance (toxic chemical emissions or production of waste).

- **Climate Wise Program**: We denoted 3 studies with mixed results. Brouhle et al. (2012) evidence an increase in environmental innovation for participants, but only for low levels of R&D firms. Morgensten et al. (2007) show that participation in Climate Wise led to a slight (3–5%) increase in electricity costs that vanished after two years and to a slight (4–8%) decrease in electricity costs that persisted for at least three years. Finally, Welch et al. (2000) find no reduction in CO2 emissions amongst participants.

- **Strategic Goals Program**:[54] Brouhle et al. (2009) evidence that participants display little if any additional reduction in toxic chemical emission.

We can conclude that, on the one hand, some programs seem indeed to lead to additional efforts (ISO14001, 33/50 Program) while it is less evident for others (Climate Wise Program, SGP).

These partial results suggest that in some cases CSR indeed result in additional social

[50] Babakri et al. (2004), Dasgupta et al. (2000), King et al. (2005), Melnyk (2002), Potoski and Prakash (2005), Russo (2009), Szymanski and Tiwari (2004).
[51] Barla (2007).
[52] Khanna and Damon (1999), Gamper-Rabindran (2006), Sam et al. (2009).
[53] Vidovic and Khanna (2007).
[54] The SGP was launched by the EPA in 1997. It encourages companies "to go beyond environmental compliance" through the achievement of seven specific goals: 50% reduction in water usage, 25% reduction in energy use, 90% reduction in organic TRI releases, 50% reduction in metals released to water and air (as reported to TRI), 50% reduction in land disposal of hazardous sludge, 98% metals utilization, and reduction in human exposure to toxic materials in the facility and surrounding community.

and environmental results while sometimes it does not. Further research is crucial to confirm whether CSR can be additional or not and to identify the conditions under which it is achievable.

# 4 Conclusion

In this chapter, we have reviewed the literature in order to assess CSR profitability and to perform a socio-environmental evaluation of CSR. More specifically we have aimed at answering the three following questions: Is CSR profitable and how? Are firms really implementing the responsible practices that they communicate about? Do firms responsible practices really go beyond mandatory requirements?

The economic literature identifies five mechanisms that can increase firms' profits:

1. Some consumers may accept to buy ethical or green products at a higher price or boycott 'dirty' corporations.

2. Potential employees may be more attracted to responsible firms; Employees may work in a more productive way, or for lower wages, in responsible firms.

3. self-regulating firms may face an increased productivity of their environmental inputs.

4. Cost of capital may be reduced for responsible firms, either because there exist socially responsible investors or because they perform better than other firms.

5. CSR may also impact the legislative process by preventing it, influencing the toughness of the chosen mandatory standards or reducing the frequency of inspections by authorities.

Even though only a dozen of studies exist about the labor market mechanism, it looks like a promising one. Authors' findings clearly indicate that there may be a real profit opportunity for self-regulating firms, even though its magnitude for a given corporation highly depends on the distribution of its workers' ethical and environmental values. In particular, empirical evidence suggest that this would probably be a stronger opportunity for firms whose workforce encompasses an important proportion of highly skilled workers.

The other mechanism that also apparently impacts positively firms is taking place on the capital market. It seems that investors are not indifferent to firms social and environmental behaviors and, more precisely, that they value it. Nevertheless, overall they value it little and the impact on firms profits is not striking. A finer understanding of the mechanisms on the capital market and the conditions under which they can be profit enhancing is needed.

On the contrary, despite the attractivity of the idea that an increased productivity of environmental inputs can increase financial performance, empirical evidence almost unanimously prove the contrary. Presumably, the initial costs born by the firm to increase its environmental performance (research & development, investment in new technologies, etc.) outweigh any potential benefits drawn from an increased productivity.

At first glance, increased profits through the offering of ethical and green products seems possible because numerous studies evidence the existence of a substantial price premium for such products. Nevertheless, these higher prices are not confronted nor with the increase in costs born by the firms to supply such goods, neither with the actual demand size that firms could meet. It is thus impossible to conclude about the real profitability of this mechanism.

Similarly, for methodological reasons, the total lack of empirical evidence forbid us to conclude about the efficiency of the mechanism targeted at public authorities.

Finally, it is important to keep in mind that CSR is a credence good for individual stakeholders, they can not ascertain the veracity of firms' claims neither by search or experience. There is a need for specialized agents (extra-financial rating agencies, NGOs, labeling institutions, etc.) that have sufficient resources to observe firms' self-regulatory activities and to convey this information to stakeholders. Without them, stakeholders are not able to reward firms self-regulation efforts through any the mechanisms described above and CSR can not be a profit opportunity.

As for the socio-environmental evaluation of CSR, it is very difficult to provide any clear conclusions because, despite the uttermost importance of this issue for the public debate, there is a strong lack of empirical evidence on both questions that we have addressed. There are anecdotal evidence about greenwashing, mostly provided by non academic actors, that imply that it is indeed a matter of concern, but wide and systematic analysis of this phenomenon do not yet exist. Similarly, partial results suggest that in some cases CSR indeed result in additional social and environmental results while it sometimes does not, but further research is crucial to confirm whether CSR is be additional or not and to identify the conditions under which such it is achievable.

# Chapter 2

# Incentives for Quality in Friendly and Hostile Informational Environments

## English summary

We develop a simple lemons model with endogenous quality where disclosure is quality-dependent. The distinctive feature of the analysis is to contrast friendly informational environments, in which quality is more often disclosed when it is high than when it is low, and hostile environments, in which the converse holds. Differences are clear-cut: Hostile environments give rise to a bandwagon effect across sellers, which can lead to multiple equilibria. In contrast, friendly environments create free riding among sellers, which always induces a unique equilibrium. Comparative statics results are also contrasted. A key notion is that incentive provision is relatively better when the informational environment targets less expected evidence. The results shed new light on several insights of the literature on statistical discrimination, collective reputation and quality certification.

## Résumé français

Nous développons un simple modèle de "lemons" avec un niveau de qualité endogène et une révélation de la qualité dépendante de ce niveau. La caractéristique distinctive de notre analyse est de contraster des environnements informationnels amicaux, dans lesquels la qualité est plus souvent révélée quand elle est haute que lorsqu'elle est basse, et des environnements hostiles, où c'est le contraire. Les différences sont nettes. Les environnements hostiles donnent lieu à un effet d'entraînement entre vendeurs, qui peut entraîner des équilibres multiples. En comparaison, les environnements amicaux permettent aux vendeurs d'adopter des comportements de passagers clandestins et cela entraîne toujours un équilibre unique. Les comparatives statiques offrent elles aussi des résultats contrastés. Une notion clé est que l'incitation à produire de la qualité haute est plus forte quand l'environnement informationel est biaisé dans le sens où on ne s'y attend pas. Ces résultats apportent une lumière nouvelle sur plusieurs intuitions existant dans

la littérature sur la statistique, la réputation collective et la certification de la qualité.

# 1 Introduction

It is widely recognized that buyers may have less information about certain product attributes than the seller, implying severe inefficiencies. Since the publication of Akerlof's paper in 1970, this lemon problem has been explored in the economics literature from many different perspectives. A crucial yet widely overlooked aspect is that, in reality, the asymmetry of information between buyer and seller may be more or less severe depending on the level of product quality, and that this has important consequences for the sellers' incentives to provide high quality.

Consider the example of the reliability of a vehicle. Reliability basically means the absence of breakdowns. Hence, when a car actually breaks down, its quality is disclosed.[1] But if nothing occurs, it does not mean that the vehicle will not break down in the future. In this case, experience leads consumers to be more informed about the quality of poorly reliable products than about that of reliable ones.

In a similar vein, anti-doping tests in sport can only identify a limited set of performance-enhancing drugs and it is common knowledge that certain doping substances are undetectable. As a result, tests are able to provide evidence of doping, but they fail to ascertain its absence. That is, they can only uncover low quality.

In other circumstances, consumers receive more informative signals when quality is high rather than low. For example, a movie award selectively signals a high-quality product. But the consumers are left uncertain about the quality of non-awarded movies: they could be good or bad, given that not all good movies receive a prize. Academic publication is another illustration: prestigious journals (mostly) include good papers, but a fraction of the unpublished manuscripts are excellent–notably those under review/revision

[1] Reliability is to a large extent an experience good attribute in that it is revealed over time after the purchase. But this does not prevent quality disclosure to influence demand either through repeated transactions or through reputation effects. Alternatively, the potential buyer can test the vehicle for a short period and obtain some information–mostly in case of a breakdown.

at good journals.

Those examples suggest that the quantity of information about a good or service available to a buyer or, more generally, to a user, frequently varies with the level of product quality, and can be biased in either direction. In some cases, the informational environment is *friendly* - more information is available on high quality - while in others, the environment is *hostile* - more information is available on low quality. Note that these differences could be either related to the very nature of the evaluation technology (reliability, doping) or to the selective disclosure of evaluation results (movie awards, academic publications). To clarify things on the notions of friendly and hostile environments, one can think of the monitoring technology as a generic supervisor, whose attitude is either friendly, in which case it has a tendency to put more emphasis on good news (perhaps by withholding bad news), or hostile, in which case more emphasis is put on bad news. An unbiased supervisor is referred to as neutral.

In this paper, we develop a simple model that accommodates all those types of environments *at the same time* and we seek to identify their impacts on quality provision. It depicts a continuum of agents who are willing to produce and sell a good whose quality can either be low, at no cost, or high, at a cost varying across agents. Each agent first chooses the level of quality, which is imperfectly observable by the potential buyer. Once the agent has selected its level of quality, some monitoring occurs, which discloses quality with a probability $g$ if quality is high (”good news”) and probability $b$ if quality is low (”bad news”). After monitoring, the buyer updates its beliefs about quality and decides to purchase or not the good.

It is easy to see that differences between $g$ and $b$ are all but neutral. In particular, when the buyer receives no news after monitoring, belief updating is totally different: when the informational environment is friendly ($g > b$), the buyer knows that monitoring filters out high quality. Therefore receiving no news about a product leads him to become more

pessimistic about its quality. When the informational environment is hostile ($g < b$), no news conversely improves its belief because he is aware that monitoring filters out low quality.[2]

The model describes how agents' quality investments interact with these belief revisions. We find clear-cut differences between hostile and friendly environments. Hostile environments give rise to a bandwagon effect among agents, which leads beliefs to be self-fulfilling in some cases, so that there may be multiple equilibria. The intuition is the following: when the prior is very optimistic, supplying high quality is an equilibrium because hostile monitoring would easily reveal the agents supplying low quality. But when the buyer's prior belief is pessimistic, incentives to increase quality are limited because it is relatively difficult to ascertain high quality, and there is few hope for the seller to prove wrong a pessimistic belief.

In contrast, friendly environments create a form of free riding across agents, which always induces a unique equilibrium. Comparative statics results are also contrasted. In particular, we show that moving from a hostile to a friendly environment increases the average equilibrium quality when the resources allocated to monitoring–reflected in the sum $g + b$ in our model–is low. The reverse is true for a higher $g + b$. The idea is that little information induces little reward to high quality firms and thus little incentive to produce high quality. Knowing that, the buyer is pessimistic about a good's quality on which he did not receive any feedback. It is more effective to go against the buyer's belief by increasing the probability to get the full reward when producing high quality. Also, reducing the cost of quality (in a precise sense which is defined in the paper) increases quality in any case, but less when the environment is friendly.

This paper is obviously far from being the first to explore the impact of informational

[2]Our model has the minimal information structure to capture those aspects in a tractable way, but it is still richer than most standard models of disclosure or imperfect monitoring with endogenous incentives.

environments on quality provision.[3] The distinct feature of our analysis is to build a model that highlights the key differences between hostile and friendly environments. In previous works, this dimension, albeit crucial as we point out, is either left implicit as in works on statistical discrimination (Coate and Loury , 1993) and collective reputation (Tirole, 1996); or the focus is on friendly environments as in the literature on quality disclosure which looks at the incentives for firms to voluntarily disclose quality and for certifiers[4] to provide unbiased certification about quality (for a recent survey, see Dranove and Jin, 2010). Importantly, the information structure we study is closely related to those in MacLeod (2007) and Board and Meyer-ter-Vehn (2010). In the former, two types of situations are studied: the case of a "normal good", for which breakdowns are relatively rare, and "innovative goods", for which breakthroughs constitute rare events. Those two situations pertain to the actual result of effort, not to the information structure. In the latter paper, the setting is dynamic and news arrive according to a Poisson process. Either there is no information revealed, or, with Poisson arrival, the information structure "ticks". Whether a tick is good or bad news depends on whether high or low quality ticks more often. Hence the signal combination can be either no news/good news or no news/bad news, and the interpretation of a tick depends on the parameters.

Our work is also related to the model of Kamenica and Gentzkow (2011). They study the design of informational environment, a term we borrow from their paper. However in their analysis the state of the world is fixed, and unknown to all agents, while in our model quality is endogenous (and the seller is obviously aware of the quality he has cho-

[3]Recent important works on the information structures of lemons problems include Sarath (1996), Levin (2001) and Kessler (2001). However they assume exogenous quality, hence the incentive dimension at the heart of our analysis is absent.

[4]There exists an important literature on information intermediaries, Lizzeri (1999) is among the seminal papers. Those analyses underline that certifiers somehow bias the informational environment towards friendliness. While we do abstract from strategic auditors here, it is an interesting connection to make in future works. The role of NGOs and firm monitoring is typically a setting of interest, in which the literature is still at an early stage (Lyon, 2010; Lyon and Maxwell, 2011). In such context, models are still needed that account for both friendliness and hostility in information revelation. Typically, environments in which "bad cops" NGOs such as Greenpeace are a majority generate a hostile informational environment.

sen). They are interested in how the informational environment influences an ex-post action, while our focus is on how (ex-ante) moral hazard is influenced by the informational environment.[5]

This allows to derive new results and to provide new insights on existing results–such as the interpretation of multiple equilibria. In the penultimate section of the paper, we relate our results to different branches of the literature.

The structure of the paper is straightforward. Section 2 presents the model. In sections 3 and 4, we characterize equilibria and conduct comparative statics exercises. In Section 5, we explore how our results relate to the existing literature. In particular, we apply our analysis to three research streams: statistical discrimination, collective reputation and quality certification. Section 6 gathers final comments.

# 2 The model

We consider a game with a continuum of agents (firms, sellers) that each produces one unit of a good which quality is imperfectly observable by a representative buyer (consumer, user). The quality variable $a$ is binary, with $a \in \{0,1\}$, and set by each agent. Choosing $a = 0$ costs the agent nothing , whereas choosing $a = 1$ costs $c$, which is heterogenous across agents, and follows a cumulative distribution $F(c)$ and density $f(c)$ on $[\underline{c};\overline{c}]$. $F$ and $f$ are common knowledge, but each agent privately observes the realization

[5]There is also a connection with inspection games (see Avenhaus et al., 2002, for an overview). In an inspection game, the inspectee chooses to comply (at some cost) or not with a previously agreed course of action, and the inspector can conduct an inspection (a statistical test) to verify compliance. The inspector strategy is hence related to the informational environment we model. There are many difference with our setting, though, the most crucial difference being that inspection games are simultaneous, while our setting is sequential. Other related questions appear in the literature on information systems of agency problems Relatedly, the literature on monitoring in principal-agent models (e.g. Dye, 1986; Fagart and Sinclair-Desgagné, 2007) studies the design of ex-post audits (of effort) that are contingent on some realized outcomes. In our case, there is always a one-dimensional signal, and further conditional audit can not be conducted. In addition, our assumptions on commitment are different from the ones made in those two strands of the literature.

of his cost.[6]

The buyer's (marginal) willingness to pay for the high quality ($a = 1$) is $\Delta$ and it is 0 for the low quality ($a = 0$). For the sake of clarity, most of the analysis is carried under the normalization $\underline{c} = 0$ and $\bar{c} = \Delta = 1$, which is qualitatively unimportant.[7] Finally, we also assume that the agents can fully extract the buyer surplus, which first gives the most incentives to the agents and second makes welfare analysis transparent. Under these assumptions, if quality were perfectly observable, the social optimum would be attained since all agents would choose $a = 1$ and sell the good at price $\Delta$, since $\Delta \geq c$ for all $c$.

The buyer receives an imperfect signal $s \in \{H, L, \varnothing\}$ on the quality of the good supplied by each agent. The monitoring technology is asymmetric: the signal is generated according to the probabilities $g = \Pr[s = H|a = 1]$ and $b = \Pr[s = L|a = 0]$ where $g$ can be less or higher than $b$. With probability $1 - g$ when $a = 1$ and with probability $1 - b$ when $a = 0$, an empty signal $\varnothing$ is generated. We interpret this outcome as "no news", in the sense that it carries no evidence, but this signal may of course still lead to Bayesian interpretation. Overall, the information structure is thus one of partial hard information since signals $H$ and $L$ are evidence of the actual quality.[8]

The sequence of events, illustrated in figure 1, is as follows:

- **Stage 1:** The seller privately learns his cost $c$, then chooses $a$.
- **Stage 2:**
  - If the seller has chosen $a = 1$, the buyer learns this "good news", $s = H$, with probability $g$. With probability $(1 - g)$, no information is disclosed ($s = \varnothing$).

[6] Equivalently, we model an agent with unknown cost and unobservable effort.

[7] We show in the appendix that the results go smoothly through when some agents have high costs, and never choose $a = 1$. The insights also carry through the case where some agents have negative costs. These agents could be interpreted either as intrinsically motivated by high quality production, or equivalently, as "honest" agents always producing the reference high quality. Finally, an important assumption that we also discuss regards the willingness to pay for low quality being 0.

[8] We show in Appendix A how to extend our result to imperfect signals of quality, i.e. to a soft information structure.

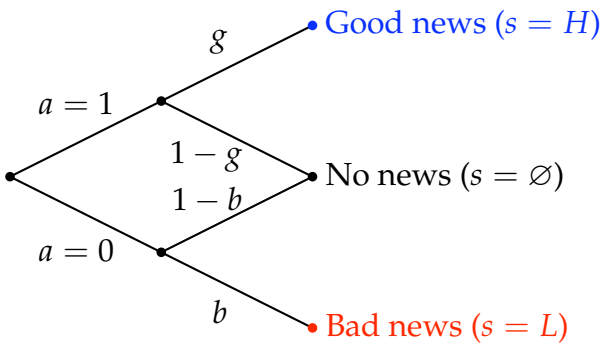



Figure 1: Information structure.

  - If the seller has chosen $a = 0$, the buyer learns this "bad news", $s = L$, with probability $b$. With probability $(1 - b)$, no information is disclosed ($s = \varnothing$).

- **Stage 3:** The buyer transfers $\Delta$ to the seller if $s = H$, but gives nothing for a bad news $s = L$. When no evidence is generated about the real action, the buyer forms a belief $\mu = \Pr[a = 1|s = \varnothing]$. Given this belief, the seller only receives the conditional expected value $\mu\Delta$, which is the willingness to pay of the uninformed buyer.[9]

Our main goal is to study the equilibrium distribution of qualities depending on the informational environment $(g, b)$ and the cost distribution $F$.

The setting is sufficiently abstract to capture many real-world situations. The buyer can represent consumers in a final market with vertical differentiation in which endogenous quality is not perfectly observable (e.g., environmental attributes in a market with green consumers), a firm hiring employees whose intrinsic productivity resulting from

[9]There are several extensive forms that corresponds to this payment scheme. A natural situation would be one in which the seller makes a take-it-or-leave-it price offer. The timing of this offer deserves some careful treatment. It could be made ex-ante, before the seller learns its costs and thus consists of a conditional price (on the signal to be realized), ex-post, once information is public, or at an interim stage, once the seller knows his type, but before the public signal is known. For the interim case, one could assume that the offer is made before or after quality is chosen. Ex-ante and ex-post offers clearly lead to the same outcome. In turn, interim offers are more subtle to analyze, as prices can be used as signalling devices (e.g. Milgrom and Roberts, 1986; Bagwell and Riordan, 1991). We take the view that the price offers should be renegotiation proof (or alternatively offers are made ex-post), having in mind that the seller could simply withdraw its product in case the offer is not profitable ex-post. Then equilibria are pooling and the extensive form does not matter. Another straightforward way to justify this payoff structure is that there are at least two identical buyers interested in the product, and that they bid once information is public.

past efforts is uncertain, lenders in a capital market, etc. The model can also apply to non-market situations such as school testing: A teacher needs to grade students whose level of performance is not always observed for monitoring resources are limited. In this context, $\Delta$ represents the grade for a student whose level of performance $a = 1$ is disclosed, 0 is the grade for a student whose observed performance is $a = 0$. Finally, $\mu\Delta$ is the grade of students for whom the teacher has no information.

# 3 Equilibrium analysis

## 3.1 Incentives and cutoff equilibria

We adopt the notion of perfect Bayesian equilibrium where each agent chooses its best reply to the market's belief and the buyer's belief is consistent with the quality set by the different types of agents.

Let $\Pi(a, c)$ be the expected payoff of an agent with cost $c$. For a given belief $\mu$, the possible expected payoffs are: $\Pi(1, c) = [g\Delta + (1 - g)\mu]\Delta - c$ and $\Pi(0, c) = (1 - b)\mu\Delta$. The agent then chooses $a = 1$ whenever[10] $\Pi(1, c) \geq \Pi(0, c)$, which translates into:

$$c \leq g\Delta - (g - b)\mu\Delta \tag{2.1}$$

Note that the last RHS term $(g - b)\mu\Delta$ of this incentive constraint is negative if $g > b$ and positive if $g < b$. That is, if the environment is friendly, an increase in the belief $\mu$ reduces incentives to increase quality whereas it raises incentives in a hostile environment. We will show below that this asymmetry has fundamental consequences on equilibria.

An almost immediate consequence of the incentive constraint is that all the Bayesian equilibria of the game have a cutoff structure, namely they will all be characterized by a

[10]Mixing by one type is here unimportant given it has zero weight, and we assume that the unique indifferent type chooses $a = 1$ over $a = 0$.

cost threshold below which agents choose quality 1 and above which the others choose quality 0. A first lemma states this formally.

**Lemma 1** *All Bayesian equilibria of the game described above are cutoff equilibria. Namely, any equilibrium is characterized by some $c^*$ such that all firms with $c \leq c^*$ choose $a = 1$ and all firms with $c > c^*$ choose $a = 0$. When $0 < c^* < 1$, this cutoff satisfies:*

$$c^* = g\Delta - (g - b)\mu^*\Delta. \tag{2.2}$$

*The corresponding equilibrium belief $\mu^*$ is consistent with the cutoff $c^*$ according to Bayesian updating:*

$$\mu^* = \frac{(1-g)F(c^*)}{(1-g)F(c^*) + (1-b)(1-F(c^*))} \tag{2.3}$$

**Proof.** Consider an equilibrium of the game in which the beliefs of the stakeholder upon receiving no news is some $\mu^*$, and suppose that there exists a $\hat{c}$ such that $\hat{c} \leq g\Delta - (b-g)\mu^*\Delta$. Then for all $c \leq \hat{c}$ we have $c \leq \hat{c} \leq g\Delta - (b-g)\mu^*\Delta$, hence the best reply to the equilibrium belief $\mu^*$ of all types below $\hat{c}$ is to choose $a = 1$. Similarly, if some type chooses $a = 0$, then all types above also choose $a = 0$. This establishes that an equilibrium is characterized by a cutoff $c^*$, possibly the extreme types 0 or 1. Bayesian revision obtains from the fact that the fraction of high quality firms is $F(c^*)$, the mass of types lower than the cutoff. ■

Given that an equilibrium is entirely characterized by a cost cutoff $c^*$, we will often refer directly to $c^*$ as an equilibrium, the corresponding belief being unequivocally obtained using (2.3).

## 3.2 Existence and stability

To investigate existence and stability properties of the equilibria we first use a standard fixed point representation of the problem. We also impose $\Delta = 1$ to simplify notations. Combining (2.3) and (2.2) yields:

$$c^* = \Phi(c^*) \quad \text{with} \quad \Phi(c) \equiv \frac{g(1-b) - (g-b)F(c)}{(1-b) - (g-b)F(c)} \tag{2.4}$$

The function $\Phi$ is continuous with $\Phi(0) = g$ and $\Phi(1) = b$. Hence existence of equilibrium is ensured by the intermediate value theorem. Note that there may exist multiple equilibria, an issue we study below.

Finally, we introduce the following notion of stability, that we will use later:

**Definition 1** *An equilibrium $c^*$ is stable if $\left|\frac{d\Phi}{dc}(c^*)\right| < 1$, and unstable (a tipping point) otherwise.*

This is a standard definition[11] which amounts to say that a stable $c^*$ is an attractive fixed point. Underlying this notion is the idea of dynamic adjustment: In the case where small perturbations induce a little deviation of the equilibrium from $c^*$ to $c^* + \varepsilon$, the cost cutoff will return to $c^*$ if the equilibrium is stable. This occurs if the iterated function sequence $c^* + \varepsilon$, $\Phi(c^* + \varepsilon)$, $\Phi(\Phi(c^* + \varepsilon))$,... converges to $c^*$. Alternatively, the equilibrium is unstable if small perturbations lead to a diverging dynamic adjustment towards a nearby stable equilibrium. $c^*$ is then said to be a tipping point since a downward deviation $c^* - \varepsilon$ and an upward deviation $c^* + \varepsilon$ lead to two different stable equilibria.

[11] A closely-related version of stability is considered in Jackson and Yariv (2007), which does not require $\Phi$ to be continuously differentiable. In our framework, both turn out to be equivalent. See also Brock and Durlauf (2001).

## 3.3 Equilibrium characterization

For the remainder of the exposition, we work with an alternative representation of equilibria which will prove useful for obtaining in particular stability and comparative statics result. Another advantage of this representation is that it allows straightforward graphical interpretations, and makes the equilibrium fraction of high quality apparent. When $g \neq b$, we suggest rewriting Equation (3.3) as follows:

$$F(c^*) = M(c^*|g,b) \quad \text{with} \quad M(c|g,b) \equiv \frac{(g-c)(1-b)}{(g-b)(1-c)} \tag{2.5}$$

The function $M$ will be referred to as the *monitoring curve*, as it contains the data pertaining to the informational environment, and is independent of the distribution $F$. We derive some properties that will be used in the subsequent analysis:

**Lemma 2** *(Properties of the monitoring curve)*

1. *The monitoring curve $M(.|g,b)$ is decreasing and concave in a friendly environment.*
2. *The monitoring curve is increasing and convex in a hostile environment.*
3. *The monitoring curve consists in a vertical line (i.e. not a function) in a neutral environment (when $g = b$).*

**Proof.** Note that $\frac{\partial M}{\partial c} = -\frac{(1-b)(1-g)}{(g-b)(1-c)^2}$ and $\frac{\partial^2 M}{\partial c^2} = -\frac{2(1-b)(1-g)}{(g-b)(1-c)^3}$. Both derivatives are negative (resp. positive) when $g < b$ (resp. $g > b$). When $g = b$, the function $\Phi(c)$ is constant, equal to $g(= b)$, which translates into the monitoring curve being a vertical line. ■

With these properties of the monitoring line in mind, it is easier to translate the stability properties in our framework. In particular, they translate into easy comparisons of the density function and the slope of the monitoring line.

