Effect of Consumer Awareness on Corporate Social Responsibility Under Asymmetric Information

Xiaomeng Guo[1], Guang Xiao[2], and Fuqiang Zhang[3]

[1] *Faculty of Business, Hong Kong Polytechnic University, xiaomeng.guo@polyu.edu.hk*

[2] *Faculty of Business, Hong Kong Polytechnic University, guang.xiao@polyu.edu.hk*

[3] *Olin Business School, Washington University in St. Louis, fzhang22@wustl.edu*

Abstract

This paper studies the interaction between a firm and consumers under the consideration of corporate social responsibility. The firm can be either socially responsible or socially irresponsible; however, the consumers cannot observe the firm's exact type, which is private information. The firm can try to signal its type through pricing and other information sharing mechanisms (e.g., issue sustainability reports and obtain third-party certifications). We find that due to the existence of asymmetric information, increasing consumer awareness of corporate social responsibility does not necessarily help promote responsible corporate behaviors. Specifically, when a larger fraction of consumers become socially concerned or when the consumers have stronger willingness to reward (punish) the responsible (irresponsible) firm, the responsible firm could be worse off whereas the irresponsible firm could be better off. This is because the seemingly attractive trend in consumer behavior will affect the responsible firm's signaling cost as well as its equilibrium strategy (separating vs. pooling). In addition, we find that improving the signaling accuracy will always benefit the responsible firm but may or may not hurt the irresponsible firm. Our results suggest that addressing the information asymmetry issue is the key in promoting corporate social responsibility. In particular, concerned parties should first exert efforts to create transparency in firms' sustainability practices before making investments to educate consumers and influence their purchasing behaviors.

Key words: Corporate social responsibility, consumer behavior, information asymmetry, signaling

1 Introduction

Corporate social responsibility (CSR) has become an increasingly important and challenging issue faced by companies in today's business world. According to *Financial Times*, corporate social responsibility is "a business approach that contributes to sustainable development by delivering economic, social and environmental benefits for all stakeholders". Recognizing its vital role for long-term success, many firms have attempted to integrate CSR into their business strategies. Examples include Starbucks' commitment to the environment through the "Shared Planet" program (Starbucks, 2006), and Patagonia's initiative to ensure that products are produced under safe, fair, and humane working conditions throughout its supply chain (Patagonia, 2014). At the same time, however, scandals caused by irresponsible and incompetent corporate behaviors abound in recent media. For instance, Greenpeace (2011) profiles the serious problem of toxic water pollution that results from the release of hazardous chemicals by the textile industry in China. The report also identifies the global brands that source from the polluting Chinese textile suppliers, including Abercrombie & Fitch, Calvin Klein, Converse, H&M, etc. These companies have implemented neither comprehensive chemical management policies nor clear measures to restrict the release of hazardous substances into water beyond local regulations. In 2013, the clothing factory building collapse near Dhaka, Bangladesh, killed as many as 800 workers, mainly due to the lack of safety requirements from brand-name retailers such as Walmart, which sources from these garment factories (Husock, 2013). This negligence of corporate responsibilities has not only led to serious social and environmental consequences, but has also drawn wide public attention and generated discussion about firms' roles in society.

There has been growing consumer awareness of corporate social responsibility across the globe. In 2015, Cone Communications and Ebiquity jointly conducted a survey of 9,709 consumers in nine of the largest countries in the world by GDP[1] to learn about their perceptions and behaviors related to CSR. The survey indicates that the majority of consumers have strong accountability to address social and environmental issues and are primed for participation in CSR efforts with the understanding that corporate should do more than just making a profit (CONE-Communication/Ebiquity, 2015). Specifically, the survey finds that about 31% of the participants would reward a company for operating responsibly (e.g., pay a higher price), 19% would punish a company with irresponsible behavior (e.g., boycott the product), 40% would equally reward or punish a company based on how it operates, and the rest of 10% would not care about CSR in making their decisions. The consumers' CSR awareness and strong intentions regarding reward/punishment provide firms with incentives and opportunities to behave responsibly and collaborate with consumers to push the boundaries of corporate responsibility for a better society. For example, after launching the "Shared Planet" program, Starbucks began using multi-media channels (e.g., online websites and sustainability reports) to publicize the details of the program and educate consumers on the importance of CSR to the whole community (Starbucks, 2006). As estimated by De Pelsmacker et al. (2005), 10% of consumers are

[1] The nine countries are United States, Canada, Brazil, United Kingdom, Germany, France, China, India, and Japan.

willing to pay a small premium for fair traded coffee. Such multi-media advertising and education could help Starbucks to better serve the consumer segment with CSR awareness.

Even though consumers' growing concern about CSR would incentivize firms to behave responsibly, there are potential barriers that prevent consumers from observing whether a firm is truly responsible. First, CSR is a broad concept and covers many aspects of a firm's business activities; as a result, consumers may have limited information regarding a firm's responsibility type. Second, even though many firms voluntarily disclose their sustainability activities, there is great variation in their reports due to the lack of consistent reporting standards (see Boersma, 2013, for an example of disclosure discrepancies in Australia). Finally, since consumers are unable to verify firms' practices, they may not fully believe the information disclosed by firms (Ghosh and Galbreth, 2013). The survey above shows that 48% of participants view companies with blind optimism, i.e., assuming they operate as responsibly as possible, unless they hear otherwise; whereas, 52% assume a company is not acting responsibly until they see or hear evidence to prove otherwise (CONE-Communication/Ebiquity, 2015). Consequently, it is possible that consumers may treat an irresponsible firm as responsible and vice versa. Such information asymmetry may distort consumers' willingness to reward or punish firm activities, and thus lead to ambiguous implications of consumer responsibility awareness for firms.

Many firms have realized the importance of creating transparency about their social responsibility practices (Kraft et al., 2016). There are a variety of ways for a firm to communicate its CSR efforts with consumers. A firm may try to apply for certifications to manifest its CSR standards, such as ISO 26000, SA8000, or Fairtrade certification (see Chen and Lee, 2016). A firm may also use public reports (e.g., the Global Reporting Initiative) and media coverage to advertise its sustainability activities. For example, as indicated by Chen and Slotnick (2015), following California law's requirement regarding supply chain disclosure on slavery, Safeway issues public reports to disclose its supplier requirements, supplier inspections, etc. In 2005, after the scandal of using child labor, Nike was the first company in its industry to create transparency by disclosing information about its contract factories, including detailed pay scales, and working conditions in those factories (Newell, 2015).

The government may also play a significant role in promoting transparency of firms' CSR practices. In fact, around the world governments are introducing laws that require, encourage, and support sustainability reporting (UNEP, 2013). According to a study by Hauser Center for Nonprofit Organizations (Lydenberg, 2012), at least twenty countries across six continents require or strongly encourage sustainability reports or similar disclosure, and more than a dozen countries mandate varying levels of corporate sustainability disclosure. More countries are at the crossroads of adopting stricter regulations and developing new solutions to improve sustainability reporting.

Motivated by the above observations, in this paper we study the interaction between a firm and consumers under the presence of social responsibility consideration and asymmetric information. The purpose is to obtain a better understanding of the CSR issue and provide useful insights for both the firm and the government. Specifically, we focus on the following research questions. First, how does consumer awareness affect the firm's pricing decision and profit performance? Over time the number of socially concerned consumers has been growing across the globe. We would like to examine the implication of this trend for different types of firms. Second, is it always desirable to induce consumers to reward socially responsible firms while punishing irresponsible firms? Intuitively, more pressure from the consumer side (e.g., higher reward/punishment levels for responsible/irresponsible firms) would encourage firms to pursue socially responsible practices. Third, what is the impact of increasing the signaling accuracy? As discussed above, the firm may use various mechanisms to signal its responsibility type to consumers. The answer to this question may help the firm understand the value of creating transparency in sustainability practices; it may also provide guidance for the government when making policies to improve sustainability reporting.

To address the above questions, we develop a signaling game framework, in which a monopoly firm sells a product with exogenous quality to the market. The firm has private information about its CSR type, i.e., either responsible or irresponsible. A fraction of consumers have social responsibility awareness and would reward (punish) the responsible (irresponsible) firm. Although the firm's CSR type is not observable to consumers, the firm may signal its type through pricing and other communication mechanisms. Due to the previously discussed reasons, the firm's signal is imperfect and may suffer from the type I and type II errors. Take the certification signaling mechanism for example: A type I error means that a socially responsible firm fails to obtain the certification while a type II error means that a socially irresponsible firm obtains the certification. After observing the firm's price and signal, consumers make purchasing decisions based on their updated belief about the firm's type.

We use the case with symmetric information as a benchmark (i.e., the consumers can observe the firm's type). We find that under symmetric information, increasing the fraction of socially responsible consumers or enhancing their reward/punishment levels always benefits the responsible firm and hurts the irresponsible firm, which is quite intuitive. Therefore, in an ideal world without information asymmetry, to promote responsible corporate behaviors, both the government and responsible firms should exert effort to educate consumers so a larger portion of the population is aware of CSR and is willing to increase financial reward/punishment levels.

However, the introduction of asymmetric information complicates the problem because socially oriented consumers can no longer observe the firm's type and apply their punishment/reward accordingly. In this case, the irresponsible firm has incentives to mimic the responsible firm whereas the responsible firm may need to signal its true type at a cost. We highlight the main findings from the asymmetric information case as follows. First, the firm's profit is not monotonic in the fraction of socially responsible consumers, regardless of the firm's type. In particular, increasing the fraction of socially responsible consumers may benefit the irresponsible firm but hurt the responsible firm under asymmetric information. This unexpected result can be explained as follows. On one hand, as the fraction of socially responsible consumers increases, the irresponsible firm will have a stronger incentive to mimic the responsible firm; as a result, the responsible firm may need to incur a higher signaling cost to separate itself from the irresponsible firm. On the other hand, if the two types of firm pool together, the consumers may punish the pooling firm as they are unable to distinguish between the two types. In addition, the firm's equilibrium pricing strategy may change according to the fraction of responsible consumers as well (it may switch between a separating equilibrium and a pooling equilibrium).

Second, for the socially responsible firm, its profit always increases in the consumers' reward level; but interestingly, its profit always decreases in the consumers' punishment level. In contrast, the profit of an irresponsible firm is not monotonic in the reward/punishment levels. In particular, a more severe punishment may actually improve the profit of the socially irresponsible firm, which is counterintuitive. Again these results are driven by the signaling cost and the fact that the firm's equilibrium pricing strategy may change as the reward/punishment levels vary. The above two results indicate that without transparent information regarding firms' type, one should be careful about promoting the awareness of CSR and inducing higher reward/punishment levels from consumers, as they may hurt the socially responsible firm while benefiting the socially irresponsible firm. Due to the existence of information asymmetry, haphazard efforts to promote CSR by influencing consumer behavior may lead to the opposite of the desired outcome.

Third, a higher signaling accuracy always benefits the socially responsible firm, and it may either hurt or benefit the irresponsible firm. Intuitively, a socially irresponsible firm wants to hide its true type and thus should be worse off under a higher signaling accuracy. Contrary to the intuition, we find that the socially irresponsible firm sometimes prefers a higher signaling accuracy for the following reason. When the type II error probability (i.e., the irresponsible firm obtains the certification) is high, the irresponsible firm may benefit from accurate signaling because once it passes the certification, it would be more difficult for consumers to discern its true type. Note such a result hinges upon the existence of the type II error. It suggests that if the government wants to mandate information disclosure from firms (e.g., require firms to file sustainability reports or go through the certification process), it should carefully design the disclosure process to ensure accurate signaling; particularly, minimizing the type II error would make the information disclosure policy more effective because it reduces the chance of benefiting the irresponsible firms.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the model. We characterize the equilibrium outcomes for the symmetric and asymmetric information cases in Sections 4 and 5, respectively. Then we conduct sensitivity analysis to investigate the impact of key model parameters in Section 6. Section 7 concludes the paper. All proofs and additional discussions are presented in Appendices A-C.

2 Literature Review

There has been an emerging literature on managing social responsibility in recent years (see Lee and Tang, 2017, for a discussion of more opportunities in this field). This research stream mainly focuses on designing and evaluating firms' strategies to induce socially responsible behaviors in value chains. Such strategies include, but are not limited to, sourcing, process improvement, auditing, etc. Agrawal and Lee (2015) study the effect of sourcing policies on the adoption of a sustainable production process in a bilateral supply chain. Huang et al. (2015) consider a three-tier supply chain and investigate the downstream manufacturer's control vs. delegation decisions on the upstream supplier's social responsibility improvement. Cho et al. (2016) analyze various mechanisms to prevent the use of child labor. Plambeck and Taylor (2016) study the mechanisms to induce suppliers to comply with social responsibility standards. Chen and Lee (2016) design incentive contracts to motivate suppliers to behave responsibly. Chen et al. (2016) investigate the supplier-certification collusion issue. Iyer and Singh (2017) discuss the incentives for firms to seek voluntary product safety certifications. Kim (2015) studies a firm that randomly experiences violations and designs the inspection policies to induce self-reporting. From both the firm's and the NGO's points of views, Kraft et al. (2013a,b) discuss how to provide incentives for competing firms to remove potentially toxic substances from their products. All of the above papers center around designing operational mechanisms/incentives to prevent irresponsible behaviors under the potential presence of information asymmetry between the value chain members. Our work differs from the above stream in that we consider a monopoly firm selling to consumers with different awareness levels of social responsibility. Under information asymmetry with respect to the firm's social responsibility type, we adopt the signaling game methodology to investigate the impact of consumer responsibility concerns on the firm's pricing decision and the associated profit implications. Our results suggest that, contrary to the conventional wisdom, increasing consumers' social responsibility awareness (via advertising or education) may backfire, which hurts the responsible firm but benefits the irresponsible firm.

Guo et al. (2016) investigate the impact of consumers' social responsibility concern on a firm's sourcing and pricing decisions. There are two suppliers, one responsible and one irresponsible. It has been found that more socially responsible consumers may drive the firm to source more from the irresponsible supplier. Their results are driven by the cost and reliability tradeoffs. Different from Guo et al. (2016) where the firm's sourcing decision is endogenous and is public information, we consider the situation when the firm's social responsibility type is exogenously given but not observable to consumers. We show that due to information asymmetry, more socially responsible consumers may hurt the socially responsible firm and benefit the socially irresponsible firm; also, as consumers' punishment level for irresponsibility increases, the socially responsible firm may be worse off and the socially irresponsible firm may be better off. Our work complements Guo et al. (2016) by providing a comprehensive discussion on how information asymmetry alters the conventional wisdom regarding CSR.

There is an extensive literature that studies the impact of CSR activities on firm profitability and market value. Mixed results have been found on the relation between firm value and CSR activities (see Berman et al., 1999, for a review of the literature). Thus it is not clear whether the costs of pursuing CSR can always outweigh the benefits. There are papers that look deeper into when and how CSR activities help improve firm profitability. For example, Sen and Bhattacharya (2001) identify important company-specific and individual-specific factors that affect how CSR actions influence consumer reactions. Du et al. (2010) discuss strategies on how to more effectively communicate firms' CSR practices to stakeholders, including consumers. Servaes and Tamayo (2013) show that high advertising intensity helps improve the impact of CSR activities on firm value, because advertising increases consumer awareness of the firm, including its CSR actions. The majority of these studies adopt the empirical approach. In fact, there is a lack of theoretical studies that help understand the mechanism underlying how consumer awareness of CSR affects firm profitability, especially with the presence of asymmetric information. Our paper helps to bridge this gap in the literature.

Finally, our paper adopts the signaling game framework, which has many applications to various operations and marketing settings. Lai et al. (2011) study the channel-stuffing phenomenon and show that firms might strategically inflate reported sales by stuffing the downstream channels with inauthentic orders when there is asymmetric information and short-term interest in the stock price. Lai et al. (2012) and Lai and Xiao (2016) further study the signaling of demand uncertainty to investors within supply chain and single firm settings, respectively. Li et al. (2014) examine the supplier encroachment issue under asymmetric information and show that with the ability to encroach, the supplier may prefer to sell to either a better informed or a totally uninformed reseller in different scenarios. In the marketing literature, Moorthy and Srinivasan (1995) analyze a firm's strategy to signal quality via money-back guarantees. Desai and Srinivasan (1995) study a franchiser's effort to signal demand through two- or three- part tariffs. Desai (2000) investigates a manufacturer's demand signaling to downstream retailers through multiple messages. Jiang et al. (2014) analyze the pricing strategies and the market outcome in service markets where the provider has private information about its own type and consumer's condition. Guo and Jiang (2016) study a firm's cost signaling strategy using price and quality to consumers with fairness concerns. Jiang et al. (2016) focus on demand signaling within a distribution channel under wholesale price contract and investigate the impact of forecast accuracy on firms' demand information sharing. Jiang and Yang (2017) analyze a two dimensional signaling game with experience goods and study the impact of consumer information sharing on a firm's pricing and quality decisions. Our work differs from the above literature by considering the signaling of firm responsibility type to consumers; the purpose is to understand the effect of consumers' CSR awareness under information asymmetry.

3 Model

We consider a single firm who sells a product to consumers. There are two possible types for the firm: either socially responsible (denoted as R) or not responsible (denoted as N). The R-type firm conducts business in a responsible way, e.g., it respects environmental standards, sources from sustainable suppliers, and takes care of the social issues in its community. In contrast, the N-type firm does not behave in a responsible way and may break social or environmental standards. The R-type firm incurs a unit production cost c_R while the N-type firm incurs a unit production cost c_N. We assume $0 \leq c_N < c_R$, i.e., the N-type firm enjoys a cost advantage. Nature decides the firm's type, which is private information and not observable to consumers. The firm is R-type with probability γ and N-type with probability $1-\gamma$, where $\gamma \in [0,1]$. Since this research focuses on corporate social responsibility, we assume that the product quality q is exogenously given and independent of the firm's type. The firm can set the retail price p.

In reality, a firm may exert efforts to communicate its type to consumers. For example, the firm may apply for a certification from a reputable third-party agent (e.g., the International Organization for Standardization, or ISO), utilize media channels to advertise its activities, or issue corporate sustainability reports to the public. For expositional convenience, we will focus on the example where the firm may apply for an ISO certification; however, the model setting and results also apply to general situations where the firm can use any of the above means to signal its type to consumers. In the base model, we assume there is no cost for the certification; the case with a positive certification cost will be discussed in Section 6.3. The signal can potentially reveal the firm's type, but the outcome of the signal is not always perfectly accurate. This could be due to random factors in the certifying process (e.g., it may require complex statistical tests). As a result, two types of errors may occur: The R-type firm may not pass the certification process (type I error), while the N-type firm may pass the process (type II error).

Specifically, the signal S has two possible values: $S = 1$ and $S = 0$, corresponding to the case that the firm successfully obtains the certification and the case that the firm fails to do so, respectively. Let ρ_R, $\rho_N \in [0, 1]$ measure the accuracy of the signal, then the signaling mechanism can be constructed as follows. Assume the R-type firm can successfully obtain the certification with probability $Pr(S = 1|R) = \gamma + \rho_R(1-\gamma)$ and may fail with probability $Pr(S = 0|R) = (1-\rho_R)(1-\gamma)$; and the N-type firm may obtain the certification with probability $Pr(S = 1|N) = (1 - \rho_N)\gamma$ and may fail with probability $Pr(S = 0|N) = (1 - \gamma) + \rho_N\gamma$.[2] Note that $Pr(S = 1|R) \geq Pr(S = 1|N)$ always holds, which indicates that the R-type firm is always more likely to obtain the certification than the N-type firm. One can easily derive $Pr(R|S = 1) = \frac{\gamma+\rho_R(1-\gamma)}{1+(\rho_R-\rho_N)(1-\gamma)}$, $Pr(R|S = 0) = \frac{\gamma(1-\rho_R)}{1+\gamma(\rho_N-\rho_R)}$, $Pr(N|S = 1) = \frac{(1-\gamma)(1-\rho_N)}{1+(\rho_R-\rho_N)(1-\gamma)}$, and $Pr(N|S = 0) = \frac{1-\gamma(1-\rho_N)}{1+\gamma(\rho_N-\rho_R)}$. Let $\gamma_S := Pr(R|S)$ denote the updated probability of the firm being R-type for a given signal $S \in \{0, 1\}$. When $S = 1$, $\gamma_1 := Pr(R|S = 1)$; when $S = 0$, $\gamma_0 := Pr(R|S = 0)$.

