TI 2026-019/VIII
Tinbergen Institute Discussion Paper

Tier count in traffic-light eco-labels

Erik Ansink[1]
Frederic Klapwijk[2]

1 Vrije Universiteit Amsterdam, Tinbergen Institute

2 Vrije Universiteit Amsterdam

Tier count in traffic-light eco-labels

March 31, 2026

Abstract

We study whether the number of tiers in a traffic-light eco-label affects consumer welfare. We develop a model in which consumers observe price, quality, and the label tier, but infer a product's environmental quality ("greenness") imperfectly. In that setting, adding a tier can improve the information content of the label while still lowering welfare near newly created cutoffs. We then use a discrete choice experiment on trash bags with two-tier and three-tier traffic-light labels to test the behavioral ingredients of the model. Estimated treatment effects in the choice experiment show that adding a middle tier makes respondents more likely to choose the highest label tier, increases price sensitivity, and reduces the use of price as a cue for greenness. We interpret these results as evidence that tier count changes both the choice response to the highest label tier and price-based inference. A structural interpretation also yields a positive estimate of the price-proxy parameter, and the three-tier choice data imply a positive but modest separation between the yellow and green tiers under transparent normalizations.

Keywords: Eco-labels; traffic-light labels; consumer inference; environmental disclosure; consumer welfare
JEL Codes: D12; D83; L15; Q58

Data Availability Statement: The data and code underlying this study will be deposited in a public repository upon acceptance of the manuscript. During the review process, they are available from the authors upon request.

1 Introduction

The number of tiers in a traffic-light label is a policy choice. It is often treated as if adding tiers merely refines information, but that conclusion does not follow when consumers also infer greenness from price and other product cues. This paper studies that design problem as a question of environmental information policy: when does finer categorical disclosure improve choices, and when can it backfire? Traffic-light style schemes are common in both food and environmental settings, ranging from front-of-pack nutrition labels to energy ratings. Real-world schemes also differ in the number and structure of tiers, and those structures are revised over time. For example, EU energy labels were rescaled in 2021 from the old A+++ to D format back to a simpler A to G scale, and the revised EU tyre label uses five classes rather than seven (Crosetto et al., 2016; Carlsson et al., 2022).

The starting point is simple. Consumers observe a product's price, quality, and label tier, but not its true greenness. They therefore form beliefs using both the formal label and other observable cues. In such an environment, an additional tier can make the label more precise while also creating sharper tier boundaries. If consumers overreact to those tier boundaries, especially for products close to a newly introduced cutoff, the new tier can worsen choices rather than improve them. The overall welfare effect is therefore ambiguous (Allcott and Taubinsky, 2015; Lusk and Marette, 2012; Mason, 2011). The paper's central claim is that tier count is a welfare-relevant policy variable, not that finer disclosure is generically harmful. More broadly, the mechanism rests on the idea that consumers use observable signals such as price to infer otherwise hard-to-observe product attributes (Roth and Diamantopoulos, 2009; Kirmani and Rao, 2000; Cagé and Rouzet, 2015).

A simple example illustrates the mechanism. Suppose two products are nearly identical in true greenness, with greenness scores of 0.18 and 0.22, and that the greener product is slightly more expensive. A coarse two-tier scheme may place both products in the same tier, so the consumer chooses the cheaper option. Under a three-tier scheme, however, the new cutoff may be located at 0.20, implying that the product at 0.22 may jump into a visibly higher tier while the one at 0.18 remains in the lower tier. Holding prices fixed, if that extra label distinction is large enough to raise the perceived utility of the greener product above that of the cheaper one, the consumer switches. Yet the true greenness difference may still be too small to justify the price premium in realized utility terms. In that case, finer disclosure lowers realized utility even though it raises perceived utility.

The paper makes two main contributions. First, it offers a theory in which adding a tier can lower realized consumer welfare near a newly created cutoff. The mechanism is not that finer labels become less informative on average, but that sharper tier boundaries can move beliefs by more than they move true information. Second, it brings that mechanism to the data with a two-tier versus three-tier choice experiment on trash bags. The empirical analysis is designed to test the mechanism rather than to estimate welfare directly, and a structural interpretation recovers the price-proxy parameter and normalized tier gaps under transparent normalizations.

That theory-to-empirics link is straightforward. If adding a middle tier makes the highest tier more behaviorally salient, respondents should become more likely to choose it. If the extra tier reduces reliance on informal market cues, price should matter less as a signal of greenness and more through its direct role in choice. The experiment is built to test exactly those margins: the choice response to the highest label tier, the role of price in choice, and the use of price as a cue for greenness. The results support that mechanism: adding a middle tier increases the choice response to the highest label tier and reduces the use of price as a cue for greenness. This implies that label revisions should be evaluated not only by how much information they add, but also by how they reshape behavior at newly created tier boundaries.

This paper speaks most directly to the literature on how label design shapes consumer response. The central issue is not only how consumers respond to labels, but whether tier count is itself a meaningful policy variable in environmental label design. Existing work shows that traffic-light and symbolic labels can affect choices in food and environmental settings (Ellison et al., 2014; Drescher et al., 2014; Hallstein and Villas-Boas, 2013; Li et al., 2018; Muller et al., 2019; Taillie et al., 2024). It also shows that effects depend on format, salience, and complexity (Kortelainen et al., 2016; Crosetto et al., 2016; Carlsson et al., 2022; Marette et al., 2019; Thøgersen et al., 2024), while proliferating eco-label structures can create consumer confusion (Brécard, 2014). Related work on eco-label gradation and multi-tier label design shows that the number and structure of tiers can affect equilibrium outcomes (Li and van 't Veld, 2015; Fischer and Lyon, 2019; Nadar and Ertürk, 2021). In related empirical work, multi-tier quality labels also shape consumer perceptions of product quality (Costanigro et al., 2019). What remains underdeveloped is the design margin emphasized here: whether increasing the number of tiers can distort choices near newly created cutoffs, with corresponding local welfare consequences.

The remainder of the paper proceeds as follows. Section 2 develops the theory, including a brief strategic-firm extension. Section 3 describes the experimental design and its implications for the empirical tests. Section 4 presents the results. Section 5 concludes.

