Signaling under Bilateral Uncertainty: The Case of Green Labels

Dolores Garrido*, Ana Espinola-Arredondo†, and Felix Munoz-Garcia‡
School of Economic Sciences
Washington State University
Pullman, WA 99164

May 27, 2019

Abstract

This paper examines the role of firms' uncertainty about consumers' environmental concerns in the emergence of greenwashing. We consider a signaling model where a firm (either green or brown) chooses whether to acquire a green label to signal its type to consumers (either green or brown). Uncertainty stems from two sources: (i) an uninformed consumer; and (ii) an uninformed firm. We examine under which conditions information dissemination arises under bilateral uncertainty. We show that penalties from greenwashing and the proportion of green consumers play a substitutive role in the existence of a separating equilibrium in which the green innovation is revealed. We also identify situations in which pooling equilibria promoting greenwashing are sustained. We find that bilateral uncertainty, compared to the case in which only the firm is informed (unilateral uncertainty), induces more greenwashing. Suggesting that educational programs become more relevant in this setting.

KEYWORDS: Signaling game; bilateral uncertainty; green label; greenwashing.
JEL CLASSIFICATION: D81, D82, L15, Q50.

*323D Hulbert Hall, WSU, Pullman, WA 99164. E-mail: m.garridogarcia@wsu.edu.
†111C Hulbert Hall, WSU, Pullman, WA 99164. E-mail: anaespinola@wsu.edu.
‡103H Hulbert Hall, WSU, Pullman, WA 99164. E-mail: fmunoz@wsu.edu.

1 Introduction

Consumers are becoming more environment-conscious when making a purchase decision. For instance, Cone Communications (2013) reports that almost half of US consumers search for environmental information when deciding what good to buy, being Millennials the cohort taking this lead.[1] Hence, firms have more incentives to highlight their environmental practices with the hope to attract environment-conscious consumers. For example, US sales of food and beverages, household and personal care products with sustainable attributes grew four times larger than sales of conventional products from 2014 to 2017. By 2021, sustainability sales are expected to make up 25% of total store sales (Nielsen, 2018). Environmental characteristics of products are, however, not publicly observable to consumers. Green firms (i.e., using an environment-friendly production process) usually make green claims in the product's label to overcome these information asymmetries. The number of green labels has considerable increased in the last fifteen years.[2]

The current outlook of green labels to satisfy consumers' green demands constitutes a ready-made breeding ground for greenwashing in the form of false or misleading green labels. Terrachoice (2010) found that 32% of a total of 5,296 home and family green products in the US and Canada used false labels. Greenwashing is not exempt of risk for firms; consumers, environmental activists and non-governmental organizations (NGOs), as well as policy makers, play and important role to deter firms from doing this practice (Lyon and Maswell, 2011; Lyon and Montgomery, 2015; Berrone et al., 2017).[3] However, the scope of action of stakeholders is still very limited as a consequence of a lax and uncertain regulation (Lorance, 2010; Delmas and Burbano, 2011; European Commission, 2014).

Our paper examines how uncertainty between a firm and a consumer affects the emergence of greenwashing under the presence of a penalty system. We consider a model where a firm (either green or brown) decides whether to signal the use of a environment-friendly production process to a consumer (either green or brown). The consumer observes the firm's action (i.e., label or no label) but does not observe the firm's environmental practices during the production of the good, and responds buying or not buying. We first focus on the case of unilateral uncertainty, in which only the consumer is the uninformed player. We consider the consumer always value more the green product, independent of her type; we restrict our attention to two types of signals (i.e., green label or no label).[4] Second, we analyze how a second layer of uncertainty (i.e., uninformed firm which is not able to observe the type of consumer who receives the signal) affects our equilibrium results. In addition, we consider different price regions to point out similarities and draw contrasts

[1] According to a Nielsen study (2018), Millennials are more likely than BabyBoomers to pay more for products with sustainable ingredients (90% vs. 61%).

[2] Today's world picture of green labels shows more than 450 options (almost half of which are present in the US) under different schemes (public vs. private, mandatory vs. voluntary) and degree of sustainability strictness (Gruere, 2013; Ecolabel Index, 2019).

[3] Seele and Gatti (2017) consider this risk and update the definition of greewashing as "co-creation of an external accusation toward an organization with regard to presenting a misleading green message."

[4] For simplicity, we assume that the green label can include either a third-party certification or a self-reported claim.

in equilibrium results.

In the unilateral case, we show that an informative (separating) equilibrium in which the green firm acquires a green label and the brown firm does not label can be sustained under larger conditions with the presence of a green consumer. This equilibrium is supported by a labeling cost which is significantly different for a green than for a brown firm, and a strict eco-labeling program and/or an effective greenwashing punishment system, deterring the brown firm from using false labels. We analyze how results change under different price ranges, and find that higher prices reduce the use of green labels by the green firm, thus limiting information transmission. We also show the existence of an uninformative (pooling) equilibrium promoting greenwashing. In the case of a green consumer, greenwashing emerges if the price of the good is sufficiently high to compensate the cost of labeling and expected penalties the brown firm might face. In the case of a brown consumer, we show that this uninformative equilibrium does not arise unless this type of consumer is willing to sue the brown firm from doing greenwashing, as the green consumer is assumed to. In other words, under the presence of a brown consumer greenwashing is less likely to arise. However, it hinders information about the environmental characteristics of a product given that the green firm has less incentives to acquire a green label, and thus provide more information to uninformed consumers.

The results from the bilateral case show that an extra layer of uncertainty favors the emergence of greenwashing in a context in which the proportion of environment-conscious consumers (i.e., the probability of facing a green consumer) is relatively high. In this setting greenwashing can be reduced if penalties for making a false label are substantial, otherwise a heterogeneous population of green and brown consumers is the only mechanism to deter greenwashing. These findings suggest that policies aimed to motivate both green and brown consumers to detect greenwashing, as well as to structures to facilitate complaints towards this fraudulent practice, can help to reduce greenwashing. They can play an important role on the dissemination of information and the reduction of greenwashing.

Related Literature. The practice of greenwashing has increasingly been analyzed from different perspectives in last years (see Lyon and Montgomery, 2015 for an extensive literature review). Previous studies of this phenomenon are mainly focused on the role of prices and signaling costs as mechanisms to facilitate information transmission. Mahenc (2008) finds that prices can act as signals to green consumers if the green product is more costly to produce than the brown. Volle (2017) analyzes the distortions created by price signaling and show the emergence of an uninformative equilibrium promoting greenwashing when prices are the only signal under consideration. Hamilton and Zilberman (2006) and Mahenc (2017) find that certification costs effectively reduce greenwashing if they are large enough to deter the brown firm from mimicking the green firm labeling strategy.

The role of penalty costs to discourage false signaling is, however, less explored (Connelly et al., 2011). Lyon and Maxwell (2011) introduce this factor in the form of the threat of an activist. This paper highlights the key role of penalties for deterring the practice of greenwashing, thus promoting information transmission. It seeks to obtain a better understanding of the importance of penalties

by analyzing how they act as complements or substitutes when firms face uncertainty about the proportion of environment-concerned consumers.

Our paper also contributes to signaling theory by introducing a second layer of uncertainty. The bilateral or two-sided uncertainty to examine conditions shaping relations between firms and consumers have mainly focused on bargaining situations (Chatterjee and Samuelson, 1983; Cramton, 1992) and adverse selection scenarios (Gale, 2001). However, to the best of our knowledge, none of these previous studies analyze the role of labels. The consideration of bilateral uncertainty allow us to analyze to what extent a context of increasing environmental concerns affects firm's strategic behavior, and how this situation promotes of reduces greenwashing in combination of signaling and penalty costs.

The next section describes the model. Section 3 and 4 examine and discuss the set of equilibria under unilateral and bilateral uncertainty, respectively. Section 5 concludes.

2 Model

Consider a signaling game between a firm and a consumer. The firm sells a good at a given price, $p > 0$, and the firm can be of two types, either green (G) if it has implemented an environment-friendly production process, or brown (B) if it uses a dirty technology, which is private information to the firm. Production costs are type-dependent, $C_i > 0$, where $i \in \{G, B\}$ denotes the firm's type, which allows for $C_G \geq C_B$ or $C_G < C_B$.[5] The firm has two labeling strategies: to acquire a green label (L) or not (NL).[6] A green label is costly to the firm; it has a type-dependent cost (per unit sold) of $L_i > 0$, where $i \in \{G, B\}$; and $L_B \geq L_G > NL = 0$. The cost of labeling can coincide for the brown and the green firm in the case of non-certified or self-reported green labels, and is greater for the brown firm in the case of a third-party certified label. The firm can face an extra cost, $K \geq 0$, in the form of an expected penalty from greenwashing, which is increasing in the probability of being found liable and in the amount of the penalty.[7]

If the green firm acquires a green label, its payoff is $p - C_G - L_G$ if the consumer buys, while the brown firm receives $p - C_B - L_B - K$. If firm i acquires a green label and the consumer does not buy, its payoff becomes $-C_i - L_i$. If the firm does not label, its payoff is $p - C_i$ if the consumer buys; and $-C_i$ if she does not. In addition, $p > L_i$, so both types of firm have incentives to acquire a green label.

The consumer can be of two types, either green (G) or brown (B). The green consumer cares more about the environmental impact of her consumption than the brown consumer. Let V_i^j denotes the consumer's valuation of the good, where $i \in \{G, B\}$ indicates the firm's type, and $j \in \{G, B\}$ the consumer's type. If the consumer buys the product, her payoff is $V_i^j - p$, and zero otherwise. We assume $V_G^j > V_B^j$ for every j, that is, the valuation of a green good is higher for both the green

[5] The cost of implementing the production process is sunk.

[6] The acquisition of a green label is assumed not to be mandatory. Otherwise, labels could not act as a signal.

[7] This cost can be interpreted as the legal and other potential costs for the brown firm if, after doing greenwashing, it is sued by the consumer, an organization or any other third party, and is declared guilty.

and brown consumer. However, the green consumer exhibits a higher valuation for the green good than the brown consumer (i.e., $V_G^G > V_G^B$). For simplicity, we consider that only the brown firm can be found liable when using a false green label when facing a green consumer. In other words, the brown consumer, given that she cares less about the environment, is not willing to spend resources in suing a brown firm after purchasing a false green-labeled good. We further relax this assumption.

We consider four possible price regions, $p \in \{A, B, C, D\}$, for the consumer to make purchasing decisions as shown in Figure 1. For simplicity V_G^B is normalized to zero. If the price of the good is relatively low (i.e., $p \in \{A\}$), when the green consumer observes a green (brown) good, she purchases (does not purchase) since $V_G^G > p$ ($V_B^G = 0 < p$, respectively); in contrast, the brown consumer purchases the brown good, i.e., $V_B^B > p$, but also buys the green good since it is sufficiently inexpensive, i.e., $V_G^B > p$. If the good is sold at a price in region B, both types of consumer only buy the green good. At $p \in \{C\}$, the green consumer still buys the green good, but the brown consumer does not buy given that this price is higher than the valuation she assigns to the good. Finally, no type of consumer buys, either the green or brown good, if it is sufficiently expensive (i.e., $p \in \{D\}$).