As ρ_R increases, both $Pr(S = 1|R)$ and γ_1 increases, which means the R-type firm is more likely to obtain the certification (i.e., the type I error is reduced) and a firm who has the certification is more likely to be R-type. As ρ_N increases, both $Pr(S = 1|N)$ and γ_0 decrease, which means the N-type firm is less likely to obtain the certification (i.e., the type II error is reduced) and a firm who does not have the certification is more likely to be N-type. Therefore, the accuracy of the certification signal increases in both ρ_R and ρ_N. Only when $\rho_R = \rho_N = 1$, the certification process is perfectly accurate and both type I and type II errors no longer exist, i.e., the certification signal can perfectly reveal the firm's type ($Pr(R|S = 1) = 1$ and $Pr(N|S = 0) = 1$). When $\rho_R = \rho_N = 0$, the certification is completely useless and uninformative, i.e., the signal S has no effect on the prior probability of the firm's type ($Pr(R|S = 1) = Pr(R|S = 0) = \gamma$ and $Pr(N|S = 1) = Pr(N|S = 0) = 1 - \gamma$).

[2] An alternative way to model the certification errors is directly assuming that $Pr(S = 1|R) = \rho_R$ and $Pr(S = 1|N) = 1 - \rho_N$, where $\rho_R, \rho_N \in [0, 1]$ and $\rho_R \geq 1 - \rho_N$, which is a generalization of the certification model used in Iyer and Singh (2017). All of our results continue to hold under this alternative assumption.

Notice that since sending the signal is costless, each type of firm will find it optimal to send the signal. To see this, consider the following two possible situations: First, the consumers do not know whether the firm chooses to send the signal or not, and not sending the signal is equivalent to $S = 0$ from their perspective (i.e., the firm will not obtain the certification if it does not apply). Second, the consumers can observe whether the firm tries to send the signal or not and let $S = \emptyset$ denote the case that the firm does not send the signal. In the former situation, it is easy to verify that both types of firm will find it optimal to send the signal. In the later situation, one can show that the R-type firm always finds it optimal to send the signal and, thus, a firm who does not send the signal ($S = \emptyset$) must be N-type (i.e., $Pr(R|S = \emptyset) = 0$ and $Pr(N|S = \emptyset) = 1$). Given this, it is trivial to see that the N-type firm also prefers to send the signal. Therefore, in the following discussion, we omit the case where the firm does not send the signal, which is always off the equilibrium.

The total number of consumers is normalized to one. Each consumer's valuation for the product is θq, where θ represents the consumer's willingness-to-pay for quality. The consumers are heterogeneous in θ and we assume that θ is uniformly distributed on $[0, 1]$. There are two types of consumers: A fraction τ of consumers care about social responsibility (denoted as R) and others do not (denoted as N). An N-type consumer's utility of purchasing the product is given by $u_N = \theta q - p$, regardless of the firm's type. A R-type consumer's utility of purchasing from the R-type firm is $u_{RR} = \theta q - p + \delta_R$ and that from the N-type firm is $u_{RN} = \theta q - p - \delta_N$, where $\delta_R \geq 0$ represents the socially responsible consumers' reward for social responsibility and $\delta_N \geq 0$ represents the punishment for irresponsibility. The consumers do not know the firm's true type, but they can observe the signal $S \in \{0, 1\}$ and price p. Thus an N-type consumer's expected utility is the same as u_N; an R-type consumer's expected utility can be written as

$$u_R(\mu(R|p, S)) = \mu(R|p, S)u_{RR} + (1 - \mu(R|p, S))u_{RN} = \theta q - p + (\mu(R|p, S)\delta_R - (1 - \mu(R|p, S))\delta_N), \quad (1)$$

where $\mu(R|p, S) \in [0, 1]$ is the consumer's updated belief about the probability of the firm being R-type.

The sequence of the events is as follows. First, the firm's type $j \in \{R, N\}$ is determined by a random draw. Then the j-type firm chooses whether to send a signal S by applying for certification.[3] The realized signal will be $S = 1$ with $Pr(S = 1|j)$ and $S = 0$ with $Pr(S = 0|j)$. After the signal S is revealed, the firm sets the retail price p. After observing the signal S and the price p, each consumer updates her belief about the firm's type and decides to buy if and only if her expected utility is non-negative. Table 1 summarizes the notations used in this paper. We make two additional

[3] As mentioned previously, since there is no cost for sending the signal, both types of firm will choose to send.

assumptions for ease of exposition: *Assumption 1.* $0 \leq \delta_R \leq \min\{q - c_R, (q + c_N)/2\}$, and *Assumption 2.* $0 \leq \delta_N \leq q$. These assumptions imply that the R-type consumers' reward and punishment levels are not too high, which helps rule out some trivial cases in analysis. They are not restrictive in practice. For example, De Pelsmacker et al. (2005) estimate that 10% of consumers are willing to pay a small premium for fair traded coffee. Also the qualitative results in this paper will hold even if we relax these assumptions.

4 Symmetric Information

We first analyze the benchmark with symmetric information, i.e., the firm's type is observable to consumers. In this case, the signal S is irrelevant. We use "ˆ" over a variable to indicate the symmetric information case. For any given p, each type of firm's profit is as below:

$$\hat{\pi}_N(p) = \begin{cases} (p - c_N)(\tau(1 - \frac{p+\delta_N}{q}) + (1-\tau)(1 - \frac{p}{q})), & \text{if } p \leq q - \delta_N \\ (p - c_N)(1-\tau)(1 - \frac{p}{q}), & \text{if } q - \delta_N < p \leq q \\ 0, & \text{if } p > q \end{cases}, \quad \text{and} \tag{2a}$$

$$\hat{\pi}_R(p) = \begin{cases} (p - c_R)(\tau + (1-\tau)(1 - \frac{p}{q})), & \text{if } p \leq \delta_R \\ (p - c_R)(\tau(1 - \frac{p-\delta_R}{q}) + (1-\tau)(1 - \frac{p}{q})), & \text{if } \delta_R < p \leq q \\ (p - c_R)\tau(1 - \frac{p-\delta_R}{q}), & \text{if } q < p \leq q + \delta_R \\ 0, & \text{if } p > q + \delta_R \end{cases}. \tag{2b}$$

Each type of firm chooses the price p to maximize its profit. Proposition 1 characterizes the j-type ($j \in \{R, N\}$) firm's optimal price decision and the corresponding profit under symmetric information.

Proposition 1. *Consider the symmetric information case where the firm type is observable to consumers.*

(a) The N-type firm's optimal price decision and the corresponding profit are given by:

$$\hat{p}^*_N = \begin{cases} \frac{q + c_N - \delta_N \tau}{2}, & \text{if } \delta_N < \frac{(q - c_N)(1 - \sqrt{1-\tau})}{\tau} \\ \frac{q + c_N}{2}, & \text{otherwise} \end{cases}, \quad \text{and} \tag{3a}$$

$$\hat{\pi}^*_N = \begin{cases} \frac{(q - c_N - \delta_N \tau)^2}{4q}, & \text{if } \delta_N < \frac{(q - c_N)(1 - \sqrt{1-\tau})}{\tau} \\ (1-\tau)\frac{(q - c_N)^2}{4q}, & \text{otherwise} \end{cases}; \tag{3b}$$

Table 1: Summary of Notations

Notation	Description
q	Quality of product
p	Price of product
θ	Consumers' willingness-to-pay for quality, follows $U[0,1]$
γ	Prior probability of the firm being socially responsible
τ	Fraction of consumers concerned about social responsibility
c_R	R-type firm's unit production cost
c_N	N-type firm's unit production cost, $0 \leq c_N < c_R$
δ_R	R-type consumer's reward for the R-type firm, $\delta_R \geq 0$
δ_N	R-type consumer's punishment for the N-type firm, $\delta_N \geq 0$
ρ_R	Measure of the type I error, i.e., the higher the ρ_R, the higher the probability for the R-type firm to obtain the certification
ρ_N	Measure of the type II error, i.e., the higher the ρ_N, the lower the probability for the N-type firm to obtain the certification
u_N	N-type consumer's utility of purchasing the product
u_{RR}	R-type consumer's utility of purchasing from the R-type firm
u_{RN}	R-type consumer's utility of purchasing from the N-type firm
π_R	R-type firm's profit
π_N	N-type firm's profit
S	Firm's signal, $S = 0$ or 1, indicating whether the firm obtains the certification or not
γ_S	$\gamma_S = Pr(R\|S)$, i.e., the posterior probability of the firm being R-type given signal $S \in \{0,1\}$;
$\mu(R\|p, S)$	R-type consumers' posterior belief about the probability of the firm being R-type after observing signal S and price p

(b) The R-type firm's optimal price decision and the corresponding profit are given by:

$$\hat{p}_R^* = \begin{cases} \frac{q+c_R+\delta_R\tau}{2}, & \textit{if } \delta_R \le \delta_3 \\ \delta_R, & \textit{if } \delta_3 < \delta_R \le \delta_2 \textit{ and } c_R \le \frac{q\left(\tau+\sqrt{\tau}-2\right)^2}{\tau^2-5\tau+4} \\ \frac{q+c_R-c_R\tau}{2(1-\tau)}, & \textit{if } \delta_2 < \delta_R \le \delta_1 \textit{ and } \tau \le \frac{1}{2} \textit{ and } c_R \le \frac{q\left(\frac{2\tau^2-7\tau+4}{\tau-1}+4\sqrt{(1-\tau)\tau}\right)}{5\tau-4} \\ \frac{q+c_R+\delta_R}{2}, & \textit{otherwise} \end{cases}, \textit{ and} \tag{4a}$$

$$\hat{\pi}_R^* = \begin{cases} \frac{(q-c_R+\delta_R\tau)^2}{4q}, & \textit{if } \delta_R \le \delta_3 \\ (\delta_R-c_R)(\tau+(1-\tau)(1-\frac{\delta_R}{q})), & \textit{if } \delta_3 < \delta_R \le \delta_2 \textit{ and } c_R \le \frac{q\left(\tau+\sqrt{\tau}-2\right)^2}{\tau^2-5\tau+4} \\ \frac{(q-c_R(1-\tau))^2}{4q(1-\tau)}, & \textit{if } \delta_2 < \delta_R \le \delta_1 \textit{ and } \tau \le \frac{1}{2} \textit{ and } c_R \le \frac{q\left(\frac{2\tau^2-7\tau+4}{\tau-1}+4\sqrt{(1-\tau)\tau}\right)}{5\tau-4} \\ \tau\frac{(q+\delta_R-c_R)^2}{4q}, & \textit{otherwise} \end{cases}, \tag{4b}$$

where $\delta_1 = \max\{c_R + \frac{q-c_R(1-\tau)}{\sqrt{\tau(1-\tau)}} - q, \frac{2\sqrt{(\tau-1)\left(c_R^2(\tau-1)-c_Rq(\tau-2)+q^2(\tau-1)\right)}-\tau(c_R+q)+2(c_R+q)}{4-3\tau}, \frac{q-c_R}{\sqrt{\tau}}\}$, $\delta_2 = \min\{\frac{q+c_R(1-\tau)}{2(1-\tau)}, \frac{2\sqrt{(\tau-1)\left(c_R^2(\tau-1)-c_Rq(\tau-2)+q^2(\tau-1)\right)}-\tau(c_R+q)+2(c_R+q)}{4-3\tau}\}$, *and* $\delta_3 = \min\{\frac{q+c_R}{2-\tau}, \frac{q-c_R}{\sqrt{\tau}}\}$.

Proposition 1 can be explained as follows. When the R-type consumers' punishment for the N-type firm is not too high (i.e., $\delta_N < \frac{(q-c_N)(1-\sqrt{1-\tau})}{\tau}$), the N-type firm will set a relatively low price to serve both types of consumers; otherwise the N-type firm will set a relatively high price to only serve the N-type consumers. When the R-type consumers' reward for the R-type firm is large (i.e., $\delta_R > \delta_1$), the R-type firm will set a relatively high price to only serve the R-type consumers; when the R-type consumers' reward is low (i.e., $\delta_R < \delta_3$), the R-type firm will set a relatively low price to serve a portion of the R-type and N-type consumers; otherwise, the R-type firm will set an intermediate price to serve all R-type consumers and part of N-type consumers. One can easily verify that $\hat{p}_N^* < \hat{p}_R^*$, i.e., the N-type firm always charges a lower price than the R-type firm.

Next, we investigate how the fraction of socially responsible consumers and their reward/punishment level affect the firm's optimal price and profit. The result is given in Proposition 2.

Proposition 2. *For the symmetric information case, it can be shown:*

(a) The N-type firm's optimal price and profit are decreasing in the fraction of R-type consumers (i.e., τ) and their punishment level (i.e., δ_N).

(b) The R-type firm's optimal price and profit are increasing in the fraction of R-type consumers (i.e., τ) and their reward level (i.e., δ_R).

Under symmetric information, the N-type firm is always worse off when there are more socially responsible consumers or when they exhibit stronger punishment behaviors. In contrast, the R-type firm always benefits from a larger socially responsible consumer base and their stronger reward behaviors. As a consequence, both the R-type firm and the government should focus on raising consumers' CSR awareness and educating them to enhance their reward/punishment activities to effectively warn the firms with non-ethical and irresponsible behaviors. However, the effectiveness of the above strategies depends critically on the assumption that consumers know the firm's responsibility type. Without such information transparency, as we will show next, these strategies may not work as desired.

5 Asymmetric Information

In this section, we analyze the situation where the firm's type is not observable to consumers, i.e., there is asymmetric information between the firm and consumers. Since an R-type consumer's willingness-to-pay for the R-type firm is higher than that for the N-type firm, the R-type firm has an incentive to reveal its true type whereas the N-type firm has an incentive to mimic the R-type firm. The R-type consumers form their belief about the firm's type based on the signal S and price p, and their expected utility is then given by Equation (1) in the previous section. In the following subsections, we analyze the Perfect Bayesian Equilibria (PBE) for the asymmetric information case. There are two types of equilibria: separating and pooling. In a separating equilibrium, different types of firms choose different prices, from which consumers can infer the firm's type. In a pooling equilibrium, both types of firm choose the same price and thus the consumers cannot infer the firm's type and their posterior belief is determined by the signal S (i.e., $\mu(R|S) = \gamma_S$). There is a continuum of Perfect Bayesian Equilibria depending on the consumer's off-equilibrium-path beliefs about the firm's type. We will refine the multiple equilibria by using the *lexicographically maximum sequential equilibrium* (LMSE) concept, which is formalized by Mailath et al. (1993) and has been applied in various marketing and operations management studies (see, e.g., Jiang et al. 2016; Guo and Jiang 2016; Schmidt et al. 2015, etc.).[4]

In our setting, the LMSE refinement selects the unique and the most profitable outcome for the R-type firm. The only separating equilibrium that could survive the LMSE refinement is the

[4]We remark that the Intuitive Criterion, introduced by Cho and Kreps (1987), is another widely adopted refinement tool in signaling game. However, as discussed in the related literature (see Mailath et al., 1993; Schmidt et al., 2015; Guo and Jiang, 2016; Jiang et al., 2016; Schmidt and Buell, 2017, etc.), the Intuitive Criterion sometimes eliminates plausible pooling equilibrium and may select a separating equilibrium with unrealistic off-equilibrium beliefs. In Appendix C, we will discuss the Intuitive Criterion as the equilibrium refinement tool and demonstrate that the main driving forces in our model are robust.

least-cost separating outcome from the R-type firm's perspective. The only pooling equilibrium that could be the LMSE is the R-type firm's most efficient pooling equilibrium among all pooling outcomes. If the least-cost separating equilibrium provides a higher profit for the R-type firm than the most-efficient pooling equilibrium, then the LMSE refinement selects this separating equilibrium as the unique outcome; otherwise, the most-efficient pooling equilibrium is the unique LMSE. The underlying intuition for such refinement can be explained as follows: Given that the R-type firm has a strong incentive to reveal its true type, if the R-type firm finds the pooling outcome more profitable than the separating outcome, the pooling outcome should be selected and vice versa. The detailed technical analysis and proofs are given in the appendix.

In the rest of this section, we first characterize the unique separating equilibrium and the unique pooling equilibrium that could be the LMSE in Section 5.1 and Section 5.2, respectively. Then we discuss which type of equilibrium will be selected by the LMSE refinement in different parameter regions in Section 5.3.

5.1 Separating Equilibrium

In a separating equilibrium, the two types of firm choose different prices, from which the consumers can infer the firm's type after observing the price. The unique separating equilibrium that could be the LMSE is the least-cost separating equilibrium as characterized in Lemma 1.

Lemma 1. *Under asymmetric information, the least-cost separating equilibrium must satisfy: (a) the N-type firm chooses price $p^*_{N,sep} = \hat{p}^*_N$ and its profit is $\pi^*_{N,sep} = \hat{\pi}^*_N$; (b) the R-type firm chooses price*

$$p^*_{R,sep} = \begin{cases} \hat{p}^*_R, & \textit{if } c_N \leq \bar{c}_N \\ \tilde{p}_{R,sep}, & \textit{otherwise} \end{cases}, \tag{5}$$

and its profit is

$$\pi^*_{R,sep} = \begin{cases} \hat{\pi}^*_R, & \textit{if } c_N \leq \bar{c}_N \\ \tilde{\pi}_{R,sep}, & \textit{otherwise} \end{cases}, \tag{6}$$

where $\bar{c}_N$, $\tilde{p}_{R,sep}$, and $\tilde{\pi}_{R,sep}$ are given in the appendix.

In the least-cost separating equilibrium, the N-type firm's price and profit are the same as those under symmetric information. When the N-type firm's production cost is lower than a threshold (i.e., $c_N \leq \bar{c}_N$), it is more profitable for the N-type firm to charge a low price rather than to mimic the R-type firm's high pricing strategy ($\hat{p}^*_R > \hat{p}^*_N$). Thus, the R-type firm can choose its first-best pricing strategy as in the symmetric-information case and costlessly separate himself from the N-type (i.e., $p^*_{R,sep} = \hat{p}^*_R$). We call this type of equilibrium "costless separating". In contrast, when the N-type firm is not very cost-efficient (i.e.,$c_N > \bar{c}_N$), it has a strong incentive to mimic the R-type firm so that it can take advantage of the R-type consumers' high willingness-to-pay. Thus, in order to credibly separate from the N-type firm, the R-type firm has to deviate from the first-best pricing strategy and incur a signaling cost; we call such an equilibrium "costly separating". The R-type firm can achieve the least-cost separation by choosing a higher-than-first-best price (i.e., $p^*_{R,sep} = \tilde{p}_{R,sep} > \hat{p}^*_R$) to ensure that the N-type firm finds it unprofitable to mimic. One can easily verify that $\tilde{\pi}_{R,sep} < \hat{\pi}^*_R$ due to the signaling cost. Next, we investigate how the key parameters, τ, δ_R, and δ_N, affect the firm's price and profit in the least-cost separating equilibrium. The results are given in Lemma 2.

Lemma 2. *In the least-cost separating equilibrium: (a) The R-type firm's price is increasing in τ, δ_N and δ_R; its profit is increasing in δ_R, decreasing in δ_N, but not monotonic in τ; (b) the N-type firm's price and profit are decreasing in τ and δ_N, and are independent of δ_R.*

Recall from the symmetric information case, the R-type firm's profit increases in the fraction of R-type consumers (i.e., τ) and their reward level (i.e., δ_R), and it is independent of R-type consumers' punishment level for the N-type firm (i.e., δ_N). However, under asymmetric information, the R-type firm's profit might even decrease in τ and δ_N in the least-cost separating equilibrium. As more consumers become socially responsible or their punishment level for the N-type firm increases, the N-type firm has a stronger incentive to mimic the R-type firm. Hence, the R-type firm has to incur a higher signaling cost to separate from the N-type firm, which leads to a lower profit.