2 Theory

2.1 Model Setup

This section develops a deliberately simple consumer-side model of traffic-light label design. The model is designed to isolate one question: when a regulator increases the number of label tiers, does the added categorical detail improve choices, or can it distort beliefs enough to reduce consumer welfare? The key ingredients are minimal. Products differ in quality, price, and greenness. By quality we mean the product's ordinary consumption or performance quality, while greenness refers to the product's environmental quality. Consumers observe quality and price, but do not observe greenness directly. Instead, they observe the assigned label tier. The traffic-light label is therefore not a fourth product attribute, but an observed signal derived from greenness through a regulator's tiering rule. Greenness is often difficult for consumers to assess directly, while a regulator or certifier may be able to classify products using verified environmental information. The label then serves as a simplified public signal of an otherwise hard-to-observe attribute. Beliefs about greenness combine the formal label with cues derived from observed quality and price. This setup is sufficient to show why tier count is a meaningful policy variable and why more detailed labeling need not translate mechanically into better decisions. Modeling quality and greenness separately is important for that purpose. Quality captures the product's direct consumption value, whereas greenness is the environmental attribute the label is meant to summarize. The possibility that consumers value both, while observing only the first directly, is what creates the scope for eco-label design to matter.

The model is intentionally local. Rather than solving a full market equilibrium, it studies what happens when a coarse label partition is refined by the introduction of a middle tier. This is best understood as an information-design problem embedded in a vertically differentiated market, with attention restricted to products near the newly created cutoff. That focus is enough to isolate the paper's mechanism: adding an extra tier can shift beliefs by more than it changes underlying information. It also yields a simple sufficient-condition welfare result and the comparative statics that guide the empirical analysis.

We model the impact of a newly created tier by studying a pair of products located on opposite sides of the new cutoff. We denote products' quality, price, and greenness by q, p, and g, respectively. We refer to the pair of products by L and H, with $g_H > g_L$. Product H is the greener product. Let $\Delta g = g_H - g_L > 0$. Likewise, we write $\Delta q = q_H - q_L$, and $\Delta p = p_H - p_L$. We do not impose sign restrictions on Δq or Δp in the setup; those are introduced only where needed for particular comparative statics.

Assume utility is additively separable and linear in quality, price, and greenness, so that realized utility for consumer i from product j is

$$U_{ij} = \kappa_i q_j + \gamma_i g_j - p_j, \tag{1}$$

with $\kappa_i > 0$ and $\gamma_i > 0$. Then consumer i's realized utility difference between buying H and buying L is

$$\Delta_i \equiv U_{iH} - U_{iL} = \kappa_i \Delta q + \gamma_i \Delta g - \Delta p. \tag{2}$$

If $\Delta_i > 0$, the greener product H is the realized-utility maximizing choice for consumer i, and vice versa. However, consumers do not observe Δg directly. They infer it from the label and from cues derived from observed quality and price.

We proceed by introducing two labeling schemes that differ in the number of tiers: N versus $N + 1$. Let the coarser labeling scheme be an N-tier scheme in which both products receive the same label. The corresponding perceived utility difference is

$$\hat{\Delta}_i^N = \kappa_i \Delta q + \gamma_i \left[\lambda_p \Delta p + \lambda_q \Delta q\right] - \Delta p, \tag{3}$$

where λ_p and λ_q capture the use of quality and price as proxies for greenness. This reduced-form proxy channel follows the same basic idea as broader work in which consumers infer hard-to-observe product attributes from other observable cues (Roth and Diamantopoulos, 2009; Kirmani and Rao, 2000; Cagé and Rouzet, 2015). Now consider a finer $(N+1)$-tier scheme in which the new cutoff places H in a higher tier than L. The corresponding perceived utility difference is

$$\hat{\Delta}_i^{N+1} = \kappa_i \Delta q + \gamma_i \left[\Delta\mu + \lambda_p \Delta p + \lambda_q \Delta q\right] - \Delta p. \tag{4}$$

Compared with (3), $\Delta\mu$ is the only new term. Let $\mu_k^{N+1} \equiv \mathbb{E}[g \mid k]$ denote expected greenness in label tier k under the finer partition. Let k_L and k_H denote the tiers assigned to products L and H under the finer scheme, with k_H higher than k_L. Then $\Delta\mu \equiv \mu_{k_H}^{N+1} - \mu_{k_L}^{N+1} > 0$ is the increase in expected greenness implied by moving from the lower to the higher tier. This is the only channel through which the extra tier affects beliefs in the model. Consumers nevertheless continue to combine that formal tier signal with the same price and quality cues under both schemes.

Let j_i^N and j_i^{N+1} denote the options chosen by consumer i under the N-tier and $(N+1)$-tier schemes, respectively. For either scheme $s \in \{N, N+1\}$, define consumer welfare as expected realized utility from the chosen option, $W^s = \mathbb{E}\left[U_{ij_i^s}\right]$. More tiering can make the label more informative without improving the choices consumers make. We therefore compare schemes using realized utility rather than informational precision alone.

2.2 Welfare Comparison

The main concern arises near the newly created cutoff. When two products are close in true greenness, a label-implied increase in expected greenness can trigger switching toward the higher-tier product. In this local two-product setup, the only possible choice change is from L to H: the finer scheme raises perceived greenness of H relative to L, but never the reverse. Let the set of switchers under the finer scheme be

$$\mathcal{S}(\Delta\mu) = \left\{ i : \hat{\Delta}_i^N \leq 0 < \hat{\Delta}_i^{N+1} \right\}. \tag{5}$$

A mistake occurs when that switch lowers realized utility. The set of such consumers is

$$\mathcal{M}(\Delta\mu) = \{ i : i \in \mathcal{S}(\Delta\mu) \text{ and } \Delta_i < 0 \}. \tag{6}$$

These are consumers who choose L under the coarse scheme, switch to H under the finer scheme, and are made worse off by doing so.

The proposition below states the total-welfare condition in words. Because prices and qualities are fixed across the two schemes, welfare changes only for consumers in the switcher set $\mathcal{S}(\Delta\mu)$. The subset $\mathcal{M}(\Delta\mu)$ consists of switchers who are made worse off by the new tier. Total consumer welfare therefore falls when the consumers in $\mathcal{S}(\Delta\mu)$ are, on average, made worse off by switching, so that losses among members of $\mathcal{M}(\Delta\mu)$ outweigh any gains among the remaining switchers. More formally, consumers outside $\mathcal{S}(\Delta\mu)$ choose the same product under both schemes, so their realized utility is unchanged. Consumers inside $\mathcal{S}(\Delta\mu)$ switch from L to H, so their realized-utility change is exactly $\Delta_i = U_{iH} - U_{iL}$. The sign of $W^{N+1} - W^N$ is therefore pinned down by the average realized gain among switchers, which proves the following proposition.

Proposition 1. *If $\mathcal{S}(\Delta\mu)$ is non-empty and $\mathbb{E}[\Delta_i \mid i \in \mathcal{S}(\Delta\mu)] < 0$, then $W^{N+1} < W^N$.*

The proposition is deliberately local. It does not claim that more tiers always reduce welfare. It identifies a sufficient condition under which they do so: the new tier boundary causes enough consumers near the cutoff to switch toward the higher-tier product even though, on average, that switch does not justify the price premium.