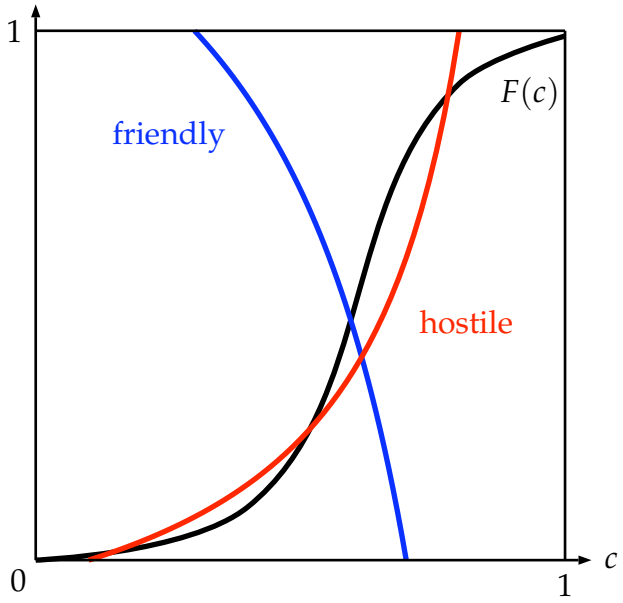



Figure 2: Equilibrium representation in the $(c, F(c))$-space.

**Lemma 3** *(Stability criteria)*

1. *In a friendly environment, the equilibrium is stable if* $f(c^*) < -\frac{\partial M}{\partial c}$.
2. *In a hostile environment* ($g < b$) *an equilibrium is stable if* $f(c^*) < \frac{\partial M}{\partial c}$ *and unstable otherwise.*

**Proof.** The stability condition is $\left|\frac{\partial \Phi}{\partial c}\right| < 1$, and the definition of $\Phi$ gives $\frac{\partial \Phi}{\partial c} = f(c^*)\left(\frac{\partial M}{\partial c}\right)^{-1}$. The conclusions obtain using the previous lemma. ■

Using this representation, an equilibrium cutoff is such that the graph of the cumulative distribution intersects the auditing line, as pictured on figure 2[12]. When the monitoring curve is decreasing, i.e. in a friendly environment, the equilibrium is necessarily unique, since $F$ is increasing. In turn, there might be multiple intersections between $F$ and $M$ when $g < b$. The lemma says that stable equilibria correspond to crossings of $M$ by $F$ from above.

We summarize the analysis of this section in a first proposition:

[12] In the symmetric case $g = b$, $M$ actually consist of a vertical line at $c = g = b = c^*$.

**Proposition 1** *(Characterization of equilibria)*

1. *In a neutral environment ($g = b$), there exists a unique equilibrium with $c^* = g = b$.*

2. *In a friendly environment ($g > b$), there exists a unique perfect Bayesian equilibrium defined by (3.3). Moreover the cutoff is such that $b \leq c^* \leq g$.*

3. *In a hostile environment ($g < b$), there might be multiple equilibria defined by (3.3), and they are all such that $g \leq c^* \leq b$. Moreover the lowest and the highest equilibrium are stable.*

The first part of the proposition corresponds to the traditional symmetric monitoring technology ($g = b$). In this case, the equilibrium cutoff $c^* = g = b$ does not depend on the distribution $F$. Still, the fraction of agents choosing high quality is of course $F(g)$. But there is no strategic effect between the different types of agents.

The major result lies in the existence of multiple equilibria in hostile environments (and only in hostile environments). To understand what drives this result, note first that the equilibrium belief (2.3) increases with the number of agents opting for high quality, $F(c^*)$. Hence, quality choice generates a positive externality as equilibrium payoffs of all agents increase with the belief, whether they choose $a = 1$ or not.

This externality has different impacts on quality choice in the two informational environments. To explain why, one needs to turn back to the RHS of the incentive constraint (2.1). In particular, the last term $(g - b)\mu\Delta$ is negative if $g > b$ and positive if $g < b$. This means that:

- if the environment is friendly, the externality passing through the belief reduces incentives to increase quality in a friendly environment (a free riding effect);

- if the environment is hostile, it provides more incentives (a bandwagon effect).

Loosely speaking, quality choices of different types of agents are strategic substitutes in a friendly environment. They are strategic complement in a hostile environment. As a result, the beliefs tend to be self-fulfilling in the latter case. Consider for instance the extreme case where $g = 0$ and $b = 1$. If the belief is $\mu = 0$, the buyer basically thinks the agents will never choose high quality, and this belief can never be proven wrong because $g = 0$. Agents have thus no interest in raising quality since they will never get any premium for that. We get just the opposite result if $\mu = 1$: every agent has no choice but to provide high quality as low quality would always be revealed (since $b = 1$).

For the sake of applications, our equilibrium characterization has two corollaries pertaining to equilibrium multiplicity, the first one regarding the information structure and the second one regarding the cost distribution.

**Corollary 1** *A necessary condition for multiple equilibria is that the informational environment is hostile.*

This is typically important for models of statistical discrimination, that are discussed in the application section. It is also linked to a discussion on multiple equilibria in compliance model. Compliance by definition creates a hostile environment. However, while hostility is necessary for multiplicity in our framework, it is not sufficient. In particular, we provide a condition for uniqueness on the distribution function:

**Corollary 2** *A sufficient condition for the equilibrium to be unique irrespective of the information structure is that the cost distribution is concave and $g > 0$. When $g = 0$, $c^* = 0$ is also an equilibrium, but it is unstable.*

Hence in particular if the distribution of costs is uniform the equilibrium is unique.

# 4 Comparative Statics

We now investigate how changes in informational environment or in cost distribution influence equilibria.

## 4.1 Changes in informational environment

How do variations in $g$ and $b$ affect the equilibrium? It is easy to obtain the following comparative statics:

**Proposition 2** *All else equal, the equilibrium cutoff (and hence the fraction of agents opting for high quality),*

1. *increases with the quality of information when the equilibrium is stable:* $\frac{\partial c^*}{\partial g}, \frac{\partial c^*}{\partial b} \geq 0$.
2. *decreases with the quality of information when the equilibrium is a tipping point:* $\frac{\partial c^*}{\partial g}, \frac{\partial c^*}{\partial b} < 0$.

**Proof.** Differentiating the equilibrium condition (2.5) with respect to $g$ and $b$ one obtains:

$$\frac{\partial c^*}{\partial g} = \frac{\frac{\partial M}{\partial g}}{f(c^*) - \frac{\partial M}{\partial c}} \text{ and } \frac{\partial c^*}{\partial b} = \frac{\frac{\partial M}{\partial b}}{f(c^*) - \frac{\partial M}{\partial c}}$$

In friendly environments, we know that $f(c^*) - \frac{\partial M}{\partial c} \geq 0$ as $\frac{\partial M}{\partial c} = -\frac{(1-b)(1-g)}{(g-b)(1-c)^2} < 0$. Then we differentiate $M$ with respect to $g$ and $b$ : $\frac{\partial M}{\partial g} = \frac{(1-b)(c-b)}{(1-c)(g-b)^2}$ and $\frac{\partial M}{\partial b} = \frac{(1-g)(g-c)}{(1-c)(g-b)^2}$. $\frac{\partial M}{\partial g}, \frac{\partial M}{\partial b} \geq 0$ follows from $b \leq c^* \leq g$ (Proposition 1). In hostile environments, results from $\frac{\partial M}{\partial g}, \frac{\partial M}{\partial b} > 0$ (because $g < c^* \leq b$) and $f(c^*) < \frac{\partial M}{\partial c}$ (resp. $f(c^*) \geq \frac{\partial M}{\partial c}$) when the equilibrium is stable (resp. unstable). ■

The first part of the proposition is very intuitive: better information creates higher incentives to supply high quality. In contrast, the second could seem awkward at first

glance. But but it is not if one adopts a dynamic view of the equilibrium. As explained above, a tipping point is a level of $c$ at which any small upward deviation leads to a higher stable equilibrium, and any downward deviation leads to a lower equilibrium. Hence, the lower the tipping point, the higher the chance to shift to a higher equilibrium. The overall message remains thus the same: better information improves quality.

As the focus of the paper is the contrast between hostile and friendly informational environments, it sounds appropriate to see how the relative weight of friendly and hostile monitoring influences the equilibrium. A straightforward strategy for looking at this issue consists in solving the following optimization program:

$$\max_{g,b} \quad c^*(g,b) \quad \text{subject to} \quad g+b=q \tag{2.6}$$

where $q$ represents the overall quantity of information which is less than 1 because monitoring resources are limited for instance.[13] One example of interpretation would be the following: there is a limited number of independent imperfect tests (or questions), measured by $q$, that can be performed on quality. Each test has equal power, and can identify either low quality or high quality. The problem is then to choose the balance between the different type of tests to be run. Another example would be a committee with a given size $q$. Each committee member is imperfectly informed on some specific elementary aspect of quality, but some members have preferences in favor of the agent (they are friendly), other have preferences against the agent (they are hostile). Then the probabilities $g$ and $b$ depend linearly on the composition of the committee, known in advance by the agent. We establish a new proposition for such a scenario:

**Proposition 3** *Suppose $F$ is concave and monitoring resources are limited, so that $g+b \leq q$. Then :*

[13] Another modeling route would be to assume explicit monitoring costs. At any rate, it is not a priori clear what shape such a cost function should take.

1. *The optimal informational environment is extreme:* $g^* = 0$ *or* $g^* = q$.

2. *The optimal informational environment is purely friendly (hostile) if* $F(\frac{q}{2-q}) < (>)\frac{1}{2}$.

**Proof.** See the Appendix. ■

The overall message is clear: friendliness becomes preferable when the amount of information $q$ decreases and/or when the weight of low-cost agents becomes less important ($F$ is lower at $q/(2-q)$). The intuition is as follows: a low $q$ entails that little information is disclosed about firms' activities. As a result, little reward is granted to high quality firms and incentives for quality are limited. Knowing that, the Bayesian buyer holds a pessimistic belief that one agent about whom he does not receive any feedback is producing high quality. Similarly he is pessimistic when $F$ is low as it implies high quality is likely to be too costly for the agent.

In this context, Proposition 3 tells us it is more effective to go against the buyer's belief. That is, to motivate agents to provide high quality by increasing the probability to get the full reward when producing high quality. And conversely when the buyer is optimistic about the level of high quality supply he faces (because $q$ is high and/or $F$ is low).

## 4.2 Change in costs

In this section we study more generally the consequences of changes in costs. More specifically, we consider a decrease in the sense of First-Order Stochastic Dominance (FOSD): a distribution $F$ first-order stochastically dominates a distribution $G$ when $F(c) \leq G(c)$ for all $c$.

We establish a new proposition.

**Proposition 4** *Consider a decrease in costs in the sense of FOSD:*

1. *The fraction of high quality supplied in equilibrium increases in a friendly environment, and in any stable equilibrium in a hostile environment.*

*However,*

2. *The equilibrium cutoff decreases in a friendly environment.*

3. *The equilibrium cutoffs of stable equilibria increase in a hostile environment. Except for the higher equilibrium, some equilibria may disappear after FOSD shifts of $F$.*

*Hence a decrease in costs makes hostile environments relatively more efficient at inducing incentives than friendly environments.*

**Proof.** Consider a family of distributions $F(c;\theta)$ with $\frac{\partial F}{\partial \theta} > 0$, so that, if $\theta' > \theta$, $F(.;\theta)$ first-order stochastically dominates $F(.,\theta')$. From (2.5) follows:

$$\frac{\partial c^*}{\partial \theta} = -\frac{\frac{\partial F}{\partial \theta}}{f(c^*) - \frac{\partial M}{\partial c}} \tag{2.7}$$

Hence $\frac{\partial c^*}{\partial \theta} > 0$ if $g > b$ as we know that $f(c^*) - \frac{\partial M}{\partial c} > 0$ in this case. $\frac{\partial c^*}{\partial \theta} < 0$ if $g < b$ follows from the fact that $f(c^*) - \frac{\partial M}{\partial c} < 0$ for stable equilibria in hostile environments.

Then we have

$$\frac{dF(c^*(\theta);\theta)}{d\theta} = f(c^*)\frac{\partial c^*}{\partial \theta} + \frac{\partial F}{\partial \theta}$$

Plugging (2.7) in this expression leads to

$$\frac{dF(c^*(\theta);\theta)}{d\theta} = -\frac{\frac{\partial M}{\partial c}\frac{\partial F}{\partial \theta}}{f(c^*) - \frac{\partial M}{\partial c}}$$

which is positive as $f(c^*) - \frac{\partial M}{\partial c}$ and $\frac{\partial M}{\partial c}$ have opposite signs in both environments for stable equilibria. ■

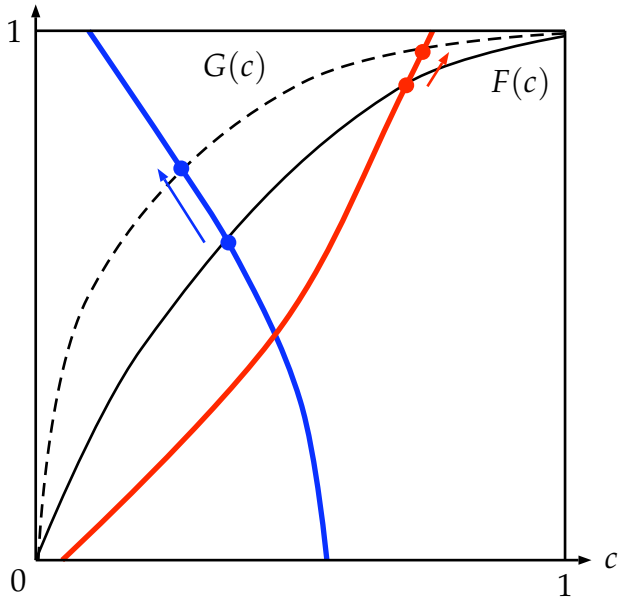



Figure 3: Decrease in costs (FOSD).

Figure 3 provides a representation of these results: a decrease in costs in the sense of first-order stochastic dominance is just a shift of the cost distribution from $F$ to $G$. Figure 4 illustrates this point in a special case where the two environments initially performs equally well with $F$. When costs decrease, i.e. when the distribution becomes $G$, with $G > F$, the hostile environment induces higher average quality.

The fact that decreasing the cost of quality raises high quality supply is not surprising. But why is the impact less important in friendly environments? Remember the contrast we pointed out previously between the two environments. In friendly environments, rising the buyer's belief has a negative impact on one agent's incentive to choose high quality, resulting in a free riding effect. This attenuates the positive effect of increased supply of high quality on incentives. In hostile environments, a self reinforcing mechanism turns agent quality choices into strategic complements, which exacerbates the positive effect of increased supply of high quality.

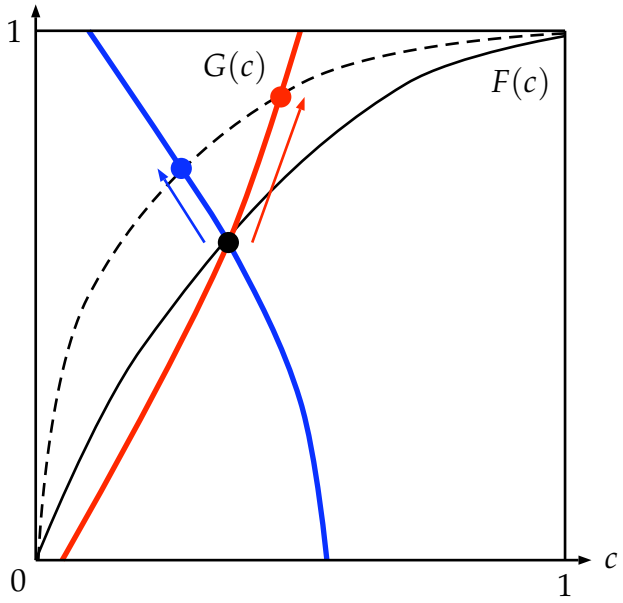



Figure 4: Comparative advantage.

## 4.3 News revealed in Equilibrium

In our model, more monitoring is always better (be it because equilibrium quality is higher or tipping points are lower), and the previous sections have studied what type of monitoring was most efficient under different circumstances. This section aims at deriving some testable implications. One aspect is important in terms of empirical work: what is usually observed consists only of a news stream. The underlying behavior is of course not directly observed, and the monitoring activity and its nature are not easily observed. Hence, suppose for instance that one only observes a news stream. The question then is the following: wan can be inferred from the news stream? What can be said on the intensity of monitoring and on the incentive effect of monitoring–even when assuming that the underlying technology or types of the firm is fixed? How does a technology shock translates in terms of the mix of good and bad news? It turns out that such questions have no straightforward answers as we show in this section. To that end, we investigate how *equilibrium* news provision varies with improvements of the informa-

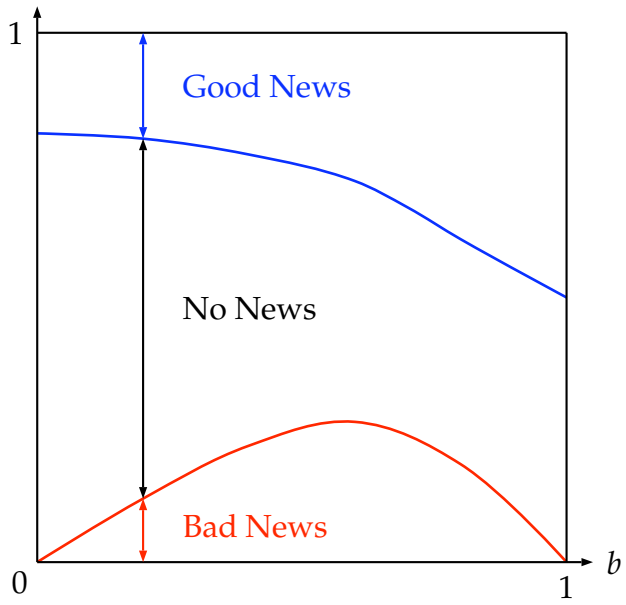



Figure 5: Equilibrium News against $b$ (uniform distribution, $g = 1/2$).

tional environment. In other words, we study the properties of $G(g,b) \equiv gF(c^*(g,b))$ and $B(g,b) = b[1(F(c^*(g,b))]$, the quantity of good news in equilibrium, and bad news, respectively. While the quantity $N(g,b) = 1 - G(g,b) - B(g,b)$, i.e. the relative absence of news, is theoretically well defined, it obviously poses empirical difficulties, and do not focus on it.

In order to make the analysis transparent, we assume that the distribution $F$ is concave, so that the equilibrium is always unique. It is enough for our point to study this case. We obtain the following proposition.

**Proposition 5** *Suppose $F$ is concave. Then:*

1. *The equilibrium quantity of good news is increasing in $(g,b)$.*

2. *The equilibrium quantity of bad news is decreasing in $g$. The profile of bad news is non-monotone in $b$ since $B(g,0) = B(g,1) = 0$, and the effect of $b$ is ambiguous in the interior.*

**Proof.** Good news trivially increase in both cases because more information gives more incentives, hence does $gF(c^*)$. Bad news decrease with $g$ since there are less low

quality in equilibrium, and it is detected with constant probability $b$. In turn, one has $\frac{\partial B}{\partial b} = 1 - F(c^*(g,b)) - bf(c^*(g,b))\frac{\partial c^*}{\partial b}$, which in general has an ambiguous sign, except for extreme $b$'s. ■

The main point in the proposition is that the effect of an increase in hostile monitoring on equilibrium stream of bad news is in general ambiguous. This is shown on figure 5 for the case of the uniform distribution with $g = \frac{1}{2}$.

One can also wonder what a specific increase of monitoring, namely neutral monitoring, implies in terms of the mix of good and bad news revealed in equilibrium. Let $m = g = b$ represents the intensity of monitoring. In this case, one has $G(m) \equiv B(m,m) = mF(m)$ and $B(m) = m(1 - F(m))$, since then $c^*(m,m) = m$ from the first proposition. It is immediate that $G(m)$ is increasing, but $B$ again is non-monotonic (in $m$). It is hump-shaped if and only if $\frac{d}{dm}[mF(m)] > 1$.

As come to a change in costs, the effect on news streams is rather intuitive. We summarize it in the next proposition, the straightforward proof is left to the reader.

**Proposition 6** *Suppose that $F$ is concave. All else equal, if costs decrease in the sense of First-Order Stochastic Dominance:*

1. *The quantity of equilibrium good news increases and the quantity of equilibrium bad news decreases.*

2. *The quantity of news generated in equilibrium $G(g,b) + B(g,b)$ increases in a friendly environment and decreases in a hostile environment.*

This result might be helpful in identifying the nature of an informational environment in case an exogenous shock on the costs is available. Suppose for the sake of illustration that a new technology appears that would help reduce CO2 emissions of cars, such as the hybrid engine. Then whether the total news flow on adoption of this new technology by

firms' fleet is increasing or decreasing when the technology spreads would indicate if the scrutiny is rather hostile or friendly. In a hostile environment, the news flow on the adoption of hybrid cars by firms would overall decrease, while it would overall increase in a friendly environment as the technology develops. Since it is sometimes hard to disentangle empirically whether a news is good or bad, this result helps in that it allows drawing inferences only on the basis of whether the topic is popular or not at a given date.

# 5 Applications and relation to the literature

## 5.1 Statistical discrimination

Models of statistical discrimination are used to explain group inequality. In the literature originated from the seminal contribution of Arrow (1973), average group differences are endogenously derived in equilibrium without assuming any ex ante exogenous differences between groups.[14] A prominent example is the model by Coate and Loury (1993) which describes the interaction between employers and groups of workers whose individual productivity is imperfectly observed. Discrimination then amounts to the existence of multiple equilibria, which implies that ex ante identical groups could exhibit different average level of qualifications ex post (and thus different wages). Our model sheds a new light on the multiple equilibria issue by giving sensible conditions for the equilibrium to be unique, and relates the multiplicity problem to the informational environment.

To explain this, we now put our analysis in a discrimination context. The agents are potential workers which belong to a given sexual or racial group. Individual agent's level of qualification can either be low ($a = 0$) or high ($a = 1$) and is endogenously chosen by making an investment $c$ in human capital for $a = 0$ and 0 for $a = 0$. The buyer is

[14] In contrast to Phelps (1972), where ability is on average different in different groups.

an employer who intends to hire workers. If the worker is qualified, the employer gets a return $\Delta$ and 0 if it is not. The problem is that the employer does not observe $a$ when making hiring decisions. But it can rely on a test which reveals the level of qualification with a probability $g$ if the agent is qualified and $b$ if it is not.[15] As we continue to assume that the agent extracts the total expected employer's return[16], the employer offers the wage $\Delta$ when the test reveals the agent is qualified, 0 when $a = 0$ is disclosed and $\mu\Delta$ when the test is not conclusive where $\mu$ is the probability assigned by the employer that the tested agent is actually qualified. Under these assumptions, Proposition 1 replicates the standard result of the statistical discrimination literature that there might multiple equilibria.

The literature on statistical discrimination does not convey any clear messages over the conditions under which discrimination occurs. Proposition 1 in Coate and Loury (1993, p. 1126) does give a necessary and sufficient condition for multiple equilibria, but the condition is not interpretable (and they do not actually try to interpret it). In contrast, our model invites to focus on the role of technology used to evaluate worker productivity in the labor market and we show that multiplicity can only occur in hostile environments. This may generate novel policy implications on the design of evaluation tests and procedures. They need to be more friendly in the sense we give to friendliness in this paper. More specifically, our results suggest that in hostile environments with discrimination:

- increasing $g$ is more efficient than reducing $b$ for it increases the average productivity (Proposition 2) and can eliminate discrimination.

- If resources is limited so that it is not possible to increase $g$ without reducing $b$, reducing discrimination is in line with increasing average qualification when $g + b$ is

[15] The literature uses a different information structures with soft information.

[16] Although not realistic in a labor market context, this assumption does not influence qualitatively the results.

low as suggested by Proposition 3. That is, when evaluation is difficult for intrinsic reasons or because resources are limited.

Our framework also gives insights on the effectiveness of other policy approaches. For instance, subsidizing education investments of discriminated groups (decreasing $c$ in our model) is less effective when evaluation hostility is not too severe (Proposition 4). At this stage, these points are not thoroughly established but they give interesting new directions of research.

## 5.2 Collective reputation

An interesting application of our model consists in reinterpreting it as a collective reputation model.[17] Firms with different types (costs) are pooled into a single group. While sometimes the results of quality investment by a given firm turns out to be individually observable (with probability $g$ or $b$, depending on whether the investment is high or low), it can escape pooling with other firms for which no signal is available. The collective reputation effect hence arises when no information is revealed, in which case all firms obtaining the null signal are treated equally, i.e. under the same pricing umbrella, and receive in equilibrium a price of $\mu^*\Delta$.

The seminal contribution in this area has been made by Tirole (1996) who describes a principal who contracts with an agent only if is sufficiently confident that the agent will not engage in corrupt activities. This a model in which, at each date $t$, the principal is matched with a new agent. There is imperfect information about the agent's past behavior: with a probability $b$, the principal knows that the agent has been found corrupt at least once. In addition the probability $b$ is higher when the agent has cheated more in the

[17] One could also think of it as a one player reputation problem. However some interpretations would require , especially as regards stability and externality across agents' type. See Bar-Isaac and Tadelis (2008) for a recent survey on reputation, framed precisely in a buyer/seller setup.

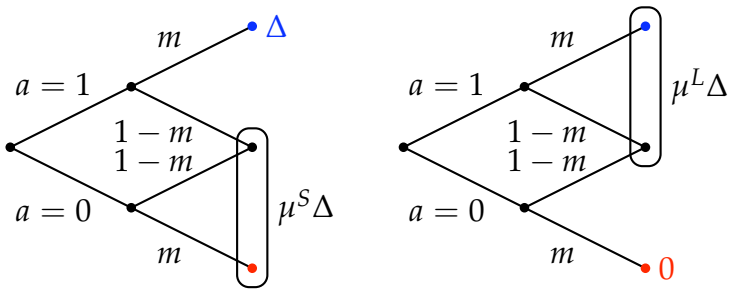
past. As a result, the agent trades-off the current benefit of corruption and the loss in reputation. The model then looks at the steady states of the model. Tirole (1996, proposition 1, p. 9) shows the existence of multiple steady states stemming from the dynamic complementarity between past and future reputations, implying that corruption may persist or, conversely, may be maintained at a low level. Our analysis shows that these results rest on the hostility of the informational disclosure mechanism: the principal can only learn something when the agent has been corrupt. Hostility implicitly originates here from the fact that the quality variable is conformity with a rule. Being found non compliant initiates a judicial process, which selectively discloses information to the general public on low quality.[18]

Note that, as Levin (2009) points out, the collective reputation model can be interpreted as a dynamic version of standard statistical discrimination models. Blume (2005); Baron (2006) provides a bridge between the two approaches by considering the dynamic version of a canonical statistical discrimination model.

## 5.3 Quality disclosure and certification

The model can also provide novel insights on quality certification (for a survey of this literature, see Dranove and Jin, 2010). Consider for instance the case of a product label which selectively signals high quality. The certifier decides to grant the label on the basis of a private evaluation, obtained with a given inspection ability. More precisely, assume that the certifier owns an imperfect but symmetric monitoring technology: with probability $m < 1$, it is fully informative on $a$, while with probability $1 - m$ it generates an empty signal. The monitoring technology is thus a priori neutral: it either uncovers the true quality or discloses nothing. A first possible labeling policy is *strict*. Under the strict

[18]Obviously, the process can also exonerate the defendant, but this outcome is arguably less frequent, and not accounted for in Tirole (1996).