5.2 Pooling Equilibrium

In a pooling equilibrium, both types of firm choose the same price, hence the consumer's posterior belief about the firm's type only depends on the signal S (i.e., $\mu(R|S)$). The unique pooling equilibrium that could be the LMSE is the most-efficient pooling equilibrium as given in Lemma 3.

Lemma 3. *In the most-efficient pooling equilibrium, for a given signal $S \in \{0, 1\}$, both types of firm* *choose the same price $p^*_{pool,S}$ and the j-type firm earns a profit $\pi^*_{j,pool,S}$:*

$$p^*_{pool,S} = \begin{cases} p_L(\bar{\delta}_S), & \text{if } \bar{\delta}_S < 0 \\ p_H(\bar{\delta}_S), & \text{if } \bar{\delta}_S \geq 0 \end{cases}, \tag{7a}$$

$$\pi^*_{R,pool,S} = \begin{cases} \pi_{R,L}(\bar{\delta}_S), & \text{if } \bar{\delta}_S < 0 \\ \pi_{R,H}(\bar{\delta}_S), & \text{if } \bar{\delta}_S \geq 0 \end{cases}, \tag{7b}$$

$$\pi^*_{N,pool,S} = \begin{cases} \pi_{N,L}(\bar{\delta}_S), & \text{if } \bar{\delta}_S < 0 \\ \pi_{N,H}(\bar{\delta}_S), & \text{if } \bar{\delta}_S \geq 0 \end{cases}, \tag{7c}$$

where $\bar{\delta}_S = \gamma_S\delta_R - (1-\gamma_S)\delta_N$ is the R-type consumers' expected punishment/reward, and the expressions of $p_i(\bar{\delta}_S)$ and $\pi_{j,i}(\bar{\delta}_S)$, $i = H, L$ and $j = R, N$, are given in the appendix.

For a given signal $S \in \{0, 1\}$, in the most-efficient pooling equilibrium, the consumers' posterior belief is the same as the prior belief conditional on signal S, i.e., the firm is R-type with probability γ_S and N-type with probability $1 - \gamma_S$. If $\bar{\delta}_S < 0$, then the R-type consumers would punish a pooling firm; otherwise they would reward a pooling firm. One can verify that the pooling price is always lower than the R-type firm's first-best price under symmetric information (i.e., $p^*_{pool,S} < \hat{p}^*_R$), but could be higher or lower than the N-type firm's first-best price $\hat{p}^*_N$. Next, we investigate how the key model parameters, τ, δ_R, δ_N, and γ_S, affect the firm's price and profit in the most-efficient pooling equilibrium.

Lemma 4. *In the most-efficient pooling equilibrium, for a given signal S: (a) Each type of firm's price and profit are decreasing in δ_N, but increasing in both γ_S and δ_R; (b) each type of firm's price and profit are decreasing in τ when $\bar{\delta}_S < 0$ and increasing in τ when $\bar{\delta}_S > 0$.*

In the most-efficient pooling equilibrium, as the R-type consumers' expected rewards/punishment (i.e., $\bar{\delta}_S$) increases, the pooling firm can charge a higher price and earn more profit. $\bar{\delta}_S$ is higher as the R-type consumers' reward level for the R-type firm (i.e., δ_R) increases, their punishment level for the N-type firm (i.e., δ_N) decreases, or the consumers' belief about the probability of the firm being R-type (i.e., γ_S) increases. Thus in the pooling equilibrium, each type of firm's price and profit decrease in δ_N, but increase in γ_S and δ_R.

Part (b) of Lemma 4 indicates that in the most-efficient pooling equilibrium, as the fraction of R-type consumers (i.e., τ) increases, each type of firm is worse off if the R-type consumers would punish the firm on expectation (i.e., $\bar{\delta}_S < 0$) and is better off if the R-type consumers would reward the firm on expectation (i.e., $\bar{\delta}_S > 0$). Recall that under symmetric information, as τ increases, the R-type firm is always better off and the N-type firm is always worse off. In the pooling outcome, however, the R-type consumers may punish the R-type firm (when $\bar{\delta}_S < 0$) or reward the N-type firm (when $\bar{\delta}_S > 0$) since they are unable to infer the firm's true type. This leads to the interesting result that the R-type firm might be worse off while the N-type firm might be better off as τ increases. Note that the rationale that the R-type firm's profit may decrease in τ in the pooling equilibrium is different from that in the separating equilibrium. The latter is due to the fact that more R-type consumers may cause a higher signaling cost for the R-type firm.

5.3 Equilibrium Outcome

We have characterized both the least-cost separating equilibrium and the most-efficient pooling equilibrium. Now, we apply the LMSE refinement tool to select the unique equilibrium outcome. Proposition 3 specifies the conditions under which the outcome is separating or pooling.

Proposition 3. *Under asymmetric information, there exists a unique $\gamma^* \in [0, 1]$, such that: (a) The unique LMSE outcome is the least-cost separating equilibrium as characterized in Lemma 1 given that the signal is $S \in \{0, 1\}$ and $\gamma_S < \gamma^*$; (b) the unique LMSE outcome is the most-efficient pooling equilibrium as characterized in Lemma 3 given that the signal is $S \in \{0, 1\}$ and $\gamma_S > \gamma^*$.*

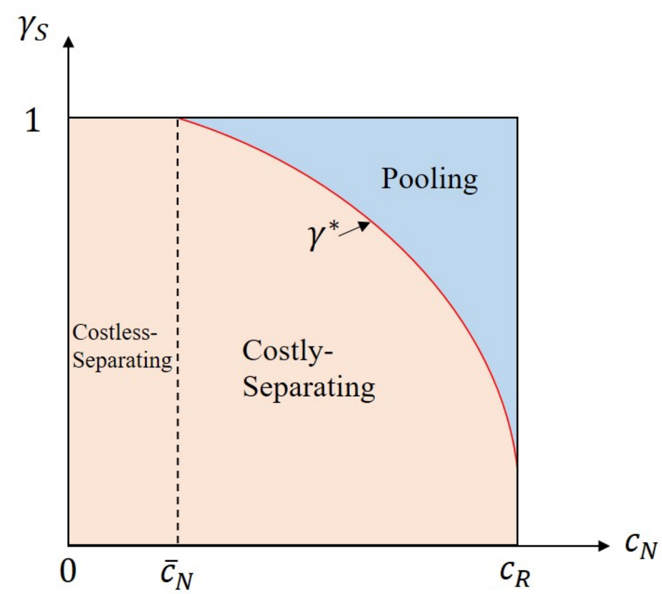


Figure 1: Illustration of the Equilibrium Outcome under Asymmetric Information

Proposition 3 fully characterizes the final equilibrium outcome for any given signal S. When the R-type consumers' belief about the probability of the firm being R-type conditional on the signal

S is high (i.e., $\gamma_S > \gamma^*$), then their expected reward/punishment for a pooling firm (i.e., $\bar{\delta}_S$) is more significant. Thus, the R-type firm finds the pooling strategy more profitable than incurring a signaling cost to separate from the N-type firm, which leads to the pooling outcome. On the other hand, when γ_S is low (i.e., $\gamma_S < \gamma^*$), the R-type consumers' expected reward/punishment for a pooling firm (i.e., $\bar{\delta}_S$) is less significant, which makes the pooling strategy less profitable for the R-type firm, thus leading to the separating outcome. In addition, in the separating region, there are two possible outcomes: costless separating and costly separating, as discussed in Section 5.1. When the N-type firm is cost-efficient (i.e., $c_N \leq \bar{c}_N$), the costless separating arises as the N-type firm sets a low price and sells more products instead of mimicking the R-type firm. When the N-type firm is not cost-efficient (i.e., $c_N > \bar{c}_N$), the costly separating arises as the N-type firm has a strong incentive to mimic the R-type firm, who needs to incur a signaling cost to separate. Figure 1 shows the equilibrium outcome defined by the values of consumers' belief conditional on signal S (i.e., γ_S) and the N-type firm's production cost c_N.[5]

5.3.1 Ex-Ante Profit Before Signal

We have characterized the equilibrium outcome for any given signal S, and the j-type firm's *ex-post* profit in the equilibrium is given by

$$\pi^*_{j,S} = \begin{cases} \pi^*_{j,sep} & \text{if } \gamma_S < \gamma^* \\ \pi^*_{j,pool,S} & \text{if } \gamma_S > \gamma^* \end{cases}. \tag{8}$$

Before the signal is revealed, the j-type firm's signal could be either $S = 1$ (e.g., it successfully obtains the certification) with probability $Pr(S = 1|j)$ or $S = 0$ (i.e., it fails to obtain the certification) with probability $Pr(S = 0|j)$. Thus, the j-type firm's *ex-ante* expected profit in the equilibrium is given by

$$\mathbb{E}[\pi^*_j] = Pr(S = 1|j) \cdot \pi^*_{j,1} + Pr(S = 0|j) \cdot \pi^*_{j,0}. \tag{9}$$

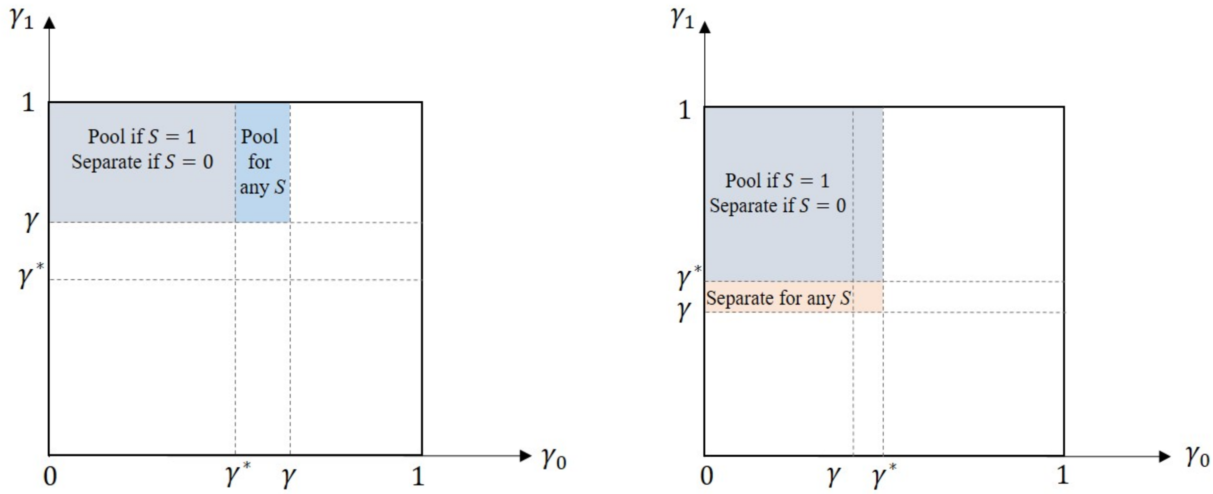
We present the ex-ante equilibrium outcome in Corollary 1 based on consumers' belief about the firm's type conditional on the signal $S = 1$ and the signal $S = 0$ (i.e., γ_0 and γ_1). Note that hereafter we refer to separating as the least-cost separating and pooling as the most-efficient pooling.

Corollary 1. *(a) If $\gamma_0 < \gamma_1 < \gamma^*$, the equilibrium outcome is separating for any signal $S \in \{0, 1\}$.*

[5] We remark that the threshold γ^* is independent of the certification signal S, but it has a complicated expression involving c_N, c_R, δ_N, δ_R, and τ. In fact, we find that γ^* is decreasing in c_N and thus Proposition 3 could also be written in terms of c_N. That is, there exists a c^*_N such that the unique LMSE outcome is the least-cost separating equilibrium if $c_N < c^*_N$ and is the most-efficient pooling equilibrium otherwise.

(b) If $\gamma^* < \gamma_0 < \gamma_1$, the equilibrium outcome is pooling for any signal $S \in \{0, 1\}$.

(c) If $\gamma_0 < \gamma^* < \gamma_1$, the equilibrium outcome is pooling when the signal $S = 1$ and is separating when $S = 0$.



(a) $\gamma > \gamma^*$ (b) $\gamma < \gamma^*$

Figure 2: Illustration of the Ex-Ante Equilibrium Outcome

Figure 2 depicts the results of Corollary 1. Note that γ is the consumers' prior belief of the firm's type before the signal S is revealed and one can easily verify that $\gamma_0 \leq \gamma \leq \gamma_1$. When both γ_1 and γ_0 are small enough (i.e., smaller than γ^*), then the R-type will separate himself from the N-type firm regardless of the realized signal S; when both γ_1 and γ_0 are large enough (i.e., greater than γ^*), then the N-type firm is able to pool with the R-type firm regardless of the realized signal S; otherwise, the equilibrium will depend on the realized signal S, i.e., the equilibrium outcome is pooling if $S = 1$ and is separating if $S = 0$. In the last case, if the N-type firm can obtain the certification, then it is able to pool with the R-type firm; and if the R-type firm does not obtain the certification, then it will deviate to the high pricing strategy to separate himself from the N-type firm.

6 Impact of Key Parameters

In this section, we conduct comparative statics analysis to understand how the key model parameters affect the firm's profit under asymmetric information. More specifically, our analysis will focus on the impact of (1) the fraction of social responsible consumers (i.e., τ); (2) the responsible consumers' reward/punishment level (i.e., δ_R and δ_N); and (3) the accuracy of the signal (i.e., ρ_R and ρ_N).

The purpose is to derive useful insights into how information asymmetry challenges and alters the conventional understanding of corporate social responsibility.

It is worth noting that all the comparative statics analysis in this section is on the entire feasible range of a given parameter. That is, as the chosen parameter changes, the final equilibrium type may change accordingly, i.e., our analysis endogenizes such equilibrium changes.

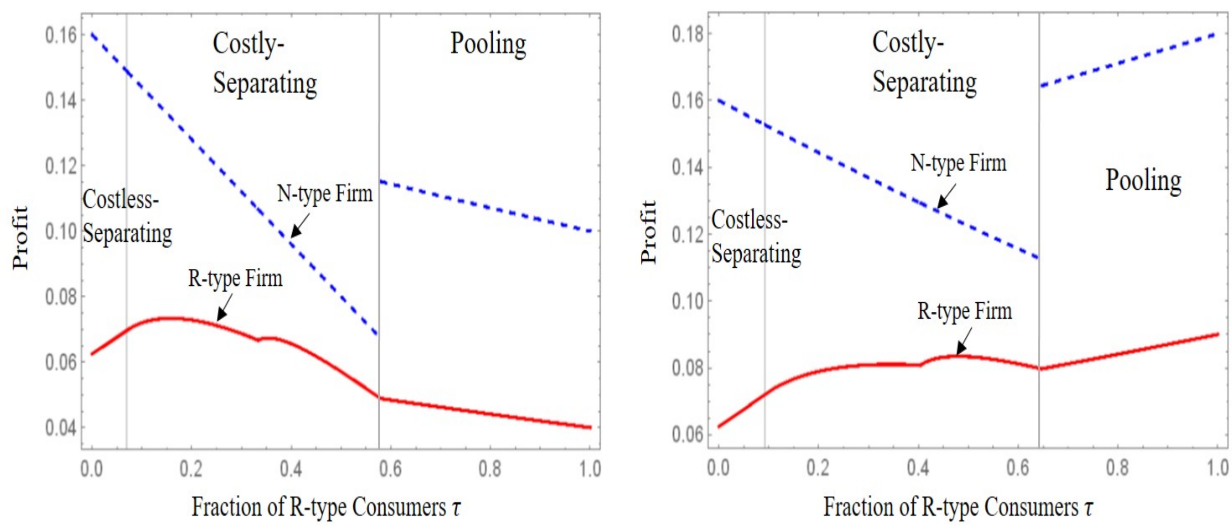
6.1 Fraction of R-type Consumers

First, we analyze the effect of the fraction of the R-type consumers (i.e., τ) on the equilibrium outcome. Intuitively, one might think that as the total number of socially responsible consumers increases, the N-type firm should be worse off and the R-type firm should be better off since these consumers would reward the R-type firm while punishing the N-type firm. This is true when the consumers can observe the firm's type (i.e., in the symmetric information case) or the signal is perfect (i.e., $\rho_N = \rho_R = 1$). However, when the firm's type is not observable to consumers and the signal is not perfect, this intuition may not hold. The formal result is given in Proposition 4. Note that Proposition 4 holds for both the ex-post profit given the signal $S = 1$ or 0 (see Equation 8) and the ex-ante profit before the signal S is revealed (see Equation 9).

Proposition 4. *For each type of firm, both the ex-ante and ex-post profits are not monotonic in* τ.

Within each equilibrium region, the R-type firm might be worse off as more consumers become socially responsible. As shown in Lemma 2, in the separating equilibrium region, as more consumers become aware of socially responsibility, the N-type firm has a stronger incentive to mimic the R-type firm, and thus, the R-type firm has to incur a higher signaling cost to separate from the N-type firm, which hurts the R-type firm. When the signaling cost is too high for the R-type to separate, the pooling equilibrium will arise. According to Lemma 4, for the pooling equilibrium region, if the R-type consumers punish a pooling firm (i.e., $\bar{\delta}_S < 0$), then more socially responsible consumers would also make the R-type firm suffer. Combining the above arguments explains the result that the R-type firm's profit may decrease in τ within each equilibrium region. As τ increases, the equilibrium type may switch from separating to pooling or from pooling to separating, which also leads the non-monotonicity of the R-type firm's profit with respect to τ.

Intuition suggests that more socially responsible consumers should make the N-type firm worse off. However, in the pooling equilibrium region, if the R-type consumers would reward a pooling firm (i.e., $\bar{\delta}_S > 0$), then the N-type firm could take advantage of the R-type consumers' high willingness-to-pay and benefit from more R-type consumers, as shown in Lemma 4. Even when the R-type consumers would punish a pooling firm (i.e., $\bar{\delta}_S < 0$), there also exists a situation where the N-type firm could benefit from a higher τ. As τ increases, the equilibrium type may switch from separating to pooling since the R-type has to incur a higher separating cost to prevent the mimicking from the N-type. When τ increases from the separating region to the pooling region, the N-type enjoys a higher profit no matter $\bar{\delta}_S > 0$ or $\bar{\delta}_S < 0$. This is because the R-type consumers' punishment level for the N-type firm in the separating outcome is always stronger than that for the N-type firm in the pooling outcome (i.e., $-\delta_N < \bar{\delta}_S$). These arguments together explain the result that the N-type firm's profit may increase in τ. Figure 3 shows how each type of firm' ex-post profit changes in τ. In addition, each type firm's ex-ante profit is not monotonic in τ, either, as both $Pr(S = 1|j)$ and $Pr(S = 0|j)$ are independent of τ, $j = R, N$. We remark that each type firm's ex-ante or ex-post price is also non-monotonic with respect to τ.



(a) Ex-Post Profit Given Signal S and $\bar{\delta}_S < 0$ (b) Ex-Post Profit Given Signal S and $\bar{\delta}_S > 0$

Figure 3: Effect of τ on Firm's Ex-Post Profit
Note $q = 1$, $\gamma_S = 0.5$, $c_R = 0.5$, $c_N = 0.2$, and $\delta_R = 0.4$; $\delta_N = 0.6$ in Figure 3 (a); $\delta_N = 0.2$ in Figure 3 (b).