2.3 Comparative Statics

The next result shows how the size of the mistake region identified in (6) varies with the model parameters. Define the switch threshold for consumer i in terms of the label-implied greenness gap. Setting $\hat{\Delta}_i^{N+1} = 0$ in equation (4) and solving for $\Delta\mu$ yields

$$\Delta\mu_i^* = \frac{\Delta p - \kappa_i \Delta q}{\gamma_i} - \lambda_p \Delta p - \lambda_q \Delta q. \tag{7}$$

Consumer i switches from L under the coarse scheme to H under the finer scheme whenever $\Delta\mu > \Delta\mu_i^*$.

Equation (7) has a simple interpretation. The term $(\Delta p - \kappa_i \Delta q)/\gamma_i$ is the greenness advantage that product H would need in order to justify its price-quality tradeoff in realized utility terms. The proxy terms $\lambda_p \Delta p$ and $\lambda_q \Delta q$ reduce that threshold because they make consumers infer part of the greenness advantage of H from observed price and quality even before the extra tier is introduced.

The sign assumptions in the next proposition are only those needed to make the comparative statics directional. The case $\Delta p > 0$ is the natural baseline in which the higher-tier product is also more expensive, whether because greener production is costlier or because the better tier supports a price premium; in that case, stronger price-based proxy use raises the likelihood of switching. The additional assumption $\Delta q \geq 0$ picks out the directional case in which consumption quality points the same way. Correspondingly, the natural prior is that λ_p is positive, whereas the sign of λ_q is application specific and not pinned down here. If instead $\Delta q < 0$, the quality cue points away from H, so stronger quality-based proxy use can dampen switching rather than reinforce it.

Proposition 2. *Suppose $\Delta p > 0$ and $\Delta q \geq 0$. Holding $(\Delta p, \Delta q)$ fixed, the measure of the mistake region $\mathcal{M}(\Delta\mu)$ weakly increases with $\Delta\mu$, weakly increases with λ_p and λ_q, and weakly decreases with the true greenness gap Δg.*

Proof. By (7), consumer i switches under the finer scheme if and only if $\Delta\mu > \Delta\mu_i^*$. Since

$$\Delta\mu_i^* = \frac{\Delta p - \kappa_i \Delta q}{\gamma_i} - \lambda_p \Delta p - \lambda_q \Delta q, \tag{8}$$

it follows immediately that $\Delta\mu_i^*$ is weakly decreasing in λ_p when $\Delta p > 0$ and weakly decreasing in λ_q when $\Delta q \geq 0$. Hence larger λ_p or λ_q weakly enlarge the set of switchers. A larger $\Delta\mu$ also weakly enlarges the set of switchers directly.

The mistake region can be written as

$$\mathcal{M}(\Delta\mu) = \{i : \Delta\mu > \Delta\mu_i^* \text{ and } \gamma_i \Delta g < \Delta p - \kappa_i \Delta q\}, \tag{9}$$

where the second inequality is equivalent to $\Delta_i < 0$ by (2). This representation shows that larger $\Delta\mu$, λ_p, and λ_q weakly expand the mistake region through the switching condition, while a larger true greenness gap Δg weakly shrinks it through the realized-utility condition. The realized-utility condition does not depend on $\Delta\mu$, λ_p, or λ_q, so those parameters affect $\mathcal{M}(\Delta\mu)$ only through the switching condition. This proves the proposition. □

The intuition is straightforward. A larger label-implied greenness gap makes the higher tier more attractive, and stronger proxy use reinforces that effect when price and quality both point toward H. By contrast, a larger true greenness gap makes switching less likely to be a mistake, because the higher-tier product is then more likely to justify its price-quality tradeoff in realized utility terms.

These results motivate the empirical analysis introduced below. The choice experiment varies the number of tiers, so it can test whether introducing a middle tier changes how respondents trade off the highest label tier, price, and quality in choice, and whether respondents use price as a cue for greenness. The empirical analysis therefore focuses on reduced-form treatment contrasts in the theory's main behavioral margins: price-based inference, the relative appeal of the highest label tier, and the role of price in choice.

The model suggests where treatment effects are most likely to appear. Holding the underlying product menu fixed, the behavioral effect of adding a tier should be stronger when consumers place more weight on perceived greenness and when non-label cues such as price are more informative about environmental quality. The same logic can in principle extend to quality differences, but the model does not require price and quality to play equally important roles in every application. These considerations motivate the empirical focus on how the added tier changes response to the highest label tier, the role of price in choice, and stated price-based inference. The choice experiment therefore speaks to the mechanism, not to the welfare sign, which also depends on how firms and products are positioned around the new cutoff.

2.4 Strategic-Firm Extension

The theory above holds quality and price fixed in order to isolate the consumer-side consequences of changing the label structure. A natural extension allows firms to respond strategically to the placement of label cutoffs. For the purposes of this paper, the key question is not to solve a full pricing game, but to ask how such responses would shift the primitives that drive Propositions 1 and 2. In the notation of the main model, endogenous firm behavior matters only insofar as it changes the realized greenness gap Δg, the price gap Δp, or the mass of products located close to the relevant cutoff (for conciseness, we abstract from separate quality adjustment here).

This perspective immediately clarifies how strategic behavior can amplify or mitigate the consumer-side mechanism. It amplifies the main result if firms target cutoffs without much improving true greenness, because that compresses the distribution of products around the threshold and makes small Δg more common. It also amplifies the result if qualifying for a better label allows firms to charge higher prices, because a larger Δp makes induced switching harder to justify in realized utility terms. By contrast, strategic behavior mitigates the result if firms respond by making genuine greenness improvements, since a larger Δg shrinks the mistake region in Proposition 2. These comparative-static implications do not require a separate equilibrium theorem; they follow directly from the role of Δg and Δp in (2) and (7).

Existing evidence suggests that amplification is plausible. In agricultural markets, firms respond strategically to sustainability disclosure and product grading, which supports the idea that label design can affect supply-side choices as well as consumer demand (Villas-Boas et al., 2020). In adjacent labeling settings, threshold-based categories also appear to generate discontinuous responses around rating boundaries (Sejas-Portillo et al., 2025). Those findings are consistent with cutoff targeting and therefore with an amplification channel in the present framework. Theoretical work on eco-labeling also points the same way. Amacher et al. (2004) show in a vertically differentiated duopoly that eco-labeling affects firms' environmental-quality choices and price competition, while Fischer and Lyon (2014), Li and van 't Veld (2015), and Fischer and Lyon (2019) show that label design and multitier structures can change equilibrium incentives, albeit in settings with strategic sponsors or competing labels. Those are not the institutional assumptions used here, but they provide a standard reason to expect that endogenous firm behavior may reinforce the consumer-side mechanism. At the same time, the same strategic responses could mitigate the consumer-side downside if firms improve true greenness enough to justify the higher tier. The paper therefore treats endogenous firm behavior as a policy-relevant extension that likely shifts the magnitude of the welfare effect, but not as a channel identified by the present experiment.