Strict policy Lenient policy

Figure 6: Labeling policies.

policy, the certifier grants the label only if $a = 1$. The other possible rule is a *lenient* policy, which consists in labeling products only if the certifier does not observe $a = 0$. Suppose finally that firms first choose the quality of their product, then apply for certification, and that consumers observe only whether a product is certified or not. That is, labeling is a public coarsening of the (private) information of the certifier, as shown on Figure 6 We also assume for simplicity that the certifier is truthful. Under the strict policy, consumers will be sure that the quality of a labeled product is high. Under the lenient policy, consumers remain uncertain about labeled products quality,[19]. In turn, a lenient policy informs fully on non-labeled products, which are necessarily of low quality.

Our model predicts very different equilibrium qualities under the two labels:

- When the certifier uses a strict rule, the agent's expected profit opting for high quality is $\Pi(1, c) = [m + (1 - m)\mu^S]\Delta - c$ and $\Pi(0, c) = \mu^S\Delta$. Hence, the incentive constraint is $c \leq m(1 - \mu^S)\Delta$. Hence this it corresponds to a purely friendly environment with $g = m$ and $b = 0$.

- When the certifier uses a lenient rule, we have $\Pi(1, c) = \mu^L\Delta - c$ and $\Pi(0, c) = (1 - m)\mu^L\Delta$, implying the following incentive constraint $c \leq m\mu^L\Delta$. This corresponds to

[19] Harbaugh et al. (2011) deal with a related problem in which the it is the quality of monitoring inside the label that is uncertain. We consider here that the potential consumers know whether the label has a strict or a lenient policy.

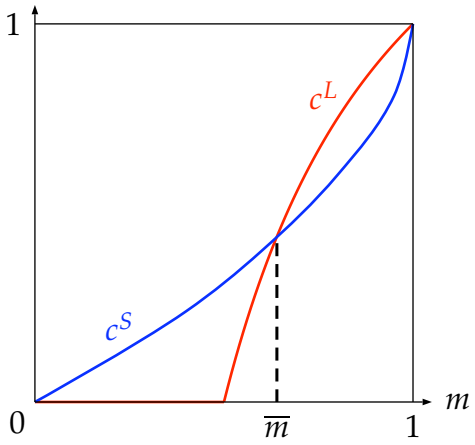



Figure 7: Equilibrium cutoffs under the two policies (uniform distribution).

a purely hostile environment with $g = 0$ and $b = m$.

Strict labels signal only high quality, and are thus associated with higher product prices. In turn, non-labeled products when the label is strict are of mixed quality, and their equilibrium price is intermediate. Lenient labels, on the other hand, filter out bad products ones, hence they allow the consumers to identify perfectly poor quality–hence non-labeled product trade at the lowest price. Indeed, prices for both labeled *and* non-labeled products are higher when the label is strict. What matters for overall incentives is the gap between the price of a labeled product and the price of a non-labeled product, and which gap is bigger depends on the quality of information $m$ and the cost distribution. A first immediate consequence of Proposition 1 is that with lenient labels, there may be multiple equilibria. The label can reach recognition and create strong incentives, or fall flat, and create only mild differences between labeled and non-labeled products.

Assume now that the cost distribution is concave, so that there is a equilibrium stable equilibrium under both policies (Corollary 2). Proposition 3 suggests the following policy implications: strict labels are socially preferable when evaluation is difficult (low $m$) and/or when the cost of high quality tends to be relatively high. Our model relates this to the friendliness or hostility that labeling does generate. Figure 7 illustrates this in the

case of the uniform distribution.

# 6 Conclusion

The model developed in this paper tackles an issue which has been thoroughly explored in the literature: the impact on sellers' incentives of asymmetric information on the quality they supply. We provide an original dichotomy of informational environments, between friendly environments, in which high quality is more easily identified by buyers than low quality, and hostile environments, in which this is the opposite.

We show that whether the informational environment is friendly or hostile makes a considerable difference. In particular, hostility gives rise to a bandwagon effect across agents which can generate in some circumstances multiple equilibria, which can not occur in friendly environments. The generic insight we uncover is that it is more effective to rely on the monitoring technology which can lead to the most substantial revisions of the buyers' equilibrium prior. For instance, hostility provides higher incentives for quality than friendliness when the cost of quality is low (which implies that buyers are optimistic over the average level of quality). Similarly, friendliness is more effective in increasing quality when information is poor, as this means little incentives to provide quality and thus pessimistic beliefs.

# 7 Appendix: Omitted Proofs

## 7.1 Proof of Proposition 3

We prove the points of the proposition in turn. Notice that since $F$ is concave the equilibrium is unique, up to the unstable $c^* = 0$ equilibrium in the limit case $g = 0$. Since for an arbitrarily small $g$ this equilibrium however disappear, we will ignore it, keeping in mind

that a strict $g = 0$ could be understood as arbitrarily small. Figure 8 is meant to illustrate the argument of the proof (note that it represents a case where a fully hostile environment is optimal).

First, we show that for any given $y \in [0,1]$, the highest $c$ such that $y = M(c, g, b)$ and $g + b = q$ belongs to a monitoring curve with either $g = b$ or $g = q$. Consider the inverse (in $c$) of the monitoring curve: $M^{-1}(y|g,b) = \frac{(g-b)y-g(1-b)}{(g-b)y-(1-b)}$. Along the budget constraint, we aim to maximize $M^{-1}(y, g, q-g)$. But $\frac{d^2}{dg^2}M^{-1}(y|g, q-g) = \frac{(2(2-q)^2y(1-y))}{(1+g-q+y(q-2g))^3)}$ which is always positive since $0 \leq y \leq 1$ and $0 \leq g \leq q < 1$. Hence $M^{-1}$ is convex, and its maximum is reached for an extreme $g$ for any given $y$. Now, the maximal equilibrium $c^*(g, q-g)$ has to be on a monitoring curve, by definition of $M$. Suppose it is not on an extreme monitoring curve. Then there exists an extreme monitoring curve which as we have seen is on the right of $c^*(g, q-g)$. Then since $F$ is increasing, the intersection of $F$ and this extreme monitoring curve has to be be on the right of $c^*(g, q-g)$, which implies that the corresponding equilibrium has higher average quality. This proves the first point: any optimal equilibrium has to be on one of the two extreme curves.

Second, consider now the point of intersection of these two extreme monitoring curves. In terms of $c$, it solves $M(c|q,0) = M(c|0,q)$. Since the first monitoring curve is decreasing and the second one is decreasing, the intersection is unique, i.e. there exists a unique $\hat{c}$ solving the equation, which is equal to $\hat{c} = \frac{q}{2-q}$. This implies that $M^{-1}(y|q,0) =$ is higher than $M^{-1}(y|0,q) =$ if and only if $y \leq F(\hat{c})$. This implies that the intersection of $F$ with an extreme monitoring curve happens either with $M(c|q,0)$ if $F(\hat{c})$ or with $M(c|0,q)$ otherwise, which yields the criterion in the second point.

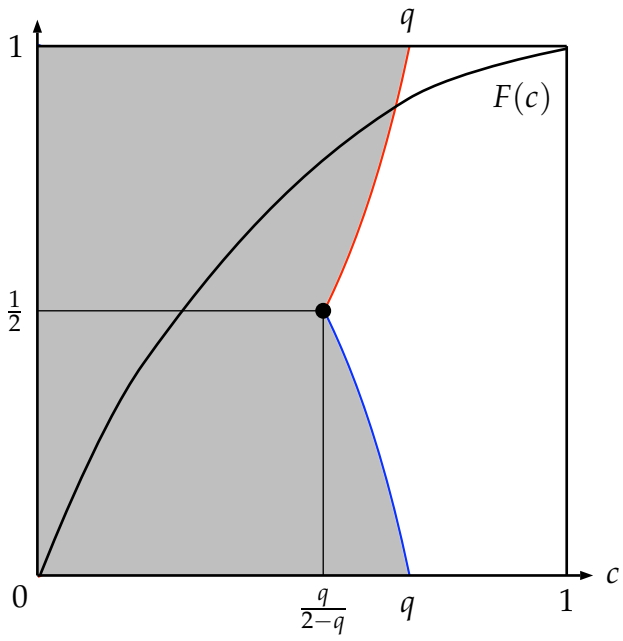



Figure 8: Illustration of the proof of proposition 3.

# Chapter 3

# The Informational Role of Non Governmental Organizations to Induce Self-Regulation: Cheering the Leaders or Booing the Laggards?

## English summary

Non-governmental organizations (NGOs) play a key role in creating incentives for firms to develop a Corporate Social Responsibility (CSR) policy by disclosing publicly self-regulatory corporate efforts. Their informational behavior is heterogeneous: some NGOs mostly disclose information on firms that do not behave responsibly (e.g., Greenpeace). Others are specialized in revealing firms that are socially or environmentally responsible (e.g., the Marine Stewardship Council). We develop a model describing the interactions between a NGO, a continuum of firms and a representative stakeholder, who values positively CSR, to explain what drives the NGO communication choice and its impact on the level of self-regulation.

We show that the NGO specializes in equilibrium: depending on the size of its budget, it either chooses to cheer the leaders or to boo the laggards. We extend the model to the case with multiple NGOs. We also introduce the possibility of NGO-corporate partnerships and derive policy implications.

## Résumé français

Les organisations non gouvernementales (ONGs), en exposant publiquement les efforts d'auto-régulation des entreprises, jouent un rôle clé dans leur incitation à développer des politiques de Responsabilité Sociale et Environnementale (RSE). Leurs comportement informationnel est hétérogène: certaines ONGs révèlent surtout des informations sur les entreprises irresponsables (e.g. Greenpeace). D'autres se spécialisent dans la mise en avant d'entreprises vertueuses (e.g. le Marine Stewardship Council). Nous développons un modèle qui décrit les interactions entre une ONG, un continuum d'entreprises et une partie prenante représentative en faveur de la RSE afin d'expliquer ce qui conduit les choix de communication des ONGs et leurs impacts sur le niveau d'auto-régulation.

Nous montrons que l'ONG se spécialise à l'équilibre: selon l'importance de son bud-

get, elle choisit d'acclamer les meilleurs ou de huer les mauvais élèves. Nous complétons notre analyse en envisageant la présence de multiples ONGs. Nous introduisons aussi la possibilité pour les ONGs de former des partenariats avec les entreprises et en dérivons des recommandations.

# 1 Introduction

Firms frequently take socially and environmentally friendly actions not required by law, thereby privately providing public goods, or voluntarily internalizing externalities. For example, the oil company Total finances the rehabilitation of the Anggana mangrove forest in the Mahakam Delta in Indonesia. The chemical and pharmaceutical company Bayer develops programs to promote employees gender diversity. Komatsu[1] seeks to address employee mental health concerns. The Coca-Cola Company has implemented a comprehensive corporate policy, including quantified objectives regarding packaging recycling, water stewardship, energy conservation, etc. Albeit ill-defined, the concept of Corporate Social Responsibility (CSR) is a convenient umbrella term to designate such activities.

As self-regulatory efforts are arguably costly, the reasons why corporations are willing to self-regulate have been extensively explored in the economic literature. Many works stress the fact that some stakeholders are willing to reward CSR leading firms or, alternatively, to punish laggards. For example, some consumers may accept to buy their products at a higher price or boycott 'dirty' corporations (Bagnoli and Watts, 2003; Arora and Gangopadhyay, 1995). Employees may work in a more productive way, or they may accept lower wages in environmentally- or socially-responsible firms (Brekke and Nyborg, 2008). Socially responsible investors may propose capital to leading firms at a reduced cost (Heinkel et al., 2001).

The problem is that self-regulation is a credence good whose benefit is impossible for an individual consumer, employee or investor to ascertain. It provides the rationale for the existence of specialized actors that have sufficient resources to observe self-regulatory activities of individual firms and to convey this information to the stakeholders. Without such a monitoring, stakeholders are not able to reward individual efforts or punish laziness, thereby reducing self-regulation in the first place.

[1] A multinational corporation that manufactures construction, mining, and military equipment.

This paper focuses on NGOs. Beside labeling organizations or rating agencies, they now play a key role in monitoring and communicating over CSR corporate efforts (Feddersen and Gilligan, 2001).

The starting point of the paper is the observation that, in the real world, NGOs have heterogeneous informational behaviors. Some NGOs are specialized in the transmission of good news: they certify that a firm – or a product – is socially or environmentally responsible. An illustration is the Marine Stewardship Council that rewards sustainable fishing by certifying about 12 percent of the world catch intended to human consumption. Other NGOs mostly send bad news: Greenpeace is the emblematic example of an organization that fiercely conducts campaigns against firms that they do not deem to behave responsibly. Lie in between neutral NGOs which send both good and bad news, as for example WWF which alternates between cooperative and offensive behavior toward firms depending on their compliance with their demands. In the following, we refer to the NGOs which cheer CSR leaders as good cops whereas bad cops are NGOs which boo the laggards. Note that this heterogeneity may reflect selective monitoring - the NGO focuses on the observation of positive activities - or selective disclosure - the NGO chooses to communicate only some of its monitoring results.

In this paper we seek to understand when and why NGOs choose between sending good news or bad news. Another objective of the paper is to study how differences in NGO communication shape the firms' incentives to self regulate.

To answer these questions, we build a simple model with asymmetric information that describes the interaction between a continuum of firms that may individually decide to self-regulate ($a = 1$) or not ($a = 0$), a stakeholder who is willing to pay for self-regulation and a NGO that seeks to maximize the average level of self-regulation under a budget constraint and that provides the stakeholder with imperfect information on $a$. We also extend the analysis to a setting with multiple NGOs.

Information disclosure by the NGO is asymmetric. More specifically, the NGO observes action and conveys this information to the stakeholder with a probability $g$ when $a = 1$ (good news) and with a probability $b$ when $a = 0$ (bad news). The parameters $g$ and $b$ are endogenously selected by the NGO which can decide to be neutral ($g = b$), a good cop ($g > b$) or a bad cop ($g < b$). We use the model to investigate how the NGO disclosure probabilities $(g, b)$ shape firms' behavior and then to characterize the probabilities chosen in equilibrium by the NGO.

We show that the NGO decides to specialize: If its budget is limited, it chooses to be a good cop ($g > 0$ and $b = 0$). When more resources are available, it opts for being a bad cop ($b > 0$ and $g = 0$). The intuition is that it is more effective to go against the stakeholder expectations: A low budget limits disclosure probabilities, implying low firms' incentives to self-regulate. Being aware of this, the stakeholder holds pessimistic beliefs on the expected level of self-regulation. By filtering good firms out, a good cop improves the belief about firms whom she receives no news about.

We then extend the analysis by assuming multiple NGOs which non cooperatively select their technology. It gives birth to a coordination problem whereby NGO choices are biased towards friendliness. We also study the potential role of a welfare-maximizing regulator and we introduce the opportunity for firms to form partnerships with NGOs.

The economic literature on self-regulation has already explored the role of NGOs and activists in triggering corporate socially- and environmentally-responsible investments. For instance, "Good Cop/Bad Cop" is the title of a recent book edited by Lyon (2010) which contrasts the heterogeneity of NGO strategies towards business. But none of the contributions in this book develops a theoretical analysis of what drives such strategies. There exist theoretical works which specifically deal with bad cops. For instance, Baron (2001) and Lyon and Maxwell (2011) investigate the impact of NGOs which are able to penalize firms which do not make sufficient environmental or social efforts. Others deal

with good cops, and in particular with the role of NGOs in product labeling (ecolabels, fair trade, etc.). For example, Bottega and De Freitas (2009) compare two certifiers: a NGO and a for-profit organization. None of these papers deal with the different types of NGO in a unified framework as we do in this paper.

The paper is structured as follows. We present the base model with a single NGO in the next section. We extend to the case with multiple NGOs and we address regulation issues in the third section. In the fourth section, we analyze the role of corporate-NGO partnership. We summarize the results and conclude in the last section.

# 2 The base model

## 2.1 Assumptions

We consider a continuum of firms that may self-regulate or not. Each firm makes a binary decision: $a = 1$ if it self-regulates, $a = 0$ if it carries on with business-as-usual. Self-regulation costs $c$, which is heterogeneous across firms and uniformly distributed over $[0;1]$. Assuming that $c$ is positive is in line with the very nature of self-regulation: it consists in improving social and environmental performance beyond the business-as-usual. That is, once all actions at negative cost have been implemented. The assumption that $c \leq 1$ is irrelevant qualitatively. It simplifies the notations. Uniformity of distribution ease interpretation as it allows obtaining closed-form expressions. It also leads to rule out multiple equilibria which can arise in hostile informational environments (see Fleckinger et al., 2012).

We assume the existence of a stakeholder who has a (marginal) willingness to pay $w_1 \geq 1$ for self-regulation and $w_0 \leq 0$ for business-as-usual.[2] But the stakeholder is not able to observe $a$ on her own and bases her decision to reward or not the firms on the

[2] At some point, we normalize $w_1 - w_0$ to 1, which does not alter our results qualitatively

information provided by the NGO. More specifically:

- In the case where a firm chooses $a = 1$, the NGO discloses the value of $a$ with a probability $g$. With a probability $1 - g$, no information is generated about the firm's action. The stakeholder believes the news (hard information). We rule out any potential concerns about the NGO credibility.

- In the case where $a = 0$, the disclosure probability is $b$. With probability $1 - b$, she receives no news.

If the stakeholder learns that $a = 1$, she transfers $w_1$ to that firm. Hence, we assume that the firm is able to fully extract the stakeholder's surplus.[3] Conversely, she punishes the firm with a negative transfer $w_0$ when she learns that $a = 0$. When she receives no news about the action of a given firm, she forms a belief $\mu$ that the firm self-regulates and transfers $\mu w_1 + (1 - \mu)w_0$ to the firm. The stakeholder is sophisticated in that she relies on Bayes' rule to form her belief.

The disclosure technology is thus fully described by the probabilities $(g, b)$. These are endogenously chosen by the NGO, which can decide to be neutral $(g = b)$, a good cop $(g > b)$, or a bad cop $(g < b)$. However, the NGO has limited resources which prevent to obtain and/or convey perfect information about $a$. More specifically, we introduce the assumption that $g + b \leq \alpha$.[4]

The timing of the game is as follows:

1. The NGO selects the disclosure probabilities $(g, b)$.

2. Each firm decides to self-regulate $(a = 1)$ or not $(a = 0)$.

[3] We could assume that the firms could only extract a share, say $\lambda$, of the surplus. This would not affect qualitatively the results.

[4] As we will see below the linearity of this resource constraint simplifies the analysis by generating simple corner solutions. But relaxing this assumption would not qualitatively alter the results.

3. The NGO discloses the value of $a$ with probabilities $g$ and $b$ if $a = 1$ and $a = 0$, respectively.

4. The stakeholder transfers $w_0$, $w_1$ or $\mu w_1 + (1-\mu)w_0$ depending on the messages received.

Note that, if information on $a$ were perfect, all firms would self-regulate, which is socially optimal as $c \leq w_1$ for all $c$. Our analysis explore how information asymmetry prevents from reaching this first best outcome. In this respect, the fact that the NGO has limited resources is crucial. Without this hypothesis, the NGO would provide perfect information ($g = b = 1$), inducing self-regulation by all firms.

We now proceed in two steps. First, we characterize firms' response to a given monitoring technology $(g, b)$. Second we identify the technology selected by the NGO which seeks to maximize CSR under the constraint $g + b = \alpha$.

## 2.2 Firm's choice

Consider a firm of type $c$. Its expected payoff is $\Pi(a = 1) = gw_1 + (1-g)\left(\mu w_1 + (1-\mu)w_0\right) - c$ under self-regulation while its payoff is $\Pi(a = 0) = bw_0 + (1-b)\left(\mu w_1 + (1-\mu)w_0\right)$ otherwise. Hence, the firm self-regulates if:

$$c \leq (w_1 - w_0)((g - (g-b)\mu)$$

An almost immediate consequence of this incentive constraint is that the CSR equilibrium is characterized by a cost threshold below which firms self-regulate and above which they do not. Let $c^*$ denote that threshold and normalize $w_1 - w_0 = 1$ to simplify notations. We have

$$c^* = g - (g-b)\mu \tag{3.1}$$

The belief that is consistent in the Bayesian sense with this cutoff is then:

$$\mu^* = \frac{(1-g)c^*}{(1-g)c^* + (1-b)(1-c^*)} \tag{3.2}$$

The two equations (3.1) and (3.2) define the Bayesian equilibrium.

It will prove more convenient to adopt a standard fixed point representation of the equilibrium to investigate equilibrium existence properties. Combining (3.1) and (3.2), the cutoff equilibrium satisfies:

$$c^* = \Phi(c^*) \quad \text{where} \quad \Phi(c) \equiv \frac{g(1-b) - (g-b)c}{(1-b) - (g-b)c} \tag{3.3}$$

We establish properties of $\Phi$ that will be useful in the following:

**Lemma 4** *1)* $\Phi(0) = g$, $\Phi(1) = b$. *2)* $\Phi$ *is increasing if* $g > b$ *and decreasing if* $g \leq b$. *3)* $\Phi$ *is concave.*

**Proof.** 1) Obvious. 2) Differentiation of $\Phi$ yields $\frac{d\Phi}{dc}(c) = -\frac{(g-b)(1-g)(1-b)}{[1-b-(g-b)c]^2}$, which is positive if $g \leq b$ and negative otherwise. 3) We have $\frac{d^2\Phi}{dc^2}(c) = \frac{(g-b)^2(1-g)(1-b)}{[(g-b)c-(1-b)]^3}$, which has the same sign as that of $(g-b)c-(1-b)$. This expression is obviously negative when $g-b \leq 0$. When $g-b > 0$, it is maximized in $c = 1$ and $(g-b)c-(1-b) = -(1-g) < 0$ in this case. ■

Then we use these properties to draw the function $\Phi$ in Figure 1. In this graph, any equilibrium corresponds to the intersection of $\Phi$ with the 45 degree line. When $g > b$ (see Figure 1, on the left) there exists a unique equilibrium as $\Phi$ is decreasing, starts above the 45 degree line ($\Phi(0) = g > 0$) and ends up below ($\Phi(1) = b < 1$).

In the case where $g < b$, $\Phi$ is increasing. If $g > 0$, we have $\Phi(0) = g > 0$, meaning that $\Phi$ starts above the 45 degree line in $c = 0$. It is the case depicted in Figure 1, on the

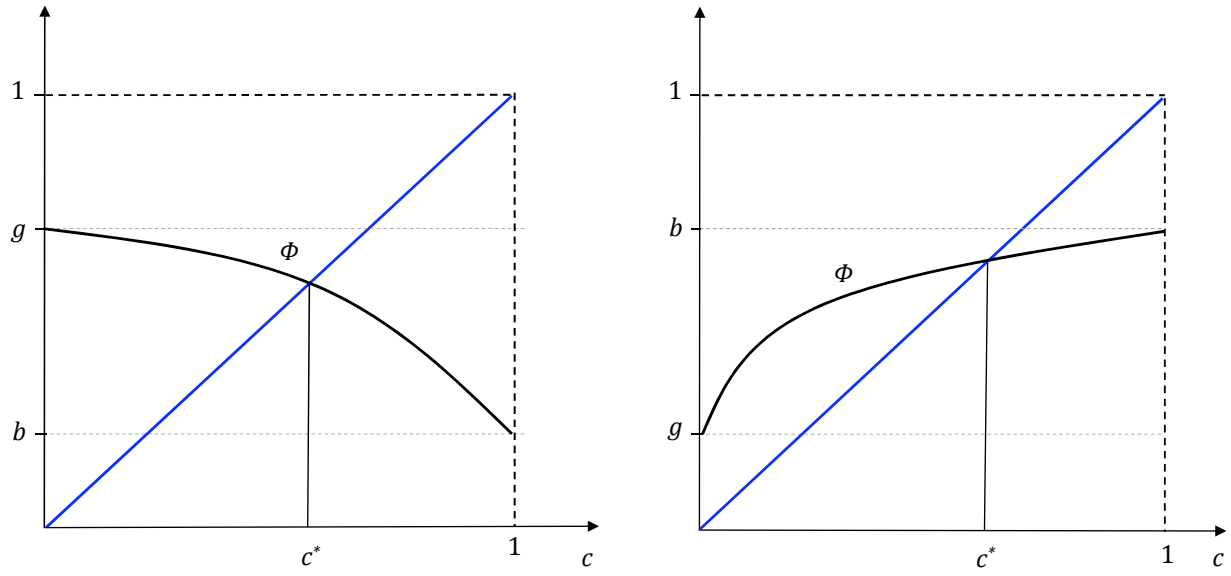



Figure 1: **Good cop** $(g > b)$ **versus bad cop** $(g < b)$ **equilibria**

right. Property 1 and the intermediate value theorem ensures existence. The equilibrium is also unique: as $\Phi$ crosses the 45 degree line from above in $c^*$, concavity of $\Phi$ ensures it will never cross it for higher values of $c$.

Things are slightly more complicated when $g = 0$. To begin with, $c = 0$ is always an equilibrium ($\Phi(0) = 0$). If $d\Phi/dc(0) \leq 1$, which is equivalent to $b \leq \frac{1}{2}$, this equilibrium is unique for $\Phi$ is concave as argued before. When $d\Phi/dc(0) > 1$, there will be a second equilibrium as $\Phi$ is higher than $c$ when deviating upward from zero.

We now summarize our findings.

**Proposition 7 (Existence)** *There always exists a (subgame) equilibrium which is defined by the condition (3.3). But the equilibrium is not necessarily unique. More precisely:*

- *When $g > 0$, there exists a unique equilibrium $c^*$ which is strictly positive. $c^*$ is defined by (3.3).*
- *When $g = 0$ and $b \leq \frac{1}{2}$, the unique equilibrium is $c^* = 0$.*

- *When $g = 0$ and $b > \frac{1}{2}$, there exist two equilibria. The first is $c^* = 0$ and the second is strictly positive.*

In the following, we rely on simulations to establish some of our results in order to avoid cumbersome calculations. We thus need an explicit expression of $c^*$ in the case where it differs from zero.

**Lemma 5** *Equilibrium with a strictly positive amount of CSR is defined as follows:*

- *When $g = b$, we have $c^* = b = g$.*
- *When $g \neq b$ :*

$$c^* = \frac{1}{2(g-b)}\left(g - 2b + 1 - \sqrt{(g-2b+1)^2 - 4g(1-b)(g-b)}\right)$$

**Proof.** See in Appendix. ■

## 2.3 NGO choice

Our objective is now to characterize the disclosure probabilities $(g^*, b^*)$ that will be selected by the NGO, given the firms' response. Formally, the NGO solves:

$$\max_{g,b} c^*(g,b) \text{ subject to } g + b = \alpha \tag{3.4}$$

where $c^*$ is the equilibrium defined by Proposition 1.

The possibility that the NGO chooses $g^*$ and $b^*$ such that $c^* = 0$ is excluded since it precisely seeks to maximize $c^*$.[5] We can thus concentrate on the equilibrium characterized in Lemma 2.

[5]This also applies to the zero equilibrium when it coexists with the positive equilibrium in the case where $g = 0$ and $b > 1/2$: as soon as $g$ deviates upward from zero with an infinitesimal amount, this equilibrium disappears.