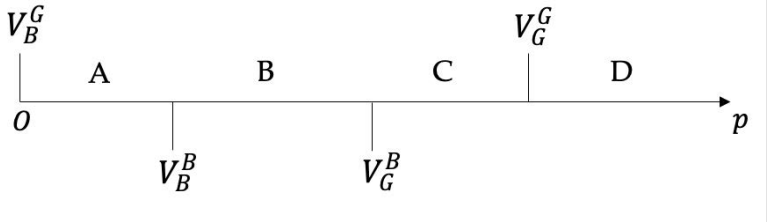

Figure 1. Price regions

The next lemma presents the equilibrium behavior in the complete information setting where both the consumer and the firm can perfectly observe the other player's type.

Lemma 1 (Complete information). *Under complete information, both types of firm do not label; and the consumer buys the good if her valuation is higher than the price at which it is sold.*

Intuitively, under complete information, the true type of the firm is known. Thus, no type of firm needs to spend resources on acquiring a green label, since purchasing decisions are based solely on the price of the good but are unaffected by labeling.

In the following sections we show how the presence of uncertainty between the firm and the consumer leads one or both types of company to acquire a green label to induce purchases.

3 Signaling under Unilateral Uncertainty

We first focus the analysis on the case in which the consumer is the only uninformed player (unilateral uncertainty), and then examine the case in which both the consumer and firm are uninformed (bilateral uncertainty). The time structure of the game under unilateral uncertainty is the following:

1. Nature selects the firm's type (either green or brown).
2. The firm privately observes its type, and decides whether or not to acquire a green label.
3. The consumer does not know the firm's type, but assigns a prior probability q to the firm being green. Upon observing the firm using a green label (no label), she updates her beliefs to $\mu \equiv prob(G|L)$ ($\gamma \equiv prob(G|NL)$, respectively), where $\mu, \gamma \in [0, 1]$. The consumer then responds buying or not buying the good.

We examine conditions under which both separating and pooling Perfect Bayesian Equilibria (PBE) arise. All equilibria presented survive Cho and Kreps' (1987) Intuitive Criterion. Proofs are relegated to the appendix.

3.1 Green Consumer

We first analyze the case in which the firm faces a green consumer, and then that in which the firm faces a brown consumer.

Proposition 1. *In the signaling game under unilateral uncertainty with a green consumer, the separating strategy profile in which the green firm acquires a green label and the brown firm does not label can be sustained as a PBE if and only if* $p \leq L_B + K$*; when* $p \in \{A, B, C\}$*; the consumer buys after observing a green label, and she does not buy otherwise.*

In this separating equilibrium, the green firm has incentives to acquire a green label. However, the brown firm does not have incentives to mimic this strategy since the cost of adopting a green label, which embodies L_B and the expected penalty K, exceeds the market price, $p \in \{A, B, C\}$. When both types of firm face the same labeling cost $L_B = L_G$, this informative equilibrium can still be sustained if the penalty from greenwashing, K, is sufficiently high. However, when L_B increases, firms become more asymmetric in their labeling costs, thus expanding the parameter values under which this PBE exists. Hence, the emergence of this informative equilibrium might be promoted by a strict and reliable eco-labeling program (making L_B significantly higher than L_G) and/or an effective greenwashing punishment system (high K). In addition, as price increases, moving from region A to B, and C, this separating equilibrium is less likely to arise; higher prices reduce the use of green labels. In a setting with an extremely high price (i.e., $p \in \{D\}$) this separating equilibrium cannot arise given that the green firm does not have incentives to label as the consumer does not buy the good; thus blocking the transmission of the environmental characteristics of goods.

The opposite separating equilibrium, in which only the brown firm acquires a label, cannot be supported for any price region given that the brown firm would be revealing its type, thus inducing the green consumer to not buy. She only buys when no label is in place since she interprets that the unlabeled good is coming from a green firm. In such a situation, a brown firm always earns a higher payoff from not labeling its goods.

Proposition 2. *In the signaling game under unilateral uncertainty with a green consumer, the following pooling PBEs can be sustained:*

1. *Both types of firm acquire a green label if and only if* $p \geq L_B + K$*; the green consumer only buys if equilibrium beliefs satisfy* $q \geq \frac{p}{V_G^G} \equiv \overline{q}^G(p)$ *and off-the-equilibrium beliefs are* $\gamma < \overline{q}^G(p)$*; when* $p \in \{A, B, C\}$.

2. *Both types of firm do not label;*

   (a) *the green consumer only buys after observing no label if equilibrium beliefs satisfy* $q \geq \overline{q}^G(p)$*, under any off-the-equilibrium beliefs* $\mu \in [0, 1]$*; when* $p \in \{A, B, C\}$.

   (b) *the green consumer does not buy if and only if* $p \geq L_B + K$*;*

      i. *if equilibrium beliefs satisfy* $q < \overline{q}^G(p)$ *and off-the-equilibrium beliefs are* $\mu < \overline{q}^G(p)$*; when* $p \in \{A, B, C\}$;

      ii. *for all equilibrium and off-the-equilibrium beliefs* $q, \mu \in [0, 1]$ *when* $p \in \{D\}$.

The first pooling equilibrium, in which both types of firm follow a labeling strategy, can be sustained if the price compensates the cost of labeling and expected penalty from greenwashing for the brown firm. In this case, the green consumer only buys after observing a green label since her beliefs of facing a green firm are high (i.e., $q \geq \overline{q}^G(p)$). Notice that, *ceteris paribus*, as the price of the good increases, cutoff $\overline{q}^G(p)$ increases, (i.e., $\overline{q}^G(A) < \overline{q}^G(B) < \overline{q}^G(C)$), thus shrinking the parameter conditions under which this uninformative equilibrium arises. That is, a higher price requires the green consumer has stronger beliefs (i.e., higher q) about the green label being used by a green firm to induce the brown firm to assume the risk of greenwashing. However, this result also suggests that markets of luxury goods are a breeding ground for greenwashing in contexts in which green firms are likely and the cost of greenwashing (i.e., $L_B + K$) is low.[8]

In the second pooling equilibrium no type of firm acquires a label. The green firm does not have incentives to assume the cost of labeling given that consumer's beliefs (i.e., a green firm is likely) lead her to purchase the unlabeled good (case 2a). This case can emerge in contexts in which the green nature of firms is well known by consumers (i.e., strong private brand reputation) or the environment-friendly production process is widespread.[9] In case 2b, where the green firm

[8] An example of such market is the case of luxury cars.

[9] In The Netherlands and Norway only 26% and 37% of consumers, respectively, look for environmental information relative to a 50% of consumers (on average) in other European countries (European Commission, 2014). This could be explained as consumers assuming that most firms use green technologies.

is unlikely (i.e., $q < \overline{q}^G(p)$), the green consumer does not purchase the good. As a consequence, neither the green nor the brown firm has incentives to acquire a label. In this case, a higher price increases the probability of this uninformative equilibrium to arise.

3.2 Brown Consumer

We now analyze the case of a brown consumer. Recall that unlike the green consumer, the brown consumer shows a positive valuation for the brown good (i.e., $V_B^B > 0$), and purchases it if it is sufficiently cheap (i.e., $p \in \{A\}$). Lemma 2 and proposition 3 present equilibria results under the assumption that this type of consumer is not willing to spend resources on suing the brown firm for doing greenwashing (i.e., expected penalty K is not in place). Corollary 1 and 2 relax this assumption and show results when the brown consumer she is able to detect greenwashing and initiate a complaint, as a result of which the brown firm might face a greenwashing penalty, $K \geq 0$.

Lemma 2. *In the signaling game under unilateral uncertainty with a brown consumer, no separating strategy profile can be sustained as a PBE if the brown consumer does not have the capacity to sue the brown firm from doing greenwashing.*

Intuitively, the actions from the green and the brown firm reveal their types to the brown consumer. Hence, the firm does not have incentives to follow a type-dependent strategy, in which it labels (does not label) its goods when it is green (brown, respectively). Instead, both types of firm have incentives to follow the same strategy. At a relatively low price (i.e., $p \in \{A\}$), the brown consumer buys the good regardless of the message she observes; thus the green firm does not have incentives to label it goods given the opposite action guarantees the purchase by this type of consumer, by saving the labeling cost. If the price of the good exceeds the valuation she assigns to the brown good but it is lower enough to make her buy the green good (i.e., $V_G^B > p$ when $p \in \{B\}$), the brown firm has incentives to mimic the labeling strategy of the green firm at no risk of being caught from greenwashing; in such a case the brown consumer is not willing to make a greenwashing complaint. Finally, if the price is sufficiently expensive to make the brown consumer to not purchase the green good (i.e., $p \in \{C, D\}$), labeling would only add extra cost and zero benefit to the green firm.

In summary, with a brown consumer, an informative equilibrium in which the environmental characteristics of the green good are communicated cannot arise. However, as stated in the next corollary, a separating equilibrium in which the firm follow a type-dependent strategy might arise if the brown consumer has motivations to sue the brown firm from doing greenwashing.

Corollary 1. *When the brown consumer is willing to sue the brown firm for doing greenwashing, the separating strategy profile in which the green firm acquires a green label and the brown firm does not label can be sustained as a PBE if and only if $p \leq L_B + K$; when $p \in \{B\}$; the consumer buys after observing a green label, and she does not buy otherwise.*

Corollary 1 shows that an type-dependent informative equilibrium can arise with a brown consumer only if the brown firm faces the risk of being sued by this type of consumer. In such a case, a symmetric condition to the setting with a green consumer is required for this separating equilibrium to arise. That is, the labeling cost and/or greenwashing penalty exceed the price the brown firm earns; deterring it to mimicking the green firm. However, unlike the setting with a green consumer, the transmission of information is restricted to a low-intermediate price range, $p \in \{B\}$; that is a price sufficiently inexpensive to make the brown consumer buy, the green good but not cheap enough to also purchases the brown good.

Proposition 3. *In the signaling game under unilateral uncertainty with a brown consumer, the following pooling PBEs can be sustained:*

1. *Both types of firm acquire a green label; the brown consumer buys if equilibrium beliefs satisfy* $q \geq \frac{p-V_B^B}{V_G^B-V_B^B} \equiv \overline{q}^B(p)$, *and off-the-equilibrium beliefs are* $\gamma < \overline{q}^B(p)$*; when* $p \in \{B\}$.
2. *Both types of firm do not label;*
   (a) *the brown consumer buys;*
      i. *for all equilibrium and off-the-equilibrium beliefs* $q, \mu \in [0,1]$; *when* $p \in \{A\}$;
      ii. *if equilibrium beliefs are* $q \geq \overline{q}^B(p)$ *for any off-the-equilibrium beliefs* $\mu \in [0,1]$*; when* $p \in \{B\}$*;*
   (b) *the brown consumer does not buy;*
      i. *if equilibrium beliefs satisfy* $q < \overline{q}^B(p)$, *and off-the-equilibrium beliefs are* $\mu < \overline{q}^B(p)$; *when* $p \in \{B\}$;
      ii. *for all equilibrium and off-the-equilibrium beliefs* $q, \mu \in [0,1]$; *when* $p \in \{C, D\}$.