2.5 Structural Interpretation

The empirical analysis below remains primarily reduced form, but the experimental design also permits a limited structural interpretation. The survey included an item to probe proxy inference. This item varied prices of two similar products while holding the displayed label and product quality fixed. As such, it can be used to recover the price-proxy parameter λ_p under an explicit normalization. The quality-proxy parameter λ_q is not targeted by the design in the same way, so the structural recovery in Section 4 focuses on the price channel. That exercise will show whether the price-proxy mechanism in Proposition 2 is quantitatively present in the data and how much separation the three-tier design creates between the yellow and green tiers under a transparent normalization.

3 Experimental Design

The theory delivers a local mechanism: changing tier count can alter tier valuation, the role of informal cues such as price, and therefore the quality of consumer choices near newly created cutoffs. The choice experiment (CE) tests whether the underlying behavioral margins move when the number of tiers changes. The empirical design is therefore used to ask whether a finer label changes the response to the highest label tier and the extent to which respondents rely on price as a cue for greenness.

The empirical setting is a stated-preference discrete choice experiment on trash bags with "enviro-score" labels. The enviro-score is a stylized traffic-light eco-label that is presented as summarizing a product's overall environmental impact over its lifecycle, including production, transport, use, and disposal. It is designed to mimic the visual logic of real traffic-light labels while remaining generic enough for experimental control. Trash bags are used because they form a relatively stripped-down product category: unlike food or other experience goods, they are less likely to trigger strong prior beliefs about taste, health, or brand quality, and they offer few salient non-environmental attributes beyond quality and price. That makes them useful for isolating how respondents interpret the traffic-light signal itself rather than a bundle of pre-existing associations.

On each choice card, respondents were asked which roll of trash bags they would buy from the offered alternatives, with the option to select the opt-out if they would purchase none of them. Appendix Figure 2 shows an example choice card. The CE design and reporting follow standard stated-preference practice in emphasizing transparent attribute construction, an opt-out option, sample and implementation detail, and careful interpretation of what the task can identify (Johnston et al., 2017).

A CE is useful for this purpose because it lets us vary tier count while holding the product domain and the competing price-quality trade-offs fixed, making it possible to isolate how respondents react to the informational structure of the label. Across treatments, the CE varies three product attributes: label, quality, and price. Quality takes two levels, regular and extra thick; in the CE, product quality is thus captured by bag thickness. Price takes three levels, low, medium, and high. Price is respondent-specific because each respondent first reported the price closest to what their household normally pays for a roll of trash bags, after which the experimental price levels were mapped into respondent-specific dollar amounts. In particular, the reported reference price becomes the middle price level p_2, while the low and high levels are set to $p_1 = 0.75p_2$ and $p_3 = 1.5p_2$. The key treatment varies the number of label tiers. One group of respondents saw a two-tier traffic-light label, while the other saw a three-tier version with an intermediate yellow tier. In the theory, that added tier creates a new boundary between adjacent label classes. In the CE, it provides the key variation for testing whether adding a middle tier changes the choice response to the highest label tier and reliance on price as a cue for greenness.

The CE contains 32 choice sets per treatment, split into four blocks of eight, with one block assigned to each respondent. The selected design maximized D-efficiency subject to the restriction that no alternative strictly dominated another. An opt-out option was added uniformly as a fourth alternative on every card. In the analysis data, label and quality attributes are therefore defined for the three active products and set to “none” for the opt-out. Importantly, thickness is varied independently of the label, so the experiment does not impose any fixed relationship between product quality and greenness. Price also varies within the CE, but unlike quality, its joint distribution with label reflects the design restriction that strictly dominating alternatives are excluded.

The survey also included a follow-up question designed to probe proxy inference. Respondents were asked which of two otherwise similar products had higher greenness when price differed. We code whether a respondent uses price as a proxy at all, and separately whether the more expensive product is explicitly judged greener. This question is important because the theory's proxy-use channel operates through beliefs rather than through product choice alone. The survey contained four versions of this prompt, varying the displayed label color and bag thickness while keeping the paired products identical except for price. The follow-up question was asked only after the CE choice tasks; Appendix Figure 3 shows an example.[1] Appendix Figure 4 summarizes the core survey flow used in the analysis.

The CE provides reduced-form tests of the theory's mechanism. The empirical analysis focuses on three questions: (i) whether consumers use non-label cues, especially price, to infer greenness, (ii) whether adding an intermediate tier makes respondents more likely to choose the highest label tier, holding other attributes fixed, and (iii) whether adding an intermediate tier changes the relative importance of the label and price in choice. These questions map directly to the treatment contrasts in the data. We therefore estimate choice-set fixed-effects models for product choice and use the follow-up question as a direct test of price-based inference. The empirical objective is not to recover attribute-level willingness to pay, but to estimate treatment contrasts in the behavioral margins highlighted by the theory.

The CE was implemented in Qualtrics and fielded from July 25 to July 29, 2020 to U.S. respondents through Amazon Mechanical Turk. The study received IRB/ethics approval.[2] The study was fielded before the recent surge in generative-AI-assisted survey responding, which likely reduces that particular data-quality concern. Using MTurk offers a practical way to recruit a large online sample quickly and to randomize respondents into otherwise identical survey instruments, but it also comes with standard limitations: the sample is not population representative, and the purchase task is hypothetical rather than incentive-compatible in a literal market sense. For the current paper, that trade-off is acceptable because the empirical objective is to recover treatment contrasts in a common decision environment rather than actual purchase commitments. Respondent recruitment applied basic platform-level eligibility filters. The initial batches required MTurk Masters status; later batches used a minimum 95% approval rate and at least 50 completed prior MTurk tasks. To prevent respondents from entering both treatment arms, the two treatment surveys were never live at the same time, and subjects were tagged afterwards to prevent re-entry.

Mean completion time in the cleaned sample is 290 seconds, or about 4.8 minutes, and compensation was $0.50 upon completion; slightly above average for MTurk tasks

[1] The CE also included several basic response-quality checks. Before the main choice tasks, respondents had to complete a control item that required them to identify the option with the highest enviro-score before continuing. Choosing the opt-out option triggered a short follow-up asking why none of the products would be chosen. In addition, the first choice card was repeated later in the sequence, which allows us to construct a simple internal-consistency check.

[2] Approval was granted by the SBE Research Ethical Review Board at Vrije Universiteit Amsterdam, reference number `SBE7/16/2020jak500`.

Table 1: Respondent summary by treatment

Statistic	2-tier	3-tier
Female (%)	45.3	39.6
Age 40+ (%)	33.8	42.2
Income 50k+ (%)	50.5	55.2
College+ (%)	74.9	64.3
Respondents	311	308

of comparable type and length. The raw survey files contain 638 respondents in total. Cleaning is minimal: we retain respondents who reached at least 99% survey progress and drop incomplete responses. This removes 19 respondents, 8 in the two-tier treatment and 11 in the three-tier treatment. The cleaned sample contains 619 respondents: 311 in the two-tier treatment and 308 in the three-tier treatment. The resulting long-format dataset contains 13,467 alternative-by-choice-set observations used in the main fixed-effects specifications.

4 Results

Table 1 summarizes respondent characteristics by treatment. The two treatment groups are broadly similar, as expected under random assignment, though the three-tier sample is somewhat older and less educated than the two-tier sample. These differences are modest and do not alter the basic impression of broadly comparable treatment groups. The basic response-quality checks reported in footnote 2 also do not indicate major differential concerns across treatments: the repeated-card inconsistency rate is essentially identical in the two groups (25.1% versus 25.0%), and protest-type opt-out responses differ only modestly between them (6.1% versus 2.9%).

In the remainder of this section, we first use choice-set fixed-effects models on the active product alternatives to recover how the three-tier treatment changes trade-offs among label, price, and bag thickness within a choice set. Next we use the follow-up question to recover the price-proxy channel more directly. Finally, we provide a limited structural interpretation that estimates the price-proxy parameter and links the reduced-form evidence back to the theory.

4.1 Treatment effects

For the product-choice analysis, each choice card yields a product-level outcome equal to one if a given alternative is selected and zero otherwise. We estimate choice-set fixed-effects models on the active product alternatives to recover how the three-tier treatment changes trade-offs among label, price, and bag thickness within a choice set. Our baseline estimating equation is

$$
\begin{aligned}
Y_{ijc} = \alpha_{ic} &+ \beta_1 \log(\text{price}_{ijc}) + \beta_2 \text{thick}_{ijc} + \beta_3 \text{yellow}_{ijc} + \beta_4 \text{green}_{ijc} \\
&+ \delta_1 \left(\text{green}_{ijc} \times \text{three-tier}_i\right) + \delta_2 \left(\text{thick}_{ijc} \times \text{three-tier}_i\right) \\
&+ \delta_3 \left(\log(\text{price}_{ijc}) \times \text{three-tier}_i\right) + \varepsilon_{ijc},
\end{aligned} \tag{10}
$$

where Y_{ijc} is an indicator that respondent i chooses product j on card c, α_{ic} is a respondent-card fixed effect, and three-tier$_i$ indicates assignment to the three-tier treatment. Equation (10) is estimated by OLS as a linear probability model with respondent-card fixed effects and respondent-level clustered standard errors. The interaction terms are causally identified by random assignment to the two-tier versus three-tier CE treatment. Because α_{ic} absorbs all card-specific constants, only within-card attribute differences identify the coefficients. Accordingly, the main coefficients of interest are the interaction terms δ_1, δ_2, and δ_3, which measure how the effects of the green label, thickness, and price change in the three-tier CE relative to the two-tier CE. The baseline coefficients β_1 through β_4 are still needed because they define the corresponding attribute effects in the omitted two-tier treatment, but they are not the paper's main focus. The paper is primarily interested in treatment contrasts, not in the absolute level of attribute responses within a single treatment.

The main choice specifications based on equation (10) are estimated on the active product alternatives only, excluding the opt-out option. This exclusion drops 4,952 of the 19,808 alternative-by-choice-set observations in the long data, or 25 percent. We do so because the opt-out is not a product alternative with directly comparable label, price, and thickness attributes, whereas the mechanism tests concern how respondents trade off those attributes across the offered products. As discussed in footnote 3, the main treatment patterns are nevertheless robust to simple checks that incorporate the opt-out margin. Standard errors are clustered at the respondent level.

Table 2 reports three nested choice specifications. The first includes log price, an indicator for extra thickness, and dummies for yellow and green labels. The second adds interactions between the three-tier treatment and the green-label and extra-thick indicators. The third also interacts log price with the three-tier treatment. The intermediate specification is useful because it shows how the treatment contrast for the highest label tier changes once treatment-specific label and quality effects are admitted, but before also allowing price sensitivity to differ across treatments. The richest specification then adds the price interaction to test whether part of the treatment contrast instead operates through a change in the role of price in choice. The two main coefficients of interest are therefore the interaction between the green label and the three-tier treatment and the interaction between log price and the three-tier treatment. The CE also reveals how

Table 2: Choice-set fixed-effects models

	(1) Baseline FE	(2) Treatment FE	(3) Treatment FE + price slope
Log price	−0.034	−0.051**	−0.002
	(0.026)	(0.026)	(0.037)
High quality	0.001	0.022	0.004
	(0.012)	(0.017)	(0.019)
Yellow label	0.129***	0.159***	0.173***
	(0.019)	(0.020)	(0.020)
Green label	0.117***	0.087***	0.064***
	(0.019)	(0.022)	(0.024)
Log price x 3-tier			−0.103**
			(0.051)
Green x 3-tier		0.091***	0.143***
		(0.033)	(0.039)
High quality x 3-tier		−0.029	0.005
		(0.020)	(0.025)
Respondent-card FE	X	X	X
R2	0.091	0.093	0.094
Num.Obs.		13 467	

* p <0.1, ** p <0.05, *** p <0.01

respondents treat the new yellow tier within the three-tier scheme. Both red and green are observed under the two-tier and three-tier labels, while yellow appears only in the three-tier treatment. The cleanest treatment-contrast coefficients therefore remain those for attributes observed in both label structures, especially the interaction for the highest label tier. Because the pooled specification uses red as the common omitted category, the reported *Green x 3-tier* coefficient is best read as the change in the relative attractiveness of the highest label tier under the three-tier treatment. The omitted attribute categories are the red label and regular thickness, so the label coefficients are relative to red and the quality coefficient is relative to regular bags.