We first run simulations to show that there is no interior maximum. The results are displayed in Figure 2 which plots $c^*$ as a function of $g$ for different values of $\alpha$. We can see that either $c^*$ increases with $g$ when $\alpha$ is low or the curve is U-shaped with an interior minimum.

Having ruled the existence of interior solutions out, we compare the two corner solutions ($g = \alpha$ and $g = 0$) in Appendix. We obtain a new proposition.

**Proposition 8** *The NGO chooses to be a bad cop with $g = 0$ and $b = \alpha$ if $\alpha > \frac{2}{3}$. Otherwise it chooses to be a good cop ($g = \alpha$ and $b = 0$).*

**Proof.** See in Appendix. ■

Let us now discuss the conditions $\alpha \gtrless \frac{2}{3}$. A high $\alpha$ means that the NGO produces a lot of information about firms' behavior. Incentives for firms to invest in self-regulation are thus high, implying that the stakeholder holds an optimistic belief that one firm about whom she does not receive any news is actually self-regulating. In this context, it is more effective for the NGO to be hostile in order to deter firms from shirking on this optimistic state of mind. It mitigates their incentives to free ride on the lack of information. And conversely when the stakeholder is pessimistic about the level of self-regulatory efforts (because $\alpha$ is low), it is more effective to motivate firms by increasing the probability $g$ to get the full reward when self-regulating rather than the probability $b$ to get close nothing because of the stakeholder's low belief when exerting no effort. The general point here is that the best strategy consists in going against the stakeholder's a priori built upon the environment characteristics captured by the parameter $\alpha$[6].

As the NGO maximizes the average level of self-regulation and as self-regulation improves social welfare, note that:

[6]One should not give too much importance to the fact that we only obtain corner solutions for it is due to the linear budget constraint $g + b = \alpha$. Other functional forms would probably lead to interior solutions, but this would not modify qualitatively the result that the NGO does not select $g = b$ and the nature of the influence of the environment characteristics.

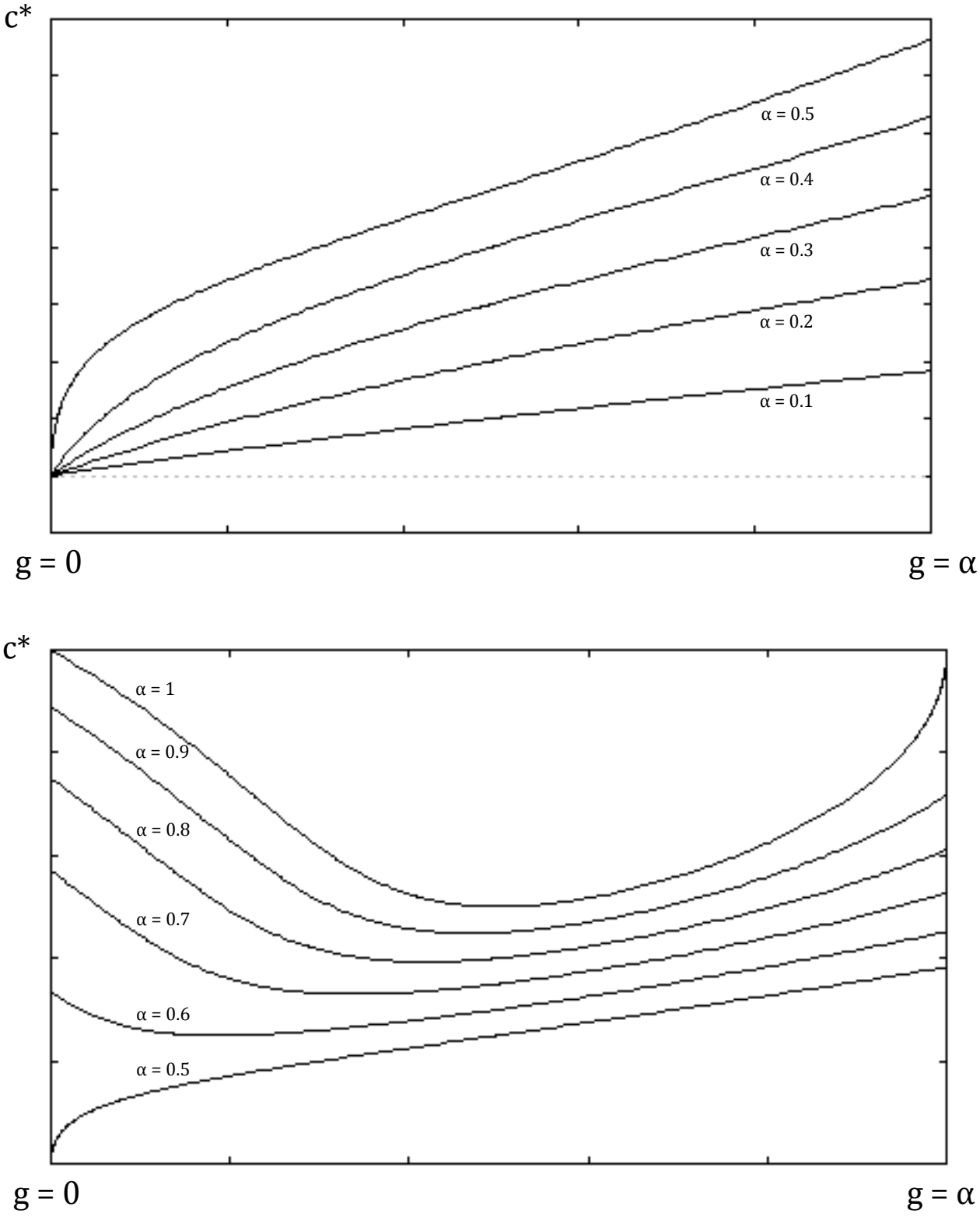



Figure 2: $c^*$ **as a function of** $g$ **for different value of** $\alpha$

**Corollary 3** *The NGO choice is socially optimal.*

# 3 Multiple NGOs

In reality, firms are under the scrutiny of several NGOs. That is why we extend the base model to account for the existence of $n$ NGOs. Each has the same budget $\alpha$ ($\alpha < 1$). Let $g_i$ and $b_i$ denote the values of $g$ and $b$ selected by the NGO $i$ ($i = 1, ..n$). Each NGO has the budget constraint $g_i + b_i = \alpha$. The overall probability $G$ that a firm with $a = 1$ is observed by at least one NGO is thus:

$$G(g_1, ..g_i, ..g_n) = g_i + (1 - g_i)\left(1 - \prod_{j \neq i}(1 - g_j)\right) \tag{3.5}$$

That is, either the firm is observed by the NGO $i$ or it does not, but at least another NGO $j \neq i$ does so.[7] Similarly, the probability that $a = 0$ is disclosed is:

$$B(b_1, ..b_i, ..b_n) = b_i + (1 - b_i)\left(1 - \prod_{j \neq i}(1 - b_j)\right). \tag{3.6}$$

The level of self-regulation is still defined by Proposition 1 except that we replace $g$ and $b$ by the aggregate probabilities $G$ and $B$.

## 3.1 Best reply

The best reply of the NGO $i$ to the $n - 1$ other NGOs strategies is the couple $(g_i, b_i)$ which solves:

$$\max_{g_i, b_i} c^*\left(G(g_1, ..g_i, ..g_n), B(b_1, ..b_i, ..b_n)\right) \text{ subject to } g_i + b_i = \alpha$$

[7] $\prod_{j \neq i}(1 - g_j)$ is the probability that none of the $n - 1$ other NGOS discloses $a = 1$. Hence, the probability that at least one discloses it is $1 - \prod_{j \neq i}(1 - g_j)$.

Note that:

$$\begin{aligned}\frac{\partial G}{\partial g_i} &= \prod_{j\neq i}(1-g_j) > 0,\\ \frac{\partial B}{\partial b_i} &= \prod_{j\neq i}(1-b_j) > 0.\end{aligned}$$

These properties simplify the problem as raising $g_i$ or $b_i$ basically means raising $G$ or $B$. The aggregate probability $G$ will be equal to $G(g_1,..,g_{j-1},0,g_{j+1}..g_n)$ if the NGO selects $g_i = 0$ and to $G(g_1,..,g_{j-1},\alpha,g_{j+1}..g_n)$ if $g_i = \alpha$. Hence, selecting $g_i$ amounts to select a value of $G$ in the interval $A_i \equiv \left[G(g_1,..,g_{j-1},0,g_{j+1}..g_n), G(g_1,..,g_{j-1},\alpha,g_{j+1}..g_n)\right]$. Consider then Figure 2. $A_i$ is a segment of the curves that are depicted in this graph. In the case where $c^*$ is increasing over the whole interval $A_i$, the NGO $i$ is better off by increasing $g_i$ as much as possible. Hence, it chooses $g_i = \alpha$. In the case where it is decreasing, $g_i = 0$. In the case where the curve is U-shaped over $A_i$, the NGO's best reply is either $g_i = 0$ or $g_i = \alpha$, depending on the precise location of $A_i$ on the curve.

The common point to every case is that the best reply is a corner choice whatever the others' strategies. The last step of this analysis consists in identifying the precise condition that defines the best of the two candidate corner solutions. We obtain the following:

**Lemma 6** *All NGOs have the same best reply function. If*

$$\frac{\alpha + \left(2 - \prod_{j\neq i}(1-g_j) - \prod_{j\neq i}(1-b_j)\right)(1-\alpha)}{2-\alpha} > \frac{\prod_{j\neq i}(1-b_j)}{\prod_{j\neq i}(1-g_j) + \prod_{j\neq i}(1-b_j)}, \tag{3.7}$$

*the best reply is to be a bad cop ($g_i = 0$ and $b_i = \alpha$). Otherwise, they prefer to be good cops ($g_i = \alpha$ and $b_i = 0$).*

**Proof.** See in Appendix. ■

## 3.2 Nash equilibrium

Having identified the best reply function of each NGO in Lemma 3, we now identify the resulting Nash equilibrium. The intermediate cases where some NGOs would opt for being a good cop and others for being a bad cop can be easily excluded. It would result in an equilibrium value of $G$ strictly higher than $G(0,..,0)$ and strictly lower than $G(\alpha,..,\alpha)$. But, given the function $c^*$ depicted in Figure 2, some NGOs would then have an incentive to deviate towards one of the corner values of $G$. We can thus concentrate the analysis on the corner cases.

Consider first the equilibrium in which all NGOs choose to be bad cops: $g_i = 0, b_i = \alpha$. Plugging these values in (3.5) and (3.6), the resulting overall probabilities are $G = 0$ and $B = 1 - (1-\alpha)^n$. Hence, the constraint (3.7) becomes:

$$1 > 2(1-\alpha)^n + (1-\alpha)^{2n-1}$$

Note that the right-hand side decreases with $\alpha$. Furthermore, the inequality is satisfied when $\alpha = 1$ and not when $\alpha = 0$. Hence, there exists a unique value of $\alpha \in (0,1)$, denoted $\alpha_{\text{lim}}$, such that:

$$1 \equiv 2(1-\alpha_{\text{lim}})^n + (1-\alpha_{\text{lim}})^{2n-1} \tag{3.8}$$

For all the values of $\alpha$ such that $\alpha > \alpha_{\text{lim}}$, there exists a Nash equilibrium where all the NGOs choose to be a bad cop.

Turning next the equilibrium where all NGOs choose to be good cops ( $g_i = \alpha, b_i = 0$), the best reply condition (3.7) is:

$$1 - (1-\alpha)^n < \frac{(1-\alpha)+1}{(1-\alpha)^{n-1}+1}$$

It is easy to see that this condition is always satisfied as the left-hand side is lower than 1

whereas the right-hand side is more than 1. We can thus conclude.

**Proposition 9** *The equilibrium depends on whether the NGOs' budget is higher or lower than a threshold $\alpha_{\text{lim}}$ defined by (3.8).*

- *If $\alpha < \alpha_{\text{lim}}$, there exists only one Nash equilibrium where all NGOs choose to be good cops.*
- *If $\alpha \geq \alpha_{\text{lim}}$, there exist two equilibria. In the first one, all NGOs are good cops. In the second, all are bad cops.*

This proposition highlights a coordination problem that arises when there are a lot of NGOs ($\alpha_{\text{lim}}$ decreases with the number of NGOs) with a confortable budget. The equilibrium is thus not always socially efficient contrary to the single-NGO case (see Corollary 1). This creates the opportunity for a welfare-improving public intervention. The next proposition presents a simple decision rule to help NGOs to select the Pareto dominant equilibrium.

**Proposition 10** 1. *If $\alpha < \alpha_{\text{lim}}$, public intervention is not necessary.*

2. *If $\alpha_{\text{lim}} \leq \alpha \leq 1 - 3^{-\frac{1}{n}}$, the regulator should help the NGOs to select the good cop equilibrium.*

3. *If $\alpha > 1 - 3^{-\frac{1}{n}}$, it should help to select the bad cop equilibrium.*

**Proof.** See in Appendix. ■

Subsidizing NGOs is another way of improving social welfare. It amounts to increasing the budget $\alpha$ available to each NGO, which in turn leads to higher disclosure probabilities $G$ or $B$. Ignoring the opportunity cost of such subsidies, welfare obviously improves as higher probabilities raise the level of self-regulation:

**Proposition 11** *Subsidizing NGOs improves social welfare as it raises disclosure probabilities and thus the level of self-regulation* ($\frac{dc^*}{dG} > 0$, $\frac{dc^*}{dB} > 0$).

**Proof.** See in Appendix. ■

# 4 Corporate-NGO partnership

In addition, in the real world firms have the opportunity to develop partnerships with NGOs. They transfer resources to some organizations, which are used to finance monitoring and reporting activities, resulting in better information on firms' behavior. In this section, we introduce in our setting the possibility for firms to enter partnerships and we examine how it influences the results obtained so far.

A NGO may commit to publicize specifically the partnering firm's social or environmental performance. For instance, the NGO grants a product label. In other cases, the firm contributes with a donation, enabling the NGO to increase its overall activities, but without targeting specifically its donors. In the following, we refer to the second scenario as a collective partnership as the increase in disclosure probabilities concerns all firms whereas the first is referred to as an individual partnership. We successively analyze these two forms.

## 4.1 Collective partnership

Under a collective partnership, a firm's contribution leads to an increase in the overall probabilities $G$ or $B$. We pose the problem as follows. In the status quo, $n$ NGOs with budget $\alpha$ have selected their communication strategy. We assume the equilibrium is socially efficient (because they were able to coordinate, eventually with a regulator's support). Hence if $\alpha < 1 - 3^{-\frac{1}{n}}$, they are all good cops. Otherwise, they are all bad cops.

We then study whether an individual firm's equilibrium profit increases with disclosure probabilities $G$ or $B$. To begin with, we establish a Lemma describing the marginal effect of $G$ and $B$ on equilibrium profits.

**Lemma 7** *We have:*

1. $\frac{d\pi(a=1)}{dB} > 0$ *and* $\frac{d\pi(a=1)}{dG} > 0$
2. $\frac{d\pi(a=0)}{dB} > 0$ *and* $\frac{d\pi(a=0)}{dG} > 0$ *if* $\alpha < 1 - 3^{-\frac{1}{n}}$ *and* $\leq 0$ *otherwise.*

**Proof.** See in Appendix. ■

We can immediately derive that:

**Proposition 12** *If the status quo is the good cop equilibrium ($\alpha < 1 - 3^{-\frac{1}{n}}$), all firms are willing to enter collective partnerships. If it is the bad cop equilibrium ($\alpha \geq 1 - 3^{-\frac{1}{n}}$), the self-regulating firms are the only potential contributors.*

This result is very intuitive: firms which do not self-regulate do not contract with bad cops. Note however that the bad cop equilibrium emerges only when the average level of CSR is high (because $\alpha \geq 2/3$), meaning that the share of firms which does not self-regulate is low in relative terms. Note also that this analysis only looks at firms' preferences, ruling out free riding issues stemming from the fact that the benefit of collective partnership is a public good. Hence, the proposition only gives a necessary condition for the existence of collective partnerships.

## 4.2 Individual partnership

We model individual partnership as follows: Each firm has the option to contract with a NGO which commits to disclose the firm's action with a probability equal to one[8] in

[8]Note that the distinction between a good cop and a bad cop is not relevant anymore when the NGO is engaged in a partnership. If the NGO is initially a good cop which increase its probability to $g = 1$. This implies that receiving no news means $a = 0$ for the stakeholder (as if $b = 1$). Conversely, if the NGO is a bad cop commiting to $b = 1$, receiving no news means $a = 1$ (as if $g = 1$).

exchange of a transfer $t$. Let $s$ denote a binary variable equal to one if the firm enters a partnership and 0 it it does not. We assume the stakeholder observes $s$. To simplify notations, we also assume $w_0 = 0$ and $w_1 = 1$ (hence, we still have $w_1 - w_0 = 1$).

In this new version of the game, a firm's strategy is the pair $(a, s)$ and its payoff is

$$\Pi(a,s) = \begin{cases} G + (1-G)\mu_1 - c & \text{if } a = 1,\ s = 0 \\ 1 - c - t & \text{if } a = 1,\ s = 1 \\ (1-B)\mu_2 & \text{if } a = 0,\ s = 0 \\ -t & \text{if } a = 0,\ s = 1 \end{cases}$$

Note that beliefs can be different when the strategy pair is $(a = 1, s = 0)$ and $(a = 0, s = 0)$. We can immediately rule out the strategy $(a = 0, s = 1)$ because it is strictly dominated by $(a = 0, s = 0)$ : it is irrational to costly reveal $a = 0$.

Consider now a firm that self-regulates $(a = 1)$. The firm has an incentive to contract with a NGO if $G + (1-G)\mu_1 - c < 1 - c - t$. Hence, if $t < (1-G)(1-\mu_1)$. This condition does not depend on $c$; it implies that either all self-regulating firms decide to signal their action or none of them does so. It simplifies the stakeholder's Bayesian updating: as soon as she observes one partnership, she knows that entering a partnership is profitable for all firms with $a = 1$. Reciprocally, all firms without partnership have thus chosen $a = 0$. In fact, as soon as there is at least one partnership, the belief $\mu_2$ is equal to 0. In this context, each firm decides to self-regulate if its payoff with self-regulation and partnership, $1 - c - t$ is positive (the payoff without self-regulation and without partnership is equal to zero). We get the classical unravelling separating equilibrium in which firms choose $(a = 1, s = 1)$ if $c \leq 1 - t$ and $(a = 0, s = 0)$, otherwise.

Let us now come back to the case where self-regulating firms do not signal their action $(t \geq (1-G)(1-\mu_1))$. As firms with $a = 0$ does not contract with a NGO, the stakeholder does not observe any partnership. It is the standard case without partnership analyzed in

the previous sections: firms choose to self-regulate if $c < c^*$ with $c^*$ defined by (3.1) and $\mu_1 = \mu^*$ defined by (3.2). We now summarize these results.

**Lemma 8** *When firms have the option to contract with a NGO such that its disclosure probability becomes one at cost $t$, two equilibria are possible:*

1. *If $t \geq (1-G)(1-\mu^*)$, none of the firms contract with a NGO and firms with type $c < c^*$ choose to self-regulate with $c^*$ defined by (3.1).*

2. *If $t < (1-G)(1-\mu^*)$, the firm of type $c \leq 1-t$ self-regulates and contract with a NGO. The others do nothing.*

In the second case, we get a separating equilibrium which fundamentally differs from the equilibrium obtained in the base model. It is easy to show that:

**Proposition 13** *The average level of self-regulation is higher when firms have the possibility to enter individual partnerships.*

**Proof.** Remember that NGOs choose to be either all good cops ($g = \alpha, b = 0$) or all bad cops ($g = 0, b = \alpha$) in the base model. If they all choose to be good cops, we have (from 3.1) that $c^* = \alpha(1-\mu^*)$. Besides, we know that $t < (1-\alpha)(1-\mu^*)$, thus $1 - t > \alpha - \alpha(1-\mu^*)$.

If they all choose to be good cops we have (from 3.1) that $c^* = \alpha\mu^*$. Besides, we know that $t < (1-\mu^*)$ which implies $1 - t > \mu^* > c^*$. ■

The proposition is very intuitive as partnerships basically increase the amount of information on self-regulation. When are then individual partnerships more likely to occur? Answering the question requires to interpret the threshold $(1-G)(1-\mu^*)$. In the bad cops equilibrium ($\alpha < 1 - 3^{-\frac{1}{n}}$), the threshold simplifies to $1-\mu^*$. When establishing Lemma 5, we have seen that $\mu^*$ was increasing with $B$. Hence, the higher the resources

available to the NGOs, the lower this threshold, and thus the less likely individual partnerships. In the good cops equilibrium, we have seen that $\mu^*$ was also increasing with $G$. Hence, the threshold is lower and the partnership less likely. Both results go in the same direction:

**Corollary 4** *The higher NGOs budget, the less likely individual partnerships.*

# 5 Conclusion

The main objective of the paper is to characterize what drives NGOs' communication strategy, contrasting good cops − NGOs which disclose information about firms which self-regulate − and bad cops − NGOs which disclose information on firms which do not. To answer these questions, we consider a model featuring a continuum of firms which can self-regulate or not, a stakeholder who is willing to reward self-regulation but who is not able to observe firms' behavior and non governmental organizations (NGOs) that induce self-regulation by imperfectly disclosing on firms' level of self-regulation.

We show that the NGO decides to specialize: if its budget is low, it chooses to be a good cop. When the budget is larger, it opts for being a bad cop. The general point here is that the best information disclosure strategy consists in going against the stakeholder's a priori built upon the environment characteristics: a high budget mean that the NGO produces much information about firms' behavior. Incentives for firms to invest in CSR are thus high, implying that the stakeholder holds an optimistic belief that one firm about whom she does not receive any news is actually self-regulating. In this context, it is more effective for the NGO to be hostile as this mitigates their incentives to free ride on the lack of information. And conversely when the stakeholder is pessimistic about the level of self-regulation (because NGOs' budget is limited), it is more effective to motivate firms by increasing the probability $g$ to get the full reward when investing in self-regulation.

We develop several extensions of the model. First, we assume multiple NGOs which non cooperatively select their technology. A coordination problem arises for the two possible equilibria -NGOs being all bad cops and NGOs being all good cops- sometimes coexist. We also introduce the option for firms to cooperate with NGOs. More specifically, we distinguish individual partnership, whereby the NGO specifically publicizes the partnering firm's behavior, and collective partnership whereby industry support leads to an increase of the aggregate disclosure probability.

Although the prime focus of the analysis is positive, it is possible to derive a general policy lesson. When the amount of resources available to NGOs is limited and/or the number of NGOs is low, everything is going smoothly and the role of a welfare-maximizing regulator is limited: There exists a single equilibrium which is socially optimal. As all NGOs are good cops, firms are willing to engage in collective or individual partnerships in order to increase the quantity of information, which in turn increases self-regulation and thus social welfare. Things get more complicated when the quantity of information is such that the social outcome is the bad cop equilibrium. In this case, the coordination problem can lead to the survival of the good cop equilibrium. It is reinforced by the fact that firms which do not self-regulate do not form partnerships anymore and self-regulating firms are less prone to enter in partnership for the benefits get smaller. To sum up, a public intervention promoting bad cops becomes increasingly useful.

# 6 Appendix

## 6.1 Proof of Lemma 2

To begin, we plug $g = b$ in (3.3) to obtain $c^*$ in the case where $g = b$. Then, solving (3.3) for $c$ yields two roots:

$$\begin{aligned} c_1 &= \frac{1}{2(g-b)}\left(g-2b+1+\sqrt{(g-2b+1)^2-4g(1-b)(g-b)}\right) \\ c_2 &= \frac{1}{2(g-b)}\left(g-2b+1-\sqrt{(g-2b+1)^2-4g(1-b)(g-b)}\right) \end{aligned}$$

When $g-b>0, c_1>1$ as this inequality is equivalent to $\sqrt{(g-2b+1)^2-4g(1-b)(g-b)} > -(1-g)$, which is obviously satisfied.

When $g-b<0$, $\sqrt{(g-2b+1)^2-4g(1-b)(g-b)} \geq |g-2b+1|$. Then we have two subcases: If $g-2b+1 \geq 0$, $|g-2b+1| = g-2b+1$, the term in brackets is obviously positive, implying $c_1 < 0$. If $g-2b+1<0$, $|g-2b+1| = -(g-2b+1)$. Hence $\sqrt{(g-2b+1)^2-4g(1-b)(g-b)} \geq -(g-2b+1)$. The term in brackets is thus positive, meaning that $c_1 \leq 0$.

## 6.2 Proof of Proposition 2

For ease of presentation, we introduce the following notations: $c_g^* \equiv c^*(\alpha,0)$ and $\Phi_g \equiv \Phi(c)$ when $g=\alpha$ and $b=0$. Hence $\Phi_g = \frac{\alpha(1-c)}{1-\alpha c}$. Similarly, $c_b^*$ is the stable equilibrium $c^*(0,\alpha)$ and $\Phi_b \equiv \Phi(c) = \frac{\alpha c}{1-\alpha(1-c)}$. Furthermore, let $c^{GB}$ denote the value of $c$ where $\Phi_g = \Phi_b$ . Straightforward calculations show that $\Phi_g(c^{GB}) = \Phi_b(c^{GB}) = \frac{\alpha}{2-\alpha}$ and $c^{GB} = \frac{1}{2}$. Then we consider two cases: $\Phi_g(c^{GB}) = \Phi_b(c^{GB}) > c^{GB}$ - which is equivalent to $\alpha > 2/3$ - and $\Phi_g(c^{GB}) = \Phi_b(c^{GB}) \leq c^{GB}$.

**Case 1:** $\alpha > 2/3$ From the intermediate value theorem (IVT) follows directly $c_g^* > c^{GB}$ : $\Phi_g(c^{GB})$ and $\Phi_g(1) = 0$ are above and below the 45 degree line, respectively. The same theorem also implies $c_b > c^{GB}$ as $\Phi_b$ is above the 45 degree line in $c^{GB}$ and ends up below in $c = 1$ ($\Phi_b(1) = \alpha < 1$).

Since $\Phi_g$ is decreasing and $\Phi_b$ is increasing (Property 2 in Lemma 1), we have the general property that $\Phi_g(c) < \Phi_b(c')$ for any $c, c' > c^{GB}$ . It is thus satisfied in the particular cases where $c = c_g^*$ and $c\prime = c_b^*$. It implies $c_g^* < c_b^*$ as $c_g^* = \Phi_g(c_g^*)$ and $c_b^* = \Phi_b(c_b^*)$.

**Case 2:** $\alpha \leq 2/3$ The IVT implies that $c_g^* \leq c^{GB}$ . We also have $c_b^* \leq c^{GB}$ because (i) $\Phi_b$ is now below the 45 degree line in $c^{GB}$ , (ii) it ends up below in $c = 0$ ($\Phi_b(0) = \alpha$) and (iii) $\Phi_b$ is concave (see footnote X). These three properties prevent $\Phi_b$ to intersect with the 45 degree line beyond $c^{GB}$. Then $\Phi_g(c) \leq \Phi_b(c\prime)$ for any $c, c\prime \leq c^{GB}$, $c_g^* \leq c^{GB}$ and $c_b^* \leq c^{GB}$ implies $c_g^* > c_b^*$ in this case.