Unlike the case with a green consumer, the first uninformative equilibrium in which both types of firm use a green label is less likely to arise when the firm faces a brown consumer; it only emerges when $p \in \{B\}$ and the brown consumer only buys the green good. In other words, the presence of a brown consumer reduces the emergence of the greenwashing. However, when the price is in the aforementioned region greenwashing is promoted under larger conditions if beliefs of facing a green firm are sufficiently high. In contrast, the second pooling equilibrium which hinders information transmission about the green characteristics of products is more likely to arise with a brown consumer.

Corollary 2. *When the brown consumer is willing to sue the brown firm for doing greenwashing, the pooling equilibria in case 1 and case 2b can only be supported as a PBE if* $p \geq L_B + K$.

Corollary 2 shows that when the brown consumer behaves similarly to the green consumer, i.e., she is willing to sue the brown firm from doing greenwashing, the PBE in which greenwashing

emerges is affected by the expected penalty, K (case 1). A severe penalty reduces the brown firm's incentives to greenwash; thus shrinking the parameter conditions under which this uninformative equilibrium arises. Similarly, the uninformative equilibrium in which both types of firm do not label and consumer's beliefs lead her to not buy the good (case 2b), requires price compensates the cost of greenwashing (i.e., $p \geq L_B + K$).

Table 1 summarizes and compares the PBEs arising under unilateral uncertainty by price region. When $p \in \{A\}$ (i.e., low price) or $p \in \{C\}$ (i.e. intermediate-high price), a type-dependent informative equilibrium arises only when the firm knows with certainty it is facing a green consumer; and information transmission about the green characteristics of goods does not occur in the case of a brown consumer. However, greenwashing also emerges when the price falls in such regions in the case of a green consumer if the brown firm finds cheap to mimic the labeling strategy of the green firm; while it does not arise with a brown consumer.

When $p \in \{B\}$ (i.e., intermediate-low price), information transmission is possible with a brown consumer if she behaves similarly to the green consumer, that is, if she is willing to sue the brown firm from doing greenwashing. Otherwise, the informative equilibrium does not emerges. In this price region, greenwashing is more likely with a brown consumer, especially if she is not willing to fill a greenwashing complaint.

Finally, under a high price (i.e., $p \in \{D\}$), greenwashing does not emerges but information transmission does not occur neither with a green nor brown consumer; none buys the good. The uninformative equilibrium in which no type of firm acquires a label emerges for all equilibrium and off-the-equilibrium beliefs.

PBE		**GREEN CONSUMER** ($j = G$)				**BROWN CONSUMER** ($j = B$)			
		$p \in \{A\}$	$p \in \{B\}$	$p \in \{C\}$	$p \in \{D\}$	$p \in \{A\}$	$p \in \{B\}$	$p \in \{C\}$	$p \in \{D\}$
$L^G NL^B$		YES $if\ p \leq L_B + K$	YES $if\ p \leq L_B + K$	YES $if\ p \leq L_B + K$	NO	NO	NO (without K) YES (with K) if $if\ p \leq L_B + K$	NO	NO
$NL^G L^B$		NO	NO	NO	NO	NO	NO	NO	NO
$L^G L^B$	$q, \gamma \geq \bar{q}^j(p)$	NO	NO	NO	Consumer does not buy regardless of message for all $q, \gamma \in [0,1]$	Consumer buys regardless of message for all $q, \gamma \in [0,1]$	NO	Consumer does not buy regardless of message for all $q, \gamma \in [0,1]$	Consumer does not buy regardless of message for all $q, \gamma \in [0,1]$
	$q \geq \bar{q}^j(p)$ $\gamma < \bar{q}^j(p)$	YES $if\ p \geq L_B + K$	YES $if\ p \geq L_B + K$	*YES* $if\ p \geq L_B + K$			YES (without K) YES (with K) $if\ p \geq L_B + K$		
	$q < \bar{q}^j(p)$ $\gamma \geq \bar{q}^j(p)$	NO	NO	NO			NO		
	$q, \gamma < \bar{q}^j(p)$	NO	NO	NO	NO	NO	NO	NO	NO
$NL^G NL^B$	$q, \mu \geq \bar{q}^j(p)$	YES	YES	YES	Consumer does not buy regardless of message for all $q, \mu \in [0,1]$	Consumer buys regardless of message for all $q, \mu \in [0,1]$	YES	Consumer does not buy regardless of message for all $q, \mu \in [0,1]$	Consumer does not buy regardless of message for all $q, \mu \in [0,1]$
	$q \geq \bar{q}^j(p)$ $\mu < \bar{q}^j(p)$	YES	YES	YES			YES		
	$q < \bar{q}^j(p)$ $\mu \geq \bar{q}^j(p)$	NO	NO	NO			NO		
	$q, \mu < \bar{q}^j(p)$	YES $if\ p \geq L_B + K$	YES $if\ p \geq L_B + K$	YES $if\ p \geq L_B + K$	YES $if\ p \geq L_B + K$	YES	YES (without K) YES (with K) $if\ p \geq L_B + K$	YES (without K) YES (with K) $if\ p \geq L_B + K$	YES (without K) YES (with K) $if\ p \geq L_B + K$

Table 1. PBEs under unilateral uncertainty

4 Signaling under Bilateral Uncertainty

In this setting, both the firm and the consumer are uninformed about each others' type. As in the unilateral setting, the consumer, green or brown, decides to purchase the good according to her beliefs about the firm's type, i.e., $\mu, \gamma \in [0, 1]$. In addition, in this context, the firm's payoff is also expected, as it is affected by the probability of facing a green consumer, β, or a brown consumer, $(1 - \beta)$, where $\beta \in [0, 1]$. This setting seeks to capture how asymmetry of information affects the decision-making process in a context where green and brown firms coexist with green and brown consumers.

The following propositions describe how this second layer of uncertainty affects the dissemination of information and the emergence of greenwashing. We analyze parameter conditions under which separating and pooling equilibria arise when the price of the good is in a low range, i.e., $p \in \{A\}$; the green consumer buys only the green good while the brown consumer buys both types of good. In addition, we restrict our attention the setting in which only the green consumer has incentives to sue the brown firm from doing greenwashing.

Proposition 4. *In the signaling game under bilateral uncertainty, the separating strategy profile in which the green firm acquires a label and the brown firm does not label can be sustained as a PBE if and only if* $\beta \in [\beta_1(p) \equiv \frac{L_G}{p}, \beta_2(p) \equiv \frac{L_B}{p-K}]$, *where* $\beta_2(p) \leq 1$ *if* $p \geq L_B + K$; *when* $p \in \{A\}$; *the green and brown consumer buy after observing a green label.*

This separating equilibrium can be sustained under different parameter conditions as the penalty from greenwashing increases. Intuitively, a tough penalty deters the brown firm from doing greenwashing. Otherwise, if the cost of labeling for the brown firm and the greenwashing penalty are low, a heterogeneous population (composed by green and brown consumers) is needed for information dissemination. In other words, the penalty from greenwashing and/or the proportion of green consumers in the society (i.e., probability of facing a green consumer) play a substitute role for the transmission of information.

Lemma 3. *The separating strategy profile in which the green firm does not label and the brown firm acquires a green label cannot be sustained as a PBE.*

As in the unilateral case, the dissemination of information about the firm's type does not induce the brown firm to acquire a green label. If the consumer is brown, the purchase is guaranteed with a no labeling strategy. And if the consumer is green, she does not buy the green-labeled product. For example, she is able to identify the case of greenwashing and, hence, it deters her from buying the good.

Proposition 5. *In the signaling game under bilateral uncertainty, a pooling PBE in which both types of firm acquire a green label can be sustained if and only if* $\beta \geq \beta_2(p) \equiv \frac{L_B}{p-K}$, *where* $\beta_2(p) \leq 1$ *if* $p \geq L_B + K$; *and if beliefs satisfy* $q \geq \overline{q}^G(p)$, *and off-equilibrium beliefs are* $\gamma < \overline{q}^G(p)$; *when* $p \in \{A\}$.

The uninformative equilibrium in which information transmission is hindered due to the emergence of greenwashing can be sustained in situations in which both the green firm and the green consumer are sufficiently likely, and a green label is the only message that induces the green consumer to buy the good. In this case, the brown firm has greater incentives to mimic the green firm given that the cost of acquiring a green label and potential penalties from greenwashing are compensated by the price (i.e., $p \in \{A\}$), and a high likely purchase by the green consumer ($\beta \geq \beta_2(p)$). The green consumer purchases the product after observing a green label given her strong beliefs of such signal coming from the green firm ($q \geq \overline{q}^G(p)$). Note that if the probability of the brown firm being found liable and been punished from greenwashing increases at a level that this cost (per unit sold) is not compensated by the price, ($p < K$), this pooling equilibrium cannot emerge.

Proposition 6. *In the signaling game under bilateral uncertainty, a pooling PBE in which both types of firm do not label can be sustained;*

1. *the green consumer only buys after observing no label if equilibrium beliefs satisfy $q \geq \overline{q}^G(p)$, under any off-the-equilibrium beliefs $\mu \in [0,1]$; the brown consumer buys regardless of the message she observes; when $p \in \{A\}$;*

2. *the green consumer does not buy after observing no label if and only if $\beta \leq \beta_1(p) \equiv \frac{L_G}{p}$, and if equilibrium beliefs satisfy $q < \overline{q}^G(p)$, under any off-the-equilibrium beliefs $\mu \in [0,1]$; the brown consumer buys regardless of the message she observes; when $p \in \{A\}$.*

This pooling equilibrium, where both the green and brown firm do not acquire a green label, is very likely to arise under bilateral uncertainty given that the brown consumer is willing to buy the good no matter the type of firm producing it. The first case emerges in situations where the green firm is likely (high priors, $q \geq \overline{q}^G(p)$); the green consumer assigns a higher probability of the no-label message coming from a green firm; so, the green firm obtains higher profits (purchase of green and brown consumer) by saving the labeling cost. In the second case, when the probability of facing a green consumer is low, $\beta \leq \beta_1(p)$, the green firm also obtains a higher profit by not labeling; the purchase of the likely brown consumer is guaranteed.

5 Conclusion

This paper investigates the role of uncertainty in the emergence of greenwashing. We show that when the firm knows with certainty the type of consumer (either green or brown), the information transmission is supported under larger conditions. However, when a second layer of uncertainty is considered (i.e., uninformed firm) a dirty brown firm has greater incentives to greenwash by acquiring a false green label.

Preliminary results support previous research that finds that a lax and uncertain regulation is a key driver of greenwashing. In addition, we find that a context of increasing environmental concerns

(i.e., a high proportion of green consumers), such as the current state, can be a determinant factor for the emergence of this practice. In the absence of an effective enforcement system aimed to battle greenwashing, only a heterogeneous society (composed by green and brown consumers) might block this fraudulent practice. In such a case, information transmission about the characteristics of green products is favored by the uncertainty faced by both firms and consumers. These results suggest that educational policies aimed to increase environmental awareness of brown consumers are not effective for information transmission, if they are not aligned with polices that encourage the transition from dirty to green technologies and/or a severe punishment system for doing greenwashing. Instead, a policy intended to empower green consumers and other stakeholders to better identify greenwashing practices and take action against them can help to increase the pressure to dirty firms and deter them from doing greenwashing.

Results from the bilateral setting need to be extended to other price regions, and to the setting in which both the green and the brown consumer are willing to sue the brown firm from doing greenwashing.

A further extension of the model in this paper could study how the results under bilateral uncertainty change by introducing a second period in which the consumer becomes informed about the greenwashing practices of the dirty firm. This extension would add an extra component of uncertainty to the firm; in the second period the firm is uninformed about the consumer's type and about the consumer's knowledge of the presence of greenwashing.

6 Appendix: Proofs

Proof of Lemma 1

Under complete information, the green and the brown consumer can perfectly observe the firm's type. As a consequence, they buy the good if the value the assign to it is greater than the price at which it is sold, i.e., $V_i^j > p$, where $i \in \{G, B\}$ denotes the firm's type and $j \in \{G, B\}$ represents the consumer's type: A) when $p \in \{A\}$, the green consumer buys the green good while the brown consumer buys the brown and the green good; B) when $p \in \{B\}$, both types of consumer only buy the green good; C) when $p \in \{C\}$, only the green consumer buys the green good; D) when $p \in \{D\}$, no type of consumer buy the good, neither green nor brown.

Proof of Proposition 1

- Price is in region A, B, C (i.e., $p \in \{A, B, C\}$)

The green firm acquires a green label and the brown firm does not label. Upon observing a label, the green consumer updates her beliefs by Bayes' rule, inferring that the firm must be green, i.e., $\mu = 1$. No label conveys the opposite information, i.e., $\gamma = 0$. After observing a label, the green consumer buys since $V_G^G > p$; and after observing no label, the green consumer does not buy since $V_B^G = 0 < p$.

Anticipating the green consumer's response, the green firm acquires a label if and only if $p - C_G - L_G \geq -C_G$, yielding $p \geq L_G$, which holds by definition. Similarly, the brown firm does not label if and only if $-C_B \geq p - C_B - L_B - K$, implying $p \leq L_B + K$. Therefore, this separating equilibrium can be supported as a PBE if and only if $p \leq L_B + K$ when $p \in \{A, B, C\}$.

The green firm does not label and the brown firm acquires a label. Upon observing no label, the green consumer updates her beliefs by Bayes' rule, inferring that the firm must be green, i.e., $\gamma = 1$. A label conveys the opposite information, i.e., $\mu = 0$. After observing no label, the green consumer buys since $V_G^G > p$; and after observing a label, the green consumer does not buy since $V_B^G = 0 < p$.

Anticipating the green consumer's response, the green firm does not label if and only if $p - C_G \geq -C_G - L_G$, yielding $p \geq -L_G$, which holds by definition. Similarly, the brown firm acquires label if and only if $-C_B - L_B \geq p - C_B$, implying $p \leq -L_B$, which is not satisfied by definition; and the brown firm has incentives to deviate towards no label. Therefore, this separating equilibrium cannot be sustained as a PBE when $p \in \{A, B, C\}$.

- Price is in region D (i.e., $p \in \{D\}$)

The green firm acquires a green label and the brown firm does not label. Upon observing a label, the green consumer updates her beliefs by Bayes' rule, inferring that the firm must be green, i.e., $\mu = 1$. No label conveys the opposite information, i.e., $\gamma = 0$. After observing a label, the green consumer does not buy since $V_G^G < p$; and after observing no label, the green consumer does not buy since $V_B^G = 0 < p$.

Anticipating the green consumer's response, the green firm acquires a label if and only if $-C_G - L_G \geq -C_G$, yielding $L_G \leq 0$, which is not satisfied by definition; and the green firm has incentives to deviate towards no label. Therefore, this separating equilibrium cannot be supported as a PBE when $p \in \{D\}$.

The green firm does not label and the brown firm acquires a label. Upon observing no label, the green consumer updates her beliefs by Bayes' rule, inferring that the firm must be green, i.e., $\gamma = 1$. A label conveys the opposite information, i.e., $\mu = 0$. After observing no label, the green consumer does not buy since $V_G^G < p$; and after observing no label, the green consumer does not buy since $V_B^G = 0 < p$.

Anticipating the green consumer's response, the green firm does not label if and only if $-C_G \geq -C_G - L_G$, yielding $L_G \geq 0$, which holds by definition. Similarly, the brown firm acquires label if and only if $-C_B - L_B \geq -C_B$, implying $L_B \leq 0$, which is not satisfied by definition; and the brown firm has incentives to deviate towards no label. Therefore, this separating equilibrium cannot be sustained as a PBE when $p \in \{D\}$.

Proof of Proposition 2

- Price is in region A, B, C (i.e., $p \in \{A, B, C\}$)

Both types of firm acquire a label. Upon observing the equilibrium message, the green consumer cannot further update her beliefs about the firm's type, yielding $\mu = q$ in equilibrium, and $\gamma \in [0, 1]$ off-the-equilibrium. After observing a label, the green consumer buys if and only if $qV_G^G + (1-q)V_B^G - p \geq 0$, which solving for q implies $q \geq \frac{p}{V_G^G} \equiv \overline{q}^G(p)$; and after observing no label, since the off-the-equillibrium beliefs satisfy $\gamma \in [0, 1]$, the green consumer buys if and only if $\gamma V_G^G + (1-\gamma)V_B^G - p \geq 0$, which solving for γ yields $\gamma \geq \frac{p}{V_G^G} \equiv \overline{q}^G(p)$; where $V_B^G = 0$, and $p \in \{A, B, C\}$.

We analyze four different cases depending on the green consumer's beliefs:

Case 1: Consumer's beliefs are $q, \gamma \geq \overline{q}^G(p)$, that is, the green consumer buys regardless of the message she observes.

Anticipating the green consumer's response, the green firm acquires a label rather than no label if and only if $p - C_G - L_G \geq p - C_G$, yielding $L_G \leq 0$, which is not satisfied; and the green firm has incentives to deviate towards no label. Therefore, this pooling equilibrium in case 1 cannot be sustained as a PBE when $p \in \{A, B, C\}$.

Case 2: Consumer's beliefs are $q \geq \overline{q}^G(p)$, and $\gamma < \overline{q}^G(p)$, that is, the green consumer buys after observing a label, and she does not buy otherwise.

Anticipating the green consumer's response, the green firm acquires a label rather than no label if and only if $p - C_G - L_G \geq -C_G$, implying $p \geq L_G$, which holds by definition. Similarly, the brown firm acquires a label rather than no label if and only if $p - C_B - L_B - K \geq -C_B$, which is satisfied if $p \geq L_B + K$. Therefore, this pooling equilibrium in case 2 can be sustained as a PBE if and only if $p \geq L_B + K$ when $p \in \{A, B, C\}$.

Case 3: Consumer's beliefs are $q < \overline{q}^G(p)$, and $\gamma \geq \overline{q}^G(p)$, that is, the green consumer does not buy after observing a label, and she buys otherwise.

Anticipating the green consumer's response, the green firm acquires a label rather than no label if and only if $-C_G - L_G \geq p - C_G$, yielding $p \leq -L_G$, which is not satisfied; and the green firm has incentives to deviate towards no label. Hence, this pooling equilibrium in case 3 cannot be supported as a PBE when $p \in \{A, B, C\}$.

Case 4: Consumer's beliefs are $q, \gamma < \overline{q}^G(p)$, that is, the green consumer does not buy regardless of the message she observes.

Anticipating the green consumer's response, the green firm acquires a label rather than no label if and only if $-C_G - L_G \geq -C_G$, implying $L_G \leq 0$, which does not hold by definition; and the green firm has incentives to deviate towards no label. Therefore, this pooling equilibrium in case 4 cannot be sustained as a PBE when $p \in \{A, B, C\}$.

Both types of firm do not label. Upon observing the equilibrium message, the green consumer cannot further update her beliefs about the firm's type, yielding $\gamma = q$ in equilibrium, and $\mu \in [0, 1]$ off-the-equilibrium. After observing no label, the green consumer buys if and only if $qV_G^G + (1 - q)V_B^G - p \geq 0$, which implies $q \geq \frac{p}{V_G^G} \equiv \overline{q}^G(p)$; and after observing a label, since the off-the-equillibrium beliefs satisfy $\mu \in [0, 1]$, the green consumer buys if and only if $\mu V_G^G + (1-\mu)V_B^G - p \geq 0$,

which yields $\mu \geq \frac{p}{V_G^G} \equiv \overline{q}^G(p)$; where $V_B^G = 0$, and $p \in \{A, B, C\}$.

We analyze four different cases depending on the green consumer's beliefs:

Case 1: Consumer's beliefs are $q, \mu \geq \overline{q}^G(p)$, that is, the green consumer buys regardless of the message she observes.

Anticipating the green consumer's response, the green firm does not label rather than acquiring a label if and only if $p - C_G \geq p - C_G - L_G$, yielding $L_G \geq 0$, which is satisfied by definition. Similarly, the brown firm does not label rather than acquiring a label if and only if $p - C_B \geq p - C_B - L_B - K$, implying $L_B + K \geq 0$, which holds. Hence, this pooling equilibrium in case 1 can be supported as a PBE when $p \in \{A, B, C\}$.

Case 2: Consumer's beliefs are $q \geq \overline{q}^G(p)$, and $\mu < \overline{q}^G(p)$, that is, the green consumer buys after observing no label, and she does not buy otherwise.

Anticipating the green consumer's response, the green firm does not label rather than acquiring a label if and only if $p - C_G \geq -C_G - L_G$, yielding $p \geq -L_G$, which holds by definition. Similarly, the brown firm does not label rather than acquiring a label if and only if $p - C_B \geq -C_B - L_B$, implying $p \geq -L_B$, which is satisfied. Therefore, this pooling equilibrium in case 2 can be supported as a PBE when $p \in \{A, B, C\}$.

Case 3: Consumer's beliefs are $q < \overline{q}^G(p)$, and $\mu \geq \overline{q}^G(p)$, that is, the green consumer does not buy after observing no label, and she buys otherwise.

Anticipating the green consumer's response, the green firm does not label rather than acquiring a label if and only if $-C_G \geq p - C_G - L_G$, implying $p \leq L_G$, which is not satisfied; and the green firm has incentives to deviate towards a label. Hence, this pooling equilibrium in case 3 cannot be sustained as a PBE when $p \in \{A, B, C\}$.

Case 4: Consumer's beliefs are $q, \mu < \overline{q}^G(p)$, that is, the green consumer does not buy regardless of the message she observes.

Anticipating the green consumer's response, the green firm does not label rather than acquiring a label if and only if $-C_G \geq -C_G - L_G$, implying $L_G \geq 0$, which holds by definition. Similarly, the brown firm does not label rather than acquiring a label if and only if $-C_B \geq -C_B - L_B$, yielding $L_B \geq 0$, which is satisfied. Therefore, this pooling equilibrium in case 4 can be supported as a PBE when $p \in \{A, B, C\}$.

- Price is in region D (i.e., $p \in \{D\}$)

Both types of firm acquire a label. Upon observing the equilibrium message, the green consumer cannot further update her beliefs about the firm's type, yielding $\mu = q$ in equilibrium, and $\gamma \in [0, 1]$ off-the-equilibrium. After observing a label, the green consumer buys if and only if $qV_G^G + (1-q)V_B^G - p \geq 0$, which solving for q implies $q \geq \frac{p}{V_G^G} \equiv \overline{q}^G(p)$; and after observing no label, since the off-the-equillibrium beliefs satisfy $\gamma \in [0, 1]$, the green consumer buys if and only if $\gamma V_G^G + (1-\gamma)V_B^G - p \geq 0$, which solving for γ yields $\gamma \geq \frac{p}{V_G^G} \equiv \overline{q}^G(p)$; where $V_B^G = 0$, and $p \in \{D\}$. Since $p > V_G^G$, $\overline{q}^G(p) > 1$; thus $q, \gamma \in [0, 1]$ is not satisfied, implying that the green consumer does not buy regardless of the message she observes.

Anticipating the green consumer's response, the green firm acquires a label rather than no label if and only if $-C_G - L_G \geq -C_G$, yielding $L_G \leq 0$, which is not satisfied by definition; and the green firm has incentives to deviate towards no label. Therefore, this pooling equilibrium cannot be sustained as a PBE when $p \in \{D\}$.

Both types of firm do not label. Upon observing the equilibrium message, the green consumer cannot further update her beliefs about the firm's type, yielding $\gamma = q$ in equilibrium, and $\mu \in [0,1]$ off-the-equilibrium. After observing no label, the green consumer buys if and only if $qV_G^G + (1-q)V_B^G - p \geq 0$, which implies $q \geq \frac{p}{V_G^G} \equiv \overline{q}^G(p)$; and after observing a label, since the off-the-equillibrium beliefs satisfy $\mu \in [0,1]$, the green consumer buys if and only if $\mu V_G^G + (1-\mu)V_B^G - p \geq 0$, which yields $\mu \geq \frac{p}{V_G^G} \equiv \overline{q}^G(p)$; where $V_B^G = 0$, and $p \in \{D\}$. Since $p > V_G^G$, $\overline{q}^G(p) > 1$; thus $q, \mu \in [0,1]$ is not satisfied, implying that the green consumer does not buy regardless of the message she observes.

Anticipating the green consumer's response, the green firm does not label rather than acquiring a label if and only if $-C_G \geq -C_G - L_G$, yielding $L_G \geq 0$, which is satisfied by definition. Similarly, the brown firm does not label rather than acquiring a label if and only if $-C_B \geq -C_B - L_B$, implying $L_B \geq 0$, which holds by definition. Hence, this pooling equilibrium can be sustained as a PBE when $p \in \{D\}$.

Intuitive Criterion (IC)

Case 1. In this pooling PBE, both types of firm acquire a label. We analyze which type of firm, if any, has incentives to deviate towards no label. First, the green firm obtains, at most, a profit of $p - C_G$ from deviating which is higher than its equilibrium payoff of $p - C_G - L_G$; implying that this type of firm has incentives to deviate. Second, the brown firm obtains, at most, a profit of $p - C_B$ from deviating towards no label, which also exceeds its equilibrium payoff; implying that this type of firm has incentives to deviate. In summary, both types of firm have incentives to deviate towards no label, entailing that the green consumer cannot further update her off-the-equilibrium beliefs $\gamma \in [0,1]$. Therefore, this pooling PBE in Proposition 2 (case 1) survives the Cho and Krep's Intuitive Criterion.

Case 2a. In this pooling PBE, both types of firm do not label, which is responded by a purchase of the green consumer. We analyze which type of firm, if any, has incentives to deviate towards a label. First, the green firm obtains at most a profit of $p - C_G - L_G$ from deviating, which can not not exceed its equilibrium payoff of $p - C_G$. Second, the brown firm obtains at most a profit of $p - C_B - L_B - K$, which is lower than its equilibrium payoff. In summary, no type of firm has incentives to deviate towards a label, entailing that the green consumer cannot further update her off-the-equilibrium beliefs $\mu \in [0,1]$. Therefore, this pooling PBE in Proposition 2 (Case 2a) survives the Intuitive Criterion.

Case 2b. In this pooling PBE, both types of firm do not label, which is now responded with no purchase by the green consumer. We analyze which type of firm, if any, has incentives to deviate towards a label. First, the green firm obtains at most a profit of $p - C_G - L_G$ from deviating, which

exceeds its equilibrium payoff of $-C_G$; implying that this type of firm has incentives to deviate. Second, the brown firm obtains at most a profit of $p - C_B - L_B - K$, which exceed its equilibrium payoff of $-C_B$ if $p \geq L_B + K$ is satisfied. Summarizing, when $p \geq L_B + K$ holds, both types of firm have incentives to deviate towards a label, entailing that the consumer cannot further update her off-the-equilibrium beliefs $\mu \in [0, 1]$, and the pooling PBE in Proposition 2 (Case 2b) survives the Cho and Krep's Intuitive Criterion. However, when $p < L_B + K$ holds, only the green firm has incentives to deviate, leading the green consumer to update their off-the-equilibrium beliefs to $\mu = 1$, thus responding with purchasing the good, and the pooling PBE in Proposition 2 (case 2b) violates the Intuitive Criterion.

Proof of Lemma 2.

- Price is in region A (i.e., $p \in \{A\}$)

The green firm acquires a green label and the brown firm does not label. Upon observing a label, the brown consumer updates her beliefs by Bayes' rule, inferring that the firm must be green, i.e., $\mu = 1$. No label conveys the opposite information, i.e., $\gamma = 0$. After observing a label, the brown consumer buys since $V_G^B > p$; and after observing no label, she also buys since $V_B^B > p$; where $p \in \{A\}$. That is, the brown consumer buys regardless of the message she observes. Anticipating the brown consumer's response, the green firm acquires a label if and only if $p - C_G - L_G \geq p - C_G$, yielding $L_G \leq 0$, which does not hold by definition; and the green firm has incentives to deviate towards no label. Therefore, this separating equilibrium cannot be sustained as a PBE when $p \in \{A\}$.

The green firm does not label and the brown firm acquires a label. Upon observing no label, the brown consumer updates her beliefs by Bayes' rule, inferring that the firm must be green, i.e., $\gamma = 1$. A label conveys the opposite information, i.e., $\mu = 0$. After observing no label, the brown consumer buys since $V_G^B > p$; and after observing a label, she also buys since $V_B^B > p$; where $p \in \{A\}$. Anticipating the brown consumer's response, the green firm does not label if and only if $p - C_G \geq p - C_G - L_G$, implying $L_G \geq 0$, which holds by definition. Similarly, the brown firm acquires a label if and only if $p - C_B - L_B \geq p - C_B$, yielding $L_B \leq 0$, which is not satisfied by definition; and the brown firm has incentives to deviate towards no label. Hence, this separating equilibrium cannot be supported as a PBE when $p \in \{A\}$.

- Price is in region B (i.e., $p \in \{B\}$)

The green firm acquires a green label and the brown firm does not label. Upon observing a label, the brown consumer updates her beliefs by Bayes' rule, inferring that the firm must be green, i.e., $\mu = 1$. No label conveys the opposite information, i.e., $\gamma = 0$. After observing a label, the brown consumer buys since $V_G^B > p$; and after observing no label, she does not buy since $V_B^B < p$; where $p \in \{B\}$. Anticipating the brown consumer's response, the green firm acquires a label if and only

if $p - C_G - L_G \geq -C_G$, yielding $p \geq L_G$, which holds by definition. Similarly, the brown firm does not label rather than acquiring a label if and only if $-C_B \geq p - C_B - L_B$, implying $p \leq L_B$, which is not satisfied; and the brown firm has incentives to deviate towards a label. Therefore, this separating equilibrium cannot be sustained as a PBE when $p \in \{B\}$.

The green firm does not label and the brown firm acquires a label. Upon observing no label, the brown consumer updates her beliefs by Bayes' rule, inferring that the firm must be green, i.e., $\gamma = 1$. A label conveys the opposite information, i.e., $\mu = 0$. After observing no label, the brown consumer buys since $V_G^B > p$; and after observing a label, she does not buy since $V_B^B, < p$; where $p \in \{B\}$. Anticipating the brown consumer's response, the green firm does not label if and only if $p - C_G \geq -C_G - L_G$, implying $p \geq -L_G$, which holds by definition. Similarly, the brown firm acquires a label if and only if $-C_B - L_B \geq p - C_B$, yielding $p \leq -L_B$, which is not satisfied; and the brown firm has incentives to deviate towards no label. Hence, this separating equilibrium cannot be supported as a PBE when $p \in \{B\}$.

- Price is in region C, D (i.e., $p \in \{C, D\}$)

The green firm acquires a green label and the brown firm does not label. Upon observing a label, the brown consumer updates her beliefs by Bayes' rule, inferring that the firm must be green, i.e., $\mu = 1$. No label conveys the opposite information, i.e., $\gamma = 0$. After observing a label, the brown consumer does not buy since $V_G^B < p$; and after observing no label, she does not buy since $V_B^B < p$; where $p \in \{C, D\}$. That is, the brown consumer does not buy regardless of the message she observes. Anticipating the brown consumer's response, the green firm acquires a label if and only if $-C_G - L_G \geq -C_G$, yielding $L_G \leq 0$, which does not holds by definition; and the green firm has incentives to deviate towards no label. Therefore, this separating equilibrium cannot be sustained as a PBE when $p \in \{C, D\}$.

The green firm does not label and the brown firm acquires a label. Upon observing no label, the brown consumer updates her beliefs by Bayes' rule, inferring that the firm must be green, i.e., $\gamma = 1$. A label conveys the opposite information, i.e., $\mu = 0$. After observing no label, the brown consumer does not buy since $V_G^B < p$; and after observing a label, she does not buy since $V_B^B > p$; where $p \in \{C, D\}$. Anticipating the brown consumer's response, the green firm does not label if and only if $-C_G \geq -C_G - L_G$, implying $L_G \geq 0$, which holds by definition. Similarly, the brown firm acquires a label if and only if $-C_B - L_B \geq -C_B$, yielding $L_B \leq 0$, which is not satisfied; and the brown firm has incentives to deviate towards no label. Hence, this separating equilibrium cannot be supported as a PBE when $p \in \{C, D\}$.

Proof of Corollary 1

We analyze the same two separating profiles than in Proposition 2 and check that when the brown consumer is able to sue the brown firm from doing greenwashing (i.e., allowing for $K \geq 0$),

the separating equilibrium in which the *green firm acquires a green label and the brown firm does not label* can be sustained as a PBE when $p \in \{B\}$. Consumer's beliefs are the same. Anticipating the brown consumer's response, the green firm acquires a label if and only if $p - C_G - L_G \geq -C_G$, yielding $p \geq L_G$, which holds by definition. Similarly, the brown firm does not label if and only if $-C_B \geq p - C_B - L_B - K$, implying $p \leq L_B + K$. Therefore, this separating equilibrium can be supported as a PBE if and only if $p \leq L_B + K$ when $p \in \{B\}$.

Proof of Proposition 3

- Price is in region A (i.e., $p \in \{A\}$)

Both types of firm acquire a label. Upon observing the equilibrium message, the brown consumer cannot further update her beliefs about the firm's type, yielding $\mu = q$ in equilibrium, and $\gamma \in [0,1]$ off-the-equilibrium. After observing a label, the brown consumer buys if and only if $qV_G^B + (1-q)V_B^B - p \geq 0$, which solving for q yields $q \geq \frac{p - V_B^B}{V_G^B - V_B^B} \equiv \overline{q}^B(p)$; and after observing no label, since the off-the-equillibrium beliefs satisfy $\gamma \in [0,1]$, the brown consumer buys if and only if $\gamma V_G^B + (1-\gamma)V_B^B - p \geq 0$, which solving for γ implies $\gamma \geq \frac{p - V_B^B}{V_G^B - V_B^B} \equiv \overline{q}^B(p)$; where $p \in \{A\}$. Since $p - V_B^B < 0$ and $V_G^B - V_B^B > 0$, $\overline{q}^B(p) < 0$; thus $q, \gamma \in [0,1]$ is always satisfied, and the brown consumer buys regardless of the message she observes.

Anticipating the brown consumer's response, the green firm acquires a label rather than no label if and only if $p - C_G - L_G \geq p - C_G$, yielding $L_G \leq 0$, which is not satisfied by definition; and the green firm has incentives to deviate towards no label. Therefore, this pooling equilibrium cannot be sustained as a PBE when $p \in \{A\}$.

Both types of firm do not label. Upon observing the equilibrium message, the brown consumer cannot further update her beliefs about the firm's type, yielding $\gamma = q$ in equilibrium, and $\mu \in [0,1]$ off-the-equilibrium. After observing no label, the brown consumer buys if and only if $qV_G^B + (1-q)V_B^B - p \geq 0$, which solving for q yields $q \geq \frac{p - V_B^B}{V_G^B - V_B^B} \equiv \overline{q}^B(p)$; and after observing a label, since the off-the-equillibrium beliefs satisfy $\mu \in [0,1]$, the brown consumer buys if and only if $\mu V_G^B + (1-\mu)V_B^B - p \geq 0$, which solving for μ implies $\mu \geq \frac{p - V_B^B}{V_G^B - V_B^B} \equiv \overline{q}^B(p)$; where $p \in \{A\}$. Since $p - V_B^B < 0$ and $V_G^B - V_B^B > 0$, $\overline{q}^B(p) < 0$; thus $q, \mu \in [0,1]$ is always satisfied, and the brown consumer buys regardless of the message she observes.

Anticipating the brown consumer's response, the green firm does not label rather than acquiring label if and only if $p - C_G \geq p - C_G - L_G$, yielding $L_G \geq 0$, which is satisfied by definition; and the green firm does not deviate. Similarly, the brown firm does not label rather than acquiring a label if and only if $p - C_B \geq p - C_B - L_B$, implying $L_B \geq 0$, which holds by definition; and the brown firm does not deviate. Therefore, this pooling equilibrium can be sustained as a PBE when $p \in \{A\}$.

- Price is in region B (i.e., $p \in \{B\}$)

Both types of firm acquire a label. Upon observing the equilibrium message, the brown consumer cannot further update her beliefs about the firm's type, yielding $\mu = q$ in equilibrium, and $\gamma \in [0,1]$ off-the-equilibrium. After observing a label, the brown consumer buys if and only if $qV_G^B + (1-q)V_B^B - p \geq 0$, which solving for q implies $q \geq \frac{p-V_G^B}{V_G^B - V_G^B} \equiv \overline{q}^B(p)$; and after observing no label, since the off-the-equillibrium beliefs satisfy $\gamma \in [0,1]$, the brown consumer buys if and only if $\gamma V_G^B + (1-\gamma)V_B^B - p \geq 0$, which solving for γ yields $\gamma \geq \frac{p-V_G^B}{V_G^B - V_G^B} \equiv \overline{q}^B(p)$; where $p \in \{B\}$.

We analyze four different cases depending on the brown consumer's beliefs:

Case 1: Consumer's beliefs are $q, \gamma \geq \overline{q}^B(p)$, that is, the brown consumer buys regardless of the message she observes.

Anticipating the brown consumer's response, the green firm acquires a label rather than no label if and only if $p - C_G - L_G \geq p - C_G$, yielding $L_G \leq 0$, which is not satisfied; and the green firm has incentives to deviate towards no label. Therefore, this pooling equilibrium in case 1 cannot be sustained as a PBE when $p \in \{B\}$.

Case 2: Consumer's beliefs are $q \geq \overline{q}^B(p)$, and $\gamma < \overline{q}^B(p)$, that is, the brown consumer buys after observing a label, and she does not buy otherwise.

Anticipating the brown consumer's response, the green firm acquires a label rather than no label if and only if $p - C_G - L_G \geq -C_G$, implying $p \geq L_G$, which holds by definition. Similarly, the brown firm acquires a label rather than no label if and only if $p - C_B - L_B \geq -C_B$, which holds. Therefore, this pooling equilibrium in case 2 can be sustained as a PBE when $p \in \{B\}$.

Case 3: Consumer's beliefs are $q < \overline{q}^B(p)$, and $\gamma \geq \overline{q}^B(p)$, that is, the brown consumer does not buy after observing a label, and she buys otherwise.

Anticipating the brown consumer's response, the green firm acquires a label rather than no label if and only if $-C_G - L_G \geq p - C_G$, yielding $p \leq -L_G$, which is not satisfied; and the green firm has incentives to deviate towards no label. Hence, this pooling equilibrium in case 3 cannot be supported as a PBE when $p \in \{B\}$.

Case 4: Consumer's beliefs are $q, \gamma < \overline{q}^B(p)$, that is, the brown consumer does not buy regardless of the message she observes.

Anticipating the brown consumer's response, the green firm acquires a label rather than no label if and only if $-C_G - L_G \geq -C_G$, implying $L_G \leq 0$, which does not hold by definition; and the green firm has incentives to deviate towards no label. Therefore, this pooling equilibrium in case 4 cannot be sustained as a PBE when $p \in \{B\}$.

Both types of firm do not label. Upon observing the equilibrium message, the brown consumer cannot further update her beliefs about the firm's type, yielding $\gamma = q$ in equilibrium, and $\mu \in [0,1]$ off-the-equilibrium. After observing no label, the brown consumer buys if and only if $qV_G^B + (1-q)V_B^B - p \geq 0$, which implies $q \geq \frac{p-V_B^B}{V_G^B - V_B^B} \equiv \overline{q}^B(p)$; and after observing a label, since the off-the-equillibrium beliefs satisfy $\mu \in [0,1]$, the brown consumer buys if and only if $\mu V_G^B + (1-\mu)V_B^B - p \geq 0$, which yields $\mu \geq \frac{p-V_B^B}{V_G^B - V_B^B} \equiv \overline{q}^B(p)$; where $p \in \{B\}$.

We analyze four different cases depending on the brown consumer's beliefs:

Case 1: Consumer's beliefs are $q, \mu \geq \overline{q}^B(p)$, that is, the brown consumer buys regardless of the message she observes.

Anticipating the brown consumer's response, the green firm does not label rather than acquiring a label if and only if $p - C_G \geq p - C_G - L_G$, yielding $L_G \geq 0$, which is satisfied by definition. Similarly, the brown firm does not label rather than acquiring a label if and only if $p - C_B \geq p - C_B - L_B$, which is satisfied. Hence, this pooling equilibrium in case 1 can be supported as a PBE when $p \in \{B\}$.

Case 2: Consumer's beliefs are $q \geq \overline{q}^B(p)$, and $\mu < \overline{q}^B(p)$, that is, the brown consumer buys after observing no label, and she does not buy otherwise.

Anticipating the brown consumer's response, the green firm does not label rather than acquiring a label if and only if $p - C_G \geq -C_G - L_G$, yielding $p \geq -L_G$, which holds by definition. Similarly, the brown firm does not label rather than acquiring a label if and only if $p - C_B \geq -C_B - L_B$, implying $p \geq -L_B$, which is satisfied. Therefore, this pooling equilibrium in case 2 can be supported as a PBE when $p \in \{B\}$.

Case 3: Consumer's beliefs are $q < \overline{q}^B(p)$, and $\mu \geq \overline{q}^B(p)$, that is, the brown consumer does not buy after observing no label, and she buys otherwise.

Anticipating the brown consumer's response, the green firm does not label rather than acquiring a label if and only if $-C_G \geq p - C_G - L_G$, implying $p \leq L_G$, which is not satisfied; and the brown firm has incentives to deviate towards a label. Hence, this pooling equilibrium in case 3 cannot be sustained as a PBE when $p \in \{B\}$.

Case 4: Consumer's beliefs are $q, \mu < \overline{q}^B(p)$, that is, the brown consumer does not buy regardless of the message she observes.

Anticipating the brown consumer's response, the green firm does not label rather than acquiring a label if and only if $-C_G \geq -C_G - L_G$, implying $L_G \geq 0$, which holds by definition. Similarly, the brown firm does not label rather than acquiring a label if and only if $-C_B \geq -C_B - L_B$, yielding $L_B \geq 0$, which is satisfied. Therefore, this pooling equilibrium in case 4 can be supported as a PBE when $p \in \{B\}$.

- Price is in region C, D (i.e., $p \in \{C, D\}$)

Both types of firm acquire a label. Upon observing the equilibrium message, the brown consumer cannot further update her beliefs about the firm's type, yielding $\mu = q$ in equilibrium, and $\gamma \in [0, 1]$ off-the-equilibrium. After observing a label, the brown consumer buys if and only if $qV_G^B + (1 - q)V_B^B - p \geq 0$, which solving for q implies $q \geq \frac{p - V_B^B}{V_G^B - V_B^B} \equiv \overline{q}^B(p)$; and after observing no label, since the off-the-equillibrium beliefs satisfy $\gamma \in [0, 1]$, the brown consumer buys if and only if $\gamma V_G^B + (1 - \gamma)V_B^B - p \geq 0$, which solving for γ yields $\gamma \geq \frac{p - V_B^B}{V_G^B - V_B^B} \equiv \overline{q}^B(p)$; where $p \in \{C, D\}$. Since $p > V_G^B$, $\overline{q}^B(p) > 1$; thus $q, \gamma \in [0, 1]$ is not satisfied, implying that the brown consumer does not buy regardless of the message she observes.

Anticipating the brown consumer's response, the green firm acquires a label rather than no label if and only if $-C_G - L_G \geq -C_G$, yielding $L_G \leq 0$, which is not satisfied by definition; and the

green firm has incentives to deviate towards no label. Therefore, this pooling equilibrium cannot be sustained as a PBE when $p \in \{C, D\}$.

Both types of firm do not label. Upon observing the equilibrium message, the brown consumer cannot further update her beliefs about the firm's type, yielding $\gamma = q$ in equilibrium, and $\mu \in [0, 1]$ off-the-equilibrium. After observing no label, the brown consumer buys if and only if $qV_G^B + (1 - q)V_B^B - p \geq 0$, which implies $q \geq \frac{p - V_B^B}{V_G^B - V_B^B} \equiv \overline{q}^B(p)$; and after observing a label, since the off-the-equillibrium beliefs satisfy $\mu \in [0, 1]$, the brown consumer buys if and only if $\mu V_G^B + (1-\mu)V_B^B - p \geq 0$, which yields $\mu \geq \frac{p - V_B^B}{V_G^B - V_B^B} \equiv \overline{q}^B(p)$; where $p \in \{C, D\}$. Since $p > V_G^B$, $\overline{q}^B(p) > 1$; thus $q, \mu \in [0, 1]$ is not satisfied, implying that the brown consumer does not buy regardless of the message she observes.

Anticipating the brown consumer's response, the green firm does not label rather than acquiring a label if and only if $-C_G \geq -C_G - L_G$, yielding $L_G \geq 0$, which is satisfied by definition. Similarly, the brown firm does not label rather than acquiring a label if and only if $-C_B \geq -C_B - L_B$, implying $L_B \geq 0$, which holds by definition. Hence, this pooling equilibrium can be sustained as a PBE when $p \in \{C, D\}$.

Intuitive Criterion (IC)

Case 1. In this pooling PBE, both types of firm acquire a label. We analyze which type of firm, if any, has incentives to deviate towards no label. First, the green firm obtains, at most, a profit of $p - C_G$ from deviating which is higher than its equilibrium payoff of $p - C_G - L_G$; implying that this type of firm has incentives to deviate. Second, the brown firm obtains, at most, a profit of $p - C_B$ from deviating towards no label, which also exceeds its equilibrium payoff; implying that this type of firm has incentives to deviate. In summary, both types of firm have incentives to deviate towards no label, entailing that the brown consumer cannot further update her off-the-equilibrium beliefs $\gamma \in [0, 1]$. Therefore, this pooling PBE in Proposition 3 (case 1) survives the Cho and Krep's Intuitive Criterion.

Case 2a. In this pooling PBE, both types of firm do not label, which is responded by a purchase of the brown consumer. We analyze which type of firm, if any, has incentives to deviate towards a label. First, the green firm obtains at most a profit of $p - C_G - L_G$ from deviating, which does not exceed its equilibrium payoff of $p - C_G$. Second, the brown firm obtains at most a profit of $p - C_B - L_B$, which is lower than its equilibrium payoff of $p - C_B$. In summary, no type of firm has incentives to deviate towards a label, entailing that the brown consumer cannot further update her off-the-equilibrium beliefs $\mu \in [0, 1]$. Therefore, this pooling PBE in Proposition 3 (case 2a) survives the Intuitive Criterion.

Case 2b. In this pooling PBE, both types of firm do not label, which is now responded with no purchase by the brown consumer. We analyze which type of firm, if any, has incentives to deviate towards a label. First, the green firm obtains at most a profit of $p - C_G - L_G$ from deviating, which exceeds its equilibrium payoff of $-C_G$; implying that this type of firm has incentives to deviate. Second, the brown firm obtains at most a profit of $p - C_B - L_B$, which exceed its equilibrium payoff

of $-C_B$. In summary, both types of firm have incentives to deviate towards a label, entailing that the brown consumer cannot further update her off-the-equilibrium beliefs $\mu \in [0,1]$. Therefore, this pooling PBE in Proposition 2 (Case 2aii) survives the Intuitive Criterion.

Proof of Corollary 2.

We analyze the same two pooling profiles than in Proposition 3 allowing the brown consumer to sue the brown firm from doing greenwashing (i.e., $K \geq 0$) and check which of the cases identified as PBE are affected by this condition.

Both types of firm acquire a label. We check that when beliefs are $q \geq \frac{p-V_B^B}{V_G^B-V_B^B} \equiv \overline{q}^B(p)$, and off-the-equillibrium beliefs are $\gamma < q(p)^G$; this pooling equilibrium can only be sustained as a PBE if $p \geq L_B + K$ when $p \in \{B\}$.

Both types do not label. We check that this pooling equilibrium can only be supported as a PBE if $p \geq L_B + K$ for $q, \mu < \overline{q}^B(p)$ when $p \in \{B\}$, and for all $q, \mu \in [0,1]$ when $p \in \{C, D\}$.

Proof of Proposition 4

- Price is in region A (i.e., $p \in \{A\}$)

The green firm acquires a green label and the brown firm does not label. Upon observing a label, the consumer updates her beliefs by Bayes's rule, inferring that the firm must be green, i.e., $\mu = 1$. No label conveys the opposite information, i.e., $\gamma = 0$. After observing a label, the green and the brown consumer buy since $V_G^G > p$, and $V_G^B > p$, respectively. And after observing no label, the green consumer does not buy since $V_G^B = 0 < p$, and the brown consumer buys since $V_B^B > p$.

Anticipating the consumer's response, the green firm acquires a label if and only if $\beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G) \geq \beta(-C_G) + (1-\beta)(p - C_G)$, yielding $\beta \geq \frac{L_G}{p} \equiv \beta_1(p)$. Similarly, the brown firm does not label if and only if $\beta(-C_B)+(1-\beta)(p-C_B) \geq \beta(p-C_B-L_B)+(1-\beta)(C_B-L_B)$, implying $\beta \leq \frac{L_B}{p-K} \equiv \beta_2(p)$. We check that $\beta_1(p)$,$\beta_2(p) \geq 0$ since L_G, $L_B \geq 0$ hold, respectively. And we check that $\beta_1(p) \leq 1$, since $p \geq L_G$ is satisfied by definition; and $\beta_2(p) \leq 1$ if $p \geq L_B + K$. In addition, we know $\beta_1(p) \leq \beta_2(p)$ since $p \geq \frac{-L_G K}{(L_B - L_G)}$ is satisfied because by definition $p > 0$.

Hence, this separating equilibrium can be sustained as a PBE if and only if $\beta \in [\beta_1(p) \equiv \frac{L_G}{p}, \beta_2(p) \equiv \frac{L_B}{p-K}]$, where $\beta_2(p) \leq 1$ if $p \geq L_B + K$ when $p \in \{A\}$.

The green firm does not label and the brown firm acquires a label. Upon observing no label, the consumer updates her beliefs by Bayes's rule, inferring that the firm must be green, i.e., $\gamma = 1$. A label conveys the opposite information, i.e., $\mu = 0$. After observing no label, the green and the brown consumer buy since $V_G^G > p$, and $V_G^B > p$, respectively. And after observing a label, the green consumer does not buy since $V_G^B = 0 < p$, and the brown consumer buys since $V_B^B > p$.

Anticipating the consumer's response, the green firm does not label if and only if $\beta(p - C_G) + (1-\beta)(p - C_G) \geq \beta(-C_G - L_G) + (1-\beta)(p - C_G - L_G)$, yielding $\beta \geq \frac{-L_G}{p}$, which is satisfied for all $\beta \in [0,1]$. Similarly, the brown firm acquires a label if and only if $\beta(-C_B - L_B) + (1-\beta)(p -$

$C_B - L_B) \geq \beta(p - C_B) + (1 - \beta)(p - C_B)$, implying $\beta \leq \frac{-L_B}{p}$, which does not hold by definition; and the brown firm has incentives to deviate towards no label.

Hence, this separating equilibrium cannot be sustained as a PBE when $p \in \{A\}$.

Proof of Proposition 5

- Price is in region A (i.e., $p \in \{A\}$)

Both types of firm acquire a label. Upon observing the equilibrium message, the consumer cannot further update her beliefs about the firm's type, yielding $\mu = q$ in equilibrium; and $\gamma \in [0, 1]$ off-the-equilibrium. After observing a label, the green consumer buys if and only if $qV_G^G + (1-q)V_B^G - p \geq 0$, which solving for q implies $q \geq \frac{p}{V_G^G} \equiv \overline{q}^G(p)$; where $V_B^G = 0$; and the brown consumer buys if and only if $qV_G^B + (1 - q)V_B^B - p \geq 0$, which solving for q yields $q \geq \frac{p - V_B^B}{V_G^B - V_B^B} \equiv \overline{q}^B(p)$; where $p \in \{A\}$. After observing no label, since the off-the-equillibrium beliefs satisfy $\gamma \in [0, 1]$, the green consumer buys if and only if $\gamma \geq \overline{q}^G(p)$; and the brown consumer buys if $\gamma \geq \overline{q}^B(p)$. Since $p - V_B^B < 0$ and $V_G^B - V_B^B > 0$, $\overline{q}^B(p) < 0$; thus $q, \gamma \in [0, 1]$ is always satisfied, and the brown consumer always buys regardless of the message she observes.

We analyze four different cases depending on the consumer's beliefs:

Case 1: Consumer's beliefs are $q, \gamma \geq \overline{q}^G(p)$, that is, the green and the brown consumer buy regardless of the message they observe.

Anticipating the consumer's response, the green firm acquires a label rather than no label if and only if $\beta(p - C_G - L_G) + (1 - \beta)(p - C_G - L_G) \geq \beta(p - C_G) + (1 - \beta)(p - C_G)$, yielding $L_G \leq 0$, which is not satisfied; and the green firm has incentives to deviate towards no label. Therefore, this pooling equilibrium in case 1 cannot be sustained as a PBE when $p \in \{A\}$.

Case 2: Consumer's beliefs are $q \geq \overline{q}^G(p)$and $\gamma < \overline{q}^G(p)$, that is, the green consumer buys after observing a label, and she does not buy otherwise; the brown consumer buys regardless of the message she observes.

Anticipating the consumers' response, the green firm acquires a label rather than no label if and only if $\beta(p - C_G - L_G) + (1 - \beta)(p - C_G - L_G) \geq \beta(-C_G) + (1 - \beta)(p - C_G)$, which is satisfied if $\beta \geq \frac{L_G}{p} \equiv \beta_1(p)$. Similarly, the brown firm acquires a label rather than no label if and only if $\beta(p - C_B - L_B - K) + (1 - \beta)(p - C_B - L_B) \geq \beta(-C_B) + (1 - \beta)(p - C_B)$, which holds if $\beta \geq \frac{L_B}{p-K} \equiv \beta_2(p)$. We know $\beta_2(p) \geq \beta_1(p)$. Therefore, this pooling equilibrium in case 2 can be supported as a PBE if and only if $\beta \geq \beta_2(p) \equiv \frac{L_B}{p-K}$, where $\beta_2(p) \leq 1$ if $p \geq L_B + K$ when $p \in \{A\}$.

Case 3: Consumer's beliefs are $q < \overline{q}^G(p)$and $\gamma \geq \overline{q}^G(p)$, that is, the green consumer does not buy after observing a label, and she buys otherwise; the brown consumer buys regardless of the message she observes.

Anticipating the consumers' response, the green firm acquires a label rather than no label if and only if $\beta(-C_G - L_G) + (1 - \beta)(p - C_G - L_G) \geq \beta(p - C_G) + (1 - \beta)(p - C_G)$, yielding $\beta \leq \frac{-L_G}{p}$, which is not satisfied by definition; and the green firm has incentives to deviate towards no label. Hence, this pooling equilibrium in case 3 cannot be sustained as a PBE when $p \in \{A\}$.

Case 4: Consumer's beliefs are $q, \gamma < \overline{q}^G(p)$, that is, the green consumer does not buy regardless of the message she observes; the brown consumer buys regardless of the message she observes.

Anticipating the consumers' response, the green firm acquires a label rather than no label if and only if $\beta(-C_G - L_G) + (1-\beta)(p - C_G - L_G) \geq \beta(-C_G) + (1-\beta)(p - C_G)$, yielding $L_G \leq 0$, which is not satisfied by definition; and the green firm has incentives to deviate towards no label. Hence, this pooling equilibrium in case 4 cannot be sustained as a PBE when $p \in \{A\}$.

Intuitive Criterion (IC)

In this pooling PBE, both firms acquire a label. The green and the brown firm obtain a higher payoff from deviating towards no label than their equilibrium payoff. Hence, this pooling PBE in Proposition 5 survives the Cho and Krep's Intuitive Criterion.

Proof of Proposition 6

Both types of firm do not label. Upon observing the equilibrium message, the consumer cannot further update her beliefs about the firm's type, yielding $\gamma = q$ in equilibrium, and $\mu \in [0,1]$ off-the-equilibrium. After observing no label, the green consumer buys if and only if $qV_G^G + (1-q)V_B^G - p \geq 0$, which solving for q implies $q \geq \frac{p}{V_G^G} \equiv \overline{q}^G(p)$; and the brown consumer buys if and only if $qV_G^B + (1-q)V_B^B - p \geq 0$, which solving for q yields $q \geq \frac{p - V_B^B}{V_G^B - V_B^B} \equiv \overline{q}^B(p)$. After observing a label, since the off-the-equillibrium beliefs satisfy $\mu \in [0,1]$, the green consumer buys if and only if $\mu \geq \overline{q}^G(p)$; and the brown consumer buys if $\mu \geq \overline{q}^B(p)$. Since $p - V_B^B < 0$ and $V_G^B - V_B^B > 0$, $\overline{q}^B(p) < 0$; thus $q, \mu \in [0,1]$ is always satisfied, and the brown consumer buys regardless of the message she observes when $p \in \{A\}$.

We analyze four different cases depending on consumers' beliefs:

Case 1: Consumers' beliefs are $q, \mu \geq \overline{q}^G(p)$, that is, the green and the brown consumer buy regardless of the message they observe.

Anticipating the consumers' response, the green firm does not label rather than acquiring a label if and only if $\beta(p - C_G) + (1-\beta)(p - C_G) \geq \beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G)$, yielding $L_G \geq 0$, which is satisfied. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(p - C_B) + (1-\beta)(p - C_B) \geq \beta(p - C_B - L_B - K) + (1-\beta)(p - C_B - L_B)$, implying $\beta \geq \frac{-L_B}{K}$, which holds for all $\beta \in [0,1]$. Therefore, this pooling equilibrium in case 1 can be sustained as a PBE when $p \in \{A\}$.

Case 2: Consumers' beliefs are $q \geq \overline{q}^G(p)$and $\mu < \overline{q}^G(p)$, that is, the green consumer buys after observing no label, and she does not buy otherwise; the brown consumer buys regardless of the message she observes.

Anticipating the consumers' response, the green firm does not label rather than acquiring a label if and only if $\beta(p - C_G) + (1-\beta)(p - C_G) \geq \beta(-C_G - L_G) + (1-\beta)(p - C_G - L_G)$, yielding $\beta \geq \frac{-L_G}{p}$, which is satisfied for all $\beta \in [0,1]$. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(p - C_B) + (1-\beta)(p - C_B) \geq \beta(-C_B - L_B) + (1-\beta)(p - C_B - L_B)$, implying $\beta \geq \frac{-L_B}{p}$, which holds for all $\beta \in [0,1]$. Therefore, this pooling equilibrium in case 2 can be supported as a PBE when $p \in \{A\}$.

Case 3: Consumers' beliefs are $q < \overline{q}^G(p)$ and $\mu \geq \overline{q}^G(p)$, that is, the green consumer does not buy after observing no label, and she buys otherwise; the brown consumer buys regardless of the message she observes.

Anticipating the consumers' response, the green firm does not label rather than acquiring a label if and only if $\beta(-C_G) + (1-\beta)(p - C_G) \geq \beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G)$, which is satisfied if $\beta \leq \frac{L_G}{p} \equiv \beta_1(p)$. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(-C_B) + (1-\beta)(p - C_B) \geq \beta(p - C_B - L_B - K) + (1-\beta)(p - C_B - L_B)$, which holds if $\beta \leq \frac{L_B}{p-K} \equiv \beta_2(p)$. We know $\beta_1(p) \leq \beta_2(p)$. Hence, this pooling equilibrium in case 3 can be supported as a PBE if and only if $\beta \leq \beta_1(p) \equiv \frac{L_G}{p}$ when $p \in \{A\}$.

Case 4: Consumers' beliefs are $q, \mu < \overline{q}^G(p)$, that is, the green consumer does not buy regardless of the message she observes; the brown consumer buys regardless of the message she observes.

Anticipating the consumers' response, the green firm does not label rather than acquiring a label if and only if $\beta(-C_G)+(1-\beta)(p-C_G) \geq \beta(-C_G-L_G)+(1-\beta)(p-C_G-L_G)$, yielding $L_G \geq 0$, which is satisfied; and the green firm does not deviate. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(-C_B) + (1-\beta)(p - C_B) \geq \beta(-C_B - L_B) + (1-\beta)(p - C_B - L_B)$, implying $L_B \geq 0$, which holds by definition; and the brown firm does not deviate. Hence, this pooling equilibrium in case 4 can be sustained as a PBE when $p \in \{A\}$.

Intuitive Criterion (IC)

In this pooling PBE the green and the brown firm obtain a lower payoff from deviating towards a label than their equilibrium payoff. Hence, this pooling PBE in Proposition 6 (cases 1 and 2) survives the Cho and Krep's Intuitive Criterion.

References

[1] Berrone, P., Fosfuri, A., and Gelabert, L. (2017) "Does Greenwashing Pay Off? Understanding the Relationship Between Environmental Actions and Environmental Legitimacy." *Journal of Business Ethics*, 144:363-379.

[2] Chatterje, K. and Samuelson, L. (1983) "Bargaining under Two-Sided Incomplete Information: The Unrestricted Offers Case." *Operations Research*, 36(4):605-618.

[3] Cho, I.K. and Kreps, D.M. (1987) "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics,* 102(2):179-222.

[4] Cone Communications (2013) *Green Gap Trend Tracker.* Available at: http://www.conecomm.com/research-blog/2013-cone-communications-green-gap-trend-tracker (accessed Mar 4, 2019)

[5] Connelly, B.L., Certo, S.T., Ireland, R.D., and Reutzel, C.R. (2011) "Signaling Theory: A Review and Assessment." *Journal of Management,* 37(1):39-67.

[6] Cramton, P.C. (1992) "Strategic Delay in Bargaining with Two-Sided Uncertainty." *The Review of the Economic Studies*, 59(1):205-225.

[7] Delmas, M.A. and Burbano, V.C. (2011) "The Drivers of Greenwashing." *California Management Review*, 54(1):64-87.

[8] Ecolabel Index (2019). Available at: http://www.ecolabelindex.com/ (accessed March 15, 2019)

[9] European Commission (2014) *Consumer Market Study on Environmental Claims for Non-Food Products*. Directorate-General for Justice and Consumers. Available at: https://ec.europa.eu/info/sites/info/files/study_on_environmetal_claims_for_non_food_products_2014_

[10] Gale, D. (2001) "Signaling in markets with two-sided adverse selection." *Economic Theory*, 18:391-414.

[11] Gruere, G. (2013) "A Characterization of Environmental Labelling and Information Schemes." *OECD Environment Working Papers*, 2.

[12] Hamilton, S.F. and Zilberman, D. (2006) "Green markets, eco-certification, and equilibrium fraud." *Journal of Environmental Economics and Management*, 52:627-644.

[13] Lorance, A. (2010) "An Assessment of U.S. Responses to Greenwashing and Proposals to Improve Enforcement." *Hofstra Law Student Works*, Paper 3.

[14] Lyon, T.P. and Maswell, J.W. (2011) "Greenwash: Corporate Environmental Disclosure under Threat of Audit." *Journal of Economics and Management Strategy*, 20(1):3-41.

[15] Lyon, T.P. and Montgomery, A.W. (2015) "The Means and End of Greenwash." *Organization & Environment*, 28(2):223-249.

[16] Mahenc, P. (2008) "Signaling the environmental performance of polluting products to green consumers." *International Journal of Industrial Organization*, 26:59-68.

[17] Mahenc, P. (2017) "Honest versus misleading certification." *Journal of Economics & Management Strategy*, 26(2):454-483.

[18] Nielsen (2018) "Was 2018 the Year of the Influential Sustainable Consumer?." *CPG, FMCG & Retail Article,* 12-17-2018. Available at: https://www.nielsen.com/us/en/insights/news/2018/was-2018-the-year-of-the-influential-sustainable-consumer.html (accessed March 4, 2019)

[19] Terrachoice (2010) *The sins of greenwashing. Home and family edition 2010. A report of environmental claims made in North America consumer market.* Northbrook, IL: Uncerwriters Laboratories.

[20] Seele, P. and Gatti, L. (2017) "Greenwashing Revisited: In Search of a Typology and Accusation-Based Definition Incorporating Legitimacy Strategies." *Business Strategy and the Environment*, 26:239-252.

[21] Volle, A. (2017) "Why is price useless to signal environmental quality?." *FAERE Working Paper*, 2017.30.