Model 1 shows that both yellow and green raise choice probabilities relative to red, with the estimated yellow coefficient slightly larger than the green coefficient. By contrast, the coefficient on high quality is small and statistically insignificant, suggesting that thickness does not play a large role in choice even before treatment interactions are added. The level effect of price is also small, suggesting that in the two-tier CE respondents do not respond strongly to price on average. In model 2, the coefficient on *Green x 3-tier* is 0.091 and statistically significant, while the interaction between high quality and the three-tier treatment is small and statistically insignificant. Thus, moving from two to three tiers primarily changes the response to the highest label tier, not the response to thickness-based quality.

Once the price interaction is added in model 3, the coefficient on *Green x 3-tier* rises to 0.143 and remains statistically significant. This pattern suggests that allowing price sensitivity to vary across treatments strengthens rather than weakens the estimated treatment contrast on the highest label tier. The added price interaction itself is -0.103 and statistically significant, while the baseline log price coefficient in model 3 is close to zero. The negative interaction means that price has a more negative effect on choice in the three-tier treatment than in the two-tier treatment, whereas the near-zero baseline price effect implies little average price response in the two-tier CE. Taken together, these estimates imply that the three-tier CE is associated both with a stronger response to the highest label tier and with greater price sensitivity. That pattern is consistent with the theory, which predicts that adding a middle tier can change both the response to the highest label tier and the role of price in choice.[3] In the notation of Section 2, the table therefore speaks to both the formal-information margin, through the response to the highest label tier, and indirectly to the price-inference margin associated with λ_p. The follow-up question below speaks to that inference channel more directly. More broadly, Table 2 lines up with the theory in a selective way: the model allows both price and quality differences, Δp and Δq, to shape inference and switching, but in this CE the empirical mechanism appears to run mainly through the label and price margins, while the thickness-based quality margin remains comparatively muted.

4.2 Price-proxy channel

Table 2 provides revealed-choice evidence that the added middle tier changes how respondents trade off label and price in the CE. In the theory, that trade-off reflects not only the direct role of Δp in utility, but also the extent to which consumers use price itself as a cue for greenness. To dig deeper into that mechanism, we next turn to a follow-up survey question that isolates this narrower inference margin more directly. Its value is complementary to Table 2: instead of asking what respondents chose under joint price-quality-label trade-offs, it asks more directly whether price is being used as a cue for greenness. That mechanism focus is also related to evidence that eco-label effects can interact with price cues in field settings (Feuß et al., 2022).

[3]The exclusion of the opt-out from the main mechanism regressions does not appear to drive these results. At the choice-card level, the share selecting any active product rather than the opt-out is nearly identical across treatments: 83.2 percent in the two-tier CE and 83.4 percent in the three-tier CE. In a simple robustness regression, the treatment effect on choosing an active product is close to zero and statistically insignificant. Moreover, in a pooled alternative-level specification that includes the opt-out alternative as a zero-attribute baseline, the key interaction patterns remain: the *Green x 3-tier* coefficient stays positive and statistically significant, and the *log price x 3-tier* coefficient stays negative and statistically significant, while the thickness interaction remains insignificant.

Table 3: Proxy-inference outcomes by treatment

	Uses price proxy	Expensive seen greener
3-tier minus 2-tier	−0.150***	−0.045
	(0.038)	(0.033)
Shares		
2-tier (%)	42.8	23.5
3-tier (%)	27.8	19.0

Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

One version of the follow-up question (see Appendix Figure 3) asked: *"Which of the following two options do you expect to have the highest environmental quality?"* As noted in Section 3, the four variants kept that basic task fixed while varying the visual context and, in some versions, the wording. Empirically, we do not find strong evidence that the treatment contrast differs systematically across the four variants. The item is informative about mechanism: respondents who judge one product to be greener when only price differs are using price as a cue for environmental quality. We therefore estimate simple respondent-level linear-probability models in which the only regressor is an indicator for assignment to the three-tier treatment. The outcome is either an indicator that the respondent used price as any proxy for greenness, or an indicator that the more expensive product was explicitly judged greener. We code "uses price proxy" as one whenever the respondent judged either the cheaper or the more expensive product to be greener, and zero when the two products were judged equally green. Because treatment is randomly assigned, these comparisons are simple treatment contrasts.

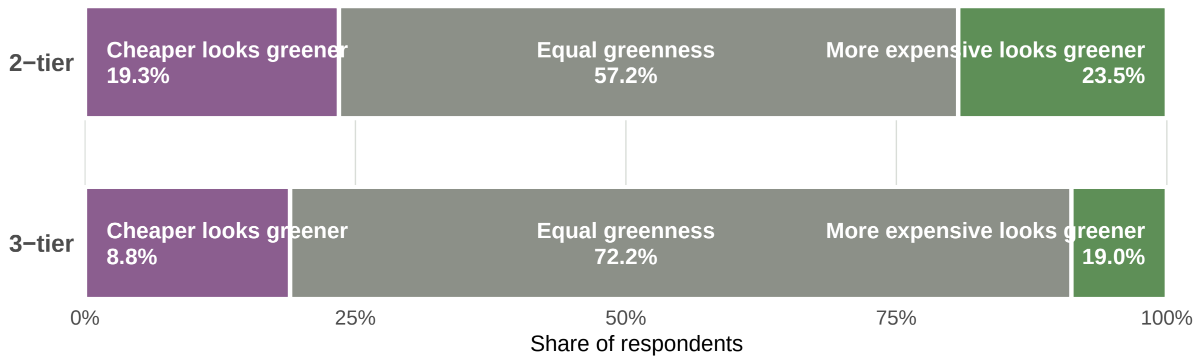

Figure 1: Perceived greenness when only price differs, by treatment.

Table 3 reports the treatment contrasts, while Figure 1 shows the full response distribution. Respondents in the three-tier treatment are materially less likely to use price as a proxy for greenness (note that in Figure 1, the purple and green blocks combined represent the share that is coded as using price as a proxy). The main result is the treatment contrast: the share using price as any proxy falls by 15 percentage points, from 42.8

percent in the two-tier treatment to 27.8 percent in the three-tier treatment. Figure 1 shows that this shift reflects declines in both types of price-based inference, with fewer respondents judging either the cheaper or the more expensive product to be greener, and a corresponding increase in the equal-greenness response. The share explicitly judging the more expensive product greener also falls, although by a smaller and statistically insignificant amount. These treatment differences indicate that finer categorical disclosure changes how respondents combine formal label information with other observed cues.

Taken together, Table 2 and Table 3 support the paper's mechanism story. The choice models show that adding a middle tier changes actual product choice by raising response to the highest label tier and changing the role of price, while the follow-up question shows directly that respondents rely less on price to infer greenness when the label is more informative. The evidence is therefore consistent with the theory's behavioral channels.