## 6.3 Proof of Lemma 3

For ease of presentation, we introduce the following notations :

$$
\begin{aligned}
g_{-i} &= \left(1 - \prod_{j \neq i}(1 - g_j)\right) \text{ and } b_{-i} = \left(1 - \prod_{j \neq i}(1 - b_j)\right) \quad c_{b_i}^* \equiv c^*(g_{-i}, b_{-i} + \alpha(1 - b_{-i})) \\
c_{g_i}^* &\equiv c^*(g_{-i} + \alpha(1 - g_{-i}), b_{-i}) \\
\Phi_{g_i}(c) &\equiv \Phi(c) \text{ with } g_i = \alpha \text{ and } b_i = 0 \\
\Phi_{b_i}(c) &\equiv \Phi(c) \text{ with } g_i = 0 \text{ and } b_i = \alpha \\
c^{GB} &= \frac{1 - b_{-i}}{2 - (b_{-i} + g_{-i})} \text{ such that } \Phi_{g_i}(c) = \Phi_{b_i}(c) \\
\Phi_{g_i}(c^{GB}) &= \frac{\alpha + (b_{-i} + g_{-i})(1 - \alpha)}{2 - \alpha}
\end{aligned}
$$

Then we follow closely the proof of Proposition 2. We know that

$$\Phi_{b_i}(0) < \Phi_{g_i}(0) \text{ and } \Phi_{b_i}(1) > \Phi_{g_i}(1), \text{ for any } g_i, g_{-i}, b_i, b_{-i}.$$

It follows that $\Phi_{b_i} < \Phi_{g_i}$ if $c < c^{GB}$ and $\Phi_{b_i} \geq \Phi_{g_i}$ if $c \geq c^{GB}$. Now we consider two subcases:

- If $\Phi_{g_i}(c^{GB}) = \Phi_{b_i}(c^{GB}) > c^{GB}$, we necessarily have $c^*_{b_i} > c^{GB}$ and $c^*_{g_i} > c^{GB}$ from the IVT. We also know that $\Phi_{b_i} > \Phi_{g_i}$ if $c > c^{GB}$ and $\Phi_{b_i}(1) > \Phi_{g_i}(1)$. Hence, $\Phi_{g_i}$ will necessarily the 45 degree line before $\Phi_{b_i}$ when departing upward from $c^{GB}$. Hence $c^*_{b_i} > c^*_{g_i}$.
- If $\Phi_{g_i}(c^{GB}) = \Phi_{b_i}(c^{GB}) < c^{GB}$, we symmetrically have $c^*_{b_i} < c^{GB}$ and $c^*_{g_i} > c^{GB}$ and $\Phi_{b_i} < \Phi_{g_i}$ if $c < c^{GB}$ and $\Phi_{b_i}(0) < \Phi_{g_i}(0)$. Hence $c^*_{g_i} > c^*_{b_i}$.

## 6.4 Proof of Proposition 4

We start characterizing the socially optimal benchmark. In the good cop equilibrium, we have $G = 1 - (1 - \alpha)^n$ and $B = 0$, whereas $G = 0$ and $B = 1 - (1 - \alpha)^n$ if all NGOs are bad cops. For ease of presentation, let $P$ denote $1 - (1 - \alpha)^n$. We thus need to compare $c^*(0, P)$ with $c^*(P, 0)$. It is exactly the comparison we have made in Proposition 2, which establishes that $c^*(P, 0) > c^*(0, P)$ if $P < 2/3$. It is equivalent to $\alpha \leq 1 - 3^{-\frac{1}{n}}$ Then we show that $\alpha_{\text{lim}} \leq 1 - 3^{-\frac{1}{n}}$, which is equivalent to $3^{-\frac{1}{n}} < 1 - \alpha_{\text{lim}}$. Hence $2(1 - \alpha_{\text{lim}})^n + (1 - \alpha_{\text{lim}})^{2n-1} > 2/3 + 3^{\left(\frac{1}{n}-2\right)}$. The left-hand side is less than 1: it decreases with $n$ and it is equal to 1 when $n = 1$. This completes the proof as $2(1 - \alpha_{\text{lim}})^n + (1 - \alpha_{\text{lim}})^{2n-1} = 1$ by definition.

## 6.5 Proof of Proposition 5

Differentiating the equilibrium condition $c^* = \Phi(c^*)$ and rearranging, we obtain:

$$\frac{dc^*}{dG} = \frac{\frac{\partial \Phi}{\partial G}}{1 - \frac{d\Phi}{dc}} \text{ and } \frac{dc^*}{dB} = \frac{\frac{\partial \Phi}{dB}}{1 - \frac{d\Phi}{dc}}$$

The stability condition is $1 - \left|\frac{d\Phi}{dc}(c^*)\right| > 0$ which implies that $1 - \frac{d\Phi}{dc} > 0$. (It is true when $\left|\frac{d\Phi}{dc}(c^*)\right| = -\frac{d\Phi}{dc}(c^*)$ and $\left|\frac{d\Phi}{dc}(c^*)\right| = \frac{d\Phi}{dc}(c^*)$). Hence we just need to look at the signs of $\frac{\partial \Phi}{dG}$ and $\frac{\partial \Phi}{dB}$, which are both positive as :

$$\frac{\partial \Phi}{\partial G} = \frac{(1-B)^2(1-c^*)}{(B-1+c(G-B))^2} \text{ and } \frac{\partial \Phi}{\partial B} = \frac{c^*(1G)^2}{(Bc-B-cG+1)^2}.$$

## 6.6 Proof of Lemma 4

**Impact of** $G$

We have

$$\frac{d\pi(a=1)}{dG} = 1 - \mu^* + (1-G)\frac{d\mu^*}{dG} \tag{3.9}$$

$$\frac{d\pi(a=0)}{dG} = (1-B)\frac{d\mu^*}{dG} \tag{3.10}$$

Rearranging (3.1) leads to $\mu^* = \frac{G-c^*}{G-B}$. Hence $\frac{d\mu^*}{dG} = \frac{1}{G-B}\left(1 - \mu^* - \frac{dc^*}{dG}\right)$. We have $\frac{dc^*}{dG} = \frac{\partial \Phi}{\partial G}\left(1 - \frac{d\Phi}{dc}\right)^{-1}$ with $\frac{\partial \Phi}{\partial G} = \frac{(1-B)^2(1-c^*)}{1-B-(G-B)c^*}$. Combining $\mu^* = \frac{G-c^*}{G-B}$ with (3.3) yields $1 - \mu^* = \frac{(1-B)(1-c^*)}{1-B-(G-B)c^*}$. Plugging this expression in $\frac{\partial \Phi}{\partial G}$ leads to $\frac{\partial \Phi}{\partial G} = \frac{(1-B)(1-\mu)}{1-B-c(G-B)}$. Hence

$$\frac{d\mu^*}{dG} = \frac{(1-\mu^*)}{(G-B)\left(1 - \frac{d\Phi}{dc}\right)}\left(1 - \frac{d\Phi}{dc} - \frac{(1-B)}{(1-B-c(G-B))}\right) \tag{3.11}$$

Then, we have $\Phi(c) = \frac{G(1-B)-(G-B)c}{(1-B)-(G-B)c}$. Inverting this function yields $c = \frac{(1-B)(G-\Phi)}{(G-B)(1-\Phi)}$. As $c = \Phi$ in equilibrium, we have

$$c^* = \frac{(1-B)(G-c^*)}{(G-B)(1-c^*)}. \tag{3.12}$$

Rearranging this equation, we get $1 - B - (G - B)c^* = (1 - B)(1 - G)/(1 - c^*)$. Substituting $1 - B - (G - B)c^*$and $\frac{d\Phi}{dc} = -\frac{(G-B)(1-c^*)^2}{(1-B)(1-G)}$in (3.11) and rearranging, we obtain

$$\frac{d\mu^*}{dG} = \frac{(1-\mu^*)}{(G-B)\left(1-\frac{d\Phi}{dc}\right)}\left(\frac{(G-B)(1-c^*)^2-(G-c^*)(1-B)}{(1-B)(1-G)}\right).$$

Rearranging (3.12) leads to $(1 - B)(G - c^*) = c^*(G - B)(1 - c^*)$. Hence

$$\frac{d\mu^*}{dG} = \frac{(1-c^*)(1-\mu^*)}{\left(1-\frac{d\Phi}{dc}\right)(1-B)(1-G)}(1-2c^*). \tag{3.13}$$

Plugging (3.13) in (3.9), we obtain

$$d\pi(a=1)/dG = \frac{(1-\mu^*)}{\left(1-\frac{d\Phi}{dc}\right)(1-G)(G-B)}\left((1-G)(c^*-B)+(1-c^*)^2(G-B)\right)$$

which is always positive: If $G > B$, $c^* - B < 0$, meaning the last term in bracket and the first are both negative. If $G < B$, they are both positive.

Turning next to the firms which do not self-regulate, we have:

$$\frac{d\pi(a=0)}{dG} = (1-B)\frac{d\mu^*}{dG}$$

which is positive if and only if $c^* < 1/2$. If we now substitute

$$c^* = \frac{1}{2(G-B)}\left(G-2B+1-\sqrt{(G-2B+1)^2-4G(1-B)(G-B)}\right)$$

in this inequality, calculations show that $c^* < 1/2$ is equivalent $G < 2/3$ if $B = 0$ and $B < 2/3$ if $G = 0$. As $B = 1 - (1-\alpha)^{n-1}$ when $G = 0$, $B < 2/3$ is equivalent to $\alpha < 1 - \frac{1}{3^{1/n}}$.

**Impact of $B$**

We have

$$\frac{d\pi(a=1)}{dB} = (1-G)\frac{d\mu^*}{dB} \tag{3.14}$$

$$\frac{d\pi(a=0)}{dB} = -\mu^* + (1-B)\frac{d\mu^*}{dB} \tag{3.15}$$

$$\frac{d\mu^*}{dB} = \frac{1}{G-B}\left(\mu^* - \frac{dc^*}{dB}\right) \tag{3.16}$$

We have $\frac{dc^*}{dB} = \frac{\partial\Phi}{\partial B}\left(1 - \frac{d\Phi}{dc}\right)^{-1}$ with $\frac{\partial\Phi}{\partial B} = \frac{c^*(1-G)^2}{(bc-B-cg+1)^2}$. Combining $\mu^* = \frac{G-c^*}{G-B}$ with (3.3) yields $\mu^* = \frac{c^*(1-G)}{1-B-(G-B)c^*}$. Hence $\frac{\partial\Phi}{\partial B} = \frac{(1-G)\mu^*}{1-B-(G-B)c^*}$. It follows that

$$\frac{d\mu^*}{dB} = \frac{\mu^*}{(G-B)\left(1-\frac{d\Phi}{dc}\right)}\left(1 - \frac{d\Phi}{dc} - \frac{(1-G)}{1-B-(G-B)c^*}\right).$$

As $1-B-(G-B)c^* = (1-B)(1-G)/(1-c^*)$ and $\frac{d\Phi}{dc} = -\frac{(G-B)(1-c^*)^2}{(1-B)(1-G)}$, we have

$$\frac{d\mu^*}{dB} = \frac{\mu^*\left((c^*-B)(1-G) + (G-B)(1-c^*)^2\right)}{(G-B)(1-B)(1-G)\left(1-\frac{d\Phi}{dc}\right)}. \tag{3.17}$$

which is always positive as $G > B$ implies $c^* - B > 0$ and conversely. Therefore $\frac{d\pi(a=1)}{dB} > 0$.

Substituting (3.17) in (3.15) and rearranging, we obtain:

$$\frac{d\pi(a=0)}{dB} = \frac{\mu^*(1-c^*)}{\left(1-\frac{d\Phi}{dc}\right)(1-B)}(1-2c^*).$$

which is positive only if $c^* < 1/2$ and thus $\alpha < 1 - \frac{1}{3^{1/n}}$.

# Chapter 4

# Should Imperfect Labels be Strict or Lenient?

## English summary

In the real world, a label is granted to a product if its quality complies with a standard. The standard can be more or less strict, implying varying shares of labeled products in the market. In this paper, we seek to characterize the level of label strictness which maximizes social welfare under endogenous product differentiation in duopoly. The model describes the interactions between two firms which endogenously choose product quality and consumers who positively but heterogeneously value quality that they cannot observe. They only observe a label which can be either lenient or strict. We assume for realism that the label is imperfect in the sense that the certification test is noisy. We find that the welfare ranking is ambiguous. The strict label dominates the lenient label if consumers' prior belief that the product is of high quality is high and/or if the certification is very noisy. This analysis allows deriving policy implications about label design.

## Résumé français

Dans le monde réel, un label est accordé à un produit si celui-ci respecte un standard. Ce standard peut être plus ou moins strict, entraînant des parts de marché différentes de produits labélisés. Dans ce papier, nous cherchons à caractériser le niveau de sévérité de label qui maximise le bien être social. Le modèle décrit les interactions entre deux entreprises qui choisissent de manière endogène le niveau de qualité de leur produit et des consommateurs pour qui la qualité, qu'ils ne peuvent pas observer, a une valeur positive mais hétérogène. Ils peuvent seulement observer un label, qui peut être laxiste ou strict. Par réalisme, nous faisons l'hypothèse que le label est imparfait dans le sens où le test de certification comprend un certain bruit. La comparaison des deux situations donne des résultats ambigus. Le label strict domine le label laxiste si les consommateurs ont un a priori positif sur le niveau général de qualité et/ou si le test de certification fait beaucoup de bruit. Cette analyse nous permet de dériver des recommandations à propos

de la conception des labels.

# 1 Introduction

An important empirical literature points out that some consumers are willing to pay more for goods and services produced under environmentally and socially friendly processes. For example, Nimon and Beghin (1999) evidence a price premium of 33.8% of the apparel price for organic cotton. Prasad et al. (2004) conducted a real life experiment in a department store and find that one consumer out of four were willing to pay up to 40% for sweatshop free athletic shoes. Ward et al. (2011) performed an online survey that reveals that consumers are, on average, willing to pay an extra $249.82–$349.30 for a refrigerator that has been awarded the American ENERGY STAR label.[1] In reaction, firms increasingly offer products that claim to be responsible. At the two dozens stores it visited both in 2009 and 2010, the consulting agency TerraChoice finds that "greener" products offering (products that claim to be "green") increased by 73% from 2,739 products in 2009 to 4,744 products in 2010.

Nevertheless, these environmental and social attributes are mostly unverifiable and are thus a credence attributes of products for consumers. Neither search nor experience can guarantee them that it was produced under responsible processes. This creates an opportunity for a third party certifier to intervene (Cason and Gangadharan, 2002) and justifies the setting up of labels that provide consumers with credible information about quality. For example, the Scandinavian Nordic Swan label is an ecolabel granted to products that meet "extremely high environmental and often climate requirements".[2] In France, the association Qualité-France created the LUCIE Label,[3] intended for organisms respecting the new ISO26000 standard, which provides guidelines for being socially responsible. The above-mentioned ENERGY STAR label is another example.

[1] ENERGY STAR label is a US label granted to products that reduce greenhouse gas emissions and other pollutants caused by the inefficient use of energy.

[2] http://www.nordic-ecolabel.org/

[3] This label is granted by the agency LUCIE, in partnership with the extra-financial rating agency VIGEO and the labeling institution AFNOR Certification: http://www.labellucie.com/

The principle of labeling may seem simple at first glance: tag firms that meet the standards and not those that do not. But in reality, they are many ways do so: How many and which environmental and social features should be considered? How tough should requirements be? How frequent should revision controls be?

In particular, there is a trade-off between choosing a strict or a lenient standard. The stricter the label is, the fewer products will be labeled, but the stronger they will be identified as high quality products and hence rewarded. The more lenient the label is, the more firms will get the label and hence have incentives to invest in higher quality in the first place, but the lesser the reward will be as being labeled won't be perceived as such a great achievement.

In this paper we aim to understand why and when one of these two types of label spurs more quality than the other, and what is optimal from a social welfare point of view.

Labels allow firms to differentiate, which has a double effect. On the one hand, it provides incentives for firms to invest and produce high quality which is welfare enhancing since without any information firms that would find profitable to invest under perfect information do not. But on the other hand, differentiation alters competition and firms can charge prices higher than their production cost and make positive profits, which can be welfare damaging. The identification of this issue in the economic literature dates back at least to Shaked and Sutton (1982).

From then on, the interaction between minimum quality standards and the level of competition, alongside its impacts on social welfare, has been an issue well acknowledged in the industrial organization literature (e.g. Ronnen (1991) or Crampes and Hollander (1995)). Our analysis contributes to this literature.

Nevertheless, while these papers assumes a perfect certification test, like most of the literature about labels (see also Ibanez and Grolleau (2008) or Bottega and De Freitas

(2009)), we model an imperfect one. To our knowledge there exist only a couple of articles that introduce an imperfect monitoring technology entailing errors of type I and II (e.g. Liang and Jensen (2007) or Mason (2009, 2011)): some low quality products are sometimes labeled while some high quality products may fail the test and remain unlabeled. We deem that this latter approach is more realistic. Indeed, many reasons can lead to mistakes in the attribution of labels. There may be measurement errors, infrequent verification controls, corruption of the labeling institution, etc.

As a result, an important feature of our paper is that we deal both with imperfect competition and imperfect certification.

There exist numerous empirical analysis of the impact of labels on the provision of high quality and results are mixed: Some authors do find that the average quality increases after the introduction of the label (e.g. Bennear and Olmstead (2008), Chen (2008), Jin and Leslie (2003) or Powers et al. (2008)). But others evidence that some sellers game the labeling system at the expense of consumers (e.g. Dranove et al. (2003), Jacob (2005), Lu (2009), or Werner and Asch (2005)), for example by reallocating efforts from unmeasured dimensions of quality to the dimensions covered by the label, without increase in overall quality. We also contribute to this debate by providing a theoretical analysis of the efficiency of different labeling processes in enhancing high quality production.

To answer our research questions, we develop a simple model of asymmetric information[4] that describes the interaction between a duopoly competing "à la Bertrand" and consumers who positively but heterogeneously value environmental and social products' attributes (also designated as high quality hereafter).[5] Firms privately decide whether to produce high quality or low quality. They are then submitted to a costless and mandatory labeling process that reveals some information to the consumers about the true level of quality of the products.

[4]Adapted from Fleckinger et al. (2012).
[5]Hence, we work with a vertical differentiation model à la Mussa and Rosen (1978).

More precisely, the monitoring technology is imperfect: it reveals perfectly the true level of product quality with a probability $m$, but with a probability $1-m$ it reveals nothing. The parameter $m$ captures the precision of the certification test. We consider two standard's levels: a lenient label who is granted even to products whose quality level is unknown, and a strict label who is granted only to products that have been confirmed to be high quality products.

Consumers do not have fully Bayesian beliefs at equilibrium. They hold a prior that a firm produces high quality and update it upon receiving the labels' signals: labeled or unlabeled. There is a recent trend in the industrial organization literature, both applied and empirical (see Spiegler (2011) for an overview), that aims at capturing some facets of bounded consumers rationality and in particular the sophistication of belief formation. This literature is mostly concerned with optimal pricing strategy to exploit behavioral biases. We take here as given some form of bounded consumers rationality which we deem realistic given the complexity of the mechanisms that a perfectly sophisticated consumer should anticipate, and assume, as in this literature, that firms take it into account.

Importantly, contrarily to others papers that study standard strictness under imperfect competition, firms have symmetric strategies at equilibrium: there is differentiation ex-post, but non ex-ante. This allows us to focus on informational insights.

We find that the welfare ranking is ambiguous, which is in line with the result in Gill and Sgroi (2012) that features a signaling game where exogenous quality monopolies choose the strictness of a quality revealing test. The strict label dominates the lenient label if consumers' prior belief that the product is of high quality is high and/or if the certification is very noisy. As for the prior, the intuition is the following: under a lenient label, the uncertainty lays on labeled firms, thus a consumer's optimistic state of mind benefits to them. Since producing high quality guarantees to be labeled, it becomes an increasingly more interesting option than not investing as the prior increases. On the contrary, a high

prior benefits to unlabeled firms under a strict label. Saving the cost of really producing high quality, which guarantees to get unlabeled, while benefiting from the optimistic consumers' state of mind becomes a more interesting option in these conditions. In conclusion, with a high prior, producing high quality is a more interesting option under a lenient label rather than a strict. And conversely.

The paper is structured as follows. We present the base model in the next section. We solve it for a lenient label and for a strict label in the third and fourth sections and compare the outcomes of the two situations in the fifth. We conclude and discuss the possible paths that we are considering to further our analysis in the last section.

## 2 The base model

We consider an economy where two firms, $A$ and $B$, sell a good of quality $q$. Each firm chooses between producing the high quality ($q = q_H$) version of the product, or the low quality ($q = q_L$) version of the good. In order to simplify notations, we assume that $q_H = 1 + \delta$ and $q_L = 1$, with $\delta \in [0;1]$. Producing high quality entails an investment cost $c$ to innovate, that is drawn independently for the two firms following a uniform distribution over $[0;1]$.[6] This is common knowledge, but each firm privately observes the realization of its cost. Producing low quality does not entail any investment. The firms face the same marginal cost of production that we normalize to 0.

Turning to the demand side, there is a continuum of heterogeneous consumers indexed by $\theta$, which captures their willingness to pay for quality. Under perfect information, a consumer of type $\theta$ buying a good of quality $q$ at a price $p$, has a surplus of

[6]Two comments: 1) Adopting a uniform distribution ensures equilibrium uniqueness and thus simplifies the mathematical analysis. It does not however change the nature of the results, see Fleckinger et al. (2012). 2) Choosing the interval $[0;1]$ entails that, under perfect information and without competition, it would always be profitable to invest and produce high quality.

$$S(\theta) = U(\theta) - p = v + \theta q - p$$

Consumers' taste $\theta$ is uniformly distributed over $[0;1]$. We assume that each consumer buys one unit of the good and that the market is fully covered, meaning that $v$ is sufficiently large so that consumers always buy one unit of the product.

Quality is not observable to consumers neither before nor after purchase and use, but consumers hold a prior $\rho$: they believe *ex ante* that a given firm produces high quality with a probability $\rho$. In addition, labels can help differentiating the different versions of the products by providing consumers with signals about the product quality: labeled ($s = l$) or non labeled ($s = \varnothing$).

Nevertheless, the monitoring process is not perfect: it reveals the true level of quality with a probability $m$, but with a probability $1 - m$ it reveals nothing. The probability $m$ reflects the reliability of the monitoring process. This parameter is common knowledge.

There is a lenient and a strict way to deal with this uncertainty, resulting in two types of label, illustrated in Figure 1:

- A **lenient label**: in this case, a firm producing high quality is always granted the label and a firm that produces low quality may also be granted the label, but only with a probability $1 - m$.
- A **strict label**: in this case, a firm producing high quality is granted the label only with a probability $m$ while a firm that produces low quality is never granted the label.

Labeling is costless and mandatory: all firms are submitted to the labelling process, at no cost.

The sequence of events is as follows:

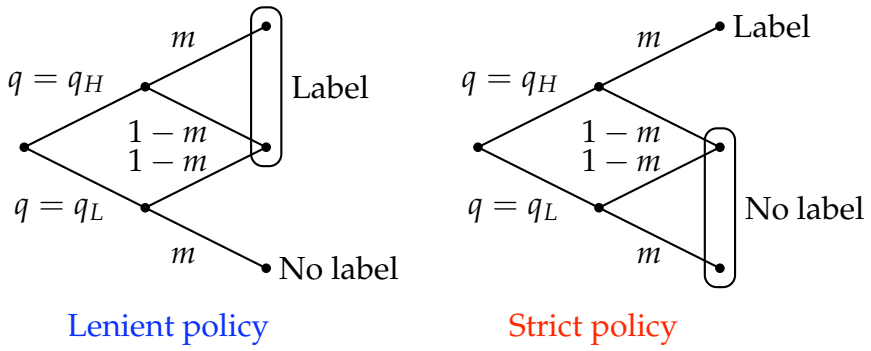



Figure 1: **Lenient and strict label information structures.**

- **Stage 1:** Firms privately learn their costs $c_A$ and $c_B$, then choose a level of quality.
- **Stage 2**: Certification occurs. Depending on firms' quality choices, label signals are sent according to the processes described above.
- **Stage 3:** Upon receiving the signals,[7] consumers update their belief about the goods' quality, following the Bayes rule.
- **Stage 4:** Firms compete "à la Bertrand".

We solve the model for each type of label separately and then compare the equilibria of the two labelling situations.

# 3 Lenient label

With a lenient label, a firm producing high quality is always granted the label and a firm that produces low quality may also be granted the label, but only with a probability $1 - m$. Consumers consider that the "non label" signal, $\varnothing$, is hard information and indicates a low quality firm. But there is uncertainty about the "label" signal, $l$: Consumers form a belief $\mu_L = \Pr[q = q_H | s = l]$ that the firm they are facing is really producing the high quality.

[7] Since there are two products, both products can get the label, none or only one out of the two.

Reasoning backward, we first solve the competition stage.

## 3.1 Competition stage

There are three types of possible labelling outcomes: both products get the label, none of them get the label and only one of them get the label. The analysis of the two first cases is straightforward: Firms compete "à la Bertrand" and the price equals the marginal cost of production, i.e. 0. In such a case, we assume that the market is divided equally between the two firms.[8]

Let's turn now to the differentiating labelling outcome. We know that with a lenient label, a firm producing high quality is always granted the label and a firm that produces low quality may also be granted the label, but only with a probability $1 - m$. Consequently, when a consumer of type $\theta$ receives the signal $s = \varnothing$, she is sure that the good is of low quality. Her surplus is simply:

$$S_{\varnothing}(\theta) = v + \theta q_L - p_{\varnothing} = v + \theta - p_{\varnothing}$$

with $p_{\varnothing}$ being the price of the unlabeled product. In turn, she is unsure about the true quality of the good when the signal is $s = l$. In this case she forms a belief $\mu_L$ that the product is of high quality, as defined above, and her surplus when consuming the labelled product is:

$$\begin{aligned} S_l(\theta) &= v + (\mu_L q_H + (1 - \mu_L) q_L)\theta - p_l \\ &= v + (\mu_L \delta + 1)\,\theta - p_l \end{aligned}$$

with $p_l$ being the price of a labeled product.

[8] This assumption matters when calculating the social welfare. Assuming equal market shares is the most neutral choice possible.