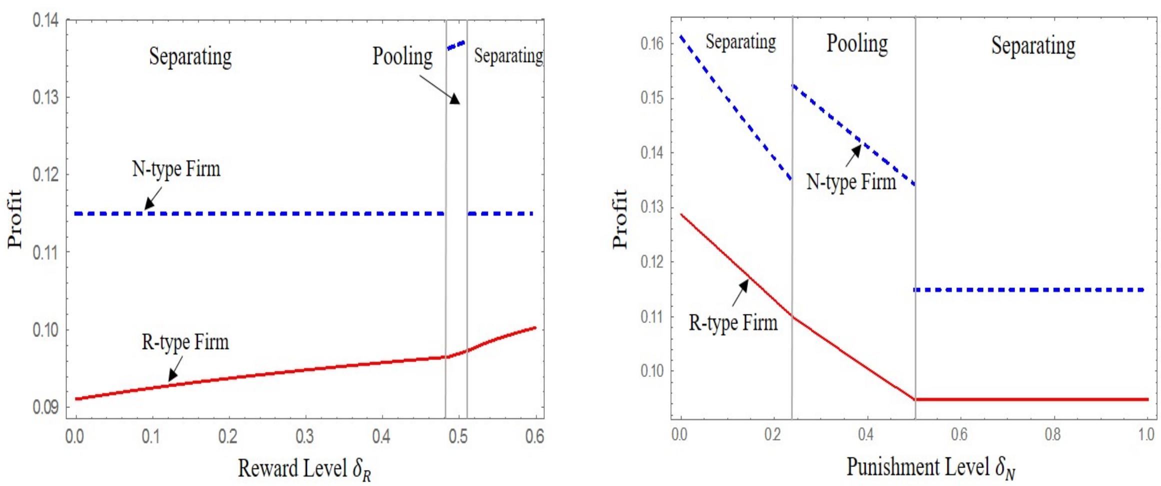
When the information is transparent or the signal is perfectly accurate, more socially responsible consumers will always benefit socially responsible firms and make socially irresponsible firms suffer. This suggests that both the government and the socially responsible firm should advocate consumers' CSR awareness and educate more consumers to be responsible. Such efforts can reward the socially responsible firms and punish socially irresponsible firms so as to pressure firms to be responsible and hence benefit the whole society.

However, in reality, information asymmetry usually exists and the signaling mechanisms are not perfect. Should the government or firms make efforts to increase the fraction of socially responsible consumers? Our results from Proposition 4 show that with information asymmetry, more consumers becoming socially responsible may hurt a socially responsible firm while benefiting a socially irresponsible firm. Consequently, without information transparency, both the government and socially responsible firms should be cautious about promoting consumers' awareness of social responsibility, since doing so may work against the desired objective. Instead, they may need to focus on eliminating the information asymmetry first.

6.2 Reward and Punishment Levels

Next, we analyze the impact of the R-type consumers' reward (δ_R) and punishment (δ_N) on the equilibrium outcome. The results are given in Proposition 5 and Figure 4. Similar to Proposition 4, we remark that Proposition 5 holds for both the ex-post profit given the signal $S = 1$ or 0 (see Equation 8) and the ex-ante profit before the signal is revealed (see Equation 9).

Proposition 5. *(a) The R-type firm's profit (both ex-ante and ex-post) is increasing in δ_R and decreasing in δ_N; (b) the N-type firm's profit (both ex-ante and ex-post) is not monotonic in δ_R and δ_N.*



(a) Effect of δ_R on Ex-Post Profit for Given S

(b) Effect of δ_N on Ex-Post Profit for Given S

Figure 4: Effect of δ_N and δ_R on Firm's Ex-Post Profit
Note $q = 1$, $\gamma_S = 0.368$, $c_R = 0.325$, $c_N = 0.197$, and $\tau = 0.287$; there is $\delta_N = 0.579$ in Figure 4 (a) and $\delta_R = 0.3$ in Figure 4 (b).

Similar to the symmetric information case, the R-type firm's profit increases in the reward level δ_R under asymmetric information. Differently, the R-type firm's profit is independent of the R-type consumers' punishment level for the N-type firm (i.e., δ_N) under symmetric information, while the

R-type firm is worse off as the punishment becomes stronger under asymmetric information. In the separating equilibrium region, as δ_N increases, the N-type firm has a stronger incentive to mimic the R-type firm and, thus, the R-type firm may have to incur a higher signaling cost to separate, which leads to a lower profit for the R-type firm. While in the pooling equilibrium region, since the R-type consumers cannot infer the firm's type, their stronger punishment will hurt both the N-type firm and the R-type firm.

Intuitively, one might think that the N-type firm should always be worse off as the consumers' punishment level increases, which holds under the symmetric information. However, interestingly, Proposition 5(b) indicates that the N-type firm may benefit from a stronger punishment when the firm's type is not observable to consumers. This result can be explained as follows. As δ_N increases, the N-type firm has a stronger incentive to mimic the R-type firm, which leads to a higher signaling cost for the R-type firm. When the signaling cost becomes too high, the R-type firm may switch from the separating strategy to the pooling strategy. Consequently, the N-type firm enjoys a higher profit as the equilibrium type switches from separating to pooling, since the R-type consumers' punishment for the N-type firm in the separating equilibrium is always stronger than that in the pooling equilibrium (i.e., $-\delta_N < \bar{\delta}_S$). Therefore, although within each equilibrium type region, the N-type firm's profit decreases in δ_N, it may increase in δ_N as the equilibrium type switches from separating to pooling. Figure 4 provides an illustrational example.

Part (b) of Proposition 5 also shows that as the consumers' rewards level for a R-type firm (i.e., δ_R) increases, the N-type firm could be either worse off or better off. Within the separating region, the N-type firm's profit is independent of δ_R; while within the pooling region, it is increasing in δ_R since the R-type consumers' willingness-to-pay for a pooling firm increases. As δ_R increases, it will also alter the R-type firm's incentive to separate or pool. The N-type firm enjoys a profit increase when the equilibrium switches from separating to pooling and suffers a profit loss when the equilibrium switches from pooling to separating. See Figure 4 for an example. The effects of δ_N and δ_R on each type firm's ex-ante profit has a similar pattern to that for the ex-post profit, since both $Pr(S = 1|j)$ and $Pr(S = 0|j)$ are independent of δ_R and δ_N, $j = R, N$. Note that each type firm's ex-ante or ex-post price is not monotonic in δ_R and δ_N, either.

Proposition 5 has the following implications. When a firm's responsibility type is observable to the public, inducing consumers to punish more on socially irresponsible activities or reward more on socially responsible activities would serve as an effective mechanism to contain non-ethical and irresponsible behaviors and promote the concept of CSR. However, in a market without information transparency, efforts spent on raising consumers' punishment toward socially irresponsible firms or their rewards' toward socially responsible firms could potentially backfire as a socially responsible firm may be hurt whereas a socially irresponsible firm may benefit from such a change. This suggests that the government and socially responsible firms should be careful regarding enhancing consumers' response to socially responsible/irresponsible activities. Information asymmetry may significantly distort the consumers' goodwills for CSR, and eliminating the information asymmetry is critical to resolving such issues.

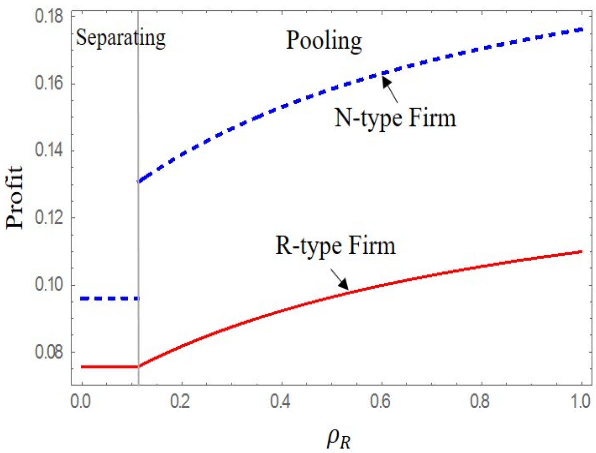
6.3 Signal Accuracy

We proceed to examine the effects of the signal accuracy. As ρ_R or ρ_N increases, the signal is more accurate for consumers to infer whether a firm is socially responsible or not. We start our analysis with the ex-post scenario (i.e., after the signal S is given) and then we analyze the ex-ante scenario (i.e., before the signal is revealed). The effects of ρ_N and ρ_R on the firm's ex-post profit for given signal S (see Equation 8) are summarized in Proposition 6.

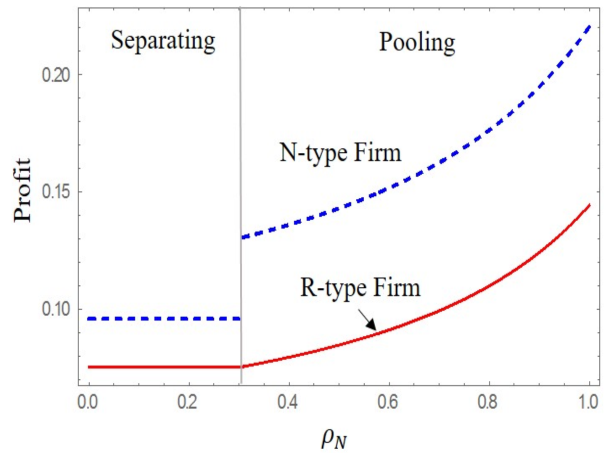
Proposition 6. *(a) Given the signal $S = 1$, the j-type firm's ex-post profit $\pi^*_{j,1}$ is increasing in ρ_N and ρ_R, for $j \in \{R, N\}$.*

(b) Given the signal $S = 0$, the j-type firm's ex-post profit $\pi^*_{j,0}$ is decreasing in ρ_R and ρ_N, for $j \in \{R, N\}$.

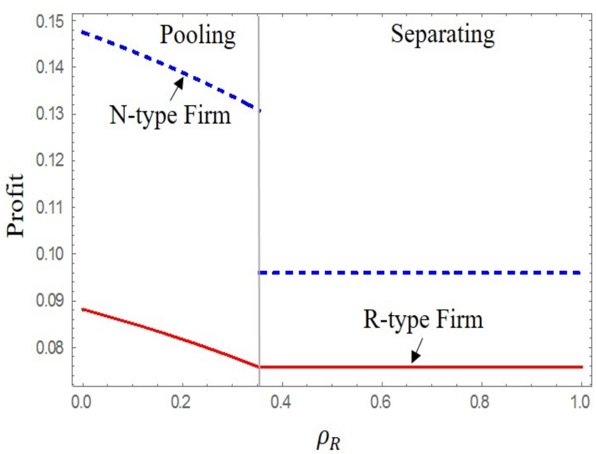
Given the signal $S = 1$, the probability of the firm being R-type (i.e., γ_1) increases in ρ_R and ρ_N; given the signal $S = 0$, the probability of the firm being R-type (i.e., γ_0) decreases in ρ_R and ρ_N. This implies that as the signal accuracy improves, the R-type consumers believe that the firm who has obtained the certification is more likely to be the R-type and the firm without the certification is less likely to be the R-type. Within the separating region, the signal is irrelevant and thus each type firm's profit is independent of ρ_N and ρ_R. While within the pooling region, as the R-type consumers' belief about the probability of the firm being R-type (i.e., γ_S) increases, their expected reward/punish level will increase. Therefore, in the pooling region, when $S = 1$ (i.e., the firm has the certification), a higher accuracy level (i.e., larger ρ_R or ρ_N) leads to a higher probability of the firm being the R-type, which benefits the firm regardless of its true type; in contrast, when $S = 0$ (i.e., the firm does not have the certification), a higher accuracy level (i.e., larger ρ_R or ρ_N) leads to a lower probability of the firm being the R-type, which hurts the firm regardless of its true type. See Figure 5 for an illustration. Note that when $S = 1$, as ρ_R or ρ_N increases, the equilibrium type may change from separating to pooling (see Figures 5(a) and (b)); when $S = 0$, the opposite happens (see Figures 5(c) and (d)).



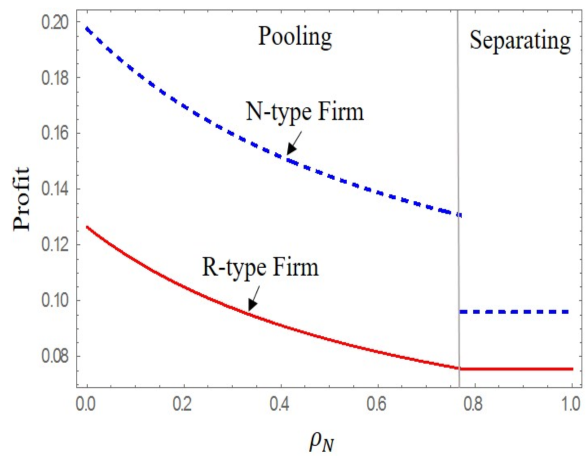
(a) Effect of ρ_R on Ex-Post Profit $\pi^*_{j,1}$



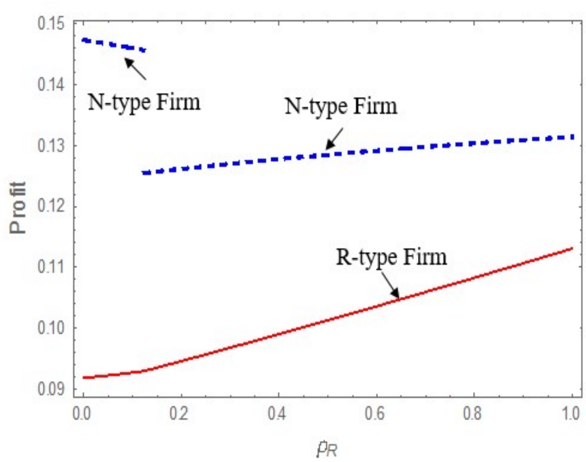
(b) Effect of ρ_N on Ex-Post Profit $\pi^*_{j,1}$



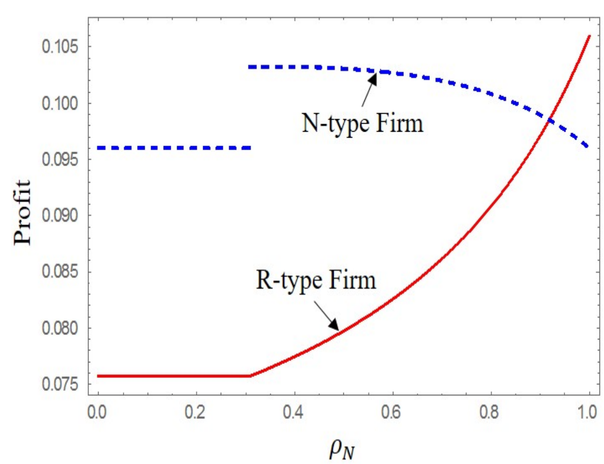
(c) Effect of ρ_R on Ex-Post Profit $\pi^*_{j,0}$



(d) Effect of ρ_N on Ex-Post Profit $\pi^*_{j,0}$



(e) Effect of ρ_R on Ex-Ante Profit $\mathbb{E}(\pi^*_j)$



(f) Effect of ρ_N on Ex-Ante Profit $\mathbb{E}(\pi^*_j)$

Figure 5: Effects of ρ_R and ρ_N on Equilibrium Outcome

Note $q = 1$, $c_R = 0.4$, $c_N = 0.2$, $\delta_N = 0.6$, $\delta_R = 0.4$, and $\tau = 0.4$ in Figure 5; $\rho_N = 0.6$ in Figure 5 (a) and (c), $\rho_N = 0.3$ in Figure 5 (e), and $\rho_R = 0.2$ in Figure 5 (b), (d), and (f); $\gamma = 0.2$ in Figure 5 (a), $\gamma = 0.3$ in Figure 5 (b) and (f), $\gamma = 0.9$ in Figure 5 (c) and (d), and $\gamma = 0.6$ in Figure 5 (e).

Proposition 6 indicates that the firm who successfully obtains the certification can benefit from the higher accuracy, and the firm who fails to get such certification will be worse off regardless of the firm's social responsibility type. That is, a socially responsible firm, who fails to obtain the certification due to the type I error, will suffer from the higher accuracy of the signal; and a socially irresponsible firm, who obtains the certification due to the type II error, will benefit from the higher accuracy of the signal. Next, we study the effects of ρ_R and ρ_N on a firm's ex-ante profit before the signal is revealed (see Equation 9). The results are summarized in Proposition 7.

Proposition 7. *(a) The R-type firm's ex-ante profit* $\mathbb{E}[\pi_R^*]$ *is increasing in* ρ_R *and* ρ_N*.*

(b) The N-type firm's ex-ante profit $\mathbb{E}[\pi_N^*]$ *is not monotonic in* ρ_R *and* ρ_N*.*

Increasing ρ_R will reduce the probability of the type I error, i.e., the R-type firm sends a signal $S = 0$ (he fails to obtain the certification); increasing ρ_N will reduce the probability of the type II error, i.e., the N-type firm sends a signal $S = 1$ (he successfully obtains the certification). In other words, as ρ_R increases, the R-type firm is more likely to obtain the certification (i.e., $Pr(S = 1|R)$ increases) and the consumers also believe that a firm with the certification is more likely to be the R-type (i.e., $\gamma_S = Pr(R|S = 1)$ increases). As ρ_N increases, the N-type firm is less likely to obtain the certification (i.e., $Pr(S = 1|N)$ decreases) and also the consumers believe that a firm with the certification is more likely to be the R-type (i.e., $\gamma_S = Pr(R|S = 1)$ increases). Consequently, the R-type firm can always benefit from the increased signal accuracy, as shown by Proposition 7(a).

One may intuit that the N-type firm should be worse off as the signal accuracy increases. However, part (b) of Proposition 7 indicates that under certain situations the N-type firm may also be better off as the signal accuracy increases. The intuition for this unexpected result is as follows. On one hand, as ρ_N increases, the N-type firm has a smaller chance to obtain the certification; and if it fails to obtain the certification (i.e., $S = 0$), its profit will be worse off for higher ρ_R or ρ_N, as explained in Proposition 6 (b). However, on the other hand, if the type II error occurs, i.e., the N-type firm obtains the certification and sends the signal $S = 1$, the N-type firm can enjoy the benefit from higher ρ_R or ρ_N, as explained in Proposition 6 (a). Therefore, combining these two counteracting effects in the N-type firm's ex-ante profit, we find that when ρ_N is not large enough (i.e., the probability of N-type obtaining the certification is high), increasing the signal accuracy may benefit the N-type firm. Such a result of Proposition 7 can be illustrated by Figure 5(e) and (f).

Although the N-type firm could benefit from increased signal accuracy, when $\rho_N = 1$ (i.e., type II error does not exist), the N-type firm is never able to get the certification and its ex-ante profit is the same as the ex-post profit when $S = 0$, which decreases in the signal accuracy ρ_R. Moreover, when the signal is perfect (i.e., $\rho_R = \rho_N = 1$), the N-type achieves the lowest profit among all regions. That is, if the type II error can be eliminated, increasing the signal accuracy will always make the N-type firm worse off.

Proposition 7 implies that socially responsible firms always have the incentive to improve the accuracy of the signal, regardless of the signal's formats, e.g., applying for a certification, issuing a public report, etc. It also suggests the government should encourage or even mandate firms to obtain the certification, and at the same time exert efforts to help improve the accuracy of the certification process. This will reward the socially responsible firms and in most cases punish the irresponsible firms. In addition, the government and socially responsible firms should be aware that the irresponsible firms may exploit the advantage of higher certification accuracy when there is a good chance to deceive the consumers (e.g., pass the certification tests). Making the certification process more reliable to reduce the type II error can help avoid this issue.

At last, we discuss the value of sending the signal $S = 1$ (i.e., obtaining a certification) and the value of sending a signal (i.e., applying for the certification). One can easily verify that $\pi^*_{j,1} \geq \pi^*_{j,0}$ for $j \in \{R, N\}$, i.e., holding the certification is always valuable for each type of firm. Let $\Delta_j = \pi^*_{j,1} - \pi^*_{j,0} \geq 0$ denote the ex-post value of having the certification; let $\mathbf{E}[\Delta_j] = \mathbb{E}[\pi^*_j] - \pi^*_{j,0} = Pr(S = 1|j) \cdot \Delta_j \geq 0$ denote the ex-ante value of applying for the certification. Proposition 8 characterizes the impact of signal accuracy on both the ex-ante and ex-post values of certification.

Proposition 8. *(a) The ex-post value of certification, Δ_j, is increasing in ρ_R and ρ_N, for $j \in \{R, N\}$;*

(b) The ex-ante value of certification for the R-type firm, $\mathbf{E}[\Delta_R]$, is increasing in ρ_R and ρ_N; the ex-ante value for N-type firm, $\mathbf{E}[\Delta_N]$, is increasing in ρ_R but is not monotonic in ρ_N;

Part (a) of Proposition 8 shows that the ex-post value of certification becomes greater for each type of firm as the accuracy increases. As the signal accuracy (i.e., ρ_R and ρ_N) increases, the consumers believe that the firm with the certification (i.e, $S = 1$) is more likely to be the R-type and the firm without the certification (i.e., $S = 0$) is less likely to be the R-type. Therefore, as shown in Proposition 6, each type of firm's ex-post profit will be higher given $S = 1$ and will be lower given $S = 0$ as the accuracy increases. Hence, having the certification becomes more valuable as the signal accuracy increases.