4.3 Structural interpretation

The follow-up question also allows a limited structural interpretation of the price-proxy channel. The reduced-form estimates above do not by themselves recover the proxy parameters, because in the main choice data the direct role of price in utility and the role of price as a cue for greenness move together. The follow-up question, however, varies only prices while holding the displayed label and bag thickness fixed, which makes it the natural place to recover the price-proxy parameter λ_p. To keep the notation close to Section 2, let the latent perceived greenness difference between the more expensive and cheaper product in that item be

$$\Delta \hat{g}_i = \lambda_p \Delta p_i + \nu_i, \tag{11}$$

where Δp_i is the respondent-specific displayed price gap and ν_i is a logistic disturbance. If respondents answer "cheaper greener," "equal," or "more expensive greener" according to thresholds on $\Delta \hat{g}_i$, then an ordered logit identifies λ_p up to the usual scale normalization on ν_i. This is just the price-proxy component of perceived greenness from (3) and (4), isolated in the follow-up question where the displayed label and thickness are held fixed. It is therefore a structural measurement equation rather than another reduced-form choice regression.

Table 4 reports the corresponding estimates. Under the ordered-logit scale normalization, the pooled estimate of λ_p is about 0.27, implying that larger displayed price gaps make respondents more likely to infer that the more expensive product is greener. This pooled estimate is the preferred summary because the treatment-specific slopes are similar in magnitude, while the interaction in a pooled specification is small and imprecisely estimated, so Table 4 reports only the pooled slope. This result fits the reduced-form evidence above: the treatment contrast in Table 3 appears to operate more through the

Table 4: Structural estimates

	Estimate
Price-proxy parameter (Eq. 9)	
λ_p (pooled)	0.270**
	(0.128)
Normalized category values (Eq. 8, 3-tier only)	
μ_Y	0.713
μ_G	0.824
$\Delta\mu_{YR} = \mu_Y - \mu_R$	0.713
$\Delta\mu_{GY} = \mu_G - \mu_Y$	0.111

Standard errors in parentheses for λ_p estimates. * p<0.1, ** p<0.05, *** p<0.01. Normalized category values impose $\mu_R = 0$ and $\gamma = 1$.

overall propensity to use price as a cue for greenness than through the marginal response to a larger price gap. The positive estimate of λ_p therefore links directly back to Proposition 2 by showing that the price- proxy channel needed for a larger mistake region is present in the data.

If one additionally imposes the normalizations $\mu_R = 0$ and $\gamma_i = 1$ for all i, which match the red omitted category in the choice regressions and fix the greenness scale, then the label coefficients from a separate three-tier conditional-logit version of equation (10) can be read as normalized tier values. Those tier-value estimates do not come from equation (11); they come from the choice-card data under the three-tier treatment alone. Using the notation introduced in Section 2, the estimates imply $\mu_Y = 0.71$ and $\mu_G = 0.82$. Under that normalization, the implied yellow-red gap, $\Delta\mu_{YR} = \mu_Y - \mu_R = 0.71$, and the green-yellow gap, $\Delta\mu_{GY} = \mu_G - \mu_Y = 0.11$. This suggests that the three-tier scheme creates a positive but modest separation between the yellow and green tiers under the maintained normalization.

The structural exercise suggests that the mechanism in Section 2 is quantitatively relevant in this CE: respondents do use price as a cue for greenness, and the three-tier scheme creates a positive but modest separation between the yellow and green tiers. At the same time, these estimates are best read as evidence on the behavioral channels through which tier count affects choice, rather than as a direct estimate of the welfare effect in the experiment.

5 Conclusion

This paper studies how tier count affects the performance of traffic-light environmental labels. The theory shows that adding a tier can sharpen formal information while also creating a new tier boundary that changes how consumers interpret the label. As a result, moving from fewer to more tiers need not improve choices. Whether it does so depends not only on the precision of the label, but also on how consumers combine label information with other observed cues.

The choice experiment and follow-up question provide reduced-form evidence on those behavioral margins. In the three-tier treatment, respondents are more likely to choose the highest label tier, become more price sensitive in choice, and are less likely to use price as a cue for greenness. Taken together, these results support the paper's mechanism: changing tier count changes both the choice response to the highest label tier and price-based inference. More tiers are not automatically better: adding a tier can increase mistakes near newly created cutoffs even when it makes the label more informative. The broader implication is that tier count is itself a meaningful policy variable in environmental label design. For regulators, the practical implication is that revisions to label structure should be pilot-tested and evaluated with particular attention to newly introduced tier boundaries rather than treated as automatically welfare-improving refinements.

The structural interpretation sharpens that conclusion. The follow-up question identifies a positive price-proxy parameter, and the three-tier choice data imply a positive but modest separation between the yellow and green tiers under transparent normalizations. Those estimates do not deliver a welfare sign for the experiment, but they show that the proxy channel in the theory is not just qualitatively present: it is quantitatively relevant in this setting.

Appendix

This appendix reproduces selected survey materials used in the analysis.

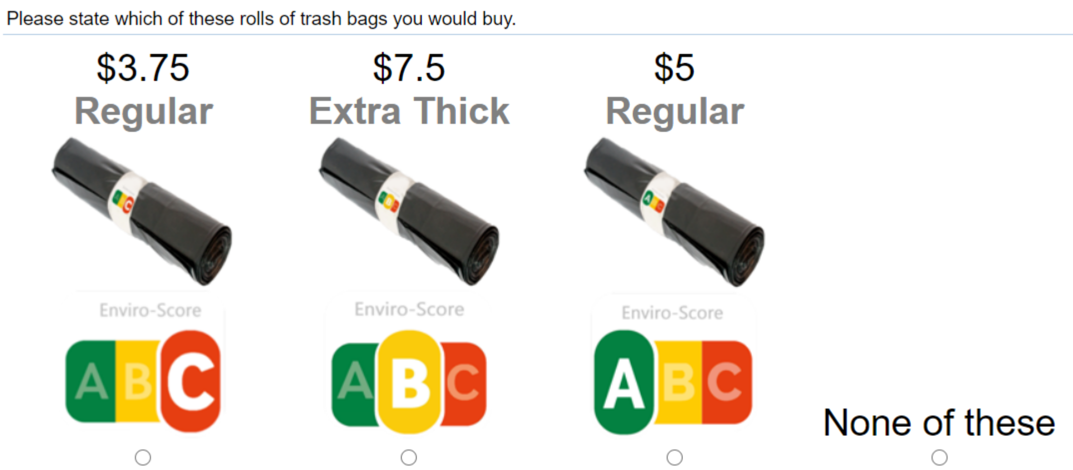

Figure 2: Example choice card.