The indifferent consumer between the two goods is of type $\hat{\theta}$ such that $S_l(\hat{\theta}) = S_\varnothing(\hat{\theta})$. Hence the threshold $\hat{\theta}$ is equal to:

$$\hat{\theta} = \frac{p_l - p_\varnothing}{\mu_L \delta}$$

Turning next to the supply side of the market, we denote $\pi_l$ the profit of the labeled firm and $\pi_\varnothing$ the profit of the unlabeled firm. The first one chooses the price $p_l$ that maximizes:

$$\pi_l = (1 - \hat{\theta}) p_l = \left(1 - \frac{p_l - p_\varnothing}{\mu_L \delta}\right) p_l$$

And the second one chooses the price $p_\varnothing$ that maximizes:

$$\pi_\varnothing = \hat{\theta} p_\varnothing = (\frac{p_l - p_\varnothing}{\mu_L \delta}) p_\varnothing$$

We can thus derive firms' profits for all labeling outcomes and sum them up in the following lemma:

**Lemma 9** *With a lenient label:*

- *When firms both get the signal label, or both get the signal unlabeled, firms have zero profits and share the market equally.*
- *When firms get different signals, the labeled firm has a market share of $\frac{2}{3}$ and its profit is $\pi_l = \frac{4\mu_L \delta}{9}$ while the unlabeled firm has a market share of $\frac{1}{3}$ and its profit is $\pi_\varnothing = \frac{\mu_L \delta}{9}$.*

**Proof.** $\frac{\partial \pi_l}{\partial p_l} = 1 - \frac{2p_l - p_\varnothing}{\mu_L \delta}$ and $\frac{\partial \pi_\varnothing}{\partial p_\varnothing} = \frac{p_l - 2p_\varnothing}{\mu_L \delta}$. Solving $\left[\frac{\partial \pi_l}{\partial p_l} = 0; \frac{\partial \pi_\varnothing}{\partial p_\varnothing} = 0\right]$ for $p_l$ and $p_\varnothing$ gives us $p_l = \frac{2\mu_L \delta}{3}$ and $p_\varnothing = \frac{\mu_L \delta}{3}$. We know that is a maximum because $\frac{\partial^2 \pi_l}{\partial p_l^2} = -\frac{2}{\mu_L \delta} < 0$ and $\frac{\partial^2 \pi_\varnothing}{\partial p_\varnothing^2} = -\frac{2}{\mu_L \delta} < 0$. ■

## 3.2 Firms' choice of quality

Consider first firm $A$. If it chooses high quality, it anticipates that it will get the label. But it is uncertain about firm $B$'s choice because it can not observe nor its cost $c_B$ neither its actions. Thus, it assigns a probability $\lambda_B$ to the possibility that firm $B$ produces high quality too. In such a case it anticipates that they both would get the label and hence null profits. Similarly, it assumes that with a probability $1-\lambda_B$ firm $B$ produces low quality. In this case, there is still a probability $1-m$ that firm $B$ would get the label and both firms make no profit, but there is also a probability $m$ that it would not get the label, and firm $A$ would be differentiated and realize the labeled profit $\pi_l$. In conclusion, the overall profit that firm $A$ expects when it produces high quality is:

$$\begin{aligned} E\Pi_A(q_H) &= \lambda_B * 0 + (1-\lambda_B)\left[m\pi_l + (1-m) * 0\right] - c_A \\ &= (1-\lambda_B)m\pi_l - c_A \end{aligned}$$

Reasoning similarly, we compute the overall profit that firm $A$ expects when it produces low quality:

$$E\Pi_A(q_L) = m\left(m\lambda_B - m + 1\right)\pi_\varnothing + m(1-m)(1-\lambda_B)\pi_l$$

Firm $A$ produces high quality if $E\Pi_A(q_H) \geq E\Pi_A(q_L)$. Namely, using Lemma 1, Firm $A$ produces high quality if its quality cost $c_A$ is such that:[9]

$$c_A \leq \left(5m\left(1-\lambda_B\right) - 1\right)\frac{\delta\mu_L m}{9} \tag{4.1}$$

[9] $E\Pi_A(q_H) \geq E\Pi_A(q_L) \Leftrightarrow (1-\lambda_B)m\pi_l - c_A \geq m\left(m\lambda_B - m + 1\right)\pi_\varnothing + m(1-m)(1-\lambda_B)\pi_l \Leftrightarrow c_A \leq \pi_l m^2\left(1-\lambda_B\right) - \pi_\varnothing m\left(m\lambda_B - m + 1\right) \Leftrightarrow c_A \leq \frac{\mu_L\delta}{9}\left[4m^2\left(1-\lambda_B\right) - m\left(m\lambda_B - m + 1\right)\right] \Leftrightarrow c_A \leq \left(5m\left(1-\lambda_B\right) - 1\right)\frac{\delta\mu_L m}{9}$

Proceeding similarly for firm $B$, we get that its incentive constraint to produce high quality in response to firm $A$'s behavior is:

$$c_B \leq (5m\,(1-\lambda_A)-1)\,\frac{\delta\mu_L m}{9} \tag{4.2}$$

We denote $c_A^*$ (resp. $c_B^*$) the cost threshold at equilibrium of firm $A$ (resp. $B$) such that if its cost is smaller than $c_A^*$ (resp .$c_B^*$) it produces high quality and conversely. Consequently, we have $\lambda_A = c_A^*$ and $\lambda_B = c_B^*$, since $c$ is uniformly distributed over $[0;1]$.

Now, to find the equilibrium value of $c_A^*$ and $c_B^*$, we need to compute consumers' belief $\mu_L$. Since they have the prior $\rho$, their equilibrium belief that the labeled firm is indeed producing high quality is:[10]

$$\mu_L^* = \frac{\rho}{1-m(1-\rho)} \tag{4.3}$$

Plugging 4.3 in 4.1 and 4.2, the equilibria is defined by the following set of equations:

$$\begin{aligned} c_A^* &= (5m\,(1-c_B^*)-1)\,\frac{\delta m}{9}\,\frac{\rho}{1-m(1-\rho)} \\ c_B^* &= (5m\,(1-c_A^*)-1)\,\frac{\delta m}{9}\,\frac{\rho}{1-m(1-\rho)} \end{aligned} \tag{4.4}$$

Solving these equations for $c_A^*$ and $c_B^*$ gives us the following proposition:

**Proposition 14** *There always exists a unique equilibrium with symmetrical behaviors for firms A and B. More precisely:*

1. *If $m < \frac{1}{5}$, both firm always choose to produce low quality, we have $c_A^* = c_B^* = 0$.*

2. *If $m \geq \frac{1}{5}$, firms choose to invest and produce high quality if their cost is below the following*

[10] $\mu_L^* = prob(q_H/\varnothing) = \frac{prob(\varnothing/q=q_H)prob(q=q_H)}{prob(\varnothing/q=q_H)prob(q=q_H)+prob(\varnothing/q=q_L)prob(q=q_L)} = \frac{\rho}{\rho+(1-\rho)(1-m)} = \frac{\rho}{1-m(1-\rho)}$

*threshold:*

$$c_L^* = \frac{\delta m \rho (5m-1)}{5\delta m^2 \rho + 9(1-m(1-\rho))}$$

**Proof.** 1) Consider firm $A$ incentive's constraint: $E\Pi_A(q = q_H/c) - E\Pi_A(q = q_L/c) = (5m\,(1-\lambda_B) - 1)\,\frac{\delta \mu_L^* m}{9} - c_A$. If $m < \frac{1}{5}$ we have $5m(1-\lambda_B) < 1$ thus $(5m\,(1-\lambda_B) - 1)\,\frac{\delta \mu_L^* m}{9} - c_A < 0$. We thus always have $E\Pi_A(q = q_H/c) < E\Pi_A(q = q_L/c)$, regardless of firm $B$'s behavior. The same is true for firm $B$, so both firms carry on with business as usual.

2) Solving for the set of equations 4.4 gives this solution. Since $m \geq \frac{1}{5}$ and $\delta m \rho(5m - 1) - 5\delta m^2 \rho - 9(1 - m(1-\rho)) = -\delta m \rho - 5\delta m^2 \rho - 9(1 - m(1-\rho)) < 0$, this solution is positive and smaller than 1. ■

# 4 Strict label

With a strict label, a firm producing high quality is granted the label only with a probability $m$ while a firm that produces low quality is never granted the label. Consumers consider that the "label" signal, $l$, is hard information and indicates a high quality firm; But there is uncertainty about the "non label" signal, $\varnothing$: Consumers form a belief $\mu_S = \Pr[q = q_H | s = \varnothing]$ that the firm they are facing is in reality producing high quality.

Again, we start with the competition stage.

## 4.1 Competition stage

As in the previous section, we just need to study the case where one firm receives the label and the other does not. We know that with a strict label, a firm producing high quality is granted the label only with a probability $m$ and that a firm that produces low quality is never granted the label. Consequently, when a consumer of type $\theta$ receives the signal $s = l$, she is sure that it is a good of high quality.

Using the same notations as before, her surplus is simply: $S_l(\theta) = v + \theta(1+\delta) - p_l$. Contrarily, she is unsure about the true quality of the good when the signal is $s = \varnothing$. In this case she forms a belief $\mu_S$ that the product is of high quality, as defined above, and her surplus to consume the unlabeled product is: $S_\varnothing(\theta) = v + \theta(\mu_S\delta + 1) - p_\varnothing$. The indifferent consumer between the two goods is of type $\hat{\theta}$ such that $S_l(\hat{\theta}) = S_\varnothing(\hat{\theta})$, hence $\hat{\theta} = \frac{p_l - p_\varnothing}{(1-\mu_S)\delta}$.

Again, the labeled firm chooses the price $p_l$ that maximizes $\pi_l = (1-\hat{\theta})p_l$ and the unlabeled firm chooses the price $p_\varnothing$ that maximizes $\pi_\varnothing = \hat{\theta}p_\varnothing$. We can thus derive firms' profits for all labeling outcomes and sum them up in the following lemma:

**Lemma 10** *With a strict label:*

- *When firms both get the signal label, or both get the signal unlabeled, firms have null profits and share the market equally.*
- *When firms get different signals, the labeled firm has a market share of $\frac{2}{3}$ and its profit is $\pi_l = \frac{4(1-\mu_S)\delta}{9}$ while the unlabeled firm has a market share of $\frac{1}{3}$ and its profit is $\pi_\varnothing = \frac{(1-\mu_S)\delta}{9}$.*

**Proof.** $\frac{\partial \pi_l}{\partial p_l} = 1 - \frac{2p_l - p_\varnothing}{(1-\mu_S)\delta}$ and $\frac{\partial \pi_\varnothing}{\partial p_\varnothing} = \frac{p_l - 2p_\varnothing}{(1-\mu_S)\delta}$. Solving $\left[\frac{\partial \pi_l}{\partial p_l} = 0; \frac{\partial \pi_\varnothing}{\partial p_\varnothing} = 0\right]$ for $p_l$ and $p_\varnothing$ gives us $p_l = \frac{2(1-\mu_S)\delta}{3}$ and $p_\varnothing = \frac{(1-\mu_S)\delta}{3}$. We know that is is a maximum because $\frac{\partial^2 \pi_l}{\partial p_l^2} = -\frac{2}{(1-\mu_S)\delta} < 0$ and $\frac{\partial^2 \pi_\varnothing}{\partial p_\varnothing^2} = -\frac{2}{(1-\mu_S)\delta} < 0$. ■

## 4.2 Firms' choice of quality

Reasoning similarly as in the lenient case and using the same notations, the overall profit that firm $A$ expects when it produces low quality is:

$$\begin{aligned} E\Pi_A(q_L) &= \lambda_B\left[m\pi_\varnothing + (1-m)*0\right] + (1-\lambda_B)*0 \\ &= \lambda_B m \pi_\varnothing \end{aligned}$$

Reasoning similarly, we compute the overall profit that firm $A$ expects when it produces high quality:

$$E\Pi_A(q_H) = \pi_l m\left(1 - m\lambda_B\right) + \lambda_B m(1-m)\pi_\varnothing - c_A$$

We can now derive firm $A$'s incentive constraint in response to firm $B$'s behavior. Namely, using Lemma 2, Firm $A$ produces high quality if its quality cost $c_A$ is such that:[11]

$$c_A \leq \frac{\delta m\left(1-\mu_S\right)\left(4-5m\lambda_B\right)}{9} \tag{4.5}$$

Proceeding similarly for firm $B$, we get that it has a similar incentive constraint than firm $A$'s. As previously, we denote $c_A^*$ (resp. $c_B^*$) the cost thresholds at equilibrium of firm $A$ (resp. $B$) and we have $\lambda_A = c_A^*$ and $\lambda_B = c_B^*$.

Since consumers have the prior $\rho$, their equilibrium belief $\mu_S^*$ that the unlabeled firm is in reality producing high quality is:[12]

$$\mu_S^* = \frac{(1-m)\rho}{1-m\rho} \tag{4.6}$$

Combining the incentives constraints and the belief, the equilibria is defined by the following set of equations:[13]

$$\begin{aligned} c_A^* &= \frac{\delta m(1-\rho)\left(4-5mc_B^*\right)}{9(1-m\rho)} \\ c_B^* &= \frac{\delta m(1-\rho)\left(4-5mc_A^*\right)}{9(1-m\rho)} \end{aligned} \tag{4.7}$$

Solving these equations for $c_A^*$ and $c_B^*$ gives us the following proposition:

[11] $E\Pi_A(q_H) \geq E\Pi_A(q_L) \Leftrightarrow \pi_l m\left(1-m\lambda_B\right) + \lambda_B m(1-m)\pi_\varnothing - c_A \geq \lambda_B m\pi_\varnothing \Leftrightarrow c_A \leq \pi_l m\left(1-m\lambda_B\right) - \pi_\varnothing \lambda_B m^2 \Leftrightarrow c_A \leq \frac{(1-\mu_S)\delta}{9}\left[4m\left(1-m\lambda_B\right) - \lambda_B m^2\right] \Leftrightarrow c_A \leq \frac{\delta m(1-\mu)(4-5m\lambda_B)}{9}$

[12] $\mu_L^* = prob(q_H/\varnothing) = \frac{prob(\varnothing/q=q_H)prob(q=q_H)}{prob(\varnothing/q=q_H)prob(q=q_H)+prob(\varnothing/q=q_L)prob(q=q_L)} = \frac{(1-m)\rho}{(1-m)\rho+(1-\rho)} = \frac{(1-m)\rho}{1-m\rho}$

[13] $c_A^* = \frac{\delta m\left(4-5mc_B^*\right)}{9}\left[1 - \frac{(1-m)\rho}{1-m\rho}\right] = \frac{\delta m(1-\rho)\left(4-5mc_B^*\right)}{9(1-m\rho)}$

**Proposition 15** *There always exists a unique equilibrium with symmetrical behaviors for firms A and B such that, for any value of m, firms choose to invest and produce high quality if their cost c is below the following threshold:*

$$c_S^* = \frac{4\delta m(1-\rho)}{9(1-m\rho)+5\delta m^2(1-\rho)}$$

**Proof.** Solving for the set of equations 4.7 gives this solution. ■

# 5 Comparison between lenient and strict labels

Having characterized the market equilibria of the lenient and strict cases in Propositions 1 and 2, we can compare the two situations. We first compare the level of average quality, i.e. the probability that firms produce high quality at equilibrium, and then consider social welfare.

## 5.1 Average quality

In both cases, firms adopt symmetric cost-threshold behaviors at equilibrium. In consequence, to compare the overall level of high quality in the two situations, we can directly compare the thresholds that we have denoted $c_L^*$ for the lenient label situation and $c_S^*$ for the strict label situation. Thus we have the following proposition:

**Proposition 16** *The threshold $c_L^*$ is increasing with $\rho$ while $c_S^*$ is decreasing with $\rho$ and we have:*

- *For $m < \frac{1}{5}$, a strict label always spurs a higher average quality than a lenient label.*
- *For $m \geq \frac{1}{5}$, a lenient label spurs a higher average quality than a strict label for high values of $\rho$ and $m$, more precisely, for $\rho > g(\delta, m)$ with $g(\delta, m) = \frac{\sqrt{625\delta^2m^4+2250\delta m^2+6480m+729}+25\delta m^2+90m}{50\delta m^2+90m}$. Conversely, a strict label spurs a higher average quality than a lenient label for $\rho < g(\delta, m)$.*

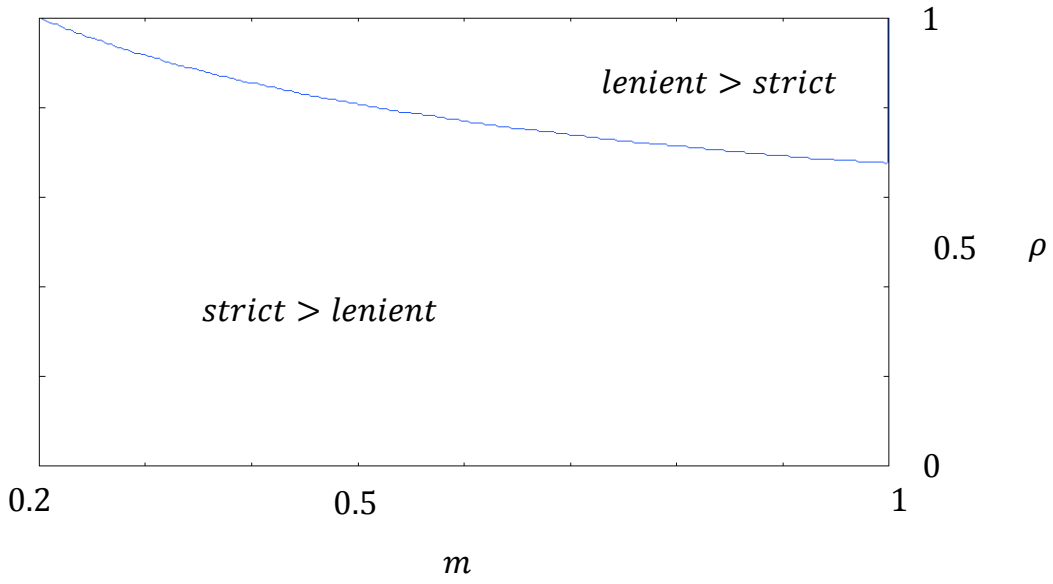



Figure 2: **Parameters' zones for which a strict label dominates a lenient label, and conversely, for** $\delta = 0.5$.

**Proof.** We have $\frac{\partial c_A^*}{\partial \rho} = \frac{9\delta m(1-m)(5m-1)}{(5\delta m^2\rho+9(1-m(1-\rho)))^2} > 0$ while $\frac{\partial c_A^*}{\partial \rho} = \frac{36\delta m(m-1)}{(9(1-m\rho)+5\delta m^2(1-\rho))^2} < 0$.

1) Obvious since in this case $c_L^* = 0$ and $c_S^* > 0$.

2) For $m \geq \frac{1}{5}$, both $c_L^*$ and $c_S^*$ are non null and we have: $c_L^* - c_S^* = \frac{\delta m\rho(5m-1)}{5\delta m^2\rho+9(1-m(1-\rho))} - \frac{4\delta m(1-\rho)}{9(1-m\rho)+5\delta m^2(1-\rho)} \Leftrightarrow c_L^* - c_S^* = \frac{\delta m(1-m)(25\delta m^2\rho^2+45m\rho^2-25\delta\rho m^2+27\rho-36)}{(5\delta m^2\rho+9(1-m(1-\rho)))(9(1-m\rho)+5\delta m^2(1-\rho))}$

Solving $c_L^* - c_S^* = 0$ gives us that $c_L^* > c_S^*$ for $\rho > g(\delta, m) = \frac{\sqrt{625\delta^2m^4+2250\delta m^2+6480m+729}+25\delta m^2+90m}{50\delta m^2+90m}$.

■

Figure 2 illustrates the values of $m$ and $\rho$ for which a strict label is more efficient than a lenient (down left zone) and for which a lenient label is more efficient than a strict (upper right zone), for a value of $\delta = 0.5$.

Let us discuss the fact that higher consumers' priors trigger higher average quality in the lenient case (higher equilibrium cost thresholds). The more optimistic the consumer was originally (high $\rho$), the higher will be its equilibrium belief $\mu_L^*$,[14] which in turn widens the gap between the labeled profit and the unlabeled profit since $\pi_l - \pi_\varnothing = \frac{\mu_L\delta}{3}$.

[14] $\frac{\partial \mu_L^*}{\partial \rho} = \frac{1-m}{(1-m(1-\rho))^2} > 0$

Intuitively, this is due to fact that, with a lenient label, the uncertainty lays on labeled firms, while unlabeled firms are definitely identified as low quality firms. Thus the consumers' optimistic state of mind benefits to labeled firms. When facing a labeled firm, a more optimistic consumer increasingly expects high quality and is willing to pay more. Since producing high quality guarantees to be labeled, it becomes an increasingly more interesting option than producing low quality as the prior increases.

The opposite happens in the strict case: While more optimistic consumers (high $\rho$) still entails a higher equilibrium belief $\mu_S^*$,[15] it now shrinks the gap between the labeled profit and the unlabeled profit, when a differentiating labelling outcome arises, since $\pi_l - \pi_\varnothing = \frac{(1-\mu_S)\delta}{3}$. It is due to the fact that, with a strict label, the uncertainty lays on the unlabeled firm, which leads the consumers' optimistic state of mind to benefit to unlabeled firms. Since producing low quality guarantees to be unlabeled and, in addition, is costless, it becomes a more interesting option than producing high quality as the prior increases.

It follows naturally that a lenient label is more efficient in triggering higher average quality for high values of $\rho$.

## 5.2 Social welfare

As explained earlier, there is no obvious reasons that a higher level of average quality would induce a higher social welfare because differentiation creates market power, which can reduce consumers' surplus. That is why we now compare the two labels in terms of social welfare.

The social welfare function is the sum of consumers' surplus and firms' profits, which falls down to consumers' utility minus the cost of quality:

[15] $\frac{\partial \mu_S^*}{\partial \rho} = \frac{1-m}{(1-m\rho)^2} > 0$

$$W = \int_0^1 S(\theta)d\theta + \Pi_A(c_A^*) + \Pi_B(c_B^*) = \int_0^1 U(\theta)d\theta - c_A^* - c_B^*$$

As a benchmark, we characterize the first best optimum under perfect information.

### 5.2.1 First best optimum

Ignoring first the cost of quality, we have:

$$W(\hat{\theta}) = \int_0^{\hat{\theta}} \theta q_L d\theta + \int_{\hat{\theta}}^1 \theta q_H d\theta = \frac{1}{2}\hat{\theta}^2 + \frac{1}{2}(1+\delta)\left(1-\hat{\theta}^2\right)$$

We first show that consumers should not consume the two qualities in the social optimum: Assume that high quality is sold on the market, meaning that the fixed cost $c$ has been incurred. As high quality is more valuable to the consumers, low quality should not be produced. Formally, this means $\hat{\theta} = 0$. Then, it may be the case that the fixed cost is too high so that the investment in high quality is not socially valuable, implying that $\hat{\theta} = 1$. It is the case if $W(0) - c < W(1)$, which is equivalent to $c > \frac{\delta}{2}$. Hence, the following lemma:

**Lemma 11** *In the first best optimum, product quality should be uniform. More specifically:*

- *When $c_A \geq \frac{\delta}{2}$ and $c_B \geq \frac{\delta}{2}$, the social optimum is that they both should produce the low quality good.*
- *When one of the two firms have a cost below $\frac{\delta}{2}$ and the other above, or when both firms have a cost below $\frac{\delta}{2}$, the social optimum is that one of them should produce the low quality good and the other do nothing.*

*In addition, the thresholds $c_L^*$ and $c_S^*$ as defined in propositions 1 and 2 are always below $\frac{\delta}{2}$. In imperfect information, with both type of label, high quality is undersupplied.*

**Proof.** 1) See above. 2) See above + in the case where both firms have their costs below $\frac{\delta}{2}$, from the social welfare point of view, only one of them should endure the cost and cover the whole market rather than both firms enduring the costs and splitting the market. 3) $\frac{\delta}{2} - c_L^* = \frac{\delta}{2} - \frac{\delta m\rho(5m-1)}{5\delta m^2\rho+9(1-m(1-\rho))} = \frac{\delta(5\delta\rho m^2+m\rho(11-10m)+9(1-m))}{2(5\delta m^2\rho+9(1-m(1-\rho)))} > 0$ and $\frac{\delta}{2} - c_S^* = \frac{\delta}{2} - \frac{4\delta m(1-\rho)}{9(1-m\rho)+5\delta m^2(1-\rho)} = -\frac{\delta(5\delta m^2(\rho-1)+m(8+\rho)-9)}{2(9(1-m\rho)+5\delta m^2(1-\rho))} > 0$ ■

This result was expected: imperfect information and competition lead to an underprovision of high quality.

### 5.2.2 Equilibrium welfare

Three scenarii are possible: both firms produce high quality, both firms produce low quality or one firm produces high quality while the other produces low quality. Consequently, the social welfare function at equilibrium writes:

$$W(c^*) = \Pr(q_H, q_H)E(U/q_H, q_H) + \Pr(q_H, q_L)E(U/q_H, q_L) + \Pr(q_L, q_L)E(U/q_L, q_L) - E(c^*)$$

We first tackle the issue of consumers' utility, by considering the scenarii in turn.

**1) If both firms' types are below** $c^*$, they both select $q = q_H$ and the consumers purchase only high quality. This event occurs with a probability $\Pr(q_H, q_H) = (c^*)^2$.

In this case, all consumers purchase high quality whatever the labeling outcome, hence the expected utility is:

$$E(U/q_H, q_H) = \int_0^1 \theta\,(\delta + 1)\,d\theta = \frac{1}{2}(1+\delta)$$

**2) If both firms' types are above** $c^*$**,** they both choose $q = q_L$ and the consumers purchase only low quality. This event occurs with a probability $\Pr(q = q_L, q = q_L) =$

$(1-c^*)^2$.

In this case, all consumers purchase low quality whatever the labeling outcome, hence the expected utility is:

$$E(U/q_L, q_L) = \int_0^1 \theta d\theta = \frac{1}{2}.$$

**3) If one firm type is below $c^*$ while the other's is above,** the former selects $q = q_H$, the latter chooses $q = q_L$ and the consumers can purchase either low or high quality. This event occurs with a probability $\Pr(q_H, q_L) = 2c^*(1-c^*)$.

In this case, one firm produces high quality while the other produces low quality: Things get more complicated as the market shares of the two types of qualities depend on the outcome of the labeling process. These processes being different for the lenient and the strict label, we need to treat them separately.

**Lenient label**

There are three possible labeling outcomes: no firm get the label $(\varnothing, \varnothing)$, one out of the two firms gets the label $(\varnothing, l)$ and both firms get the label $(l, l)$. Under a lenient label, we have the following conditional probabilities:[16]

$$\begin{aligned} \Pr(\varnothing, \varnothing \quad | \quad q_H, q_L) &= 0 \\ \Pr(\varnothing, l \quad | \quad q_H, q_L) &= m \\ \Pr(l, l \quad | \quad q_H, q_L) &= 1-m \end{aligned}$$

We can thus ignore the case $(\varnothing, \varnothing)$. In the case of *ex post* product differentiation $(\varnothing, l)$, we know from Lemma 1 that the unlabeled firm has a market share of $\frac{1}{3}$ while the labeled firm has a market share of $\frac{2}{3}$. But since we know that only one firm is indeed producing

[16]Remember that we are in the case where one firm produces high quality and the other produces low quality. For example, since a firm producing high quality always get the label, the outcome $(\varnothing, \varnothing)$ is impossible.

high quality and that it thus always gets the label, we can infer that the unlabeled firm is the low quality firm and the labeled firm is the high quality firm. Thus, the expected utility in this subcase is:

$$E(U_{Le}/q_H, q_L, (\varnothing, l)) = \int_0^{1/3} \theta d\theta + \int_{1/3}^{1} \theta(1+\delta) d\theta = \frac{4}{9}\delta + \frac{1}{2}$$

When both products are labelled $(l, l)$, we assume that market shares are equal so that $\hat{\theta} = 1/2$, each type of quality is produced in the same quantity. It gives the following expected utility:

$$E(U_{Le}/q_H, q_L, (l, l)) = \frac{1}{2}\int_0^{1} \theta d\theta + \frac{1}{2}\int_0^{1} \theta(1+\delta) d\theta = \frac{1}{4}\delta + \frac{1}{2}$$

Recollecting, the total expected utility is:

$$E(U_{Le}/q_H, q_L) = m\left(\frac{4}{9}\delta + \frac{1}{2}\right) + (1-m)\left(\frac{1}{4}\delta + \frac{1}{2}\right)$$

**Strict label**

Under a strict label, the conditional probabilities of the different labeling outcomes are:

$$\begin{aligned} \Pr(\varnothing, \varnothing \quad | \quad q_H, q_L) &= (1-m) \\ \Pr(\varnothing, l \quad | \quad q_H, q_L) &= m \\ \Pr(l, l \quad | \quad q_H, q_L) &= 0 \end{aligned}$$

We can thus ignore the case $(l, l)$. In the case of product differentiation $(\varnothing, l)$, we know from Lemma 2 that the unlabeled firm has a market share of $\frac{1}{3}$ while the labeled firm has a

market share of $\frac{2}{3}$. But since we know that only one firm is indeed producing high quality and that only such product can get the label, we can infer that the unlabeled firm is the low quality firm and the labeled firm is the high quality firm. Thus, the expected utility in this subcase is similar to the lenient case: $E(U_{St}/q_H, q_L, (\varnothing, l)) = \frac{4}{9}\delta + \frac{1}{2}$. Finally, when no products are labelled $(\varnothing, \varnothing)$, we again assume identical market shares and the expected utility is $E(U_{St}/q_H, q_L, (\varnothing, \varnothing)) = \frac{1}{4}\delta + \frac{1}{2}$.