Part (b) shows that as the accuracy increases, the ex-ante value of certification becomes higher for the R-type firm, which is due to two reasons: First, the value of having the certification (i.e., Δ_R) becomes higher; second, the probability of the R-type firm obtaining the certification (i.e.,

$Pr(S = 1|R)$) becomes higher. While for the N-type firm, the ex-ante value of the certification increases as ρ_R increases, but might decrease or increase as ρ_N increases. Similarly, as ρ_R or ρ_N increases, the value of having the certification for the N-type firm (i.e., Δ_N) is higher. However, the probability of the N-type firm obtaining the certification (i.e., $Pr(S = 1|N)$) is lower as ρ_N increases. Moreover, when $\rho_N = 1$, i.e., the type II error is completely eliminated, then $\mathbf{E}[\Delta_N] = 0$ and pursuing the certification has no value for the N-type firm.

When there is no cost for sending the signal (e.g., it is free to go through the certification process), each type of firm will choose to send the signal since $\mathbf{E}[\Delta_j] \geq 0$. However, when it is costly, the firm may or may not choose to send a signal, depending on the cost-benefit trade-off. When the probability of the N-type firm obtaining the certification is low enough (i.e, ρ_N is large enough), the ex-ante value of the certification for the N-type firm is small enough; when the probability of the R-type firm obtaining the certification is high enough (i.e, ρ_R is large enough), the ex-ante value of the certification for the R-type firm is large enough. In this situation, charging an appropriate fee for applying the certification may help prevent the N-type firm from applying, and thus make it easier for the R-type firm to separate from the N-type. This suggests that a potential way to resolve the information asymmetry issue is to impose an appropriate signaling cost for the firm (e.g., a certification fee).

The above results yield some useful implications. On one hand, the government should encourage or even mandate firms to apply for the certification, and try to tighten the standards to improve the accuracy of the process to reduce the type I error and especially the type II error. On the other hand, when it is difficult to fully resolve the type I or type II error, the government should also carefully price the certification to prevent irresponsible firms' opportunistic behaviors. Such a strategy could help restore information transparency even if the signal is not perfect. In particular, if only the R-type firm goes through the certification process while the N-type firm does not, and $\rho_R = 1$, i.e., the R-type firm can always pass the certification process, then the two types of firm can be credibly separated and consumers can identify the firm's type based on whether it holds the certification or not.

7 Conclusion

This paper studies the interaction between a firm and consumers under the presence of corporate social responsibility consideration and asymmetric information. The firm can be either the socially responsible type or the irresponsible type; however, the consumers cannot observe the firm's exact type, which is private information. The firm may attempt to reveal its type through various signaling mechanisms, but these mechanisms may not be perfectly accurate. The purpose of this research is to obtain a better understanding of the social responsibility issue and provide useful insights for both the firm and the government.

We find that information asymmetry plays a critical role in the above problem setting. Due to the existence of information asymmetry, there are some interesting findings regarding the impact of key problem parameters on the firm's pricing strategy and profit performance. Specifically, increasing consumer awareness of corporate social responsibility (i.e., increasing the fraction of socially concerned consumers) may hurt the socially responsible firm while benefiting the irresponsible firm. In addition, a stronger consumer punishment for the irresponsible firm always hurts the responsible firm, but it may lead to a higher profit for the irresponsible firm under certain conditions. Finally, the irresponsible firm may also prefer a higher signaling accuracy, which is contrary to the intuition that the irresponsible firm should try to hide its type. All these counterintuitive findings are driven by the fact that the consumers cannot observe the firm's true type, and as a result, the firm may want to reveal or mimic the responsible type through pricing and signaling. As certain problem parameters vary, they may change the signaling cost for the firm as well as the equilibrium outcome of the game between the firm and the consumers.

The findings from this research have useful implications for the government and firms. In particular, they indicate that with the presence of information asymmetry, some intuitive strategies to encourage socially responsible corporate behaviors may not work as desired. For example, simply promoting consumer awareness through advertising and educational programs may actually lead to the opposite of the intended outcome. Thus, both the government and the socially responsible firms should first exert efforts to minimize the information asymmetry regarding the firms' activities. This corroborates with the evidence that more and more firms have tried to create transparency about their production and supply chain practices (Kraft et al., 2016), and more and more governments have started to encourage or mandate sustainability reports from firms (UNEP, 2013). Indeed, by restoring information transparency, one can align the socially responsible consumers' goodwills with firms' incentives to pursue business activities that are responsible for the society and the environment.

References

Agrawal, V., D. Lee. 2015. The effect of sourcing policies on a suppliers sustainable practices. *working paper* .

Berman, S., A.. Wicks, S. Kotha, T. Jones. 1999. Does stakeholder orientation matter? the relationship between stakeholder management models and firm financial performance. *The Academy of Management Journal* **42**(5) 488–506.

Boersma, M. 2013. Mind the gap: Company disclosure discrepancies not sustainable. *The Conversation (June 13rd)* .

Chen, J., S. Slotnick. 2015. Supply chain disclosure and ethical sourcing. *International Journal of Production Economics* **161** 17–30.

Chen, L., HL. Lee. 2016. Sourcing under supplier responsibility risk: The effects of certification, audit, and contingency payment. *Management Science forthcoming* .

Chen, L., S. Yao, K. Zhu. 2016. Should brands tighten certification standards? sourcing under supplier-certifier collusion risk. *Working paper* .

Cho, I., D. Kreps. 1987. Signaling games and stable equilibria. *The Quarterly Journal of Economics* **102**(2) 179–221.

Cho, S., S. Tayur, Y. Xu. 2016. Combating child labor: Incentives and information transparency in global supply chains. *working paper* .

CONE-Communication/Ebiquity. 2015. Global csr study. *White Paper* .

De Pelsmacker, P, L Driesen, G Rayp. 2005. Do consumers care about ethics? willingness to pay for fair-trade coffee. *Journal of Consumer Affairs* **39**(2) 363–385.

Desai, PS. 2000. Multiple messages to retain retailers: Signaling new product demand. *Marketing Science* **19**(4) 381–389.

Desai, PS., K. Srinivasan. 1995. Demand signalling under unobservable effort in franchising: Linear and nonlinear price contracts. *Management Science* **41**(10) 1608–1623.

Du, S, C.B. Bhattacharya, S Sen. 2010. Maximizing business returns to corporate social responsibility (csr): The role of csr communication. *International Journal of Management Reviews* **12**(1) 8–19.

Ghosh, B., MR. Galbreth. 2013. The impact of consumer attentiveness and search costs on firm quality disclosure: A competitive analysis. *Management Science* **59**(11) 2604–2621.

Greenpeace. 2011. Dirty laundry: Unravelling the corporate connections to toxic water pollution in china. *White paper* .

Guo, R., HL. Lee, R. Swinney. 2016. Responsible sourcing in supply chains. *Management Science*

62(9) 2722–2744.

Guo, X., B. Jiang. 2016. Signaling through price and quality to consumers with fairness concerns. *Journal of Marketing Research* **53**(6) 988–1000.

Huang, L., JS. Song, R. Swinney. 2015. Managing social responsibility in multi-tier supply chains. *working paper* .

Husock, H. 2013. The bangladesh disaster and corporate social responsibility. *Forbes (May 2th)* .

Iyer, G., S. Singh. 2017. Voluntary product safety certification. *Management Science* Forthcoming.

Jiang, B., J. Ni, K. Srinivasan. 2014. Signaling through pricing by service providers with social preferences. *Marketing Science* **33**(5) 641–654.

Jiang, B., L. Tian, Y., F. Zhang. 2016. To share or not to share: Demand forecast sharing in a distribution channel. *Marketing Science* **35**(5) 800–809.

Jiang, B., B. Yang. 2017. Quality and pricing decisions of experience goods in markets with consumer information sharing. *Management Science* Forthcoming.

Kim, S. 2015. Time to come clean? disclosure and inspection policies for green production. *Operations Research* **63**(1) 1–20.

Kraft, T., F. Erhun, RC. Carlson, D. Rafinejad. 2013a. Replacement decisions for potentially hazardous substances. *Production and Operations Management* **22**(4) 958–975.

Kraft, T., L.Valdes, Y. Zheng. 2016. Supply chain visibility and social responsibility: Investigating consumers' behaviors and motives. *Working paper* .

Kraft, T., Y. Zheng, F. Erhun. 2013b. The ngo's dilemma: How to influence firms to replace a potentially hazardous substance. *Manufacturing & Service Operations Management* **15**(4) 649–669.

Lai, G., L. Debo, L. Nan. 2011. Channel stuffing with short-term interest in market value. *Management Science* **57**(2) 332–346.

Lai, G., W. Xiao. 2016. Inventory decisions and signal of demand uncertainty to investors. *Manufacturing & Service Operations Management Forthcoming* .

Lai, G., W. Xiao, J. Yang. 2012. Supply chain performance under market valuation: An operational approach to restore efficiency. *Management Science* **58**(10) 1933–1951.

Lee, HL., CS. Tang. 2017. Socially and environmentally responsible value chain innovations: New operations management research opportunities. *Management Science* Forthcoming.

Li, Z., S. Gilbert, G. Lai. 2014. Supplier encroachment under asymmetric information. *Management Science* **60**(2) 449–462.

Lydenberg, S. 2012. On materiality and sustainability: The value of disclosure in the capital markets. *White paper* .

Mailath, G., M. Okuno-Fujiwara, A. Postlewaite. 1993. Belief-based refinements in signaling games. *Journal of Economic Theory* **60**(2) 241–276.

Moorthy, S., K. Srinivasan. 1995. Signaling quality with a money-back guarantee: The role of transaction costs. *Marketing Science* **14**(4) 442–466.

Newell, A. 2015. How nike embraced csr and went from villain to hero. *Triple pundit (Jun 19th)* .

Patagonia. 2014. Patagonia 2014: The footprint chronicles URL `http://www.patagonia.com/us/footprint`.

Plambeck, E., TA. Taylor. 2016. Supplier evasion of a buyers audit: Implications for motivating supplier social and environmental responsibility. *Manufacturing & Service Operations Management* **18**(2) 184–197.

Schmidt, W., R. Buell. 2017. Experimental evidence of pooling outcomes under information asymmetry. *Management Science* **63**(5) 1586–1605.

Schmidt, W., V. Gaur, R. Lai, A. Raman. 2015. Signaling to partially informed investors in the newsvendor model. *Production and Operations Management* **24**(3) 383–401.

Sen, S., C.B. Bhattacharya. 2001. Does doing good always lead to doing better? consumer reactions to corporate social responsibility. *Journal of Marketing Research* **38**(2) 225–243.

Servaes, H., A. Tamayo. 2013. The impact of corporate social responsibility on firm value: The role of customer awareness. *Management Science* **59**(5) 1045–1061.

Starbucks. 2006. Starbucks corporation: Corporate social responsibility. *Annual Report* .

UNEP. 2013. Frequently asked questions on corporate sustainability reporting URL `https://www.globalreporting.org/resourcelibrary/GoF47Para47-FAQs.pdf`.

Appendix

This appendix is divided into three parts. Part A presents the proofs for the symmetric information case; part B presents the equilibrium analysis for the asymmetric information case; and part C discusses the Intuitive Criterion.

A. Symmetric Information

Proof of Proposition 1. We first derive the N-type firm's optimal price and then the R-type firm's optimal price when the firm's type $j \in \{R, N\}$ is observable to consumers.

A R-type consumer will purchase from the N-type firm if and only if (iff) $u_{RN} = \theta q - p - \delta_N \geq 0$, i.e., $\theta \geq (p + \delta_N)/q$; an N-type consumer will purchase from the N-type firm iff $u_N = \theta q - p \geq 0$, i.e., $\theta \geq p/q$. Thus the N-type firm's demand is given as below:

$$d_N(p) = \begin{cases} \tau(1 - \frac{p+\delta_N}{q}) + (1-\tau)(1 - \frac{p}{q}), & \text{if } p \leq q - \delta_N \\ (1-\tau)(1 - \frac{p}{q}), & \text{if } q - \delta_N < p \leq q \\ 0, & \text{if } p > q \end{cases} \cdot$$

The N-type firm's problem can be written as:

$$\max_{p \geq 0} \hat{\pi}_N(p) = \begin{cases} (p - c_N)(\tau(1 - \frac{p+\delta_N}{q}) + (1-\tau)(1 - \frac{p}{q})), & \text{if } p \leq q - \delta_N \\ (p - c_N)(1-\tau)(1 - \frac{p}{q}), & \text{if } q - \delta_N < p \leq q \\ 0, & \text{if } p > q \end{cases} \cdot$$

Solving the above optimization problem, we obtain the N-type firm's optimal price as follows:

$$\hat{p}_N^* = \begin{cases} \frac{q + c_N - \delta_N \tau}{2}, & \text{if } \delta_N < \frac{(q - c_N)(1 - \sqrt{1-\tau})}{\tau} \\ \frac{q + c_N}{2}, & \text{otherwise} \end{cases} .$$

The N-type firm's optimal profit is given by:

$$\hat{\pi}_N^* = \begin{cases} \frac{(q - c_N - \delta_N \tau)^2}{4q}, & \text{if } \delta_N < \frac{(q - c_N)(1 - \sqrt{1-\tau})}{\tau} \\ (1-\tau)\frac{(q - c_N)^2}{4q}, & \text{otherwise} \end{cases} .$$

Next, we analyze the R-type firm's problem. A R-type consumer will purchase from the R-type firm iff $u_{RR} = \theta q - p + \delta_R \geq 0$, i.e., $\theta \geq (p - \delta_R)/q$; an N-type consumer will purchase from the R-type firm iff $u_N = \theta q - p \geq 0$, i.e., $\theta \geq p/q$. Thus the R-type firm's demand is given as below:

$$d_R(p) = \begin{cases} \tau + (1-\tau)(1-\frac{p}{q}), & \text{if } p \le \delta_R \\ \tau(1-\frac{p-\delta_R}{q}) + (1-\tau)(1-\frac{p}{q}), & \text{if } \delta_R < p \le q \\ \tau(1-\frac{p-\delta_R}{q}, & \text{if } q < p \le q+\delta_R \\ 0, & \text{if } p > q+\delta_R \end{cases}.$$

The R-type firm's problem can be written as:

$$\max_{p\ge 0} \hat{\pi}_R(p) = \begin{cases} (p-c_R)(\tau + (1-\tau)(1-\frac{p}{q})), & \text{if } p \le \delta_R \\ (p-c_R)(\tau(1-\frac{p-\delta_R}{q}) + (1-\tau)(1-\frac{p}{q})), & \text{if } \delta_R < p \le q \\ (p-c_R)\tau(1-\frac{p-\delta_R}{q}), & \text{if } q < p \le q+\delta_R \\ 0, & \text{if } p > q+\delta_R \end{cases}.$$

Solving the above optimization problem, we obtain the R-type firm's optimal price decision and the corresponding profit as follows:

$$\hat{p}_R^* = \begin{cases} \frac{q+c_R+\delta_R\tau}{2}, & \text{if } \delta_R \le \delta_3 \\ \delta_R, & \text{if } \delta_3 < \delta_R \le \delta_2 \text{ and } c_R \le \frac{q\left(\tau+\sqrt{\tau}-2\right)^2}{\tau^2-5\tau+4} \\ \frac{q+c_R-c_R\tau}{2(1-\tau)}, & \text{if } \delta_2 < \delta_R \le \delta_1 \text{ and } \tau \le \frac{1}{2} \text{ and } c_R \le \frac{q\left(\frac{2\tau^2-7\tau+4}{\tau-1}+4\sqrt{(1-\tau)\tau}\right)}{5\tau-4} \\ \frac{q+c_R+\delta_R}{2}, & \text{otherwise} \end{cases}, \text{ and}$$

$$\hat{\pi}_R^* = \begin{cases} \frac{(q-c_R+\delta_R\tau)^2}{4q}, & \text{if } \delta_R \le \delta_3 \\ (\delta_R-c_R)(\tau+(1-\tau)(1-\frac{\delta_R}{q})), & \text{if } \delta_3 < \delta_R \le \delta_2 \text{ and } c_R \le \frac{q\left(\tau+\sqrt{\tau}-2\right)^2}{\tau^2-5\tau+4} \\ \frac{(q-c_R(1-\tau))^2}{4q(1-\tau)}, & \text{if } \delta_2 < \delta_R \le \delta_1 \text{ and } \tau \le \frac{1}{2} \text{ and } c_R \le \frac{q\left(\frac{2\tau^2-7\tau+4}{\tau-1}+4\sqrt{(1-\tau)\tau}\right)}{5\tau-4} \\ \tau\frac{(q+\delta_R-c_R)^2}{4q}, & \text{otherwise} \end{cases},$$

where

$$\delta_1 = \begin{cases} \frac{q-c_R}{\sqrt{\tau}}, & \text{if } c_R > \frac{q\left(\tau+\sqrt{\tau}-2\right)^2}{\tau^2-5\tau+4} \\ c_R + \frac{q-c_R(1-\tau)}{\sqrt{\tau(1-\tau)}} - q, & \text{if } \tau < 1/2 \text{ and } c_R < \frac{q\left(\frac{2\tau^2-7\tau+4}{\tau-1}+4\sqrt{(1-\tau)\tau}\right)}{5\tau-4} \\ \frac{2\sqrt{(\tau-1)\left(c_R^2(\tau-1)-c_Rq(\tau-2)+q^2(\tau-1)\right)}-\tau(c_R+q)+2(c_R+q)}{4-3\tau}, & \text{otherwise} \end{cases}, \tag{10a}$$

$$\delta_2 = \begin{cases} \frac{q+c_R(1-\tau)}{2(1-\tau)}, & \text{if } \tau < 1/2 \text{ and } c_R < \frac{q\left(\frac{2\tau^2-7\tau+4}{\tau-1}+4\sqrt{(1-\tau)\tau}\right)}{5\tau-4} \\ \frac{2\sqrt{(\tau-1)\left(c_R^2(\tau-1)-c_Rq(\tau-2)+q^2(\tau-1)\right)}-\tau(c_R+q)+2(c_R+q)}{4-3\tau}, & \text{otherwise} \end{cases}, \quad \text{and} \tag{10b}$$

$$\delta_3 = \min\{\frac{q+c_R}{2-\tau}, \frac{q-c_R}{\sqrt{\tau}}\} = \begin{cases} \frac{q+c_R}{2-\tau}, & \text{if } c_R < \frac{q\left(\tau+\sqrt{\tau}-2\right)^2}{\tau^2-5\tau+4} \\ \frac{q-c_R}{\sqrt{\tau}}, & \text{if } c_R > \frac{q\left(\tau+\sqrt{\tau}-2\right)^2}{\tau^2-5\tau+4} \end{cases}. \tag{10c}$$

This completes the proof of Proposition 1. □

Poof of Proposition 2. Taking derivative of $\hat{p}_N^*$ with respect to τ and δ_N, we can show that:

$$\frac{\partial \hat{p}_N^*}{\partial \tau} = \begin{cases} \frac{-\delta_N}{2}, & \text{if } \delta_N < \frac{(q-c_N)(1-\sqrt{1-\tau})}{\tau} \\ 0, & \text{otherwise} \end{cases} \leq 0, \text{ and}$$

$$\frac{\partial \hat{p}_N^*}{\partial \delta_N} = \begin{cases} \frac{-\tau}{2}, & \text{if } \delta_N < \frac{(q-c_N)(1-\sqrt{1-\tau})}{\tau} \\ 0, & \text{otherwise} \end{cases} \leq 0.$$

Taking derivative of $\hat{\pi}_N^*$ with respect to τ and δ_N, we can show that:

$$\frac{\partial \hat{\pi}_N^*}{\partial \tau} = \begin{cases} -\frac{2(q-c_N-\delta_N\tau)\delta_N}{4q}, & \text{if } \delta_N < \frac{(q-c_N)(1-\sqrt{1-\tau})}{\tau} \\ 0, & \text{otherwise} \end{cases} \leq 0, \text{ and}$$

$$\frac{\partial \hat{\pi}_N^*}{\partial \delta_N} = \begin{cases} -\frac{2(q-c_N-\delta_N\tau)\tau}{4q}, & \text{if } \delta_N < \frac{(q-c_N)(1-\sqrt{1-\tau})}{\tau} \\ 0, & \text{otherwise} \end{cases} \leq 0.$$

Taking derivative of $\hat{p}_R^*$ with respect to τ and δ_R, we can show that:

$$\frac{\partial \hat{p}_R^*}{\partial \tau} = \begin{cases} \frac{\delta_R}{2}, & \text{if } \delta_R \leq \delta_3 \\ 0, & \text{if } \delta_3 < \delta_R \leq \delta_2 \text{ and } c_R \leq \frac{q\left(\tau+\sqrt{\tau}-2\right)^2}{\tau^2-5\tau+4} \\ \frac{q}{2(1-\tau)^2}, & \text{if } \delta_2 < \delta_R \leq \delta_1 \text{ and } \tau \leq 1/2 \text{ and } c_R \leq \frac{q\left(\frac{2\tau^2-7\tau+4}{\tau-1}+4\sqrt{(1-\tau)\tau}\right)}{5\tau-4} \\ 0, & \text{otherwise} \end{cases} \geq 0, \text{ and}$$

$$\frac{\partial \hat{p}_R^*}{\partial \delta_R} = \begin{cases} \frac{\tau}{2}, & \text{if } \delta_R \leq \delta_3 \\ 1, & \text{if } \delta_3 < \delta_R \leq \delta_2 \text{ and } c_R \leq \frac{q\left(\tau+\sqrt{\tau}-2\right)^2}{\tau^2-5\tau+4} \\ 0, & \text{if } \delta_2 < \delta_R \leq \delta_1 \text{ and } \tau \leq 1/2 \text{ and } c_R \leq \frac{q\left(\frac{2\tau^2-7\tau+4}{\tau-1}+4\sqrt{(1-\tau)\tau}\right)}{5\tau-4} \\ \frac{1}{2}, & \text{otherwise} \end{cases} \geq 0.$$

Taking derivative of $\hat{\pi}_R^*$ with respect to τ and δ_R, we can show that:

$$\frac{\partial \hat{\pi}_R^*}{\partial \tau} = \begin{cases} \frac{\delta_R(q-c_R+\delta_R\tau)}{2q}, & \text{if } \delta_R \leq \delta_3 \\ \frac{\delta_R(\delta_R-c_R)}{q}, & \text{if } \delta_3 < \delta_R \leq \delta_2 \text{ and } c_R \leq \frac{q\left(\tau+\sqrt{\tau}-2\right)^2}{\tau^2-5\tau+4} \\ \frac{(q^2-c_R^2(1-\tau)^2)}{4q(1-\tau)^2}, & \text{if } \delta_2 < \delta_R \leq \delta_1 \text{ and } \tau \leq 1/2 \text{ and } c_R \leq \frac{q\left(\frac{2\tau^2-7\tau+4}{\tau-1}+4\sqrt{(1-\tau)\tau}\right)}{5\tau-4} \\ \frac{(q+\delta_R-c_R)^2}{4q}, & \text{otherwise} \end{cases} \geq 0, \text{ and}$$

$$\frac{\partial \hat{\pi}_R^*}{\partial \delta_R} = \begin{cases} \frac{\tau(q-c_R+\delta_R\tau)}{2q}, & \text{if } \delta_R \leq \delta_3 \\ \frac{q+c_R(1-\tau)-2\delta_R(1-\tau)}{q}, & \text{if } \delta_3 < \delta_R \leq \delta_2 \text{ and } c_R \leq \frac{q\left(\tau+\sqrt{\tau}-2\right)^2}{\tau^2-5\tau+4} \\ 0, & \text{if } \delta_2 < \delta_R \leq \delta_1 \text{ and } \tau \leq 1/2 \text{ and } c_R \leq \frac{q\left(\frac{2\tau^2-7\tau+4}{\tau-1}+4\sqrt{(1-\tau)\tau}\right)}{5\tau-4} \\ \tau\frac{(q+\delta_R-c_R)}{2q}, & \text{otherwise} \end{cases} \geq 0.$$

This completes the proof of Proposition 2. $\square$

B. Asymmetric Information

Separating Equilibrium

Proof of Lemma 1. We prove Lemma 1 in three steps. Step 1, we show that in any separating equilibrium, the N-type firm chooses price $p^*_{N,sep} = \hat{p}^*_N$. Step 2, we identify the price set Φ to which the R-type firm's separating equilibrium price $p_{R,sep}$ must belong. Step 3, among all possible separating equilibria, we find one that is most profitable (least-cost) to the R-type firm and provide a supporting belief system, μ^*_{sep}.

Step 1. Note that under asymmetric information, the N-type firm has an incentive to hide its type. In any separating equilibrium, the consumers will correctly infer the firm's type. Thus, in any separating equilibrium, the N-type firm should choose its first-best strategy $p^*_{N,sep} = \hat{p}^*_N$ (see Equation (3a)) and make a profit $\pi^*_{N,sep} = \hat{\pi}^*_N$ (see Equation (3b)) as it does under symmetric information.

Step 2. In any separating equilibrium, the N-type firm must not be able to profitably deviate to the R-type firm's strategy $p_{R,sep}$; that is,

$$\hat{\pi}^*_N \geq \pi_N(p_{R,sep}|\mu(R|p_{R,sep}, S) = 1), \tag{11}$$

where $\pi_N(p|\mu(R|p,S))$ represents the N-type firm's profit when it chooses price p and is believed to be R-type with probability $\mu(R|p,S)$ and $\pi_N(p|\mu(R|p,S) = 1)$ is given by

$$\pi_N(p|\mu(R|p,S) = 1) = \begin{cases} (p - c_N)(\tau + (1-\tau)(1 - \frac{p}{q})), & \text{if } p \leq \delta_R \\ (p - c_N)(\tau(1 - \frac{p-\delta_R}{q}) + (1-\tau)(1 - \frac{p}{q})), & \text{if } \delta_R < p \leq q \\ (p - c_N)\tau(1 - \frac{p-\delta_R}{q}), & \text{if } q < p \leq q + \delta_R \\ 0, & \text{if } p > q + \delta_R \end{cases}.$$

Let $\Phi = \{p : \hat{\pi}^*_N \geq \pi_N(p|\mu(R|p,S) = 1)\}$ denote the set of prices that are equilibrium-dominated for the N-type firm regardless of the consumers' belief (even under the most favorable belief that the firm is R-type with probability one). In any separating equilibrium, the condition for no profitable deviation by the N-type firm requires $p_{R,sep} \in \Phi$. One can show that $\hat{\pi}_N^* \geq \pi_N(p|\mu(R|p,S)=1)$ iff $p \geq \bar{p}$ or $p \leq \underline{p}$, where

$$\underline{p} = \begin{cases} \frac{1}{2}\left(c_N + \delta_R\tau + q - \sqrt{\tau(\delta_N+\delta_R)(2q-2c_N-\delta_N\tau+\delta_R\tau)}\right), & \text{if } \delta_N < \min\{\frac{(q-c_N)(1-\sqrt{1-\tau})}{\tau}, \hat{\delta}_N\} \\ \frac{q+c_N(1-\tau)-\sqrt{\tau(c_N^2(\tau-1)+2c_N\delta_N(\tau-1)+\delta_N^2(\tau-1)\tau+q^2-2\delta_N q(\tau-1))}}{2(1-\tau)}, & \text{if } \hat{\delta}_N \leq \delta_N < \frac{(q-c_N)(1-\sqrt{1-\tau})}{\tau} \\ \frac{1}{2}\left(c_N + \delta_R\tau + q - \sqrt{\tau(c_N^2+q^2+2q\delta_R-2c_N(q+\delta_R)+\delta_R^2\tau)}\right), & \text{if } \delta_N \geq \frac{(q-c_N)(1-\sqrt{1-\tau})}{\tau} \text{ and } \tau \leq \hat{\tau} \\ \frac{q+c_N(1-\tau)-\sqrt{\tau q(q(2-\tau)-2c_N(1-\tau))}}{2(1-\tau)}, & \text{if } \delta_N \geq \frac{(q-c_N)(1-\sqrt{1-\tau})}{\tau} \text{ and } \tau > \hat{\tau} \end{cases}, \tag{12a}$$

$$\bar{p} = \begin{cases} \bar{p}_1, & \text{if } \delta_N < \min\{\frac{(q-c_N)(1-\sqrt{1-\tau})}{\tau}, \frac{q-c_N-2\sqrt{\tau\delta_R(q-c_N)}}{\tau}\} \\ \bar{p}_2, & \text{if } \frac{q-c_N-2\sqrt{\tau\delta_R(q-c_N)}}{\tau} \leq \delta_N < \frac{(q-c_N)(1-\sqrt{1-\tau})}{\tau} \\ \bar{p}_3, & \text{if } \delta_N \geq \frac{(q-c_N)(1-\sqrt{1-\tau})}{\tau} \text{ and } \delta_R < \frac{(q-c_N)(1-\tau)}{4\tau} \\ \bar{p}_4, & \text{if } \delta_N \geq \frac{(q-c_N)(1-\sqrt{1-\tau})}{\tau} \text{ and } \delta_R \geq \frac{(q-c_N)(1-\tau)}{4\tau} \end{cases}, \tag{12b}$$

$$\hat{\delta}_N = \begin{cases} \frac{q-c_N-2\sqrt{(\delta_R-c_N)(q-(1-\tau)\delta_R)}}{\tau}, & \text{if } c_N < \delta_R \\ \frac{q-\delta_R}{\tau}, & \text{if } c_N \geq \delta_R \end{cases}, \tag{12c}$$

$$\hat{\tau} = \begin{cases} \frac{c_N^2+2c_N q+q^2-4c_N\delta_R-4q\delta_R+4\delta_R^2}{c_N^2+q^2-2c_N q-4c_N\delta_R+4\delta_R^2}, & \text{if } c_N < \delta_R \\ 1, & \text{if } c_N \geq \delta_R \end{cases}, \tag{12d}$$

$$\bar{p}_1 = \frac{1}{2}\left(c_N + \delta_R\tau + q + \sqrt{\tau(\delta_N+\delta_R)(2q-2c_N-\delta_N\tau+\delta_R\tau)}\right), \tag{12e}$$

$$\bar{p}_2 = \left(\sqrt{\frac{c_N^2(\tau-1) - 2c_N(\tau(\delta_N+\delta_R)+q(\tau-1)) + \tau\left(\delta_R^2-\delta_N^2\tau\right) + q^2(\tau-1) + 2q\tau(\delta_N+\delta_R)}{\tau}} + c_N + \delta_R + q\right)/2, \tag{12f}$$

$$\bar{p}_3 = (q + c_N + \tau\delta_R + \sqrt{\tau(q^2+c_N^2+2q\delta_R+\tau\delta_R^2-2c_N(q+\delta_R))})/2, \text{ and} \tag{12g}$$

$$\bar{p}_4 = \left(\sqrt{\frac{2q\tau\delta_R+\delta_R^2\tau+c_N^2(2\tau-1)+q^2(2\tau-1)-2c_N(\delta_R\tau+q(2\tau-1))}{\tau}} + c_N + \delta_R + q\right)/2. \tag{12h}$$

Thus, we have identified the price set $\Phi = \{p : p \geq \bar{p}, \text{ or } p \leq \underline{p}\}$ to which the R-type firm's separating equilibrium price $p_{R,sep}$ must belong.

Step 3. Let $p^*_{R,sep} \in \Phi$ be the R-type firm's price from the most profitable (or least-cost) separating equilibrium, i.e., $p^*_{R,sep} = arg\max_{p\in\Phi} \pi_{R,sep}(p)$, where $\pi_{R,sep}(p)$ is the R-type firm's profit in a separating equilibrium with price p and it is the same as the profit under the symmetric information, i.e., $\pi_{R,sep}(p) = \hat{\pi}_R(p)$ (see Equation (2b)). Solving the optimization problem $\max_{p\in\Phi} \pi_{R,sep}(p)$ leads to:

$$p^*_{R,sep} = \begin{cases} \hat{p}_R^*, & \text{if } c_N \leq \bar{c}_N \\ \tilde{p}_{R,sep}, & \text{otherwise} \end{cases},$$

where

$$\bar{c}_N = \max\{\bar{c}_{N,1}, \bar{c}_{N,2}\}, \tag{13a}$$

$$\bar{c}_{N,1} = c_R - \tau(\delta_N + \delta_R) - \sqrt{2\tau(\delta_R + \delta_N)(q + \delta_R\tau - c_R)}, \tag{13b}$$

$$\bar{c}_{N,2} = \frac{c_R - \sqrt{\tau\left(c_R^2 - 2c_R(\delta_R + q) - \delta_R^2(\tau - 2)\tau + q^2 + 2\delta_R q\right)} - \delta_R\tau - q\tau}{1 - \tau}, \tag{13c}$$

$$\tilde{p}_{R,sep} = \begin{cases} \bar{p}_1, & \text{if } \bar{c}_N < c_N \le \hat{\bar{c}}_N \text{ and } \delta_N < \frac{(q - c_N)(1 - \sqrt{1-\tau})}{\tau} \\ \bar{p}_2, & \text{if } c_N > \hat{\bar{c}}_N \text{ and } \delta_N < \frac{(q - c_N)(1 - \sqrt{1-\tau})}{\tau} \\ \bar{p}_3, & \text{if } \bar{c}_N < c_N \le \hat{\bar{c}}_N \text{ and } \delta_N \ge \frac{(q - c_N)(1 - \sqrt{1-\tau})}{\tau} \\ \bar{p}_4, & \text{if } c_N > \tilde{c}_N \text{ and } \delta_N \ge \frac{(q - c_N)(1 - \sqrt{1-\tau})}{\tau} \end{cases}, \tag{13d}$$

$$\hat{\bar{c}}_N = \max\{\hat{\bar{c}}_{N,1}, \hat{\bar{c}}_{N,2}\}, \tag{13e}$$

$$\hat{\bar{c}}_{N,1} = q - 2\tau\sqrt{\delta_R(\delta_N + \delta_R)} - \delta_N\tau - 2\delta_R\tau, \tag{13f}$$

$$\hat{\bar{c}}_{N,2} = \frac{q(1 - \tau) - 4\tau\delta_R}{1 - \tau}, \tag{13g}$$

and $\bar{p}_1$, $\bar{p}_2$, $\bar{p}_3$, and $\bar{p}_4$ are given in Equation (12e) to (12h)

The R-type firm's corresponding profit in the most profitable separating equilibrium is given by:

$$\pi^*_{R,sep} = \begin{cases} \hat{\pi}^*_R, & \text{if } c_N \le \bar{c}_N \\ \tilde{\pi}_{R,sep}, & \text{otherwise} \end{cases}, \tag{14}$$

where

$$\tilde{\pi}_{R,sep} = \begin{cases} (\tilde{p}_{R,sep,1} - c_R)(\tau(1 - \frac{\tilde{p}_{R,sep,1} - \delta_R}{q}) + (1 - \tau)(1 - \frac{\tilde{p}_{R,sep,1}}{q})), & \text{if } \bar{c}_N < c_N \le \hat{\bar{c}}_N \\ (\tilde{p}_{R,sep,1} - c_R)\tau(1 - \frac{\tilde{p}_{R,sep,1} - \delta_R}{q}), & \text{if } c_N > \hat{\bar{c}}_N \end{cases}. \tag{15}$$

Many belief systems can support the above separating equilibrium and one example is as follows:

$$\mu^*_{sep}(R|p, S) = \begin{cases} 1, & \text{if } p = p^*_{R,sep} \\ 0, & \text{otherwise} \end{cases}. \tag{16}$$

This completes the proof of Lemma 1. □

Proof of Lemma 2. Taking derivative of $p^*_{R,sep}$ with respect to τ, δ_N and δ_R and after some simplifications, we can show that $\frac{\partial p^*_{R,sep}}{\partial \tau} \ge 0$, $\frac{\partial p^*_{R,sep}}{\partial \delta_R} \ge 0$, and $\frac{\partial p^*_{R,sep}}{\partial \delta_N} \ge 0$. Taking derivative of $\pi^*_{R,sep}$ with respect to τ, δ_N and δ_R and after some simplifications, we can show that $\frac{\partial \pi^*_{R,sep}}{\partial \delta_R} \ge 0$, $\frac{\partial \pi^*_{R,sep}}{\partial \delta_N} \le 0$, and $\frac{\partial \pi^*_{R,sep}}{\partial \tau}$ could be either ≤ 0 or ≥ 0. We omit the details for conciseness.

On the other hand, in the separating equilibrium, the N-type firm's price and profit are the same as those under symmetric information. Thus, the proof of Lemma 2(b) follows directly from the proof of Proposition 2(a). □

Pooling Equilibrium

Proof of Lemma 3. For a given signal S, in any pooling equilibrium, the consumers' posterior belief is the same as γ_S, i.e., $\mu(R|p,S) = \gamma_S$. An N-type consumer's utility in a pooling equilibrium is $u_N = \theta q - p$. A R-type consumer's utility in a pooling equilibrium is given by $u_R = \gamma_S u_{RR} + (1-\gamma_S)u_{RN} = \theta q - p + \bar{\delta}_S$, where $\bar{\delta}_S = \gamma_S \delta_R - (1-\gamma_S)\delta_N$. Thus, in a pooling equilibrium, there are two cases to consider:

(a) if $\bar{\delta}_S \geq 0$, each type of firm's demand is given by:

$$d_{pool,S}(p) = \begin{cases} \tau + (1-\tau)(1-\frac{p}{q}), & \text{if } p \leq \bar{\delta}_S \\ \tau(1-\frac{p-\bar{\delta}_S}{q}) + (1-\tau)(1-\frac{p}{q}), & \text{if } \bar{\delta}_S < p \leq q \\ \tau(1-\frac{p-\bar{\delta}_S}{q}), & \text{if } q < p \leq q + \bar{\delta}_S \\ 0, & \text{if } p > q + \bar{\delta}_S \end{cases};$$

(b) if $\bar{\delta}_S < 0$, each type of firm's demand is given by:

$$d_{pool,S}(p) = \begin{cases} \tau(1-\frac{p-\bar{\delta}_S}{q}) + (1-\tau)(1-\frac{p}{q}), & \text{if } p \leq q + \bar{\delta}_S \\ (1-\tau)(1-\frac{p}{q}), & \text{if } q + \bar{\delta}_S < p \leq q \\ 0, & \text{if } p > q \end{cases}.$$

Therefore, in a pooling equilibrium, a j-type firm's profit is given by $\pi_{j,pool,S}(p) = (p - c_j)d_{pool,S}(p)$, for $j \in \{R, N\}$.

Next, among all pooling outcomes, we identify the most profitable pooling equilibrium for the R-type firm. The price $p^*_{pool,S}$ that maximizes $\pi_{R,pool,S}$ is given as below:

(a) if $\bar{\delta}_S \geq 0$,

$$p^*_{pool,S} = p_H(\bar{\delta}_S) = \begin{cases} \frac{q+c_R+\bar{\delta}_S\tau}{2}, & \text{if } \bar{\delta}_S \leq \delta_3 \\ \bar{\delta}_S, & \text{if } \delta_3 < \delta_R \leq \delta_2 \text{ and } c_R \leq \frac{q\left(\tau+\sqrt{\tau}-2\right)^2}{\tau^2-5\tau+4} \\ \frac{q+c_R-c_R\tau}{2(1-\tau)}, & \text{if } \delta_2 < \bar{\delta}_S \leq \delta_1 \text{ and } \tau < 1/2 \text{ and } c_R < \frac{q\left(\frac{2\tau^2-7\tau+4}{\tau-1}+4\sqrt{(1-\tau)\tau}\right)}{5\tau-4} \\ \frac{q+c_R+\bar{\delta}_S}{2}, & \text{otherwise} \end{cases},$$

where δ_1, δ_2, and δ_3 are given in Equations (10);

(b) if $\bar{\delta}_S < 0$,

$$p^*_{pool,S} = p_L(\bar{\delta}_S) = \begin{cases} \frac{q+c_R+\bar{\delta}_S\tau}{2}, & \text{if } -\frac{(q-c_R)(1-\sqrt{1-\tau})}{\tau} < \bar{\delta}_S \leq 0 \\ \frac{q+c_R}{2}, & \text{if } \bar{\delta}_S \leq -\frac{(q-c_R)(1-\sqrt{1-\tau})}{\tau} \end{cases}.$$

In the R-type firm's most profitable pooling equilibrium, the R-type and N-type firm's corresponding profits are given as below:

$$\pi^*_{R,pool,S} = \begin{cases} \pi_{R,L}(\bar{\delta}_S), & \text{if } \bar{\delta}_S < 0 \\ \pi_{R,H}(\bar{\delta}_S), & \text{if } \bar{\delta}_S \geq 0 \end{cases}, \tag{17a}$$

$$\pi^*_{N,pool,S} = \begin{cases} \pi_{N,L}(\bar{\delta}_S), & \text{if } \bar{\delta}_S < 0 \\ \pi_{N,H}(\bar{\delta}_S), & \text{if } \bar{\delta}_S \geq 0 \end{cases}, \tag{17b}$$

where

$$\pi_{R,H}(\bar{\delta}_S) = \begin{cases} \frac{(q-c_R+\bar{\delta}_S\tau)^2}{4q}, & \text{if } \bar{\delta}_S \leq \delta_3 \\ (\bar{\delta}_S - c_R)(\tau + (1-\tau)(1-\frac{\bar{\delta}_S}{q})), & \text{if } \delta_3 < \bar{\delta}_S \leq \delta_2 \text{ and } c_R \leq \frac{q\left(\tau+\sqrt{\tau}-2\right)^2}{\tau^2-5\tau+4} \\ \frac{(q-c_R(1-\tau))^2}{4q(1-\tau)}, & \text{if } \delta_2 < \bar{\delta}_S \leq \delta_1 \text{ and } \tau \leq 1/2 \text{ and } c_R \leq \frac{q\left(\frac{2\tau^2-7\tau+4}{\tau-1}+4\sqrt{(1-\tau)\tau}\right)}{5\tau-4} \\ \tau\frac{(q+\bar{\delta}_S-c_R)^2}{4q}, & \text{otherwise} \end{cases},$$

$$\pi_{N,H}(\bar{\delta}_S) = \begin{cases} \frac{(-c_R+\bar{\delta}_S\tau+q)(c_R-2c_N+\bar{\delta}_S\tau+q)}{4q}, & \text{if } \bar{\delta}_S \leq \delta_3 \\ (\bar{\delta}_S - c_N)(\tau + (1-\tau)(1-\frac{\bar{\delta}_S}{q})), & \text{if } \delta_3 < \bar{\delta}_S \leq \delta_2 \text{ and } c_R \leq \frac{q\left(\tau+\sqrt{\tau}-2\right)^2}{\tau^2-5\tau+4} \\ \frac{(c_R(\tau-1)+q)(2c_N(\tau-1)+c_R(1-\tau)+q)}{4q(1-\tau)}, & \text{if } \delta_2 < \bar{\delta}_S \leq \delta_1 \text{ and } \tau \leq 1/2 \text{ and } c_R \leq \frac{q\left(\frac{2\tau^2-7\tau+4}{\tau-1}+4\sqrt{(1-\tau)\tau}\right)}{5\tau-4} \\ \frac{\tau(-c_R+\bar{\delta}_S+q)(-2c_N+c_R+\bar{\delta}_S+q)}{4q}, & \text{otherwise} \end{cases},$$

$$\pi_{R,L}(\bar{\delta}_S) = \begin{cases} \frac{(q-c_R+\bar{\delta}_S\tau)^2}{4q}, & \text{if } -\frac{(q-c_R)(1-\sqrt{1-\tau})}{\tau} < \bar{\delta}_S \leq 0 \\ (1-\tau)\frac{(q-c_R)^2}{4q}, & \text{if } \bar{\delta}_S \leq -\frac{(q-c_R)(1-\sqrt{1-\tau})}{\tau} \end{cases}, \text{ and}$$

$$\pi_{N,L}(\bar{\delta}_S) = \begin{cases} \frac{(q+c_R+\bar{\delta}_S\tau-2c_N)(q-c_R+\bar{\delta}\tau)}{4q}, & \text{if } -\frac{(q-c_R)(1-\sqrt{1-\tau})}{\tau} < \bar{\delta}_S \leq 0 \\ (1-\tau)\frac{(q-c_R-2c_N)(q-c_R)}{4q}, & \text{if } \bar{\delta}_S \leq -\frac{(q-c_R)(1-\sqrt{1-\tau})}{\tau} \end{cases}.$$

At last, we provide a supporting belief system. Many belief systems can support the above pooling equilibrium and one example is as follows:

$$\mu^*_{pool}(R|p,S) = \begin{cases} \gamma_S, & \text{if } p = p^*_{pool,S} \\ 0, & \text{otherwise} \end{cases}. \tag{18}$$

This completes the proof of Lemma 3. □

Proof of Lemma 4. We prove Lemma 4 for $\bar{\delta}_S \geq 0$ and $\bar{\delta}_S < 0$, respectively.

(a) $\bar{\delta}_S \geq 0$: Taking derivative of $p^*_{pool,S}$ with respect to τ, γ_S, δ_R, and δ_N, we can show that $\frac{\partial p^*_{pool,S}}{\partial \tau} \geq 0$, $\frac{\partial p^*_{pool,S}}{\partial \gamma_S} \geq 0$, $\frac{\partial p^*_{pool,S}}{\partial \delta_R} \geq 0$, and $\frac{\partial p^*_{pool,S}}{\partial \delta_N} \leq 0$; taking derivative of $\pi^*_{R,pool,S}$ with respect to τ, γ_S, δ_R, and δ_N, we can show that $\frac{\partial \pi^*_{R,pool,S}}{\partial \tau} \geq 0$, $\frac{\partial \pi^*_{R,pool,S}}{\partial \gamma_S} \geq 0$, $\frac{\partial \pi^*_{R,pool,S}}{\partial \delta_R} \geq 0$, and $\frac{\partial \pi^*_{R,pool,S}}{\partial \delta_N} \leq 0$; taking derivative of $\pi^*_{N,pool,S}$ with respect to τ, γ_S, δ_R, and δ_N, we can show that $\frac{\partial \pi^*_{N,pool,S}}{\partial \tau} \geq 0$, $\frac{\partial \pi^*_{N,pool,S}}{\partial \gamma_S} \geq 0$, $\frac{\partial \pi^*_{N,pool,S}}{\partial \delta_R} \geq 0$, and $\frac{\partial \pi^*_{N,pool,S}}{\partial \delta_N} \leq 0$. The detailed expressions can be obtained via cumbersome calculations, which we omit for brevity.

(b) $\bar{\delta}_S < 0$: Taking derivative of $p^*_{pool,S}$ with respect to τ, γ_S, δ_R, and δ_N, we can show that $\frac{\partial p^*_{pool,S}}{\partial \tau} \leq 0$, $\frac{\partial p^*_{pool,S}}{\partial \gamma_S} \geq 0$, $\frac{\partial p^*_{pool,S}}{\partial \delta_R} \geq 0$, and $\frac{\partial p^*_{pool,S}}{\partial \delta_N} \leq 0$; taking derivative of $\pi^*_{R,pool,S}$ with respect to τ, γ_S, δ_R, and δ_N, we can show that $\frac{\partial \pi^*_{R,pool,S}}{\partial \tau} \leq 0$, $\frac{\partial \pi^*_{R,pool,S}}{\partial \gamma_S} \geq 0$, $\frac{\partial \pi^*_{R,pool,S}}{\partial \delta_R} \geq 0$, and $\frac{\partial \pi^*_{R,pool,S}}{\partial \delta_N} \leq 0$; taking derivative of $\pi^*_{N,pool,S}$ with respect to τ, γ_S, δ_R, and δ_N, we can show that $\frac{\partial \pi^*_{N,pool,S}}{\partial \tau} \leq 0$, $\frac{\partial \pi^*_{N,pool,S}}{\partial \gamma_S} \geq 0$, $\frac{\partial \pi^*_{N,pool,S}}{\partial \delta_R} \geq 0$, and $\frac{\partial \pi^*_{N,pool,S}}{\partial \delta_N} \leq 0$. The detailed expressions can be obtained via cumbersome calculations, which we omit for brevity.

This completes the proof of Lemma 4. □

Equilibrium Outcome

Definition of LMSE. There are two types of perfect Bayesian equilibrium (PBE) in our setting: separating and pooling. Since the PBE concept does not impose any restrictions on the out-of-equilibrium beliefs, there may exist multiple PBE. We adopt the *Lexicographically Maximum Sequential Equilibrium* (LMSE), which is proposed and formalized by Mailath et al. (1993), to refine the equilibrium outcomes. We provide the definition of LMSE for our setting below.

Definition. (Lexicographically Maximum Sequential Equilibrium) In a signaling game G, we denote the set of pure-strategy perfect Bayesian equilibria by $PBE(G)$. We denote the set of types by $\{R, N\}$ and the j-type player's payoff by $\pi_j(\sigma)$ under strategy profile $\sigma \in PBE(G)$. The strategy profile $\sigma \in PBE(G)$ *lexicographically dominates (l-dominates)* $\sigma' \in PBE(G)$ if $\pi_R(\sigma) > \pi_R(\sigma')$, or $\pi_R(\sigma) = \pi_R(\sigma')$ and $\pi_N(\sigma) > \pi_N(\sigma')$. The profile $\sigma \in PBE(G)$ is an LMSE if there does not exist $\sigma' \in PSE(G)$ that $l-dominates$ σ.

The LMSE refinement chooses the unique, most-efficient outcome from the perspective of the type that has the most incentives to reveal its identity, i.e., the R-type firm in our setting. In our model, a PBE is an LMSE if it is the R-type firm's most profitable outcome among all PBE. If there are multiple most profitable outcomes for the R-type firm, a PBE is also the N-type firm's most profitable outcome.

Lemma 1 and Lemma 3 have characterized the most profitable (least-cost) separating and most profitable (most-efficient) pooling equilibrium outcomes for the R-type firm, respectively. We can use the LMSE refinement to show that other separating and pooling equilibrium outcomes will be $l-domianted$ by the most profitable separating and pooling equilibrium outcomes; therefore other equilibria cannot be LMSE. In the following proof of Proposition 3, we use the LMSE refinement to show the unique equilibrium outcome.

Proof of Proposition 3. Let $\sigma^*_{sep} = (p^*_{R,sep}, p^*_{N,sep}, \mu^*_{sep})$ denote the least-cost separating equilibrium as characterized in Lemma 1 with the belief μ^*_{sep} given in Equation (16), and $\sigma^*_{pool} = (p^*_{pool,S}, \mu^*_{pool})$ denote the most-efficient pooling equilibrium as characterized in Lemma 3 with the belief μ^*_{pool} given in (18).

We complete the proof of Proposition 3 in five steps. Step 1, we define γ^*. Step 2, we show that when $\gamma_S < \gamma^*$, the least-cost separating equilibrium σ^*_{sep} exists. Step 3, we show that when $\gamma_S < \gamma^*$, σ^*_{sep} is the unique LMSE. Step 4, we show that when $\gamma_S > \gamma^*$, σ^*_{pool} exists. Step 5, we show that when $\gamma_S > \gamma^*$, σ^*_{pool} is the unique LMSE.

Step 1. One can easily verify that the R-type firm's profit in the least-cost separating equilibrium, $\pi^*_{R,sep}$ (see Equation (14)), is independent in γ_S, while the R-type firm's profit in the most-efficient pooling equilibrium, $\pi^*_{R,pool,S}$ (see Equation (17a)) is strictly increasing in γ_S or first independent of γ_S and then strictly increasing in γ_S. Moreover, one can verify that when $\gamma_S = 1$, $\pi^*_{R,pool} \geq \pi^*_{R,sep}$; when $\gamma_S = 0$, $\pi^*_{R,pool} < \pi^*_{R,sep}$. Thus, there exists a unique $\gamma^* \in [0,1]$ such that $\pi^*_{R,pool,S} > \pi^*_{R,sep}$ when $\gamma_S > \gamma^*$ and $\pi^*_{R,pool,S} < \pi^*_{R,sep}$ when $\gamma_S < \gamma^*$. Moreover, one can also verify that when $\gamma_S > \gamma^*$, then $\pi^*_{N,pool,S} > \pi^*_{N,sep}$.

Step 2. To show that the least-cost separating equilibrium σ^*_{sep} exists when $\gamma_S < \gamma^*$, we just need to show the following:

(1) Given the belief system $\mu^*_{sep}(R|p,S)$, the R-type firm cannot be strictly better off by deviating from strategy $p^*_{R,sep}$ to any other strategy p, i.e.,

$$\pi^*_{R,sep} \geq \max_{p \neq p^*_{R,sep}} \pi_R(p|\mu^*_{sep}(R|p,S)). \tag{19}$$

By the definition of γ^* in Step 1, we have $\pi^*_{R,sep} > \pi^*_{R,pool,S}$ when $\gamma_S < \gamma^*$. By the definition of $\pi^*_{R,pool,S}$ in Lemma 3, $\pi^*_{R,pool,S} = \max_p \pi_R(p|\mu(R|p,S) = \gamma_S)$. Since $\pi_R(p|\mu(R|p,S))$ increases in $\mu(R|p,S)$, we have $\pi^*_{R,sep} > \pi^*_{R,pool,S} = \max_p \pi_R(p|\mu(R|p,S) = \gamma_S) \geq \max_p \pi_R(p|\mu(R|p,S) = 0) \geq \max_{p \neq p^*_{R,sep}} \pi_R(p|\mu^*_{sep}(R|p,S))$. Thus, Equation (19) holds.

(2) Given the belief system $\mu^*_{sep}(R|p,S)$, the N-type firm cannot be strictly better off by deviating from strategy $p^*_{N,sep}$ to any other strategy p, i.e.,

$$\pi^*_{N,sep} \geq \max_{p \neq p^*_{N,sep}} \pi_N(p|\mu^*_{sep}(R|p,S)). \tag{20}$$

By the definition of Φ and $p^*_{R,sep}$ in the proof of Lemma 1, $\pi^*_{N,sep} \geq \pi_N(p^*_{R,sep}|\mu^*_{sep}(R|p^*_{R,sep},S) = 1)$ since $p^*_{R,sep} \in \Phi$. For any $p \neq p^*_{R,sep}$, we have $\pi^*_{N,sep} = \hat{\pi}^*_N = \max_p \pi_N(p|\mu(R|p,S) = 0) \geq \max_{p \neq p^*_{N,sep}} \pi_N(p|\mu^*_{sep}(R|p,S) = 0)$. Thus, Equation (20) holds.

Therefore, when $\gamma_S < \gamma^*$, the least-cost separating equilibrium σ^*_{sep} exists.

Step 3. In this step, we show that when $\gamma_S < \gamma^*$, the least-cost separating equilibrium σ^*_{sep} is the unique LMSE. Based on the definition of LMSE in our setting, a PBE is LMSE iff it is the R-type firm's most profitable outcome among all PBE, and, conditional on it being the most profitable for the R-type firm, it is also the N-type firm's most profitable outcome.

Given the definition of γ^* in Step 1, and the definitions of the least-cost separating equilibrium σ^*_{sep} and the most-efficient pooling equilibrium σ^*_{pool} in Lemma 1 and Lemma 3, respectively, when $\gamma_S < \gamma^*$, the least-cost separating equilibrium is the unique most profitable equilibrium outcome for the R-type firm and thus it is the unique LMSE.

Step 4. To show that the most-efficient pooling equilibrium σ^*_{pool} exists when $\gamma_S > \gamma^*$, we just need to show the following:

(1) Given the belief system $\mu^*_{pool}(R|p,S)$, the R-type firm cannot be strictly better off by deviating from strategy $p^*_{pool,S}$ to any other strategy p, i.e.,

$$\pi^*_{R,pool,S} \geq \max_{p \neq p^*_{pool,S}} \pi_R(p|\mu^*_{pool}(R|p,S)). \tag{21}$$

By the definition of $\pi^*_{R,pool,S}$ in the proof of Lemma 3, we have $\pi^*_{R,pool,S} = \max_p \pi_R(p|\mu(R|p,S) = \gamma_S)$. Since $\pi_R(p|\mu(R|p,S))$ increases in $\mu(R|p,S)$, we have $\pi^*_{R,pool,S} = \max_p \pi_R(p|\mu(R|p,S) = \gamma_S) \geq \max_{p \neq p^*_{pool,S}} \pi_R(p|\mu^*_{pool}(R|p,S) = 0)$. Thus, Equation (21) holds.

(2) Given the belief system $\mu^*_{pool}(R|p,S)$, the N-type firm cannot be strictly better off by deviating from strategy $p^*_{pool,S}$ to any other strategy p, i.e.,

$$\pi^*_{N,pool,S} \geq \max_{p \neq p^*_{pool,S}} \pi_N(p|\mu^*_{pool}(R|p,S)). \tag{22}$$

In Step 1, we have shown that when $\gamma_S > \gamma^*$, $\pi^*_{N,pool,S} > \pi^*_{N,sep} = \hat{\pi}^*_N = \max_p \pi_N(p|\mu(R|p,S) = 0) \geq \max_{p \neq p^*_{pool,S}} \pi_N(p|\mu^*_{pool}(R|p,S) = 0)$. Thus, Equation (22) holds.

Therefore, when $\gamma_S > \gamma^*$, the most-efficient pooling equilibrium σ^*_{pool} exists.

Step 5. In this step, we show that when $\gamma_S > \gamma^*$, the most-efficient pooling equilibrium σ^*_{pool} is the unique LMSE. Based on the definition of LMSE in our setting, a PBE is LMSE iff it is the R-type firm's most profitable outcome among all PBE, and if conditional on it being the most profitable for the R-type firm, it is also the N-type firm's most profitable outcome.

Given the definition of γ^* in Step 1, and the definitions of the least-cost separating equilibrium σ^*_{sep} and the most-efficient pooling equilibrium σ^*_{pool} in Lemma 1 and Lemma 3, respectively, when $\gamma_S > \gamma^*$, the most-efficient pooling equilibrium σ^*_{pool} is the unique most profitable equilibrium for the R-type firm and thus it is the unique LMSE.

Combining all the above steps completes the proof of Proposition 3. □

Proof of Corollary 1. One can easily verify that $\gamma_0 < \gamma_1$. Note Proposition 3 shows that given the signal $S \in \{0,1\}$, by applying the LMSE refinement, when $\gamma_S < \gamma^*$, the equilibrium type is the least-cost separating; when $\gamma_S > \gamma^*$, the equilibrium type is the most-efficient pooling.

Thus, if $\gamma_0 < \gamma_1 < \gamma^*$, the equilibrium type is separating when the signal $S = 0$ or $S = 1$; if $\gamma^* < \gamma_0 < \gamma_1$, the equilibrium outcome is pooling when the signal $S = 0$ or $S = 1$; if $\gamma_0 < \gamma^* < \gamma_1$, the equilibrium outcome is pooling when the signal $S = 1$ and is separating when $S = 0$. □

Proof of Proposition 4. Given the signal S, the j-type firm's *ex-post* profit in the equilibrium is given by

$$\pi^*_{j,S} = \begin{cases} \pi^*_{j,sep} & \text{if } \gamma_S < \gamma^* \\ \pi^*_{j,pool,S} & \text{if } \gamma_S > \gamma^* \end{cases}.$$

Before the signal is revealed, the j-type firm's *ex-ante* expected profit in the equilibrium is given by

$$\mathbb{E}[\pi_j^*] = Pr(S=1|j) \cdot \pi_{j,1}^* + Pr(S=0|j) \cdot \pi_{j,0}^*.$$

Recall from Lemma 2 that in the separating equilibrium, the R-type firm's *ex-post* profit is not monotonic in τ, and from Lemma 4 that in the pooling equilibrium, the R-type firm's *ex-post* profit increases in τ when $\bar{\delta}_S > 0$ and decreases in τ when $\bar{\delta}_S < 0$. Thus, the R-type firm's *ex-post* profit $\pi_{R,S}^*$ is not monotonic in τ for any signal $S \in \{0, 1\}$. One can easily verify that the R-type firm's *ex-ante* profit $\mathbb{E}[\pi_R^*]$ is not monotonic in τ, either.

Recall from Lemma 2 that in the separating equilibrium, the N-type firm's *ex-post* profit decreases in τ, and from Lemma 4 that in the pooling equilibrium, the N-type firm's *ex-post* profit increases in τ when $\bar{\delta}_S > 0$ and decreases in τ when $\bar{\delta}_S < 0$. Moreover, at the equilibrium type switching point (i.e., $\gamma_S = \gamma^*$, which is a function of τ), the N-type firm's profit in the pooling equilibrium is strictly higher than that in the separating equilibrium (i.e., $\pi_{N,pool,S} > \pi_{N,sep}^*$), so there exists a discontinuous upward jump when the equilibrium switches from separating to pooling as τ increases. Thus, the N-type firm's *ex-post* profit $\pi_{N,S}^*$ is not monotonic in τ for any signal $S \in \{0, 1\}$. One can easily verify that the N-type firm's *ex-ante* profit $\mathbb{E}[\pi_N^*]$ is not monotonic in τ, either. □

Proof of Proposition 5. Recall from Lemma 2 that in the separating equilibrium, the R-type firm's *ex-post* profit increases in δ_R and decreases in δ_N, and from Lemma 4 that in the pooling equilibrium, the R-type firm's *ex-post* profit increases in δ_R and decreases in δ_N. Moreover, the R-type firm's *ex-post* profit in the equilibrium is continuous in both δ_R and δ_N. Thus, the R-type firm's *ex-post* profit in the equilibrium $\pi_{R,S}^*$ increases in δ_R and decreases in δ_N for any signal $S \in \{0, 1\}$. Given the definition of *ex-ante* profit in Equation (9), the R-type firm's *ex-ante* profit $\mathbb{E}[\pi_R^*]$ also increases in δ_R and decreases in δ_N.

Recall from Lemma 2 that in the separating equilibrium, the N-type firm's *ex-post* profit decreases in δ_N and is independent of δ_R, and from Lemma 4 that in the pooling equilibrium, the N-type firm's *ex-post* profit increases in δ_R and decreases in δ_N. At the equilibrium type switching point (i.e., $\gamma_S = \gamma^*$, which is a function of δ_N and δ_R), the N-type firm's profit in the pooling equilibrium is strictly higher than that in the separating equilibrium (i.e., $\pi_{N,pool,S} > \pi_{N,sep}^*$), that is, there exists a discontinuous upward jump from separating to pooling. As δ_R or δ_N increases, the equilibrium type may switch from separating to pooling or from pooling to separating (see Figure 4). Thus, the N-type firm's *ex-post* profit $\pi_{N,S}^*$ is not monotonic in δ_R and δ_N. One can easily verify that the N-type firm's *ex-ante* profit $\mathbb{E}[\pi_N^*]$ is not monotonic in δ_R and δ_N, either. □

Proof of Proposition 6. Recall from Lemma 2 that in the separating equilibrium, each type firm's *ex-post* profit is independent of γ_S, and that Lemma 4 shows that in the pooling equilibrium, each type of firm's *ex-post* profit increases in γ_S. Moreover, the R-type firm's *ex-post* profit in the equilibrium is continuous in γ_S. As γ_S increases from below γ^* to above γ^*, the equilibrium type switches from separating to pooling, and there exists a discontinuous upward jump for the N-type firm's profit at $\gamma_S = \gamma^*$. Thus each type firm's *ex-post* profit in the equilibrium increases in γ_S, i.e., $\frac{\partial \pi^*_{j,S}}{\partial \gamma_S} \geq 0$ for $j \in \{R, N\}$ and $S \in \{0, 1\}$.

Note that $\frac{\partial \pi^*_{j,S}}{\partial \rho_R} = \frac{\partial \pi^*_{j,S}}{\partial \gamma_S} \cdot \frac{\partial \gamma_S}{\partial \rho_R}$ and $\frac{\partial \pi^*_{j,S}}{\partial \rho_N} = \frac{\partial \pi^*_{j,S}}{\partial \gamma_S} \cdot \frac{\partial \gamma_S}{\partial \rho_N}$ for $j \in \{R, N\}$ and $S \in \{0, 1\}$. Moreover, we have $\frac{\partial \gamma_1}{\partial \rho_R} \geq 0$, $\frac{\partial \gamma_1}{\partial \rho_N} \geq 0$, $\frac{\partial \gamma_0}{\partial \rho_R} \leq 0$, and $\frac{\partial \gamma_0}{\partial \rho_N} \leq 0$. Therefore, $\frac{\partial \pi^*_{j,1}}{\partial \rho_R} \geq 0$, $\frac{\partial \pi^*_{j,1}}{\partial \rho_N} \geq 0$, $\frac{\partial \pi^*_{j,0}}{\partial \rho_R} \leq 0$, and $\frac{\partial \pi^*_{j,0}}{\partial \rho_N} \leq 0$, for $j \in \{R, N\}$. This completes the proof of Proposition 6. $\square$

Proof of Proposition 7. The R-type firm's ex-ante profit is given by $\mathbb{E}[\pi^*_R] = Pr(S = 1|R) \cdot \pi^*_{R,1} + Pr(S = 0|R) \cdot \pi^*_{R,0} = (\gamma + \rho_R(1-\gamma))\pi^*_{R,1} + (1-\rho_R)(1-\gamma)\pi^*_{R,0}$. Taking derivative of $\mathbb{E}[\pi^*_R]$ with respect to ρ_R and ρ_N, we have:

$$\frac{\partial \mathbb{E}[\pi^*_R]}{\partial \rho_N} = (\gamma + \rho_R(1-\gamma))\frac{\partial \pi^*_{R,1}}{\partial \rho_N} + (1-\rho_R)(1-\gamma)\frac{\partial \pi^*_{R,0}}{\partial \rho_N}, \text{ and}$$
$$\frac{\partial \mathbb{E}[\pi^*_R]}{\partial \rho_R} = (1-\gamma)(\pi^*_{R,1} - \pi^*_{R,0}) + (\gamma + \rho_R(1-\gamma))\frac{\partial \pi^*_{R,1}}{\partial \rho_R} + (1-\rho_R)(1-\gamma)\frac{\partial \pi^*_{R,0}}{\partial \rho_R}.$$

Through tedious calculations, one can verify that $\frac{\partial \mathbb{E}[\pi^*_R]}{\partial \rho_N} \geq 0$ and $\frac{\partial \mathbb{E}[\pi^*_R]}{\partial \rho_R} \geq 0$. We omit the detailed expressions for brevity. Thus, $\mathbb{E}[\pi^*_R]$ increases in ρ_N and ρ_R.

The N-type firm's ex-ante profit is given by $\mathbb{E}[\pi^*_N] = Pr(S = 1|N) \cdot \pi^*_{N,1} + Pr(S = 0|N) \cdot \pi^*_{N,0} = (1-\rho_N)\gamma\pi^*_{N,1} + (1-\gamma+\rho_N\gamma)\pi^*_{N,0}$. Taking derivative of $\mathbb{E}[\pi^*_N]$ with respect to ρ_R and ρ_N, we have:

$$\frac{\partial \mathbb{E}[\pi^*_N]}{\partial \rho_N} = \gamma(\pi^*_{N,0} - \pi^*_{N,1}) + (1-\rho_N)\gamma\frac{\partial \pi^*_{N,1}}{\partial \rho_N} + (1-\gamma+\rho_N\gamma)\frac{\partial \pi^*_{N,0}}{\partial \rho_N}, \text{ and}$$
$$\frac{\partial \mathbb{E}[\pi^*_N]}{\partial \rho_R} = (1-\rho_N)\gamma\frac{\partial \pi^*_{N,1}}{\partial \rho_R} + (1-\gamma+\rho_N\gamma)\frac{\partial \pi^*_{N,0}}{\partial \rho_R}.$$

Again one can verify that $\frac{\partial \mathbb{E}[\pi^*_N]}{\partial \rho_N}$ and $\frac{\partial \mathbb{E}[\pi^*_N]}{\partial \rho_R}$ could be either positive or negative. We omit the detailed expressions for brevity. Thus, $\mathbb{E}[\pi^*_N]$ may increase or decrease in ρ_N and ρ_R. This completes the proof of Proposition 7. $\square$

Proof of Proposition 8. Taking derivative of Δ_j with respect to ρ_N and ρ_R, we have $\frac{\partial \Delta_j}{\partial \rho_N} = \frac{\partial \pi^*_{j,1}}{\partial \rho_N} - \frac{\partial \pi^*_{j,0}}{\partial \rho_N}$ and $\frac{\partial \Delta_j}{\partial \rho_R} = \frac{\partial \pi^*_{j,1}}{\partial \rho_R} - \frac{\partial \pi^*_{j,0}}{\partial \rho_R}$. Recall that in Proposition 6, we have shown that $\frac{\partial \pi^*_{j,1}}{\partial \rho_N} \geq 0$, $\frac{\partial \pi^*_{j,0}}{\partial \rho_N} \leq 0$, $\frac{\partial \pi^*_{j,1}}{\partial \rho_R} \geq 0$, and $\frac{\partial \pi^*_{j,0}}{\partial \rho_R} \leq 0$. Thus, $\frac{\partial \Delta_j}{\partial \rho_N} \geq 0$ and $\frac{\partial \Delta_j}{\partial \rho_R} \geq 0$.

Taking derivative of $\mathbf{E}[\Delta_R]$ with respect to ρ_N and ρ_R, we have $\frac{\partial \mathbf{E}[\Delta_R]}{\partial \rho_N} = (\gamma + \rho_R(1-\gamma))\frac{\partial \Delta_R}{\partial \rho_N} \geq 0$ and $\frac{\partial \mathbf{E}[\Delta_R]}{\partial \rho_R} = (\gamma + \rho_R(1-\gamma))\frac{\partial \Delta_R}{\partial \rho_R} + (1-\gamma)\Delta_R \geq 0$.

Taking derivative of $\mathbf{E}[\Delta_N]$ with respect to ρ_N and ρ_R, we have $\frac{\partial \mathbf{E}[\Delta_N]}{\partial \rho_N} = (1-\rho_N)\gamma\frac{\partial \Delta_N}{\partial \rho_N} - \gamma\Delta_N$ and $\frac{\partial \mathbf{E}[\Delta_N]}{\partial \rho_R} = (1-\rho_N)\gamma\frac{\partial \Delta_N}{\partial \rho_R} \geq 0$. One can easily verify that $\frac{\partial \mathbf{E}[\Delta_N]}{\partial \rho_N}$ could be either positive or negative. This completes the proof of Proposition 8. $\square$

C. Discussion on Intuitive Criterion

In the main paper, we have applied the LMSE concept to refine the equilibrium outcome. Another commonly used alternative is the Intuitive Criterion (Cho and Kreps, 1987). As to be shown later, the unique equilibrium that survives the Intuitive Criterion is the least-cost separating equilibrium characterized in Lemma 1 and any pooling equilibrium will be eliminated by the Intuitive Criterion. The effects of δ_N, δ_R, and τ on the equilibrium outcome are shown in Lemma 2.

As shown in Lemma 2, we know that the R-type firm's profit is increasing in the consumers' reward level (i.e., δ_R), decreasing in their punishment level (i.e., δ_N), but is not monotonic in the fraction of R-type consumers (i.e., τ), whereas the N-type firm's profit is decreasing in τ and δ_N, but is independent of δ_R. Recall from the symmetric information case, the R-type firm's profit always increases in τ and δ_R, and it is independent of δ_N. The comparison indicates that under asymmetric information, the R-type firm's profit might even decrease in τ and δ_N in the equilibrium: As more consumers become socially responsible or their punishment level for the N-type firm increases, the N-type firm has a stronger incentive to mimic the R-type firm and thus, the R-type has to incur a higher signaling cost to separate from the N-type firm, which leads to a lower profit. This indicates that the main driving forces and intuition in our model are fairly robust under different equilibrium refinement tools.

In addition, in the least-cost separating equilibrium, the consumers' posterior belief about the firm's type γ_S is irrelevant. Thus, obtaining the certification or not has no effect on the final equilibrium outcome. As a result, the ex-ante and the ex-post outcomes are identical in the least-cost separating equilibrium.

Finally, we use the following four steps to show that the least-cost separating equilibrium is the unique equilibrium that survives the Intuitive Criterion. Step 1, we show that all separating equilibria except the least-cost separating equilibrium fail the Intuitive Criterion. Step 2, we show that all pooling equilibria fail the Intuitive Criterion. Step 3, we show that the least-cost separating equilibrium exists. Step 4, we show that the least-cost separating equilibrium survives the Intuitive Criterion.

Step 1. Let $\sigma_{sep} = (p_{R,sep}, p_{N,sep}, \mu_{sep})$ denote any separating equilibrium in which the R-type firm's strategy is $p_{R,sep} \in \Phi$ and $p_{R,sep} \neq p^*_{R,sep}$ and the N-type firm's strategy is $p_{N,sep} = \hat{p}^*_N$. Let $\pi_R(\sigma_{sep})$ denote the R-type firm's profit in the equilibrium σ_{sep} and $\hat{\pi}^*_N$ is the N-type firm's profit in σ_{sep}.

If σ_{sep} survives the Intuitive Criterion, then it must satisfy the following condition: If there exists any strategy p such that $\hat{\pi}^*_N > \max_{\mu(R|p,S)} \pi_N(p|\mu(R|p,S)) = \pi_N(p|\mu(R|p,S) = 1)$, then $\pi_R(\sigma_{sep}) \geq \pi_R(p|\mu(R|p,S) = 1)$. However, in the proof of Lemma 1, we have shown that $\hat{\pi}^*_N > \pi_N(p^*_{R,sep}|\mu(R|p^*_{R,sep},S) = 1)$, and $\pi_R(\sigma_{sep}) < \pi^*_{R,sep} = \pi_R(p^*_{R,sep}|\mu(R|p^*_{R,sep},S) = 1)$, which violates the condition and thus σ_{sep} fails the Intuitive Criterion.

Step 2. Let $\sigma_{pool} = (p_{pool}, \mu_{pool})$ denote any pooling equilibrium in which each type of firm's strategy is p_{pool}. Let $\pi_j(\sigma_{pool})$ denote the j-type firm's profit in the equilibrium σ_{pool}. If σ_{pool} survives the Intuitive Criterion, then it must satisfy the following condition: If there exists any strategy p such that $\pi_N(\sigma_{pool}) > \max_{\mu(R|p,S)} \pi_N(p|\mu(R|p,S)) = \pi_N(p|\mu(R|p,S) = 1)$, then $\pi_R(\sigma_{pool}) \geq \pi_R(p|\mu(R|p,S) = 1)$.

Solving $\pi_N(\sigma_{pool}) > \pi_N(p|\mu(R|p,S) = 1)$ leads to $p > \bar{\tilde{p}}_N$ or $p < \underline{\tilde{p}}_N$; and solving $\pi_R(\sigma_{pool}) < \pi_R(p|\mu(R|p,S) = 1)$ leads to $\underline{\tilde{p}}_R < p < \bar{\tilde{p}}_R$. One can verify that $\bar{\tilde{p}}_N < \bar{\tilde{p}}_R$. Thus, there exists a small enough $\epsilon > 0$ such that $p' \equiv \bar{\tilde{p}}_R - \epsilon > \max\{\bar{\tilde{p}}_N, \underline{\tilde{p}}_R\}$. Thus, $p' > \bar{\tilde{p}}_N$ and $p' \in (\underline{\tilde{p}}_R, \bar{\tilde{p}}_R)$. When $p = p'$, we have $\pi_N(\sigma_{pool}) > \pi_N(p'|\mu(R|p',S) = 1)$ and $\pi_R(\sigma_{pool}) < \pi_R(p'|\mu(R|p',S) = 1)$, which violates the condition and thus σ_{pool} fails the Intuitive Criterion.

Step 3. Let $\sigma^*_{sep} = (p^*_{R,sep}, p^*_{N,sep}, \mu^*_{sep})$ denote the least-cost separating equilibrium as characterized in Lemma 1 with the belief μ^*_{sep} given in Equation (16). To show that the least-cost separating equilibrium σ^*_{sep} exists, we just need to show the following:

(1) Given the belief system $\mu^*_{sep}(R|p,S)$, the R-type firm cannot be strictly better off by deviating from strategy $p^*_{R,sep}$ to any other strategy p, i.e.,

$$\pi^*_{R,sep} \geq \max_{p \neq p^*_{R,sep}} \pi_R(p|\mu^*_{sep}(R|p,S)). \tag{23}$$

One can verify that $\max_p \pi_R(p|\mu(R|p,S)=0) = \begin{cases} \frac{(q-c_R-\delta_N\tau)^2}{4q}, & \text{if } \delta_N < \frac{(q-c_R)(1-\sqrt{1-\tau})}{\tau} \\ (1-\tau)\frac{(q-c_R)^2}{4q}, & \text{otherwise} \end{cases}$ and $\pi^*_{R,sep} \geq \max_p \pi_R(p|\mu(R|p,S)=0)$ always hold. Thus, Equation (23) holds.

(2) Given the belief system $\mu^*_{sep}(R|p,S)$, the N-type firm cannot be strictly better off by deviating from strategy $p^*_{N,sep}$ to any other strategy p, i.e.,

$$\pi^*_{N,sep} \geq \max_{p \neq p^*_{N,sep}} \pi_N(p|\mu^*_{sep}(R|p,S)). \tag{24}$$

By the definition of Φ and $p^*_{R,sep}$ in the proof of Lemma 1, $\pi^*_{N,sep} \geq \pi_N(p^*_{R,sep}|\mu^*_{sep}(R|p^*_{R,sep},S) = 1)$ since $p^*_{R,sep} \in \Phi$. For any $p \neq p^*_{R,sep}$, we have $\pi^*_{N,sep} = \hat{\pi}^*_N = \max_p \pi_N(p|\mu(R|p,S) = 0) \geq \max_{p \neq p^*_{N,sep}} \pi_N(p|\mu^*_{sep}(R|p,S) = 0)$. Thus, Equation (24) holds.

Therefore, the least-cost separating equilibrium σ^*_{sep} exists.

Step 4. To show that σ^*_{sep} survives the Intuitive Criterion, we only need to show the following:

(1) If there exists any strategy p such that $\pi^*_{N,sep} > \max_{\mu(R|p,S)} \pi_N(p|\mu(R|p,S)) = \pi_N(p|\mu(R|p,S) = 1)$, then $\pi^*_{R,sep} \geq \pi_R(p|\mu(R|p,S) = 1)$.

(2) If there exists any strategy p such that $\pi^*_{R,sep} > \max_{\mu(R|p,S)} \pi_R(p|\mu(R|p,S)) = \pi_R(p|\mu(R|p,S) = 1)$, then $\pi^*_{N,sep} \geq \pi_N(p|\mu(R|p,S) = 0)$.

In the proof of Lemma 1, we define $\pi^*_{R,sep} = \max_{p\in\Phi} \pi_R(p|\mu(R|p,S) = 1)$, where $\Phi = \{p : \pi^*_{N,sep} \geq \pi_N(p|\mu(R|p,S) = 1)\}$. Thus, condition (1) holds. Since $\pi^*_{N,sep} = \hat{\pi}^*_N = \max_p \pi_N(p|\mu(R|p,S) = 0)$, condition (2) also holds. Therefore, the least-cost separating equilibrium σ^*_{sep} survives the Intuitive Criterion.

Combining all the above steps confirms that the least-cost separating equilibrium is the unique equilibrium that survives the Intuitive Criterion.