Figure 3: Example follow-up question.

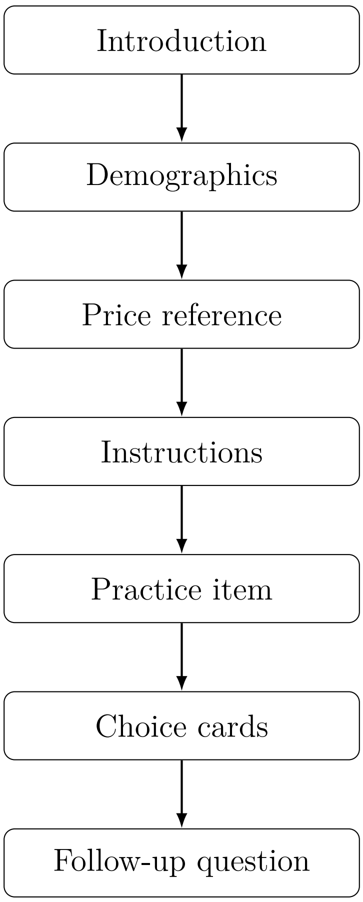

Figure 4: Core survey flow used in the analysis.

References

Allcott, H. and Taubinsky, D. (2015). Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market. *American Economic Review*, 105(8):2501–2538.

Amacher, G. S., Koskela, E., and Ollikainen, M. (2004). Environmental quality competition and eco-labeling. *Journal of Environmental Economics and Management*, 47(2):284–306.

Brécard, D. (2014). Consumer confusion over the profusion of eco-labels: Lessons from a double differentiation model. *Resource and Energy Economics*, 37:64–84.

Cagé, J. and Rouzet, D. (2015). Improving "national brands": Reputation for quality and export promotion strategies. *Journal of International Economics*, 95(2):274–290.

Carlsson, F., Kataria, M., Lampi, E., Nyberg, E., and Sterner, T. (2022). Red, yellow, or green? Do consumers' choices of food products depend on the label design? *European Review of Agricultural Economics*, 49(5):1005–1026.

Costanigro, M., Scozzafava, G., and Casini, L. (2019). Vertical differentiation via multi-tier geographical indications and the consumer perception of quality: The case of chianti wines. *Food Policy*, 83:246–259.

Crosetto, P., Muller, L., and Ruffieux, B. (2016). Helping consumers with a front-of-pack label: Numbers or colors? Experimental comparison between guideline daily amount and traffic light in a diet-building exercise. *Journal of Economic Psychology*, 55:30–50.

Drescher, L. S., Roosen, J., and Marette, S. (2014). The effects of traffic light labels and involvement on consumer choices for food and financial products. *International Journal of Consumer Studies*, 38(3):217–227.

Ellison, B., Lusk, J. L., and Davis, D. (2014). The impact of restaurant calorie labels on food choice: Results from a field experiment. *Economic Inquiry*, 52(2):666–681.

Feuß, S., Fischer-Kreer, D., Majer, J., Kemper, J., and Brettel, M. (2022). The interplay of eco-labels and price cues: Empirical evidence from a large-scale field experiment in an online fashion store. *Journal of Cleaner Production*, 373:133707.

Fischer, C. and Lyon, T. P. (2014). Competing environmental labels. *Journal of Economics & Management Strategy*, 23(3):692–716.

Fischer, C. and Lyon, T. P. (2019). A theory of multitier ecolabel competition. *Journal of the Association of Environmental and Resource Economists*, 6(3):461–501.

Hallstein, E. and Villas-Boas, S. B. (2013). Can household consumers save the wild fish? Lessons from a sustainable seafood advisory. *Journal of Environmental Economics and Management*, 66(1):52–71.

Johnston, R. J., Boyle, K. J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., Hanemann, W. M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R., and Vossler, C. A. (2017). Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists*, 4(2):319–405.

Kirmani, A. and Rao, A. R. (2000). No pain, no gain: A critical review of the literature on signaling unobservable product quality. *Journal of Marketing*, 64(2):66–79.

Kortelainen, M., Raychaudhuri, J., and Roussillon, B. (2016). Effects of carbon reduction labels: Evidence from scanner data. *Economic Inquiry*, 54(2):1167–1187.

Li, T., Kecinski, M., and Messer, K. D. (2018). Behavioural responses to science-based eco-labelling: Gold, silver, or bronze. *Applied Economics*, 50(39):4250–4263.

Li, Y. and van 't Veld, K. (2015). Green, greener, greenest: Eco-label gradation and competition. *Journal of Environmental Economics and Management*, 72:164–176.

Lusk, J. L. and Marette, S. (2012). Can labeling and information policies harm consumers? *Journal of Agricultural & Food Industrial Organization*, 10(1):1–13.

Marette, S., Nabec, L., and Durieux, F. (2019). Improving nutritional quality of consumers' food purchases with traffic-lights labels: An experimental analysis. *Journal of Consumer Policy*, 42(3):377–395.

Mason, C. F. (2011). Eco-labeling and market equilibria with noisy certification tests. *Environmental and Resource Economics*, 48:537–560.

Muller, L., Lacroix, A., and Ruffieux, B. (2019). Environmental labelling and consumption changes: A food choice experiment. *Environmental and Resource Economics*, 73(3):871–897.

Nadar, E. and Ertürk, M. S. (2021). Eco-design of eco-labels with coarse grades. *Omega*, 99:102209.

Roth, K. P. and Diamantopoulos, A. (2009). Advancing the country image construct. *Journal of Business Research*, 62(7):726–740.

Sejas-Portillo, R., Moro, M., and Stowasser, T. (2025). The simpler the better? Threshold effects of energy labels on property prices and energy efficiency investments. *American Economic Journal: Applied Economics*, 17(1):41–89.

Taillie, L. S., Wolfson, J. A., Prestemon, C. E., Bercholz, M., Ewoldt, L., Ruggles, P. R., and Hall, M. G. (2024). The impact of an eco-score label on US consumers' perceptions of environmental sustainability and intentions to purchase food: A randomized experiment. *PLOS ONE*, 19(6):e0306123.

Thøgersen, J., Dessart, F. J., Marandola, G., and Hille, S. L. (2024). Positive, negative or graded sustainability labelling? Which is most effective at promoting a shift towards more sustainable product choices? *Business Strategy and the Environment*, 33:6795–6813.

Villas-Boas, S. B., Kiesel, K., Berning, J. P., Chouinard, H. H., and McCluskey, J. J. (2020). Consumer and strategic firm response to nutrition shelf labels. *American Journal of Agricultural Economics*, 102(2):458–479.