To sum up, we end up with the exact same total expected utility as in the lenient case:

$$E(U_{St}/q_H, q_L) = m\left(\frac{4}{9}\delta + \frac{1}{2}\right) + (1-m)\left(\frac{1}{4}\delta + \frac{1}{2}\right).$$

**Expected costs**

Finally, expected costs are easily calculated since firms have the same cost threshold at equilibrium in both cases. Hence, total expected cost is

$$E(C) = 2\int_0^{c^*} cdc = (c^*)^2$$

**Total expected social welfare**

Let's wrap up and compare the two levels of social welfare. It has the same expression in the strict and lenient cases:

$$W(c^*) = \tfrac{1}{2}(c^*)^2(1+\delta) + \tfrac{1}{2}(1-c^*)^2$$
$$+2c^*(1-c^*)\left(m\left(\tfrac{4}{9}\delta + \tfrac{1}{2}\right) + (1-m)\left(\tfrac{1}{4}\delta + \tfrac{1}{2}\right)\right) - (c^*)^2$$

Using this formula in the two cases and comparing them gives us the following proposition:

**Proposition 17** • *For $m < \frac{1}{5}$, a strict label always induces a higher social welfare than a lenient label.*

- *For $m \geq \frac{1}{5}$, a lenient label induces a higher social welfare than a strict label when it induces a higher average quality ($c_L^* > c_S^*$), and conversely.*

**Proof.** We have $\frac{dW(c^*)}{dc^*} = \frac{9\delta+7\delta m-36c^*-14\delta mc^*}{36} > 0$. Consequently, whenever $c_L^* > c_S^*$ we have $W(c_L^*) > W(c_S^*)$, and conversely. ■

This tells us that higher level of quality and higher social welfare are achieved together. Remember that labels has a double effect on social welfare. On the one hand, it provides incentives for firms to invest and produce high quality which is welfare enhancing since the under provision of quality compared to the first best optimum is due to imperfect information. But on the other hand, differentiation alter competition and firms can charge prices higher than their production cost and make positive profits, which can be welfare damaging. Proposition 4 tells us that the former effect proves to be stronger than the latter.

# 6 Conclusion and Discussion

In the real world, a label is granted to a product if its level of quality is above a standard. This standard can be more or less strict, implying varying shares of labeled products in the market. In this paper, we seek to characterize the level of label strictness which maximizes social welfare.

To investigate this question, we develop a simple model that describes the interaction between two firms who can produce low or high quality and consumers who positively but heterogeneously value quality but are not able to observe products' real level of quality. They only observe the outcomes of a mandatory and free label that we first assume lenient and then strict.

Importantly, contrarily to others papers that study standard strictness under imperfect

competition firms have symmetric strategies at equilibrium: there is differentiation ex-post, but non ex-ante. This allow us to focus on informational insights.

We find that the welfare ranking is ambiguous. The strict label is more efficient in inducing high quality and social welfare than a lenient label if consumers' prior belief that the product is of high quality is high and/or if the certification is very noisy. And conversely.

As for the prior, the intuition is the following: under a lenient label, the uncertainty lays on labeled firms, thus a consumer's optimistic state of mind benefits to them. Since producing high quality guarantees to be labeled, it becomes an increasingly more interesting option than not investing as the prior increases. On the contrary, a high prior benefits to unlabeled firms under a strict label. Saving the cost of really producing high quality, which guarantees to get unlabeled, while benefiting from the optimistic consumers' state of mind becomes a more interesting option in these conditions. In conclusion, with a high prior, producing high quality is a more interesting option under a lenient label rather than a strict.

For example, since photovoltaic panels are viewed as an ecological device, our model advocates that a label intending to ensure that panels are produced with an environmentally friendly process should set lenient standards in order to induce the more investment in green production processes possible. On the contrary, given that cars are perceived as machines producing high quantities of greenhouse gases emissions, our model predicts that a label promoting low emissions models should have strict standards in order to be the most effective possible in motivating car manufacturers to invest in greenhouse gases free technologies.

These results allow us to derive policy implications. Public authorities can intervene in at least two ways. On the one hand, they can directly impact the investment in quality by setting up labels with the right level of standards' strictness: lenient for products

already considered by consumers to be of rather high quality, or strict for products that consumers regard suspiciously. On the other hand, they can indirectly impact investment in quality by increasing the precision of existing labels' monitoring technology. This can be done either by subsidizing the labeling institution, hence increasing their resources to investigate products attributes, or by legislating to ease labeling institutions' access to information (e.g. mandatory disclosure of sensible information for firms).

We are considering several possible ways to further our analysis.

Firstly, most labels are voluntary and entail some cost. Only a mandatory and subsidized label fits completely with our model. It would be a necessary robustness check to verify whether our results hold with a voluntary and costly label.

Secondly, we would like to introduce different types of labelling institutions: a for profit one that only seeks to maximize its profit and a non profit one that aim to maximize the production of high quality. They may prefer a type of label that is not the social welfare maximizing one. In particular, we expect a for profit labeling institution to be biased toward a lenient label since firms end up labeled more often, which may enhance its attractivity and hence the certifier market share. Such an extension would belong to the strand of theoretical literature about quality disclosure dedicated to certifiers that analyzes their capacity to be precise and unbiased given potential noise in the data or conflict of interests (for a literature review on this topic, see Dranove and Jin (2010) section 2.2).

Finally, we have introduced an important difference between our paper and the model it is based upon: In Fleckinger et al.'s, consumers are sophisticated and guess exactly firms' incentive constraint to produce high quality and altogether the proportion of firms that indeed produce high quality at equilibrium, which is consequently a perfect Bayesian Equilibrium. Our model entailing competition, unlike theirs, the complexity of the mechanism that a perfectly sophisticated consumer should anticipate at equilibrium turns out

to represent quite a challenge. That is why we found more realistic to limit somehow consumers discernment here, but it would be interesting to see whether our results vary with a sophisticated consumer.

# Conclusion

This thesis contributes to the economic literature about Corporate Social Responsibility (CSR) by providing a theoretical analysis of the role of third parties in creating incentives for firms to self-regulate, by conveying information about their actual behavior to their stakeholders. More precisely, I focus on understanding the differences between the impact on firms' incentives of good news, i.e., when third parties spot light on virtuous firms, and bad news, i.e., when third parties expose irresponsible firms. I will not go throughout the results in this conclusion as they are available in the introduction and in the various chapters of this dissertation. I will rather focus on policy implications and perspectives for future research.

## Policy implications

### Policy discussions acknowledge a crucial need for credible information

Firms' behavior is nowadays under an intense scrutiny and CSR has become a persasive notion in firms' communication supports. But for CSR to become a reality and not merely be a public relations tool, credible information about firms' real actions must be produced and made publicly available. This is the key challenge about CSR, and it is recognized as such in the policy arena.

An increasing number of countries set mandatory reporting standards for firms' social and environmental impacts. French was precursor in this matter with the 2001 NRE law[17] that mandates listed companies to report their social and environmental impacts, an obligation that was extended to non listed companies with the 2009 and 2010 Grenelle laws.[18]

[17]Nouvelles Régulations Economiques, or New Economic Regulations, that issue new obligations for listed companies.

[18]Grenelle laws introduce a panel of new obligations that intend to better protect the environment.

At the international level, the ISO 26000 standard was issued in 2010 by the International Standard Organization (ISO) with the objective to provide "guidance on how businesses and organizations can operate in a socially responsible way. This means acting in an ethical and transparent way that contributes to the health and welfare of society." This new norm was approved by 93% of the 99 countries and organizations members of the ISO.

Firms themselves develop strategies to improve their reporting credibility. For example, a lot of firms publicly follow internationally recognized guidelines for responsible business, such as the GRI,[19] or support global initiatives such as the Global Compact[20] or the Carbon Disclosure Project.[21] Some firms also integrate Socially Responsible Investment indexes such as the ASPI Eurozone[22] or Ethibel.[23] Another existing strategy is to team up with credible third parties that can provide firms with a morale guarantee, such as the Comité 21[24] or the Observatoire de la Responsabilité Sociétale des Entreprises[25] (see **Chapter 3**).

My thesis helps to understand how third parties, such as NGOs, labeling institutions and extra-financial rating agencies, can induce firms to self regulate through the provision of public information on their actions in the most effective way. This allows us to derive policy implications that contribute to the above-mentioned debate.

[19]The Global Reporting Initiative produces one of the world's most prevalent guidelines for sustainability reporting. It was formed by the United States based non-profits Ceres (formerly the Coalition for Environmentally Responsible Economies) and Tellus Institute, with the support of the United Nations Environment Programme.

[20]The United Nation initiative Global Compact is a principle-based framework for businesses, stating ten universally accepted principles in the area of human rights, labour, environment and anti-corruption.

[21]The Carbon Disclosure Project is an independent not-for-profit organization holding the largest database of primary corporate climate change information in the world.

[22]Advanced Sustainable Performance Indices. It gathers the 120 firms with best Vigeo extrafinancial notation.

[23]This index mixes a best-in-class approach and ethical exclusion criteria.

[24]It is a French network that was created after the Rio Summit in 1994, with the goal to implement sustainable development in the French regions. It includes firms, local authorities, NGOs, etc.

[25]The ORSE is a French network designed to study and promote socially responsible investment. It includes firms, trade unions, NGOs, etc.

## Good news or bad news?

Throughout all its results, this thesis highlights that provision of good and bad news impact very differently firms' incentives to self-regulate depending on the contexts. Extending our results from **Chapter 2** to all types of auditors, there is one important parameter that must be taken into account to weigh the relative impact of good and bad news: the overall amount of financial resources available to auditors. Namely, devoting resources to uncover bad practices is more efficient when resources are high, and conversely.

There is a range of various situations where auditors have a lot of resources and consequently provide a lot of information on firms' actions. For example, firms that deal directly with the final market are subject to more acute observation than firms that are intermediaries in the production process. Indeed, their brands and names are much more known to final consumers and it follows naturally that public attention focuses more on these firms' actions. If intermediaries end up being targeted, it is generally because they deal with firms subject to intense media attention, as it was the case for Foxconn and Apple[26].

Firms that are subject to intense scrutiny typically include extraction industries (mining, oil, gas, etc.). Clearly, this is so because if these firms fail to meet basic environmental and social standards, it can have dire consequences. Think of the 2010 BP's Deepwater Horizon massive oil spill in the Gulf of Mexico, or to Chinese coal-mines that have the reputation to be the most dangerous in the world, with an official figure of 6000 annual deaths[27].

Beyond specific sectoral characteristics of firms that are likely to be correlated with

[26] In 2009 and 2010, poor working conditions in the Foxconn factories that supply iPhones, iPads have regularly been denounced in the media. For example, in 2009, Foxconn guards were videotaped beating employees, or in April 2010, four workers committed suicide in a single month in the same factory.

[27] Officials figures reported by the NGO Chinese Labour Bulletin in its 2008 report "Bone and Blood : The Price of Coal in China", retrieved on September 20th, 2012 at http://www.clb.org.hk/en/files/File/bone_and_blood.pdf

particular scrutiny, some dimensions accross all sectors are also more sensitive. The obvious example is the emissions of greenhouse gases emissions (GHGs). In the last decades, with global warming becoming an internationally recognized issue, the pressure on firms to report and reduce their GHGs' emissions has considerably increased, even beyond potential legal requirements. One can also think of gender diversity or the level of stress in the workplace, two topics that are regularly covered in the western countries media.

In the light of our results, we can say that it would be more efficient for auditors to adopt adverse informational behavior about these types of firms or on these types of practices in order to improve firms' performance.

On the contrary, there are others contexts where information is scarce about environmental and social practices. An example is developing countries where the civil society is sometimes not well expanded and structured, resulting in little information about firms' actions. This can be due to legal impediments, a lack of financial resources available to auditors or simply to the fact that firms' CSR programs are not auditors' priority. Another case is the one of small and medium enterprises (SMEs), which represent a major share of our economy[28] and consequently have at least major social impacts. Nevertheless, they have not been targeted so far by extensive auditors investigation. Our results indicate that in these contexts where information is scarce, the few auditors at work should better be friendly.

### The case of Non-Governmental Organisations

In terms of public intervention, our results are clear and may seem surprising: as long as financial resources are low, there is no need for any intervention. This is due to firms' interest to privately finance NGOs. On the one hand, firms that would find beneficial to self-regulate under perfect information are ready to finance NGOs to allow them to

[28] As an illustration, in France, in 1997, SMEs represented 99,9% of firms, 69% of employment and 61,8% of the Gross Domestic Product while in the United States it was respectively 99,7%, 53,7% and 48 %, according to the OECD.

produce information on their actions. But on the other hand, firms that would carry on with business as usual anyway are also ready to increase NGOs' resources because it raises stakeholders' belief that the average firm is indeed responsible, a state of mind that benefits to all firms indistinctively.

Things are different when resources are high: the latter firms are not interested anymore in increasing NGOs resources because the risk that their actual behavior be exposed is now more important than the potential benefit due to a more optimistic stakeholder belief. Thus there may be room for public subsidizing of NGOs to further improve the generation of credible information. In addition, we show that the lack of coordination between NGOs may result in the persistence of friendly behaviors when it would be more efficient for them to be hostile, because resources are high. Public authorities can also improve social welfare by helping NGOs to coordinate and adopt hostile behaviors.

### The case of labels

Turning now to labels, this dissertation provides indications on the most effective level of label's strictness when dealing with product features that consumers consider with strong *a priori*, be they positive or negative. For example, European consumers are very concerned with Genetically Modified Organisms consequences on their health and on the environment. Similarly, consumers expect cars to be the source of high GHGs' emissions. Our results advise to set up labels with very strict standards in order to spur firms to adopt responsible practices on these issues in the most effective way.

On the contrary, organic vegetables and fruits are considered favorably by consumers, while sometimes their production processes are the cause of non negligible social and environmental impacts (e.g. because they are cultivated far from their final markets). Here, it would be interesting to set up a lenient label in order to single out low quality organic products.

Public authorities can also improve labels' impact on firms' incentives to provide high quality by stiring up detailed and structured public debates about the social and environmental issues at stake, in order to educate consumers and allow them to hold more sophisticated and informed opinions. We have highlighted that such debates would also prove useful more generally in hostile informational environments, where low quality is more often revealed than high quality. In these environments, there is a risk that consumers' beliefs become self fulfilling: suppose that consumers are ex ante very optimistic, supplying high quality is an equilibrium because hostile monitoring would easily reveal the agents supplying low quality. But when consumers' prior belief is pessimistic, incentives to increase quality are limited because it is relatively difficult to ascertain high quality, and there is few hope for firms to prove wrong a pessimistic belief. Consequently, intensifying the public debate in this case can deter the formation of *a priori* that would end up creating a reality.

## Perspectives

I would have liked to further my research on the differential impact of good and bad news by taking into account the following points:

- In this thesis, I have exclusively dealt with the first step in the monitoring and communication bridge that third parties create between firms' real actions and stakeholders. Namely, we have not formally explored the communication process that occurs between third parties and stakeholders and its consequences, while it is far from being as straightforward as we assume it to be.

  First, firms' communication about CSR may not be credible, but that does not mean either that all third parties are considered by stakeholders as unbiased. Many situations ground such defiance. Extra-financial rating agencies are remunerated for their services, which could induce them to be softer toward firms. Similarly, some

NGOs receive funding from firms, thus it is understandable that they should be viewed with some suspicion if they then endorse these same firms. Finally, some labeling institutions are purely and simply created by consortia of firms or by industries, which may cast doubt on the definition of the label standards as well as its attribution process.

In conclusion, good news may not be as trusted by stakeholders as bad news and thus not as efficient in inducing firms to self-regulate. It would be interesting to explore how third parties trade off between additional resources thanks to firms' donations and decreased efficiency of their communication toward stakeholders because they accepted such donations.

Second, bad news generally receive a wider attention from the media and third parties may favor them in order to gain visibility and, in the case of NGOs, donators. It would be interesting to model the competition for resources between NGOs and see how it impacts the relative provision of good and bad news.

- Finally, I would have liked to capture the phenomenon of greenwashing. Indeed, in our model bad news only mean that the firms do not self-regulate, but in reality bad news sometimes also reveal that a firm was lying about its true level of self-regulation, i.e., greenwashing. To do so, one could add a communication step in the base model. First, firms would self-regulate or not. Second, they would claim that they are self-regulating, or say nothing. And finally, third parties would reveal information, following the same scheme as before. The difference would be that bad news would sometimes be a double bad news, "the firm is irresponsible *and* was lying", which would entail a larger punition from stakeholders than for a simple bad news, "the firm is irresponsible but was not claiming to be self regulating". It is reasonable to assume that the punition would be more important in this case because

it means that the firms is untrustworthy: stakeholders and firms interact on other issues than CSR and if a firm lies about one of them, it could be lying about other issues. This loss of trust would impact the whole stakeholder-firm relationship.

# Bibliography

Akerlof, G., 1970. The Market for Lemon's: Quality Uncertainty and the Market Mechanism, *Quarterly Journal of Economics*, 84(3), 488-500.

Albinger H.S., Freeman S.J., 2000. Corporate Social Performance and Attractiveness as an Employer to Different Job Seeking Populations. *Journal of Business Ethics*, Vol.28(3): 243-253(11).

Ambec, S. and Lanoie, P., 2007. When and why does it pay to be green ? *Working Papers 200704*, Grenoble Applied Economics Laboratory (GAEL).

Arora, S. and Gangopadhyay, S., 1995. Toward a theoretical model of voluntary overcompliance. *Journal of Economic Behavior and Organization*, 28, 289-309.

Arrow, K., 1973. The Theory of Discrimination, In O. Ashenfelter and A. Rees, eds., *Discrimination in Labor Markets*. Princeton, N.J.: Princeton University Press, 3-33.

Auger, P., P.Burke, T. M. Devinney, and J. J. Louviere, 2003. What Will Consumers Pay for Social Product Features? *Journal of Business Ethics*, 42 (February): 281–304.

Avenhaus, R., von Stengel, B. and S. Zamir, 2002. Inspection games, Chapter 51, 1947-1987, in *Handbook of Game Theory*, eds. R. J. Aumann and S. Hart, North-Holland, Amsterdam.

Babakri, K.A., R. A. Bennett, S. Rao, M. Franchetti, 2004. Recycling performance of firms before and after adoption of the ISO 14001 standard. *Journal of Cleaner Production*, Vol. 12(6): 633-637.

Bagnoli, M. and Watts, S., 2003. Selling to Socially Responsible Consumers: Competition and The Private Provision of Public Goods. *Journal of Economics & Management Strategy,* 12(3), 19-445.

Bagwell, K. and Riordan, M., 1991. High and Declining Prices Signal Product Quality, *American Economic Review*, Vol. 81(1), 224-239.

Barbera, A.J. and V.D. McConnell, 1990. The Impact of Environmental Regulations on Industry Productivity: Direct and Indirect Effects. *Journal of Environmental Economics and Management*, 18, 50-65.

Bar-Isaac, H. and Tadelis, S., 2008. Seller Reputation, *Foundations and Trends in Microeconomics* 4(4), 273-351.

Barla, P., Perelman, S., 2005. Sulphur emissions and productivity growth in industrialised countries. *Annals of Public and Cooperative Economics*, 76, pp. 275-300.

Barla P., 2007. ISO 14001 certification and environmental performance in Quebec's pulp and paper industry, *Journal of Environmental Economics and Management,* Vol. 53(3): 291-306.

Baron, D.P., 2001. Private Politics, Corporate Social Responsibility, and Integrated Strategy. *Journal of Economics & Management Strategy,* 10(1): 7-45.

Baron, D.P., 2006. *Business and Its Environment,* 5th ed. Upper Saddle River, NJ: Pearson/Prentice-Hall.

Baron, D.P. and D. Diermeier, 2007. Strategic Activism and Nonmarket Strategy. *Journal of Economics & Management Strategy*, 16: 599-634.

Bartley, T. and C. Child, 2011. Movements, Markets and Fields: The Effects of Anti-Sweatshop Campaigns on U.S. Firms, 1993-2000. *Social Forces,* 90:425-451.

Batley, S.L., Colbourne, D., Fleming, P.D., Urwin, P., 2001. Citizen versus consumer: challenges in the UK green power market. *Energy Policy,* 29, 479–487.

Batley, S.L., Fleming, P.D., Urwin, P., 2000. Willingness to pay for renewable energy: implications for UK green tariff offerings. *Indoor and Built Environment,* 9, 157–170.

Bazillier, R. and J. Vauday, 2009. The Greenwashing Machine : is CSR more than Communication. *LEO Working paper.*

Bennear, L. S., and S. M. Olmstead, 2008. The Impacts of the 'Right to Know': Information Disclosure and the Violation of Drinking Water Standards. *Journal of Environmement*, 56(2):117–30.

Bennett, J., R. Blamey, J.J. Louviere, M. Morrison, 2001. Green product choice, in: J. Bennett, R. Blamey (Eds.), *The Choice Modeling Approach to Environmental Valuation*, Edward Elgar, Cheltenham, UK.

Berman, E. and L.T.M. Bui, 2001. Environmental regulation and productivity: evidence from oil refineries. *The Review of Economics and Statistics*, 83(3), 498-510.

Björner, T. B. , L. G. Hansen and C. S. Russell, 2004. Environmental labeling and consumers' choice—an empirical analysis of the effect of the Nordic Swan. *Journal of Environmental Economics and Management*, Vol.47(3), 411-434.

Blume, L., 2005. Learning and Statistical Discrimination, *American Economic Review (Papers and Proceedings)*, 95(2), 118-121.

Board, S. and Meyer-ter-Vehn, M., 2010. Reputation for Quality. *Mimeo* UCLA.

Bollino, C.A., 2009. The willingness to pay for renewable energy sources: the case of Italy with socio-demographic determinants. *The Energy Journal*, 30 (2), 81–96.

Borchers, A.M., Duke, J.M., Parsons, G.R., 2007. Does willingness to pay for green electricity differ by source? *Energy Policy*, 35, 3327–3334.

Bottega, L. and De Freitas, J., 2009. Public, Private and Nonprofit Regulation for Environmental Quality. *Journal of Economics & Management Strategy*, 18(1), 105-123.

Boyd, G.A., McClelland, J.D., 1999. The impact of environmental constraints on productivity improvement in integrated paper plants. *Journal of Environmental Economics and Management*, 38, 121–142.

Brammer S, Millington A, Rayton B., 2007. The contribution of corporation social responsibility to organizational commitment. *Int. J.Hum. Res. Manag.* 18 (10):1701-1719.

Brekke, K. and Nyborg, K., 2008. Attracting responsible employees : Green production as labor market screening. *Resource and Energy Economics*, 30(4), 509-526.

Brock, W. and Durlauf, S., 2001. Discrete Choice with Social Interactions, *Review of Economic Studies*, Vol. 68(2), 235-260.

Brouhle K., B. Graham, and D.Harrington, 2012. Innovation under the Climate Wise Program. *Working paper.*

Brouhle, K., C. Griffiths, and A. Wolverton, 2009. Evaluating the Role of EPA Policy Levers: An examination of a voluntary program and a regulatory threat in the metal finishing industry. *Journal of Environmental Economics and Management*, 57(2): 166-81.

Byrnes, B., Jones, C., Goodman, S., 1999. Contingent valuation and real economic commitments: evidence from electricity utility green pricing programmes. *Journal of Environmental Planning and Management*, 42, 149–166.

CapitalCom, the Institut Français des Administrateurs and the Observatoire sur la Responsabilité Sociétale des Entreprises, 2010. Baromètre Annuel de la Gouvernance Extra-financière. Available at http://www.capitalcom.fr/Documents/CP%20GOUVERNANCE%20EXTRA%20FI%20VF.pdf

Cason TN, Gangadharan L., 2002 Environmental labeling and incomplete consumer information in laboratory markets. *Journal of Environmental Economics and Management*, 43(1):113-134.

Cespa G., G. Cestone, 2007. Corporate Social Responsibility and Managerial Entrenchment. *Journal of Economics and Management Strategy*, 16 (3), 741-771.

Chen, M.,. 2008. Minimum Quality Standards and Strategic Vertical Differentiation: An Empirical Study of Nursing Homes.*Unpublished.*

Chih-Hai Y., Y. Tseng, and C. Chen, 2011. Environmental Regulations, Induced R&D, and Productivity: Evidence from Taiwan's Manufacturing Industries. The International Centre for the Study of East Asian Development, *Kitakyushu Working Paper* Series Vol. 2011-18.

Christainsen G. B., and R. H Haveman, 1981. The contribution of environmental regulations to the slowdown in productivity growth. *Journal of Environmental Economics and Management*, Vol. 8(4): 381-390.

Coate, S. and G.C. Loury, 1993. Will Affirmative-Action Policies Eliminate Negative Stereotypes? *American Economic Review*, 83(5), 1220-1240.

Conrad K. and D. Wastl, 1995. The Impact of Environmental Regulation on Productivity in German Industries. *Empirical Economics*, 20:615-633.

Crampes C. and A. Hollander, 1995. Duopoly and quality standards. *European Economic Review*, Vol.39(1): 71-82.

Crifo P., V. Forget and S. Teyssier, 2012. The price of unsustainability. *Working paper.*

Dasgupta S. , H. Hettige, D. Wheeler, 2000. What improves environmental compliance? Evidence from Mexican industry. *J. Environ. Econ. Manage.,* 39:39–66.

Delmas, M. A. and L. E. Grant, 2008. Eco-Labeling Strategies: The Eco-Premium Puzzle In The Wine Industry. *Working Papers 37325*, American Association of Wine Economists.

Denicolò, V., 2008. A Signaling Model of Environmental Overcompliance. *Journal of Economic Behavior & Organization,* 68, 293-303.

De Pelsmacker P., L. Driesen , G. Rayp, 2003. Are fair trade labels good business? Ethics and coffee buying intentions. *Working Paper Ghent University*, Faculty of Economics and Business Administration.

Dranove, D., and Jin, G.Z., 2010. Quality Disclosure and Certification Theory and Practice, *Journal of Economic Literature*, 48(4), 935-963.

Dranove, D., D. P. Kessler, M. McClellan, and M. Satterthwaite, 2003. Is More Information Better? The Effects of 'Report Cards' on Health Care Providers. *Journal of Political Economy*, 111(3): 555–88.

Dufour, C., P. Lanoie and M. Patry, 1998. Regulation and Productivity. *Journal of ProductivityAnalysis*, Vol.9(3): 233-247.

Dye, R., 1986. Optimal Monitoring Policies in Agency, *RAND Journal of Economics*, 17(3), 339-350.

Ek, K., 2005. Public and private attitudes towards "green" electricity: the case of Swedish wind power. *Energy Policy*, 33, 1677–1689.

Elliott, K. A. and R. B. Freeman, 2001. White Hat or Don Quixotes? Human Rights Vigilantes in the Global Economy. *Working paper No. 8102*, National Bureau of Economic Research, Cambridge, MA.

Estes, E.A. & Smith, V. K., 1996. "Price, Quality, And Pesticide Related Health Risk Considerations In Fruit And Vegetable Purchases: An Hedonic Analysis Of Tucson, Arizona Supermarkets," Journal of Food Distribution Research, Food Distribution Research Society, vol. 27(3), October.

Fagart, M.-C., and Sinclair-Desgagné, B., 2007. Ranking Contingent Monitoring Systems, *Management Science*, 53(9), 1501-1509.

Färe R., S.Grosskopf, D. Noh, W. Weber, 2005. Characteristics of a polluting technology: theory and practice. *Journal of Econometrics*, Vol.126(2): 469-492.

Feddersen, T. J. and Gilligan, T. W., 2001. Saints and Markets: Activists and the Supply of Credence Goods. *Journal of Economics & Management Strategy*, 10: 149–171.

Fleckinger P.and M. Glachant, 2011. Negotiating a voluntary agreement when firms self regulate. *Journal of Environmental Economics and Management*, 62(1): 41-52.

Fleckinger, P., Glachant, M. and Moineville, G., 2012. Incentives for Quality in Friendly versus Hostile Informational Environments. *Working paper*

Frank, R., 1996. Can socially responsible firms survive in a competitive market? In David Messick and Ann Tenbrunsel (eds.), *Codes of Conduct: Behavioral Research into Business Ethics*, New York, NY: Russel Sage Foundation. p.214-227.

Frank, R., 2003. *What Price the Moral High Ground? Ethical Dilemmas in Competitive Environments*. Princeton, NJ: Princeton University Press.

Friedman, M., 1970. The social responsibility of business is to increase its profits. *New York Times Magazine*, September 13: 32–33.

Friedman, M., 1985. Consumer Boycotts in the United States, 1970–1980: Contemporary Events in Historical Perspective. *Journal of Consumer Affairs,* 19: 96–117.

Gamper-Rabindran, S., 2006. Did the EPA's Voluntary Industrial Toxics Program Reduce Emissions? A GIS Analysis of Distributional Impacts and by-media Analysis of Substitution. *Journal of Environmental Economics and Management*, 52: 391-410.

Gil, J. M., Gracia, A. and Sanchez Garcia, M., 2000. Market Segmentation and Willingness to pay for Organic Products in Spain. *International Food and Agribusiness Management Review*, 03(2).

Gill, D. and Sgroi, D.,2011. The Optimal Choice of Pre-Launch Reviewer. *Journal of Economic Theory*, Forthcoming.

Goett, A., Hudson, K., Train, K., 2000. Customers' choice among retail energy suppliers: the willingness-to-pay for service attributes. *The Energy Journal,* 21 (4), 1–28.

Gollop, F.M. and M.J. Roberts, 1983. Environmental Regulations and Productivity Growth: The Case of Fossil-fuelled Electric Power Generation. *Journal of Political Economy*, 91(4):654-674.

Gray W. B., 1987. The Cost of Regulation: OSHA, EPA and the Productivity Slowdown. *The American Economic Review,* Vol.77(5): 998-1006.

Gray W. B. and R. J. Shadbegian, 1993. Environmental Regulation and Manufacturing Productivity at the Plant Level. *NBER Working Papers 4321*, National Bureau of Economic Research, Inc.654-674.

Gray, W.B. and R.J. Shadbegian, 1998. Environmental regulation investment timing, and technology choice. *The Journal of Industrial Economics*, XLVI(2): 235-256.

Gray, W.B. and R.J. Shadbegian, 2003. Plant vintage, technology, and environmental regulation. *Journal of Environmental Economics and Management*, 46, 384-402.

Greening D.W. and Turban D.B., 2002. Corporate social performance as a competitive advantage in attracting quality workforce. *Bus. Society*, 39:254-280.

Grösche, P., and Schröder, C., 2011. Eliciting public support for greening the electricity mix using random parameter technique. *Energy Economics* 33, 363–370.

Guilloux, G., 2006. Les produits éco-conçus – Vers une consommation durable. *Oral presentation at the Biennale du design de St-étienne*.

Hamamoto, M., 2006. Environmental regulation and the productivity of Japanese manufacturing industries. *Resource and Energy Economics*, 28, 299-312.

Hamschmidt, J. and Dyllick, T., 2006. ISO 14 001 Profitable? Yes ! But is it eco-effective?. in Schaltegger, S. and Wagner, M. (eds.), *Managing the Business Case for Sheffield*: Greenleaf Publishing, 554–568.

Hansla, A., Gamble, A., Juliusson, A., Gärling, T., 2008. Psychological determinants of attitude towards and willingness to pay for green electricity. *Energy Policy*, 36: 768–774.

Harbaugh, R., Maxwell, J. and Roussillon, B. 2011. Label Confusion: The Groucho Effect of Uncertain Standards. *Management Science*, forthcoming.

Heinkel, R., Kraus, A. and Zechner, J., 2001. The Effect of Green Investment on Corporate Behaviour. *Journal of Financial and Quantitative Analysis*, 36: 431-450.

Henion K.E., 1972. The effect of ecologically relevant information of detergent sales. *J. Market. Res.* 9 (1): 10–14.

Heyes, A., 2005. A Signaling-motive for Self-regulation in the Shadow of Coercion. *Journal of Business and Economics*, 57, 238-246.

Hiscox, M. J. and N. F. B. Smyth, 2005. Is There Consumer Demand for Improved Labor Standards? Evidence from Field Experiments in Social Labeling. *Unpublished Manuscript*. Harvard University.

Howard P.H., and P. Allen, 2008. Consumer willingness to pay for domestic 'fair trade': Evidence from the United States. Renewable *Agriculture and Food Systems*, 23(3); 235–242.

Ibanez L. and G. Grolleau, 2008. Can Ecolabeling Schemes Preserve the Environment?. *Environmental & Resource Economics, European Association of Environmental and Resource Economists*, Vol.40(2): 233-249.

Jackson, M. O. and Yariv, L., 2007. Diffusion of Behavior and Equilibrium Properties in Network Games, *American Economic Review (Papers and Proceedings)*, 97(2), 92-98.

Jacob, B. A. 2005. Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools. *Journal of Public Economics*, 89(5–6):761–96.

Jaffe B.A, S. R. Peterson, P.R. Portney and R.N Stavins, 1995. Environmental Regulation and the Competitiveness of U.S. Manufacturing: What Does the Evidence Tell Us? *Journal of Economic Literature, American Economic Association*, Vol.33(1): 132-163.

Jin, G. Z., and P. Leslie, 2003. The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards. *Quarterly Journal of Economics*, 118(2):409–51.

Johnston, R.J., Wessells, C.R., Donath, H., Asche, F., 2001. Measuring consumer preferences for eco-labeled seafood: an international comparison. *Am. J. Agric. Econ.,* 26 (1), 20 – 39.

Kamenica, E. and M. Gentzkow, 2011. Bayesian Persuasion. *American Economic Review*, Vol. 101, 2590-2615.

Kessler, A., 2001. Revisiting the Lemons Market. *International Economic Review*, Vol. 42, 25-41.

Khanna, M. and L. Damon, 1999. EPA's Voluntary 33/50 Program: Impact on Toxic Releases and Economic Performance of Firms. *Journal of Environmental Economics and Management*, 38: 1-28.

King A, Lenox M, Terlaak A., 2005. The strategic use of decentralized institutions: exploring certification with the ISO 14001 management standard. *Academy of Management,* 48: 1091–1106.

King, B. G., and S.A. Soule, 2007. Social Movements as Extra-Institutional Entrepreneurs: The Effect of Protests on Stock Price Returns. *Administrative Science Quarterly,* 52: 413-442.

Kitzmueller, M., and J. Shimshack, 2012. Economic Perspectives on Corporate Social Responsibility. *Journal of Economic Literature*, 50(1): 51–84.

Koku P. S., 2012. On the effectiveness of consumer boycotts organized through the internet: the market model. *Journal of Services Marketing*, Vol. 26(1): 20 - 26.

Koku P. S., A. Akhigbe, T. M. Springer, 1997. The Financial Impact of Boycotts and Threats of Boycott. *Journal of Business Research*, Vol.40(1): 15-20.

Koppel, H. and R., Tobias, 2012. Corporate social responsibility in the work place: Experimental evidence on CSR from a gift-exchange game. *Jena economic research papers*, No. 2011,030.

Kotchen, M.J., and Moore, M.R., 2007. Private provision of environmental public goods: household participation in green-electricity programs. *Journal of Environment Economics and Management*, 53, 1–16.

Lanfranchi J. and S. Pekovic, 2011. How Green is my Firm? Workers' Attitudes towards job, Job Involvement and Effort in Environmentally-Related Certified Firms. *Working paper.*

Lanoie, P., Laurent-Lucchetti, L., Johnstone, N. and Ambec, S.., 2011. Environmental Policy, Innovation and Performance: New Insights on the Porter Hypothesis. *Journal of Economics & Management Strategy*, Vol.20(3): 803–842.

Lanoie, P., Patry, M., Lajeunesse, R., 2005. Environmental Regulation and Productivity: New findings on the Porter Hypothesis. *Mimeo*, HEC Montréal.

Lee M., 2007. Environmental regulations and market power: The case of the Korean manufacturing industries. *Ecological Economics*, Vol.68(1–2): 205-209,.,

Lenox, M.J., and Eesley, C. E., 2009. Private Environmental Activism and the Selection and Response of Firm Targets. *Journal of Economics & Management Strategy*, Vol.18(1): 45–73.

Levin, J., 2001. Information and the market for lemons, *RAND Journal of Economics*, Vol. 32(4), 657-666.

Levin, J., 2009. The Dynamics of Collective Reputation, *The B.E. Journal of Theoretical Economics*, Contributions, vol. 9, art. 27, http://www.bepress.com/bejte/vol9/iss1/art27.

Liang J, and H. H. Jensen, 2007. Imperfect food certication, opportunistic behaviors and detection. *Selected Paper 175174* at the Annual Meeting, July 29-August 1, 2007, Portland, Oregon TN 9714, American Agricultural Economics Association (New Name 2008: Agricultural and Applied Economics Association).

Litvine, D., and Wüstenhagen, R., 2011. Helping "light green" consumers walk the talk: results of a behavioural intervention survey in the Swiss electricity market. *Ecological Economics*, 70, 461–474.

Lizzeri, A., 1999. Information Revelation and Certification Intermediaries. *RAND Journal of Economics*, 30(2): 214–31.

Loureiro, M. L. 2003. Rethinking New Wines: Implications of Local and Environmentally Friendly Labels. *Food Policy*, 28 (5–6): 547–60.

Loureiro, M. L., and S. Hine, 2002. Discovering Niche Markets: A Comparison of Consumer Willingness to Pay for Local (Colorado Grown), Organic,and GMO-Free Products. *Journal of Agricultural and Applied Economics*, 34 (3): 477–87.

Loureiro, M. L., and J. Lotade, 2005. Do Fair Trade and Eco-Labels in Coffee Wake Up the Consumer Conscience? *Ecological Economics*, 53 (1): 129–38.

Lyon, T. (Ed.), 2010. *Good Cop/Bad Cop: Environmental NGOs and Their Strategies*. RFF Press, Washington, DC.

Lyon, T., and Maxwell, J., 2008. Corporate Social Responsibility and the Environment: A Theoretical Perspective. *Review of Environmental Economics and Policy*, 1, 1-22.

Lyon, T. and Maxwell, J., 2011. Greenwash: Corporate Environmental Disclosure under Threat of Audit, *Journal of Economics & Management Strategy*, Vol. 20(1), 3-41.

Lu, S. F., 2009. Multitasking, Information Disclosure and Product Quality: Evidence from Nursing Homes. *University of Rochester Simon Graduate School of Business Administration Working Paper* FR 09-03.

Managi S., J. J. Opaluch, D. Jin and T. A. Grigalunas, 2006. Stochastic frontier analysis of total factor productivity in the offshore oil and gas industry. *Ecological Economics*, Vol.60(1): 204-215.

Margolis, J., Elfenbein, H., and Walsh, A., 2009. Does it pay to be good? A meta-analysis and redirection of research on corporate social and financial performance. *Harvard University working paper*.

Maryland (University of.), Program on International Policy Attitudes: 2000, 'Americans on Globalization: A Study of Public Attitudes'.

Marymount University Center for Ethical Concerns, 1999, 'The Consumers and Sweatshops'.

Mason, C. F., 2009. Certification of Socially Responsible Behavior: Eco-Labels and Fair-Trade Coffee, J*ournal of Agricultural & Food Industrial Organization*, Berkeley Electronic Press, vol. 7(2), pages 2.

Mason, C. F., 2011. Eco-Labeling and Market Equilibria with Noisy Certification Tests. *Environmental & Resource Economics, European Association of Environmental and Resource Economists*, vol. 48(4): 537-560.

Maxwell, J. W., and C. Decker, 2006. Voluntary environmental investment and regulatory responsiveness. *Environmental and Resource Economics,* 33: 425–39.

Maxwell, J. W., Lyon, T. P. and Hackett, S. C., 2000. Self-Regulation and Social Welfare: The Political Economy of Corporate Environmentalism. *Journal of Law and Economics*, October 2000.

McWilliams, A., and D. Siegel, 2000. Corporate social responsibility and financial performance: Correlation or misspecification? *Strategic Management Journal*, 21 (5): 603-609.

MacLeod, W. B., 2007. Reputations, Relationships, and Contract Enforcement, *Journal of Economic Literature*, 45(3), 595-628.

Melnyk S.A., Sroufe R.P., Calantone R.L., Montabon F.L., 2002. Assessing the effectiveness of US voluntary environmental programmes: an empirical study. *International Journal of Production Research,* 40: 1853–1878.

Menges, R., Schroeder, C., and Traub, S., 2005. Altruism, warm glow and the willingness-to-donate for green electricity: an artefactual field experiment. *Environmental & Resource Economics,* 31, 431–458.

Milgrom, P., 1981. Good News and Bad News: Representation Theorems and Applications. *Bell Journal of Economics*, Vol. 12(2), 380-391.

Milgrom, P., and Roberts, J., 1986. Price and Advertising Signals of Product Quality, *Journal of Political Economy*, Vol. 94(4), 796-821.

Misra, S.K., C.L. Huang. and S.L. Ott, 1991. Consumer Willingness to Pay for Pesticide-Free Produce. *Western Journal of Agricultural Economics,* Vol. 16(2): 218-227.

Montgomery, D. B. and Ramus, C. A., 2003. Corporate Social Responsibility Reputation Effects on MBA Job Choice. *Stanford GSB Working Paper* No. 1805.

Morgenstern, R. W.A. Pizer, and J. S. Shih, 2007. Evaluating Voluntary U.S. Climate Programs: The Case of Climate Wise. in Morgernstern and Pizer (eds), *Reality Check: The*

*Nature and Performance of Voluntary Environmental Programs in the United States, Europe, and Japan.* Resources for the Future: Washington D.C.

Murty M.N. , and S. Kumar, 2003. Win–win opportunities and environmental regulation: testing of porter hypothesis for Indian manufacturing industries. *Journal of Environmental Management*, Vol.67(2): 139-144.

Mussa M, and Rosen S, 1978. Monopoly and product quality. *Journal of Economic Theory* 18(2):301-317.

Nimon W. and J. Beghin, 1999. Are Eco-Labels Valuable? Evidence From the Apparel Industry. *American Journal of Agricultural Economics*, Agricultural and Applied Economics Association, Vol. 81(4), pages 801-811.

Nomura, N., and Akai, M., 2004. Willingness to pay for green electricity in Japan as estimated through contingent valuation method. *Applied Energy,* 78, 453–463.

Nyborg, K. and Zhang, T., 2011. Is corporate social responsibility associated with lower wages?. *Memorandum 01/2011*, Oslo University, Department of Economics.

Orlitsky M., Schmidt F.L. and Rynes S.L., 2003. Corporate social and financial performance: a meta-analysis. *Organization Studies*, Vol. 24(3):. 403-441.

Parsons, B., 2005. Rentabilité comparée des fermes laitières biologiques du Nord-Est. *Mimeo*, Université du Vermont.

Phelps, E., 1972. The Statistical Theory of Racism and Sexism, *American Economic Review*, 62, 659-61.

Potoski M., and A. Prakash, 2005. Covenants with weak swords: ISO 14001 and facilities' environmental performance. *Journal of Policy Analysis and Management*, Vol.24(4): 745–769.

Powers, N., A. Blackman, T.P. Lyon, and U. Narain, 2008. Does Disclosure Reduce Pollution? Evidence from India's Green Rating Project. *Resources for the Future Discussion Paper* 08-38.

Prasad M., Kimeldorf, H., R. Meyer, and I. Robinson, 2004. Consumers of the World Unite: A Market-based Response to Sweatshops. *Unpublished paper*. Department of Sociology, University of Michigan.

Pruitt, S. W., and Friedman, M., 1986. Determining the Effectiveness of Consumer Boycotts: A Stock Price Analysis of Their Impact on the Corporate Targets. *Journal of Consumer Policy*, 9 (1986): 375–387.

Pruitt, S. W., Wei, K. C. John, and White, R. E., 1988. The Impact of Union-Sponsored Boycotts on the Stock Prices of Target Firms. *Journal of Labor Research*, IX(3): 285–289.

Ramus, C. and I. Montiel, 2005. When Are Corporate Environmental Policies a Form of Greenwashing?. *Business and Society*, 44: 377-414.

Rego A., S. Leal, M. Pina e Cunha and J. Faria, 2007. How Employees' Perceptions of Corporate Citizenship Predict their Organizational Commitment. *Working Papers in Management*, Universidad de Aveiro, Area Cientìfica de Gestao.

Roe B., M.F. Teisl, A.S. Levy, and M. Russell, 2001. US consumers' willingness to pay for green electricity. *Energy Policy*, 29(11): 917–925.

Roman R.M., Hayibor S. and Agle B.R., 1999. The Relationship Between Social and Financial Performance: Repainting a Portrait. *Business & Society*, Vol. 38(1): 109-125.

Ronnen U., 1991. Minimum Quality Standards, Fixed Costs, and Competition. *The RAND Journal of Economics*, Vol. 22(4): 490-504.

Roosen. J., J.A. Fox, D.A. Hennessy. and A. Schreiber, 1998. Consumer's Valuation of Insecticide Use Restrictions: An Application to Apples. *Journal of Agricultural and Resource Economics,* Vol. 23(2):. 367-384.

Russo, M. V., 2009. Explaining the impact of ISO 14001 on emission performance: a dynamic capabilities perspective on process and learning. *Business Strategy and the Environment,* Vol.18(5): 307–319.

Sam, A., M. Khanna and R. Innes, 2009. How do Voluntary Pollution Reduction Programs (VPRs) Work? An Empirical Study of Links between VPRs Environmental Management and Environmental Performance. *Land Economics,* 4(85): 692-711.

Sarath, B., 1996. Public Information Quality with Monopolistic Sellers, *Games and Economic Behavior*, Vol. 16, 261-279.

Shadbegian R., and W. Gray, 2005. Assessing Multi-Dimensional Performance: Environmental and Economic Outcomes. *Working Papers 05-03*, Center for Economic Studies, U.S. Census Bureau.

Shaked, A. and Sutton, J., 1982. Relaxing Price Competition through Product Differentiation. *Review of Economic Studies*, Wiley Blackwell, Vol.49(1): 3-13.

Sinclair-Desgagne, B., and E. Gozlan, 2003. A theory of environmental risk disclosure. *Journal of Environmental Economics and Management*, 45:377–93.

Spiegler, R.,2011. *Bounded Rationality and Industrial Organization.* Oxford University Press.

Szymanski, M., and P. Tiwari, 2004. ISO 14001 and the reduction of toxic emissions. *Pol. Reform*, 7(1): 31–42.

Teisl, M.F., B. Roe and R.L. Hicks, 2002. Can eco-labels tune a market? Evidence from dolphin-safe labeling. *Journal of Environmental Economics and Management*, 43: 339-359.

TerraChoice, 2010. *The Seven Sins of Greenwashing*. Available at http://sinsofgreenwashing.org/findings/greenwashing-report-2010/

Tirole, J., 1996. A Theory of Collective Reputation (with applications to the persistence of corruption and to firm quality). *Review of Economic Studies*, 63, 1-22.

Turban, D. B. and D. W. Greening: 1997. Corporate Social Performance and Organizational Attractiveness to Prospective Employees. *Academyof Management Journal,* 40: 658–672.

van der Vlist, A. J. and W., Cees and F., Henk, 2007. Technical efficiency under alternative environmental regulatory regimes: The case of Dutch horticulture. *Ecological Economics*, Elsevier, Vol. 63(1): 165-173.

van Ravenswaay, E. O. and Blend, J. R., 1999. Consumer Demand For Ecolabeled Apples: Results From Econometric Estimation. *Staff Papers 11673*, Michigan State University, Department of Agricultural, Food, and Resource Economics.

Vasi I. B., and B. G. King, 2012. Social Movements, Risk Perceptions, and Economic Outcomes: The Effect of Primary and Secondary Stakeholder Activism on Firms' Perceived Environmental Risk and Financial Performance. *American Sociological Review,* Vol.77(4).

Vidovic, M. and N. Khanna, 2007. Can Voluntary Pollution Prevention Programs Fulfill Their Promises? Further Evidence from the EPA's 33/50 Program. *Journal of Environmental Economics and Management,* 53: 180-195.

Ward D. O., C. D. Clark, K. L. Jensen, S. T. Yen, and C. S. Russell, 2011. Factors influencing willingness-to-pay for the ENERGY STAR©label. *Energy Policy*, Vol.39(3): 1450-1458.

Welch, E., A. Mazur and S. Bretschneider, 2000. Voluntary Behavior by Electric Utilities:

Levels of Adoption and Contribution of the Climate Challenge Program to the Reduction of Carbon Dioxide. *Journal of Policy Analysis and Management*, 19(3): 407-425.

Werner, R. M., and D. A. Asch, 2005. The Unintended Consequences of Publicly Reporting Quality Information. *Journal of the American Medical Association*, 293(10):1239–44.

Wessells, C.R., Johnston,R.J., and Donath, H., 1999. Assessing consumer preferences for eco-labeled seafood: the influence of species, certifier and household attributes. *Am. J. Agric. Econ.* 81, 1084 – 1089.

Wu M.L., 2006. Corporate social performance, corporate financial performance, and firm size: a meta-analysis. *Journal of American Academy of Business*, 8(1): 163–171.

Yaziji, M., and Doh, J., 2009. *NGOs and Corporations: Conflict and Collaboration*. Cambridge, UK; Cambridge University Press

Yoo, S.-H., and Kwak, S.-Y., 2009. Willingness to pay for green electricity in Korea: a contingent valuation study. *Energy Policy*, 37, 5408–5416.

Zarnikau, J., 2003. Consumer demand for 'green power' and energy efficiency. *Energy Policy*, 31, 1661–1672.

# L'Economie de la Responsabilité Sociale et Environnementale de l'Entreprise: Le Rôle Informationnel des Tierces Parties

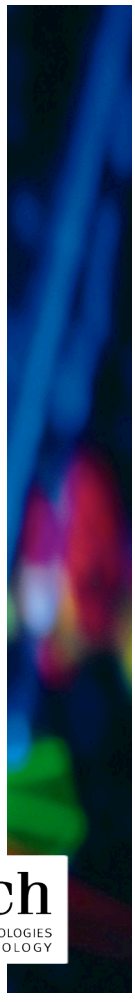
**RESUME :** Cette thèse étudie le rôle des tierces parties dans les problématiques informationnelles autour de la responsabilité sociale et environnementale des entreprises (RSE). Les entreprises revendiquent de plus en plus le fait qu'elles adoptent des pratiques responsables. Cela peut être lié au fait qu'il existe des parties prenantes (e.g. consommateurs, investisseurs) qui sont prêtes à récompenser les comportements responsables. Cependant, la véracité des revendications des entreprises est souvent impossible ou au moins difficile à certifier, ce qui peut inciter les entreprises à manipuler leur communication. Cette dissertation a pour objectif d'analyser la façon dont de tierces parties peuvent tempérer cette asymétrie d'information. Des organisations comme des agences de notations, des organisations non-gouvernementales (ONGs) ou des organismes certificateurs, peuvent examiner les actions réelles des entreprises et transmettre cette information aux parties prenantes. Ces tierces parties ont un comportement informationnel hétérogène: certaines fournissent des informations plus particulièrement sur les entreprises qui se comportent de manière irresponsable, alors que d'autres se spécialisent dans la mise en lumière des entreprises responsables. Nous développons un modèle théorique simple qui explore l'offre de qualité dans des environnements informationnels amicaux, où la qualité est révélée plus souvent lorsqu'elle est haute que lorsqu'elle est basse, et dans des environnements hostiles, où c'est l'inverse. Nous appliquons ensuite ce modèle à deux types de tierces parties: les ONGs et les labels. Nous étudions le choix de production de bonnes ou mauvaises nouvelles afin de comprendre quand et pourquoi les ONGs choisissent d'être hostiles ou amicales. Ensuite, nous cherchons à évaluer quel est le niveau de sévérité des labels qui entraîne le plus d'investissement et le bien être social le plus important.

**Mots clés :** Responsabilité Sociale et Environnementale de l'Entreprise, asymétrie d'information, révélation de la qualité, théorie des jeux, ONGs, labels

# The Economics of Corporate Social Responsibility: The Informational Role of Third Parties

**ABSTRACT :** This thesis studies the role of third parties in tempering informational issues at stake with regard to corporate social responsibility (CSR). Firms growingly claim to be adopting responsible practices, and it may be so because some socially conscious stakeholders (e.g. consumers, investors) are ready to reward such behaviors. Nevertheless, truthfulness of firms' virtue claims is often impossible or at least difficult to ascertain, which creates an incentive for firms to manipulate their communication. This dissertation intends to analyse how third parties can temper this asymmetry of information. Organisms such as rating agencies, non-governmental organisations, labeling institutions, can investigate firms' real actions and convey this information to stakeholders. These third parties have heterogeneous informational behaviors: some of them mostly disclose information on firms that do not behave responsibly while others are specialized in revealing firms that are socially or environmentally responsible. We develop a simple theoretical model that explores the provision of quality by a firm under friendly informational environments, in which quality is more often disclosed when it is high than when it is low, and hostile environments, in which the converse holds. We apply this base model to two types of third party: NGOs and labels. We endogenize the production of good and bad news in order to understand when and why NGOs choose between being hostile or friendly. Then, we seek to assess what is the more efficient level of label's strictness to spur social welfare and investment in high quality.

**Keywords** : Corporate Social Responsibility, asymmetric information, quality disclosure, game theory, NGOs, labels



MINES ParisTech

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY