Psychology & Marketing

WILEY

RESEARCH ARTICLE OPEN ACCESS

Consumers' Understanding of Energy Labels: Perception of Eco-Design With Scale Range and Color-Coding

Emil Skog[1] | Patrik Sörqvist[1,2]

[1]Department of Health, Learning and Technology, Luleå University of Technology, Luleå, Sweden | [2]Department of Building Engineering, Energy Systems and Sustainability Science, University of Gävle, Gävle, Sweden

Correspondence: Patrik Sörqvist (patrik.sorqvist@ltu.se)

Received: 6 July 2025 | **Revised:** 15 December 2025 | **Accepted:** 17 December 2025

Funding: Riksbankens Jubileumsfond, Grant/Award Number: P23-0067; Kempe foundations, Grant/Award Number: JCSMK23-0179

Keywords: consumer decisions | context | environmental impact | evaluation | judgment

ABSTRACT

Energy labels help consumers understand the environmental impact of products. This drives consumer behavior. Knowing how label features are perceived can thus have important implications for design, policy, and management. Energy labels contain different design features that convey information about the range of available energy classes. Such contextual information can help people make informed comparisons. Across three behavioral experiments, we evaluate the perception of (EU's eco-design) energy labels based on their letter scale, color-coding, and information about the scale's range of energy classes. These three features had different influences on judgments of environmental friendliness. A letter scale that begins with the alphabet (A to G) helped participants tell apart energy classes linearly. But a scale that extends beyond the alphabet (A+++ to D) led to biased and skewed responses. Color-coding and information about the scale's range of energy classes provided important contextual information that made environmental judgments more discriminating and more stable. We demonstrate novel behavioral evidence on how such contextual information leads to better understanding of energy labels, with a discussion informed by range-frequency theory, the cue-utilization framework, and the heuristics and biases framework. Implications for policy and product marketing strategies with energy labels are discussed.

1 | Introduction

Energy labels are used around the world to help consumers understand energy efficiency. The present study takes a psychological perspective to evaluate how people come to understand energy labels when evaluating environmental impact. We focus primarily on the eco-design of energy labels for electronic products within the European Union (EU), a leading institution on directives for energy labeling. The design of EU energy labels contains three key features (Figure 1): (1) A letter scale, (2) a color-coded scale, and (3) information about the scale's range (highest and lowest letters in the scale). Each of these three design features have the potential to affect how energy labels are understood, yet there is scarce behavioral evidence on how these stimulus features are perceived. We seek to evaluate how each of these three central stimulus features affect how energy labels inform judgments of environmental impact. While previous studies have investigated the effects that labels can have on consumer behavior (Majer et al. 2022; Thøgersen et al. 2024; Zhong et al. 2025), we take the novel step to investigate how different stimulus features within energy labels are perceived, to better understand their relative contributions for environmental judgments.

The design of EU energy labels has been revised over the years. In fact, each of the three key stimulus features (letters, color-coding, and information about the scale's range) have been subject to changes, or there are instances where they vary or can be modified. Regarding the use of letters in the scales, as electronic products became more energy efficient (in the 2000s

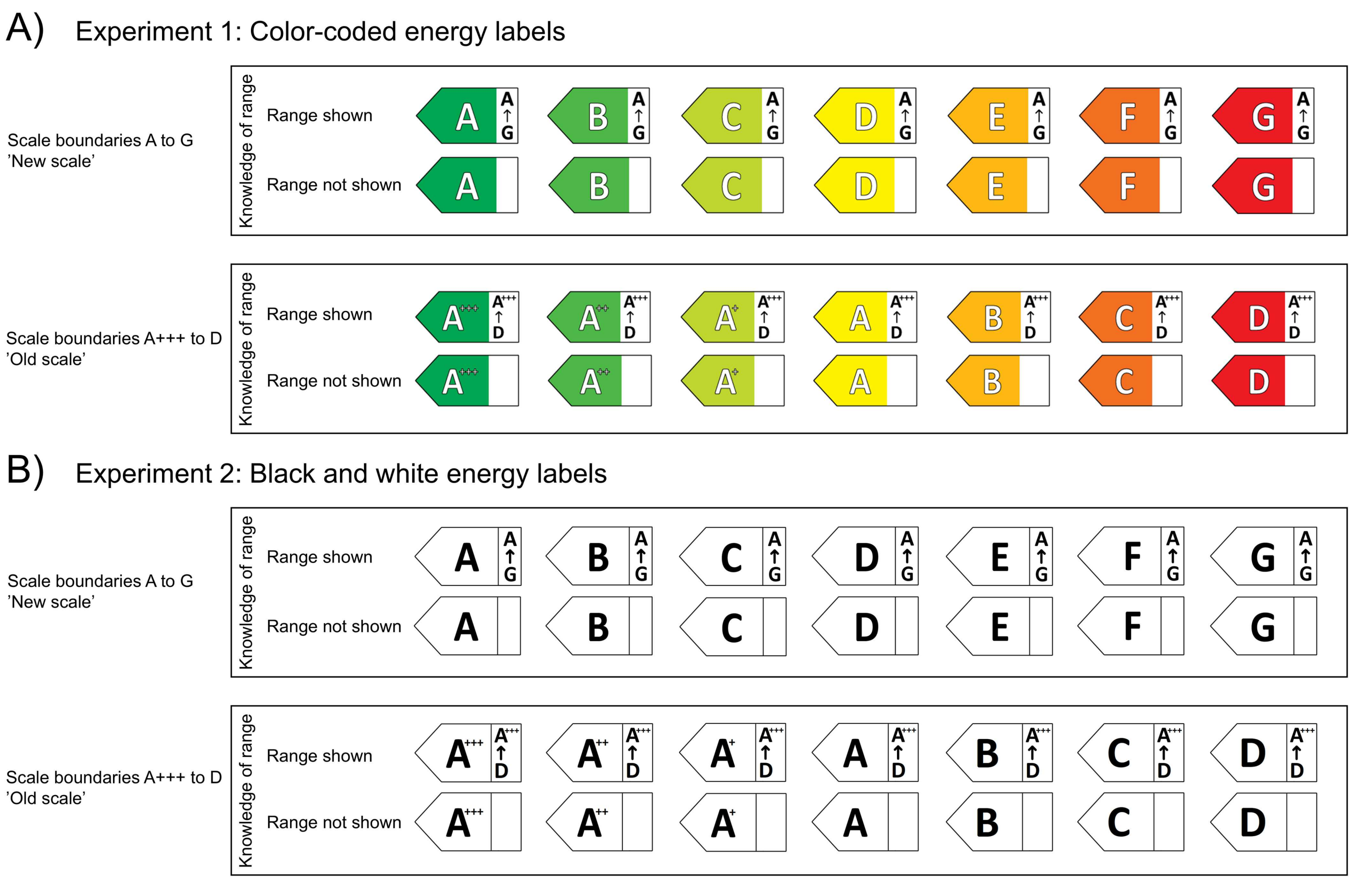

FIGURE 1 | Energy label stimuli used in Experiment (1A) and Experiment (2B). The figure panels also capture the design of each experiment, where participants were divided into four groups. Each group encountered seven energy labels, and the between-participants conditions (Knowledge of range × scale boundaries) determined which row of energy labels would be presented to a group.

compared to the 1990s), new efficient products had become more efficient than class "A" in the old scale, which represented the highest possible energy class. The solution was to extend the scale beyond class "A" by introducing the even more energy efficient energy classes "A+," "A++," and "A+++" in the scale. But it was reportedly more difficult for consumers to understand these proliferated and extended "A+" classes (e.g., Heinzle and Wüstenhagen 2012). Therefore, in a 2021 revision, the EU rescaled this "old scale" system such that "A" should typically be the highest class of energy efficiency, and "G" the lowest (Figure 1). This rescaled "new scale" system is being gradually adopted (though Internet searches by the authors revealed that ovens and other examples of electronic appliances are still visually advertised with the old "A+++" scale, in Sweden in the year 2025).

The EU's energy labels employ green-to-red "traffic light" color-coding, where green represents the most energy efficient and red represents the least energy efficient products (Figure 1A). However, even this is subject to modification as EU energy labels may be printed in black and white if color printing is not used in visual advertisement (Figure 1B). The EU has therefore made available black and white versions of the typically colored energy labels. Our study builds upon this to investigate the psychological impact of presenting energy labels in color versus in black and white.

Regarding information about the scale's range of energy classes, the EU has issued a recent notice that information about the scale's range should be included in the energy label (e.g., a label with energy class B should be accompanied by information that indicates that the range of energy classes is A to G). Range information explicitly conveys what the highest and lowest letters on the scale are (Figure 1). This new directive took effect in 2024 and offers yet another opportunity to study the psychological effects of how different stimulus features in energy labels affect judgments of environmental impact. Information about the scale's range provides an observer with a very strong cue to the scale's context, and we will soon discuss the important role that such contextual information can have in judgments.

Motivated by the fact that the use of letter scales, color-coding, and range vary in marketing contexts, the present study investigates how these scale features influence consumer's environmental judgments. We situate our studies in a theoretical background and develop hypotheses before presenting the details of the present series of experiments.

2 | Theoretical Background and Hypothesis Development

2.1 | The Managerial Importance of Environmental Judgments in Marketing and Consumer Behavior

Energy labels contribute significantly to consumers' understanding of energy efficiency, and consumers use energy labels when forming attitudes about products, within Europe (Andor et al. 2020; Sammer and Wüstenhagen 2006; Skourtos et al. 2021; Stadelmann and Schubert 2018; Thøgersen 2002) and beyond (Andika et al. 2025; Asif et al. 2023; Basiru et al. 2024; Davis and Metcalf 2016; Durmus Senyapar et al. 2024; Elmor et al. 2024; Newell and Siikamäki 2014; Wang et al. 2019). For example, Thøgersen et al. (2024) studied the effects of different kinds of eco-labels, finding that graded labels which provide more comprehensive information of available options were best at guiding consumers toward more sustainable choices. Eco-labels serve the function of revealing the environmental impact of consumer products in a comprehensive way, and labels can therefore have a large impact in shaping environmental attitudes.

One such attitude with direct managerial relevance is the product's perceived environmental friendliness. When appliances are perceived as environmentally friendly or energy efficient, demand and willingness to pay increases (Beck and Toulouse 2023; Faure et al. 2021; Pichierri and Pino 2023), especially in environmentally conscious consumers (Akroush et al. 2019; Baca and Reshidi 2025; Basiru et al. 2024; Wang et al. 2019). Moreover, Sörqvist et al. (2013, 2015) showed that eco-labels can induce a type of "green halo" effect that can enhance both environmental and nonenvironmental stimulus characteristics, further driving consumer's willingness to pay for the products. We focus specifically on how scale features of energy labels influence perceived environmental friendliness as this appears to be strong determinant in consumer choice (Silva et al. 2017; Pichierri and Pino 2023; Sörqvist et al. 2015).

2.2 | Contextual Information in Label Perception, Judgment, and Decision Making

For an observer to form a meaningful understanding of a label, it must be compared to something, as human judgments depend in large part on context (Haubensak 1992; Parducci 1965; Stewart et al. 2006). Here, context can be defined as stimuli in the environment that a target can be meaningfully compared with. And context effects are effects that arise when contextual stimuli influence judgment of a target. The context can often give a target stimulus meaning. Something can only be environmentally friendly, or energy efficient, if compared to something else (Skog et al. 2025; Skog and Sörqvist 2025; Tanner 2008; Thøgersen et al. 2024). The three design dimensions (letters, color-coding, and information about the scale's range) of EU energy labels provide a context that gives individual energy classes meaning by forming contextual points of reference for judgments of a specific energy class.

Heuristics and biases have been proposed as a theoretical framework for understanding the emergence of context effects (Tversky and Kahneman 1974). On this view, people employ heuristics—or rules of thumb—when they make judgments, which sometime end up in biases (systematic judgment tendencies, e.g., over- or underestimations) and often depart from rational choice. Heuristics are typically seen as automatic and spontaneous thought processes that make fast and largely accurate decisions in everyday life possible, without the need of elaborate and time-consuming thought processes (Gigerenzer and Goldstein 1996). According to dual-process theory (Wason and Evans 1974; Samson and Voyer 2012), human thought processes can broadly be classified as either automatic (fast, associative, reflexive, unconscious) or controlled (slow, rule based, reflective, conscious). In this framework, heuristics are often seen as part of the automatic, associative system of reasoning (Sloman 2002).

One such heuristic is "anchoring and adjustment" (Epley and Gilovich 2006; Tversky and Kahneman 1974). When making a judgment, people tend to select a starting (anchor) point and then adjust until some criterion is satisfied. If a contextually available stimulus has a high value (e.g., it is extremely environmentally friendly) and the judgment formation process utilizes this as an anchoring point, the judgment of another stimulus may be exaggerated (e.g., it is seen as very environmentally friendly) in comparison with a situation where the contextually available stimulus has a low value instead (e.g., it is extremely environmentally unfriendly). Similarly, another heuristic that can explain context effects is "availability" (Tversky and Kahneman 1974). According to this heuristic, people base their judgments on the accessibility or retrievability of information from memory. For example, when speaking to a friend about a recent restaurant visit, you might be asked whether the restaurant appeared environmentally friendly. To answer this question, you might attempt to recall vegetarian dishes from the menu as you know these are typically more environmentally friendly than their meat-based counterparts. In contexts that facilitate recall of vegetarian dishes, the restaurant will appear more environmentally friendly. A third heuristic that might be worth mentioning is "representativeness" (Tversky and Kahneman 1974). A product that appears representative of environmentally friendly objects will be judged as more environmentally friendly even if such representativeness is misleading. A hamburger dish, for example, is perceived as more environmentally friendly if served together with an "organic" apple (Gorissen and Weijters 2016). Finaly, the "affect heuristic" (Slovic et al. 2007) can play a central role in how consumers perceive environmental friendliness. On this view, people tend to place stimuli on a qualitative good-to-bad continuum, and base their judgments on this, often overlooking quantitative details that could lead to less biased responses.

2.2.1 | *Letters and Color Coding*

In our examination of the key features of EU's energy labels, we first consider how the letters and color-coding (Figure 1) can help people organize their judgments. The letters provide meaningful contextual information for how consumers should organize their judgments of different energy classes. Letters with lower ranks (e.g., "C" and "D") are intuitively associated with lower environmental friendliness than letters with higher ranks (e.g., "A" and "B") (at least to people who are familiar with the Latin alphabet [A–Z]). Past research has shown that observers find it difficult to distinguish the "A+" classes in the old scale (Heinzle and Wüstenhagen 2012). From this, we might expect a compression in the perceived environmental friendliness of the "A+" classes. Consequently, we would expect the two functions for ratings of energy classes from the two scales to be different: ratings of the old scale (A+++ to D) should follow a nonlinear, skewed, response function with dense ratings of the highest energy classes, while ratings of the new scale (A to G) should follow a more linear response function (Hypothesis 1a). Moreover, products labeled with the old scale are perceived as more sustainable than their new scale counterparts (Stasiuk and Maison 2022). We can therefore expect generally higher ratings of environmental friendliness in the old scale's energy classes (Hypothesis 1b). This could be driven by an inflated perception of environmental friendliness for the "A+" classes.

The energy labels' color-coding is based on the green-red color spectrum. This is an intuitive choice for most Western observers, where this "traffic light" system conveys that green is high, yellow is middle, and red is low in terms of energy efficiency (e.g., Thøgersen and Nielsen 2016). Hence, this "traffic light" scale communicates relative performance, a contextual cue that facilitates comparisons across consumer options (e.g., Grankvist et al. 2004; Thøgersen and Nielsen 2016). Consequently, color-coding might also provide a meaningful contextualizing base for consumers to organize their judgments, similar to that provided by letters. The color green is often associated with positive attributes such as healthy (Schuldt 2013) and environmentally friendly (Seo and Scammon 2017). Label colors can have various effects on consumers' decisions and perceptions. For example, an eco-label with a less intense green color can make a product appear more environmentally friendly—in comparison with a more intense counterpart—and by extension influence consumers' purchase intentions (Pichierri and Pino 2023). Eco-labeling can also interact with the color of the product in their effects on product perception, suggesting that perception is simultaneously shaped by visual (color) and cognitive, expectancy-based (label) inputs (Shankar et al. 2009; see also Sörqvist et al. 2013). Similarly, perception of a product's eco-friendliness can depend on the match (and mismatch) between how consumers cognitively classify a product (e.g., meat being an "environmentally harmful" product) and the color of product packaging (green signaling environmental friendliness vs. red signaling environmental harmfulness) (Sucapane et al. 2021).

A question that arises is whether colors or letters carry more weight in the decision-making process. The present study aims to discover which cue is more influential for consumers' understanding of energy labels. According to the cue-utilization framework (Cox 1962; Sullivan and Burger 1987), observers use and integrate cues in the decision-making process (Slovic 1966). An example of this integration process comes from a study by Pancer et al. (2017), who found that a green color alone, or a (noncolored) environmental label alone, has a smaller effect on sustainability perception compared to when the two cues are combined (i.e., a green eco-label). The authors proposed that this occurs because a single cue is more ambiguous than the two cues in combination. Consumers use these cues and assign them relative weight as a basis for decision making (Cox 1962). For example, "green" core attributes can dominate more peripheral attributes in product purchase decisions (Mei et al. 2025). Furthermore, cues that are fluently processed (Shah and Oppenheimer 2007) or salient (Easterbrook 1959) have a larger weight in the decision process.

Depending on the weight observers of EU's energy labels assign to letters or colors in the judgment process, the judgment of a certain energy class may vary. For example, energy class "A" has a yellow color code in the old scale (Figure 1A). If color has a high weight in the judgment process, this energy class should be seen as relatively low in environmental friendliness. Moreover, when the energy labels are presented in color rather than black and white, the perceived environmental friendliness of energy class "A" should drop. In turn, if the letter is given more weight, the environmental friendliness of this energy class may still be seen as relatively high, regardless of whether the energy labels are presented in color or black and white. Past research indicates that "traffic light" colors can have a larger effect than text-based information on a consumer's decision to make pro-environmental choices (Bengart and Vogt 2023). This could be because colors may attract more attention than their text-based counterpart (Lurie and Mason 2007). If the color-coding of energy labels has a higher weight than letters in judgments of environmental friendliness, we would expect little difference in judgments of EU's old and new scale, since their color-coding is identical (Figure 1A).

However, the color-based information in the study by Bengart and Vogt (2023) comprised a continuous visual-analog scale, whereas the text-based information comprised an exact carbon dioxide value. Other research has also shown that absolute CO_2 emission values on labels have no effect on consumer decisions (Holenweger et al. 2023), perhaps because these are much more difficult to interpret and understand than their color counterpart. This is different from the letters on the EU labels, which comprise a sequence of symbolic letters representing categorical degrees of energy efficiency. Moreover, the letters are superimposed on the color, perhaps making them a more salient cue. In this case, we might expect letters to gain more weight than colors in the decision-making process.

Taken together, if letters have a higher weight than color, then observers should perceive energy labels similarly when energy labels are presented in color as when they are presented in black and white (Hypothesis 2a). Conversely, if color has a higher weight than letters, then observers should perceive energy labels differently when energy labels are presented in color than when they are presented in black and white (Hypothesis 2b).

2.2.2 | Range Knowledge

Next, we turn to the range of the scale, a third key feature of EU's energy labels (Figure 1). According to the range-frequency theory (Parducci 1965, 1968), observers organize their responses based on the range of available stimuli. Information about the scale's range can have a strong impact on how observers organize their ratings of different energy classes. This is dependent on knowledge of the highest and lowest values (i.e., the endpoints) in the stimulus distribution. For example, an observer faced with an energy label "A" is likely to be influenced by whether they know that "A" is the highest of seven possible energy classes (as it is in the new energy scale, ranging from "A" to "G"; Figure 1) or the fourth position of the scale (as in the old energy scale, ranging from "A+++" to "D"; Figure 1). If the observer does not have this knowledge, it is more difficult to contextualize the energy class appropriately. Thus, the understanding of an energy class (e.g., "A") will depend on which scale it belongs to and the observers' knowledge of scale range (cf. Skog and Sörqvist 2025). Given knowledge of the scale's range, the observer will be able to make an informed judgment based on thorough understanding of the to-be-evaluated stimulus, in the context of the scale.

Such "range knowledge" can be provided externally, for example, by presenting information about the scale range on the energy label. For the EU energy label, scale range is shown by presenting the top and bottom letter with an arrow between them, next to the energy class (Figure 1). Even without this externally provided scale-range information, an observer can use other cues such as experience or knowledge (Birnbaum and Rouvere 2023) to intuitively organize the energy classes. Of relevance here is past knowledge of the alphabet and the "traffic light" color coding, which both can provide an intuitive idea about the scale's endpoints. There are thus multiple ways that an observer can grasp the scale's range.

The old and new energy scales are organized such that they begin with A (Figure 1); a choice where the beginning of the alphabet can be an intuitive contextual reference that is likely to convey a message of "high" energy efficiency. If no additional information about range is provided, the alphabet can still be a strong cue to how energy classes should be organized, and "A" comprises an intuitive boundary of the scale, since this is the beginning of the alphabet. In other words, even without external range information, an observer can use their knowledge about the alphabet to intuitively form an idea about the scale's starting point.

If the energy scale instead runs from an arbitrary starting point, let's say from "K" to "Q" instead of "A" to "G", observers should find it difficult to contextualize an energy class, especially if no external information about scale range is provided. If an observer is presented with a product with energy class "L" and asked to rate its environmental friendliness, ratings will probably be more organized if the observer is also told that the range runs from "K" to "Q." Without this external scale range knowledge, the observers' prior knowledge about the alphabet will be quite unhelpful. Instead, prior knowledge about color coding can be relatively helpful.

Taken together, providing observers with external information about scale range by presenting the range on the energy labels (Figure 1) should have an effect on observers' ratings as this information helps observers contextualize energy classes (Hypothesis 3a). Furthermore, external information about scale range should have a stronger effect on observers' responses when their prior experience (e.g., with color-coding or the alphabet as an organization principle) is not useful, such as when energy classes are presented in black and white (Hypothesis 3b) or when the letters run from "K" to "Q" or some other letter set that does not include the intuitive starting point of the alphabet (Hypothesis 3c). Conversely, when prior knowledge can help organize the energy classes meaningfully (e.g., color coding, from green to red), external scale range information will be less impactful.

3 | Research Gap and Study Overview

The primary aim of the present study is to evaluate how different stimulus features (letters, color-coding, and information about the scale's range) influence judgments, with a particular focus on explaining how these stimulus features provide contextual information to support judgments. While previous research has looked at how scale features such as color (Pichierri and Pino 2023) and number of categories (Thøgersen et al. 2024; Zhong et al. 2025) influence consumer perception and made direct comparisons between EU's old and new ecodesign (Stasiuk and Maison 2022), this is the first study that conducts a systematic investigation of how energy labels' stimulus dimensions interact in their effect on consumer perception. Furthermore, the critical role of scale range knowledge in consumer perception of energy labels has (to the best of our knowledge) never been explored before. In a series of three experiments, we evaluated how different stimulus/design features in the energy labels contribute to perceptions of environmental impact by manipulating the three stimulus dimensions (letters, color-coding, and information about the scale's range).

A summary of our hypotheses, supporting literature and the novel contribution of the present study is shown in Table 1. The following hypotheses were tested:

H1a. *Environmental friendliness ratings of energy classes in the old scale (A+++ to D) will follow a nonlinear, skewed, function with dense ratings of the highest energy classes, while ratings of energy classes in the new scale (A to G) will follow a more linear function (on the assumption that categorical letter cues are given more weight than color cues during judgment).*

H1b. *Environmental friendliness ratings of energy classes in the old scale (A+++ to D) will be higher than ratings of energy classes in the new scale (A to G) (on the assumption that categorical letter cues are given more weight than color cues during judgment).*

H2a. *There will be no difference in environmental friendliness ratings of energy classes presented in color and corresponding ratings of energy classes presented in black and white (on the assumption that color cues provide no additional effect beyond that which is given by the categorical letter cues during judgment).*

H2b. *There will be a difference in environmental friendliness ratings of energy classes presented in color and corresponding ratings of energy classes presented in black and white (on the assumption that color cues provide an additional effect beyond that which is given by the categorical letter cues during judgment).*

H3a. *There will be a difference in environmental friendliness ratings of energy classes depending on whether participants are informed about scale range before ratings: Range knowledge should facilitate discrimination of energy classes.*

H3b. *The effect of informing participants about scale range before ratings will be larger when energy classes are presented in black and white compared to when they are color-coded (as color-coding may provide a separate contextual cue).*

H3c. *When energy classes are based on an unintuitive scale of letters ("K" to "Q"), other contextual cues (e.g., range knowledge) will become even more important for discriminating energy classes.*

Experiment 1 and 2 investigated the effects of using two different scales; one scale ranged from letters A to G (new scale), and the other scale ranged from letters A+++ to D (old scale). These experiments tested H1a and H1b. Furthermore, in Experiment 1, all energy labels were "traffic light" color-coded versions, where the most energy efficient label was green, and the least energy efficient label was red (Figure 1A). This was changed in Experiment 2, which was an exact copy of the first experiment but with strictly black and white energy labels (Figure 1B). Differences across these two experiments can thus reveal the impact of color-coding, testing H2a and H2b.

Across all experiments, we evaluated the contextualizing effects of having knowledge of the scale's range—knowing, or not knowing, the endpoints of the scale (e.g., knowing that "A" is

TABLE 1 | Overview of hypotheses, supporting literature, and novel contribution of the present study.

Hypotheses	Key supporting literature	Novel contribution from the present study
H1a: Environmental friendliness ratings of energy classes in the old scale (A+++ to D) will follow a non-linear, skewed, function with dense ratings of the highest energy classes, while ratings of energy classes in the new scale (A to G) will follow a more linear function	Heinzle and Wüstenhagen (2012); Stasiuk and Maison (2022)	Measurement of the functions describing the relations between energy classes and perceived environmental friendliness
H1b: Environmental friendliness ratings of energy classes in the old scale (A+++ to D) will be higher than ratings of energy classes in the new scale (A to G)	Stasiuk and Maison (2022)	Measurement of the perception of individual energy classes
H2a: There will be no difference in environmental friendliness ratings of energy classes presented in color and corresponding ratings of energy classes presented in black and white	Easterbrook (1959); Shah and Oppenheimer (2007); Tversky and Kahneman (1974)	Testing how consumers assign relative weight to label cues when determining environmental friendliness
H2b: There will be a difference in environmental friendliness ratings of energy classes presented in color and corresponding ratings of energy classes presented in black and white	Bengart and Vogt (2023); Holenweger et al. (2023); Lurie and Mason (2007)	Testing how consumers assign relative weight to label cues when determining environmental friendliness
H3a: There will be a difference in environmental friendliness ratings of energy classes depending on whether participants are informed about scale range before ratings: Range knowledge should facilitate discrimination of energy classes	Parducci (1965); Skog et al. (2025); Skog and Sörqvist (2025)	Testing the role of scale range knowledge in the context of energy/eco labeling
H3b: The effect of informing participants about scale range before ratings will be larger when energy classes are presented in black and white compared to when they are color-coded	Parducci (1965); Skog et al. (2025); Skog and Sörqvist (2025)	Testing the relative importance of scale range knowledge vs. color-coding in energy/eco labels
H3c: When energy classes are based on an unintuitive scale of letters ("K" to "Q"), other contextual cues (e.g., range knowledge) will become even more important for discriminating energy classes	Parducci (1965)	Testing the relative importance of different contextualizing cues in energy/eco labels

the most energy efficient energy class, and that "G" is the least energy efficient energy class). Experiment 1 and 2 tested H3a and H3b. To complement the first two experiments, Experiment 3 used a scale with a very unintuitive range of letters (K to Q) with the aim of isolating the contextualizing effects of color-coding and knowledge of the scale's range. Thus, Experiment 3 targeted H3c, but also H3a and H3b.

4 | Experiment 1 and 2

The study centered on how different stimulus features in energy labels contribute to judgments of environmental friendliness. Experiment 1 and 2 had a high degree of methodological similarity, and the only difference was the presence (Experiment 1; Figure 1A) or absence (Experiment 2; Figure 1B) of color-coding for the energy labels. We thus provide a joint declaration of the methods of the two experiments. Both experiments focused on the differential effects of the new (A to G) and old (A+++ to D) scales of energy labels, and the effects of providing observers with knowledge of the scale's range. Our discussion of the results will first cover each experiment's results in sequence, followed by a comparative synthesis of both experiments.

4.1 | Methods

4.1.1 | Participants

Two hundred and fifty-six (123 female, 133 male, mean age = 47.5 years, SD = 13.9) participants took part in Experiment 1. Participants were recruited via Prolific (www.prolific.com) in this internet-based study. All participants were native English speakers who resided in the United Kingdom, who reported normal or corrected-to-normal vision and no color-blindness. The sample size was determined using a two-tailed a-priori power analysis in G*Power (Faul et al. 2007). We assumed that the targeted effects would be of medium size (Cohen's $d = 0.5$). This assumed effect size can be considered generic, as we had no appropriate available data to base our assumption of effect size on. With power ($1–\beta$ error probability) set to 0.80 and with an $\alpha = 0.05$, the required sample size was $N = 64$ in each group. We thus collected data from 256 participants, 64 in each group, and no participant was excluded from the analysis. Participants were compensated with £0.45 for a total study time of around 2–3 min.

Following the same procedure as above, 256 participants who had not taken part in Experiment 1 were recruited for Experiment 2 (127 female, 129 male, mean age = 46.7 years, SD = 13.9).

4.1.2 | Materials

The experiment was created using PsychoPy (version 2024.2.2, Peirce et al. 2019) and ran online with Pavlovia. Participants accessed the experiment via a hyperlink on Prolific. The experiment ran in the participants' browser on their own desktop computers, and a mouse and keyboard were required for participation.

The stimulus material presented on each trial was an image of an oven, paired with an energy label. The rest of the screen had a mid-gray background. Seven PNG images of different ovens were created with the assistance of a web-based artificial intelligence service. The ovens were similar in appearance, all had a steel-colored exterior, a dark glass window on the door, and a black electric stove top with four burners. All ovens were shown in "canonical" viewpoints. The stove tops were visible in all seven oven images. Four ovens had a front-facing profile, and three ovens were rotated about the z-axis by around 30%, revealing the side of the oven. The background in these square oven images was a neutral light gray. See Appendix A and Figure A1 for the seven oven images.

Energy labels were of the "Class arrows" type, made available by the EU in PNG format on the file exchange tool CIRCABC. In this study, we made edited versions based on these energy labels to suit our experimental manipulations. The energy labels used in Experiment 1 were color-coded (Figure 1A) and the labels used in Experiment 2 were shown in black and white (Figure 1B). The labels contained a visual indicator of the scale range, in the style of an arrow pointing from the lowest available energy class to the highest (see, e.g., the top row of energy label stimuli in Figure 1A). For our experimental manipulation, we created copies of these energy labels with removed range indicators ("Range not shown" in Figure 1). We used two scales of energy labels, defined by their letters (Figure 1). The new scale had letter scale boundaries A to G (energy classes A, B, C, D, E, F, and G), and the old scale had letter scale boundaries A+++ to D (A+++, A++, A+, A, B, C, and D).

The size of the stimulus images was standardized based on the participant's monitor height, such that an image of an oven occupied a square 40% of the participant's monitor height. Each oven image was shown in the center of the screen, and an energy label was shown immediately to the center-right of it, with no blank space between or overlapping of images. Energy labels were smaller than the oven images, occupying 10% of the monitor's height with an aspect ratio of 5:3.

4.1.3 | Design and Procedure

Both Experiment 1 and 2 were mixed factorial designs, with three factors: 2 (Knowledge of range: range shown, range not shown [between-participants factor]) × 2 (Scale boundaries: A to G, A+++ to D [between-participants factor]) × 7 (Energy class: 1st, 2nd, 3rd, 4th, 5th, 6th, 7th [within-participants factor]). The Knowledge of range factor determined whether participants were shown information about the scale's range, that is, the lowest and highest possible energy classes. Half of all participants were shown this information, and half were not shown this information (see Figure 1). Participants were randomly assigned into groups that made up the between-participants conditions. The Scale boundaries factor determined the letters used as steps on the stimulus scale, which ranged from A to G (new scale) or A+++ to D (old scale). Half of all participants encountered one scale, and the other half encountered the other scale (see Figure 1). All participants encountered seven energy classes, and this Energy class factor was the only within-participants factor. There were seven trials in the experiment, one per energy class. Participants were not forewarned about how many trials they would complete.

The study was advertised on Prolific as a study on "Energy efficiency and environmental impact." Participants provided informed consent by button press before beginning the experiment. Upon beginning the experiment, participants in all conditions encountered some instructions on what the experiment would entail, which stated: "You will judge the environmental friendliness of different ovens. All the ovens run only on electricity, and they also have built-in electric stove tops that can be used for cooking with pots and pans. Each oven also has an Energy Label that indicates how energy efficient the oven is. The energy labels divide the ovens into energy classes. The energy classes are characterized by a letter and a color. For example, energy class B is more energy efficient than energy class C." Participants pressed any key on their keyboard to continue. Instructions did not inform any participants of the color-coding in the energy labels. In Experiment 1, color-coding was always green-red in the order of best-to-worst in terms of energy efficiency (Figure 1A). In Experiment 2, there was no color-coding (Figure 1B).

The participants who were provided with knowledge of the scale's range were also told, on the initial instruction screen, that "Energy labels also display the range of possible energy classes. The least energy efficient energy class is G (or D, depending on condition), and the most energy efficient energy class is A (or A+++, depending on condition)." Participants who were provided with this foreknowledge of the scale's range encountered a second instruction screen where they were asked to confirm that they had understood that the least energy efficient energy class is G (or D, depending on condition), and the most energy efficient energy class is A (or A+++, depending on condition). Participants were to click an on-screen button with their cursor to confirm that they understood. They were then prompted to start the experiment by an additional button press.

Each trial started with a paused screen with text prompting participants to be prepared to rate the environmental friendliness of the next oven. Participants clicked the spacebar button to view and rate a new stimulus. Stimulus images of ovens and energy labels appeared in an independently random order across the trials. This random order was different for each participant, and there was no systematic pairing of ovens and energy labels. Below the stimulus images was a rating response scale that participants would click to give a response. The response scale had 11 response options. The first (leftmost) scale response option read: "Not environmentally friendly," and the eleventh (rightmost) scale response option read: "Very environmentally friendly." Trials were not limited in duration, and the stimulus was shown until a rating-scale response ended the trial. Upon responding to all seven trials, a debrief screen was shown that informed participants of the aims and expected value of the study. Participants were then finished and taken back to Prolific.

4.1.4 | *Data Availability and Transparency*

This article fully declares its hypotheses, statistical power estimates, materials, design, analyses, and results. Statistical analyses were conducted using Jamovi (version 2.3.28.0). The reported experiments were not preregistered. The data for all three experiments are publicly available via the Open Science Framework (https://doi.org/10.17605/OSF.IO/QXBER).

4.2 | Results and Discussion

Figure 2 shows the results of both experiments. Our declaration and discussion of results will first cover Experiment 1 then Experiment 2, followed by a comparative synthesis of both experiments.

4.2.1 | *Results of Experiment 1*

Each participant provided responses to seven trials, one per energy class. The top panels of Figure 2 show the results of Experiment 1. Our analyses focused on the effects and interactions of the experimental factors Knowledge of range (range shown, range not shown), Scale boundaries (A to G, A+++ to D), and how these factors interacted across the seven energy classes. To analyze the influence that these factors had on judgments of environmental friendliness, we used a repeated measures 2 (Knowledge of range: range shown, range not shown) × 2 (Scale boundaries: A to G, A+++ to D) × 7 (Energy class: 1st, 2nd, 3rd, 4th, 5th, 6th, 7th) mixed analysis of variance (ANOVA) that was adjusted with a Greenhouse-Geisser correction. The full results of the ANOVA are shown in Table 2. As expected, there was a large main effect of energy class ($p < 0.001$), and the pattern of results in Figure 2 indicates that participants easily understood that the energy labels are arranged in order, conveying different levels of environmental impact, from best to worst. This manipulation check indicates that the energy classes of both scales were comprehended. The between-subjects factors will be referred to in brief as "range," meaning the Knowledge of range factor, and "scale," meaning the Scale boundaries factor.

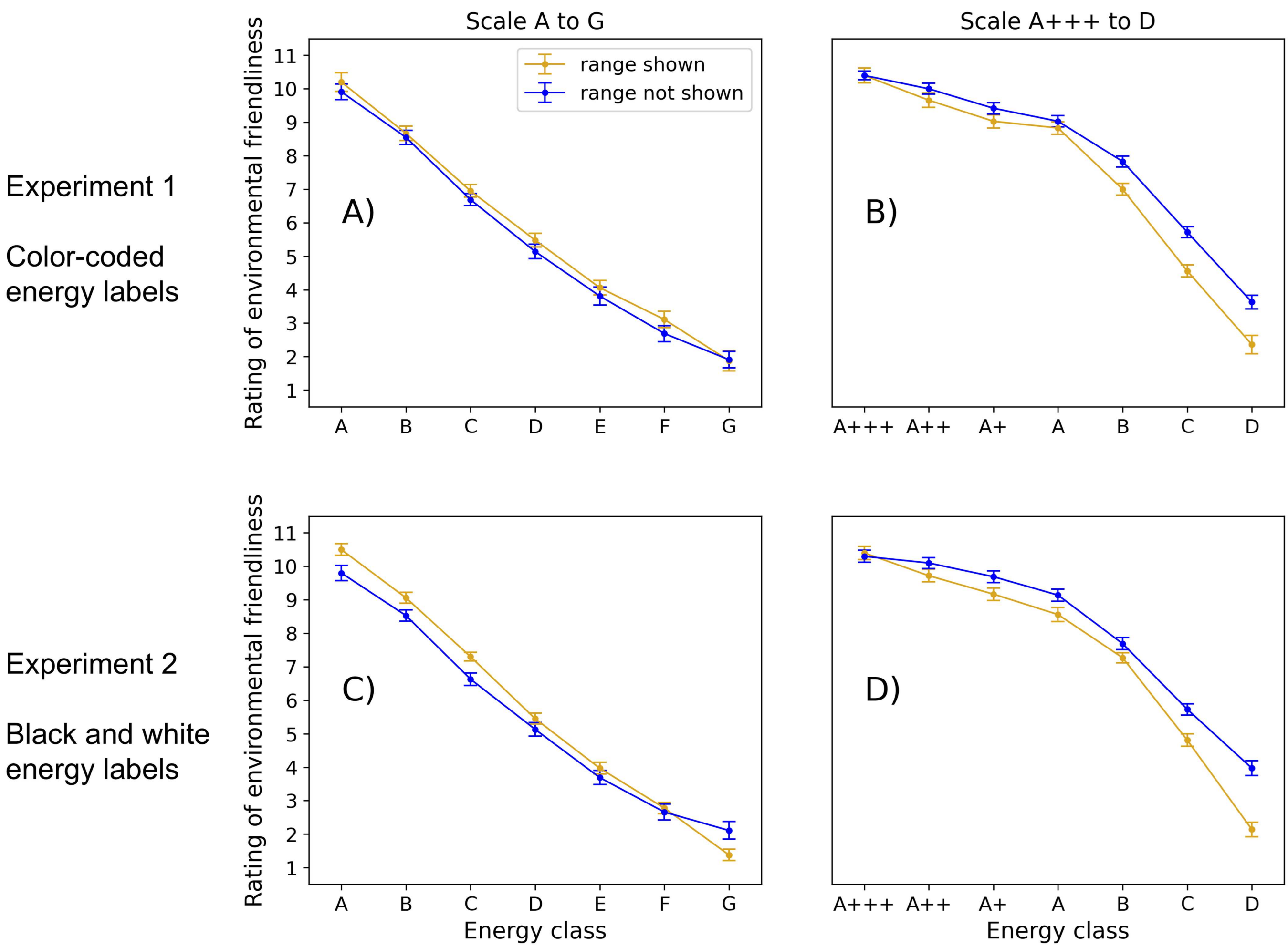

FIGURE 2 | Results of Experiment (1A and B) and Experiment (2C and D). Error bars represent the standard errors of the means.

4.2.2 | *The Letters on the Scale Matter*

There was a highly significant main effect of scale ($p < 0.001$; Table 2). The A+++ to D scale tended to produce on average higher ratings than the A to G scale (compare panels A and B in Figure 2), indicating that participants believed that ovens labeled using the A+++ to D scale were more environmentally friendly than ovens labeled using the A to G scale. This suggests that participants assigned more weight to letters than color-coding when making their judgments (confirming Hypothesis 1b), as this large difference across the scales emerged despite the two scales having identical color-coding. Letters were a dominant cue for judgments. The middle of each scale was a yellow-colored D (A to G scale), and a yellow-colored A (A+++ to D scale). These two points on the scale can be considered equal, in the sense that they are the middle (fourth) energy class with yellow color. But the average ratings in the two scale conditions differed by 3.62 points on the 11-point response scale (compare panels A and B in Figure 2). A Mann–Whitney U test confirmed the significance of this very large difference, $U(254) = 853$, $p < 0.001$. The A+++ to D scale thus tended to inflate judgments of environmental friendliness compared to the A to G scale.

Moreover, the scale factor produced a strong interaction with the energy classes ($p < 0.001$). The A to G scale produced a more stable and linear trend across energy classes than the A+++ to D scale (compare panels A and B in Figure 2). The response curve across the energy classes with the A+++ to D scale shows a 'knee' function that bends around the middle of the scale; participants were more discriminative across energy classes A, B, C, and D, and less discriminative across energy classes A+++, A++, A+, and A (Figure 2B), indicating a skewed and biased response pattern. In contrast, participants' responses were more evenly spaced across the energy classes with the A to G scale (Figure 2A). This difference across the scales was formally confirmed by a new repeated measures ANOVA that compared 2 (higher vs. lower ends of the scale) × 3 (three energy classes) × 2 (scale A to G vs. scale A+++ to D), where the key three-way interaction was significant, $F(1.71, 433.46) = 134.57$, $p < 0.001$, $\eta^2_p = 0.346$, and produced a very large effect. These behavioral results offer support for the efforts to rescale EU energy label scales towards A to G, and we recommend the phasing out of the A+++ to D scale, since it produces a lower degree of scale-step discriminability at the higher ends of the scale (the "A+" classes). We conclude that the A+++ to D scale can inflate the perceived environmental friendliness of labeled products (confirming Hypotheses 1b) and reduce an observer's ability to interpret the difference between energy classes. Especially the top "A+" classes were less distinct, confirming Hypothesis 1a.

4.2.3 | *The Role of Contextual Information*

The present study focuses on how contextual information is used in energy labeling and how such context influences environmental judgments. Without sufficient contextual information, judgments can become unreliable and unorganized (Parducci 1965; Stewart et al. 2006). We expected that the effects of having knowledge of the scale's letter range might interact across the energy classes, as the letter range knowledge could provide participants with beneficial contextual information that would help them tell apart energy classes (Hypothesis 3a). For example, knowing the endpoints of the scale (i.e., best and worst letter) should help participants to better organize their responses on the scale (Parducci 1965, 1968). But we were also open to the possibility that the green-red color-coded scale, which was used for all stimuli in Experiment 1 (Figure 1A), might provide adequate contextual information on its own, if participants assume that a green and red energy label is the most and least energy efficient, respectively. This might be revealed by no interaction between energy class × range knowledge. This latter prediction was confirmed, as indicated by no significant interaction between energy class × range ($p = 0.140$; Table 2). Furthermore, there was no three-way interaction between energy class × range × scale ($p = 0.244$), which indicates that any effects of range were not reliably engaged more with one scale than the other. This suggests that both letter scales, with color-coding, produced robust ratings that range knowledge did not significantly alter. This suggests that range knowledge did not add contextualizing information beyond that which was already given by the other stimulus characteristics—letters and colors. Thus, Hypothesis 3a was not supported by Experiment 1.

For completeness, we report that there was no main effect of range ($p = 0.101$), but that there was a significant interaction between range × scale ($p < 0.001$). This interaction is of less interest to us, but it shows that seeing the range tended to increase ratings in the A to G scale (Figure 2A) but lower ratings in the A+++ to D scale (Figure 2B). A supplementary analysis treated the self-reported demographic variables age as a covariate and sex as a between-participants factor. This revealed

TABLE 2 | Results of a repeated measures ANOVA for the full Experiment 1.

Factor	*df*	*F* value	*p* value	η^2_p
Energy class	2.36, 594.55	841.68	< 0.001 ***	0.770
Range	1, 252	2.71	0.101	0.011
Scale	1, 252	342.12	< 0.001 ***	0.563
Energy class × range	2.36, 594.55	1.92	0.140	0.008
Energy class × scale	2.36, 594.55	40.80	< 0.001 ***	0.139
Range × scale	1, 252	13.65	< 0.001 ***	0.051
Energy class × range × scale	2.36, 594.55	1.41	0.244	0.006

Note: Energy class reflects the Energy class factor with seven levels. Range reflects the Knowledge of range factor with two levels. Scale reflects the Scale boundaries factor with two levels.
***$p < 0.001$.

no effects of age, but sex produced a three-way interaction with energy class × range × sex, $F(2.4, 595.35) = 3.09$, $p = 0.037$, $\eta^2_p = 0.012$. Female participants were more influenced by range than male participants. Range helped female participants make a greater distinction between energy classes, compared to when no range was shown (energy class × range for only female participants, $F(2.59, 311.14) = 4.19$, $p = 0.009$, $\eta^2_p = 0.034$), in partial support for Hypothesis 3a. This is consistent with previous findings which have suggested that females can be more sensitive to contextual cues than males (Meyers-Levy and Maheswaran 1991; Meyers-Levy and Sternthal 1991). However, the lack of a basic interaction between energy class × range ($p = 0.140$; Table 2) suggests that the effects of range were unreliable and should be tested with different experimental parameters. The role of range knowledge will be examined further in both following experiments. We now turn to discuss the results of Experiment 2, which studied the effects of range when there was no color-coding in the energy labels. Later, we declare Experiment 3 which took the stimulus manipulation one step further in a study of the effects of range with a very unintuitive letter scale, in a factorial design which also evaluated the effects of color versus no color.

4.2.4 | *Results of Experiment 2*

As for Experiment 1, our analyses focused on the effects and interactions of the mixed 2 (Knowledge of range: range shown, range not shown) × 2 (Scale boundaries: A to G, A+++ to D) × 7 (Energy class: 1st, 2nd, 3rd, 4th, 5th, 6th, 7th) design, which was analyzed with a corresponding repeated measures ANOVA that was adjusted with a Greenhouse-Geisser correction. The full results of this ANOVA are shown in Table 3, and the data from Experiment 2 are visualized in the bottom panels of Figure 2.

The effects of the two scales (A to G vs. A+++ to D) replicated the results previously seen and discussed in Experiment 1; there was a main effect of scale ($p < 0.001$) and an interaction across scale × energy classes ($p < 0.001$). Again, we see the same tendency in results where the new A to G scale produced a more linear trend in ratings across energy classes (Figure 2C), whereas ratings of the old A+++ to D scale tended to be less discriminating in the upper ends of the scale than in the lower ends of the scale (Figure 2D), indicating biased and skewed responses. This again confirms Hypotheses 1a and 1b and reinforces our suggestion that the A+++ to D scale is worse than the A to G scale, as it can impair an observer's ability to tell apart scale steps, and it can cause an inflation in perceived environmental friendliness.

4.2.5 | *The Effect of Range Knowledge—Without Color-Coded Energy Labels*

Experiment 2 was a direct replication of Experiment 1 except energy labels were presented in strictly black and white, with no color-coding (Figure 1). Experiment 2 thus serves to re-evaluate Hypothesis 3a without any influence of color-coding in energy labels. We expected that the influence of range might be stronger in the absence of color (Hypothesis 3b), which should lead to a significant interaction with range across the energy classes. Color was suspected of having provided "range-like" contextual information in Experiment 1, thus the removal of color-coding may increase the participants' reliance on range information, as range hypothetically becomes a relatively more important contextual cue in the absence of color-coding (Hypothesis 3b). We were interested in the possible context effect provided by color-coding, as, according to EU directives, energy labels may be printed in black and white. The black and white energy labels thus exist in real world applications, and as far as we are aware, there exists no available evidence on the behavioral effects of removing color from energy labels.

The primary aim of Experiment 2 was to investigate whether knowledge of range would have an impact on ratings of energy classes, now that color was removed from the energy labels. Range did have a significant interaction across energy classes ($p < 0.001$, Table 3) without color (supporting Hypothesis 3a and 3b). This contrasts with the results of Experiment 1 (which used color), where this interaction failed to reach significance at $p = 0.140$ (Table 2). This suggests that color likely provided contextual information, where participants could organize their ratings from high-low environmental friendliness based on green-red color in the energy labels (supporting Hypothesis 3b). Consistent with Hypothesis 3b, range knowledge was more impactful for organizing rating responses when color-coding was not available. We interpret this, as a more general rule, to mean that one contextual cue becomes relatively more impactful when there are fewer other contextual cues available.

This has applied implications, as we demonstrate behavioral evidence that information about the scale's range is even more important in cases when energy labels are printed in black and white. EU energy labels may be printed in black and white in

TABLE 3 | Results of a repeated measures ANOVA for the full Experiment 2.

Factor	*df*	*F* value	*p* value	η^2_p
Energy class	2.44, 615.62	1182.61	< 0.001***	0.824
Range	1, 252	3.06	0.081	0.012
Scale	1, 252	378.13	< 0.001***	0.600
Energy class × range	2.44, 615.62	10.41	< 0.001***	0.040
Energy class × scale	2.44, 615.62	61.08	< 0.001***	0.195
Range × scale	1, 252	17.85	< 0.001***	0.066
Energy class × range × scale	2.44, 615.62	0.38	0.724	0.002

Note: Energy class reflects the Energy class factor with seven levels. Range reflects the Knowledge of range factor with two levels. Scale reflects the Scale boundaries factor with two levels.
***$p < 0.001$.

visual advertisement when color printing is not used. Without color, information about the scale's range can be of special utility for understanding energy labels. Evidence that responses were more organized with range knowledge can be seen in the details of the bottom panels of Figure 2. The interaction maximizes at the extreme ends of the scales (1st and 7th energy classes; Figure 2C,D), where participants who were not shown range information (blue data) rated the ends of the stimulus scales (x-axis of Figure 2C,D) as farther from the ends of the response scale (y-axis of Figure 2C,D) compared to participants who were shown range information (gold data). This was seen at all ends of the energy classes in Figure 2C,D, except for the "A+++" class (Figure 2D), where both groups appear to have been at a similar response ceiling. This indicates that range knowledge helped participants to tell apart the energy classes, supporting Hypothesis 3a and 3b. A supplementary analysis on demographic variables found no effects of age, and sex did not significantly alter the effects of range knowledge.

4.2.6 | *Synthesis of Experiment 1 and 2 and Cross-Experiment Analysis*

However, as can be seen when comparing the top and bottom panels of Figure 2, representing the results of Experiments 1 and 2, respectively, the two experiments produced very similar overall results. Judgments tended to be dominated by the letters more so than color-coding or range knowledge (confirming Hypothesis 2a, and disconfirming Hypothesis 2b). Therefore, we formally tested if the effects of having color-coded energy labels versus having black and white energy labels were reliable in a cross-experiment analysis (with comparisons across Experiment 1 and 2). This analysis covers all panels of Figure 2. Here, we focused on the range × energy class × color (Experiment) three-way interaction, for each of the two scales separately (A to G and A+++ to D). The first analysis evaluated the differential effects of range × energy class × color for the A to G scale (cf. Figure 2A,C), which found no significant interaction, $F(2.13, 536.29) = 1.12$, $p = 0.331$, $\eta^2_p = 0.004$. The second analysis evaluated this three-way interaction for the A+++ to D scale (cf. Figure 2B,D), and similarly found no interaction, $F(2.33, 586.62) = 1.19$, $p = 0.310$, $\eta^2_p = 0.005$. This suggests that, while the range × energy class interaction was significant without color-coding ($p < 0.001$; Experiment 2; Table 3) but not with color-coding ($p = 0.140$; Experiment 1; Table 2), the effect of color on range × energy class was statistically unreliable and should be evaluated further (thus undermining the support for Hypothesis 3b). The lack of a three-way interaction, where color did not significantly alter the range × energy class interaction with these two scales, further suggests that the letters in both scales are very strong cues that participants rely on to organize their rating responses. Our Experiment 3 therefore used a far less intuitive letter scale to isolate the effects of color-coding and range knowledge on energy class ratings.

5 | Experiment 3

Experiment 3 was like the first two experiments, but it used a very unintuitive letter scale (K to Q), instead of the more intuitive and typical scales A to G or A+++ to D. In both of these traditional scales, participants are likely to assume that "A" (and its variants with added "+" signs) represents a good energy rating, as the beginning of the alphabet provides an intuitive "high point." However, if the letter "K" represents the best available energy class, participants are not likely to believe that the "K" class is a good energy rating unless there are some other forms of contextualizing cues that can inform participants of the scale's boundaries (Hypothesis 3c). We expected that range knowledge (Hypothesis 3a) and color-coding (Hypothesis 3b) can provide these types of contextual cues. Further, the previous two experiments have shown that, while range and color-coding can have some impact on judgments, the impact is relatively small; most of the response variance is explained by the letters. Therefore, with an unintuitive letter scale, we expected the effects of range knowledge and color to be stronger in Experiment 3 than in the first two experiments (Hypothesis 3c). Experiment 3 also examined whether range knowledge and color-coding provide equally strong contextual cues. We expected that range knowledge would provide a stronger cue to the scale's context than the color-coding, since information about the range of letters is a more explicit design feature in the energy labels than information about the color-coding.

5.1 | Methods

5.1.1 | *Participants*

Keeping consistent with the first two experiments, we recruited the same number of 256 participants ($N = 64$ in each group) via Prolific (134 female, 122 male, mean age = 44.2 years, SD = 14.2). No participant had previously taken part in Experiment 1 or 2. Two participants were replaced owing to computer errors and missing data. All participants were native English speakers who resided in the United Kingdom, who reported normal or corrected-to-normal vision and no color-blindness. We followed the same ethical considerations as described under Experiment 1 and 2. Participants gave informed consent and were debriefed upon completing the study. Participants were compensated with £0.45 for a total study time of around 2–3 min.

5.1.2 | *Materials, Design and Procedure*

Unless noted here, the materials, design, and procedure were identical to Experiment 1. Experiment 3 was a mixed factorial design with experimental factors: 2 (Knowledge of range: range shown, range not shown [between-participants factor]) × 2 (Color-coding: color, black and white [between-participants factor]) × 7 (Energy class: K, L, M, N, O, P, Q [within-participants factor]). Participants were randomly assigned into one of the four groups that made up the between-participants conditions.

We created new energy labels that were of the same "Class arrow" style as in the previous two experiments, but where the letters were changed to K, L, M, N, O, P, and Q, and the range information about the letters was an arrow pointing from Q to K. These seven letters were drawn from the middle of the alphabet, rather than from the beginning of the alphabet (cf. Experiment 1 and 2) to achieve a less intuitive range of letters. These new energy labels, and how they varied across the four between-participants conditions, are shown in Figure 3.

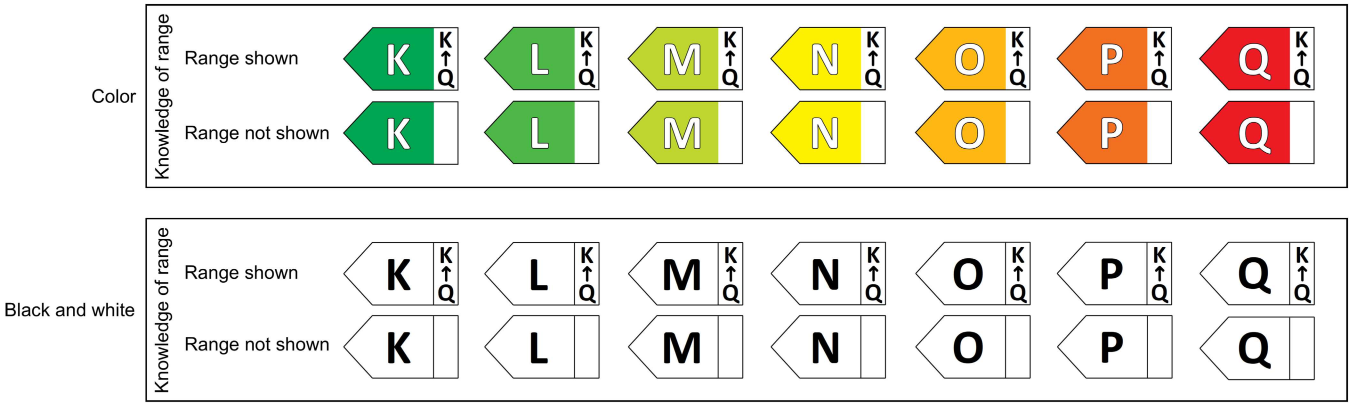

FIGURE 3 | Energy label stimuli used in Experiment 3. The figure is organized like Figure 1, but the four between-participants conditions were Color-coding (color, black and white) × knowledge of range (range shown, range not shown).

For colored energy labels, we used the typical "traffic light" color-coding. The software and stimuli were otherwise unchanged from Experiment 1, and the new energy labels had the same size, shape, and position on the participants' monitors as in the previous experiments. Instructions were edited to state that "alphabetical order reveals energy efficiency classes, where letters that appear earlier in the alphabet are more energy efficient," rather than the previous version of the instructions which gave energy classes "B" and "C" as examples of energy classes.

5.2 | Results and Discussion

Results of Experiment 3 are shown in Figure 4. Our analyses focused on the effects and interactions of our mixed factorial design: 2 (Knowledge of range: range shown, range not shown) × 2 (Color-coding: color, black and white) × 7 (Energy class: K, L, M, N, O, P, Q). The study focused on how these conditions influenced participants' ratings of environmental friendliness. The conditions were analyzed with a repeated measures ANOVA that was adjusted with a Greenhouse-Geisser correction. The full results of this ANOVA are shown in Table 4.

Hypothesis 3c predicted that range knowledge and color-coding would have a significant impact on ratings with a less informative letter scale (K to Q). The pattern of results shown in Figure 4 supports this prediction. The main effect of range ($p < 0.001$; Table 4) shows that participants were reluctant to rate the energy classes as high in environmental friendliness unless they were given range knowledge (compared blue and gold data in Figure 4). Range interacted with the energy classes ($p < 0.001$), showing that participants discriminated the energy classes far more with range knowledge than without range knowledge (Figure 4). The main effect of range and its interaction across energy classes supports Hypothesis 3a. Hypothesis 3c was also supported because the effects of range were far larger with an unintuitive letter scale (Figure 4) than with the traditional letter scales used in the first two experiments (Figure 2).

Importantly, color also significantly interacted with the energy classes ($p < 0.001$; Table 4), which reveals that the color-coding also impacted how participants organized their judgments of energy classes (Figure 4; supporting Hypothesis 3b). In Experiment 3, both range knowledge and color-coding helped participants to make sense of, and to tell apart, the energy classes. This shows that both range knowledge and color-coding can provide important contextual information. But we note that these context effects provided by range and color were strongest when the letters were less useful as a cue (cf. Figures 2 and 4). This supports Hypothesis 3c but also strengthens our conclusion that the letters were a dominant cue for judgments.

While both range knowledge and color-coding helped participants to tell apart the energy classes, the effect of range knowledge was greater than the effect of color-coding (Figure 4), as shown by a three-way interaction ($p = 0.001$; Table 4). This indicates that the explicit range knowledge was even more effective than color-coding for participants to discriminate energy classes, even though both provided impactful context.

A supplementary analysis investigated whether color-coding had any impact when range was shown (a comparison of the gold data across Figure 4A,B). This revealed that there was no interaction between color × energy class ($F < 1$, $p = 0.868$), indicating that color-coding can be superfluous when the range is known. This, taken together with the results of the full ANOVA (Table 4), shows that range knowledge was especially important when energy labels were printed in black and white, without color-coding. A supplementary analysis on demographics found no effects of age, and sex did not significantly alter the contextualizing effects of color or range knowledge.

Without either range knowledge or color-coding, and with an unintuitive letter scale, participants made relatively little discrimination of energy classes (see blue data in Figure 4B). This condition provided scarce contextual information, where participants had very few reference points when judging energy classes. Following our predictions, judgments depended on some contextualizing information about the to-be-rated stimulus scale. Experiment 3 demonstrates that range knowledge or color-coding can provide highly impactful contextual information in the absence of any other intuitive scale cue (i.e., letters that are matched with the beginning of the alphabet).

6 | General Discussion

The present study investigated the perception of three stimulus/design features in EU "Class arrow" energy labels, and how these contribute to judgments of environmental friendliness. These stimulus features were: (1) The letter scale, (2) the color-

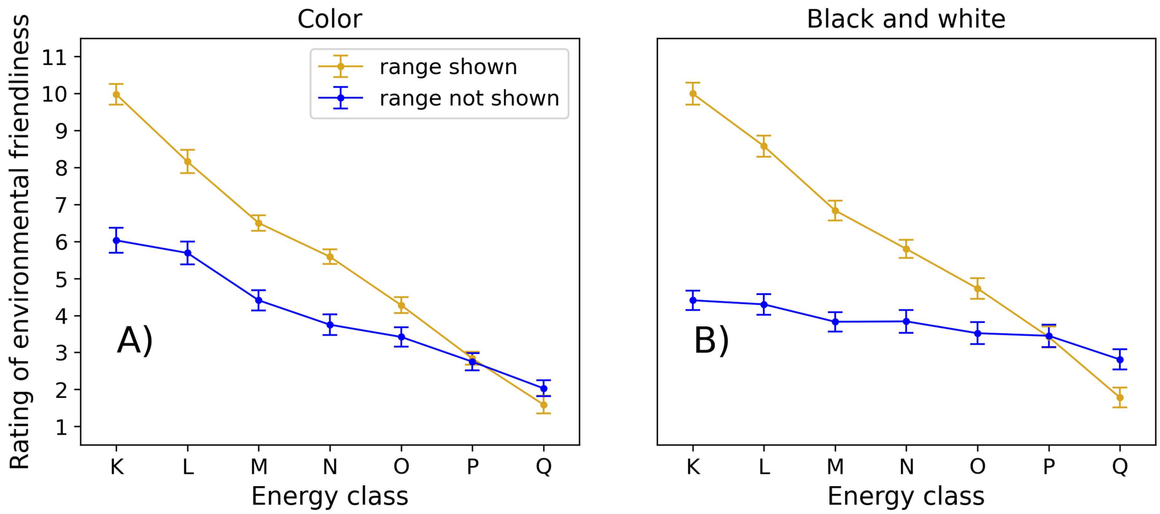

FIGURE 4 | Results of Experiment 3. Error bars represent the standard errors of the means.

TABLE 4 | Results of a repeated measures ANOVA for the full Experiment 3.

Factor	*df*	*F* value	*p* value	η^2_p
Energy class	3.66, 922.45	343.65	< 0.001***	0.577
Range	1, 252	101.10	< 0.001***	0.286
Color	1, 252	0.02	0.899	0.000
Energy class × range	3.66, 922.45	83.28	< 0.001***	0.248
Energy class × color	3.66, 922.45	6.00	< 0.001***	0.023
Range × color	1, 252	2.62	0.107	0.010
Energy class × range × color	3.66, 922.45	4.70	0.001**	0.018

Note: Energy class reflects the Energy class factor with seven levels. Range reflects the Knowledge of range factor with two levels. Color reflects the Color-coding factor with two levels.
**$p < 0.1$; ***$p < 0.001$.

coding, and (3) information about the scale's range. The results of three experiments inform several conclusions, which will be discussed in detail below. First, participants' judgments are strongly driven by which letters are present on a scale. This manifests in a high degree of similarity in the ratings of the "A+" classes of the old scale (confirming H1a), and in the generally higher ratings of the energy classes of the old scale (confirming H1b). Second, letters are given greater weight than color-coding in the formation of judgments (confirming H2a, disconfirming H2b). Third, color-coding and information about the scale's range provide contextual information which can help participants discriminate energy classes, especially when the set of letters is unintuitive (partially supporting H3a and H3b, and supporting H3c). Thus, having some contextual information is more important when other contextual cues are diminished. This again confirms that judgments were dominated by the letters. We examine the results of our three experiments and consider their implications for the design of energy labels.

6.1 | Theoretical Implications

In this section, we discuss the theoretical implications of our findings in relation to the cue-utilization framework (Cox 1962; Slovic 1966; Sullivan and Burger 1987), the heuristics and biases framework (Tversky and Kahneman 1974), and range-frequency theory (Parducci 1965). The discussion is divided into two sections, covering the three key design features of EU's ecodesign energy labels (categorical letters, color-coding, and range information).

6.1.1 | *Categorical Letters and Color-Coding*

The results of our three experiments, taken together, suggest that the categorical letters of EU's energy labels is a stimulus cue that is given much weight in the perception formation process. Importantly, color-coding does not receive the same weight. This can explain several of our findings, such as why the "A+" classes were poorly discriminated, and why products with energy classes from the old scale were perceived as generally higher in environmental friendliness than products with energy classes from the new scale. The smaller discriminations between the "A+" classes (Heinzle and Wüstenhagen 2012) and generally higher environmental friendliness ratings of the old scale (Stasiuk and Maison 2022) have been observed before. We suggest that these biases can be explained by the fact that lesser weight is given to the color-coding cue than to the categorical letter cue.

The relative unimportance of color-coding might appear surprising given past research showing effects of color in eco-labels. Color-based information can sometimes have stronger influence on consumer behavior than text-based information (Bengart and Vogt 2023; Holenweger et al. 2023), perhaps because color is a cue that attracts attention (Lurie and Mason 2007) and salient and fluently processed cues tend to gain more weight in the perception-decision process (Easterbrook 1959; Shah and Oppenheimer 2007). But our results show that the categorical letter cue is the more dominant feature in EU's ecodesign energy labels. The letters are superimposed on the colors and presented with strong visual contrast against the background (Figure 1). Furthermore, the letters can arguably be processed fluently and associatively without reflective thought, stimulating heuristic thinking such as "representativeness" (Tversky and Kahneman 1974) or "affect-based" responses (Slovic et al. 2007).

Basing judgments heuristically on letters' representativeness or affect (on a good-to-bad continuum), without much thought about other cues, is arguably a fast and frugal way of making a fairly accurate judgment of a product's environmental friendliness. For example, participants seem to have understood letters at the beginning of the alphabet as representative of a "high point," indicating a high degree of environmental friendliness. This can explain several of our findings, such as why participants made smaller discriminations of energy classes above class A (A+++, A++, and A+) than below class A (B, C, and D), as seen with ratings of the old scale (A+++ to D) in Experiment 1 and 2. Participants were relatively insensitive to the differences between the A+ classes because A was already considered environmentally friendly. The fact that the new (A to G) scale led to more appropriate discriminations of energy classes strengthen the interpretation that participants intuitively anchored their responses to the beginning of the alphabet (Experiment 1 and 2). It can also explain the results of Experiment 3, where the most energy efficient class was represented by the letter K. Here, participants assigned low ratings to all energy classes in the absence of range knowledge or color-coding. Participants were likely leaving room on the response scale for potential early-alphabet stimuli (like A, B, and C), which can explain the low ratings of letters K to Q. Note that contextual information guided a radical adjustment of response strategies for energy classes K to Q (see below for a discussion on the role of contextual information).

6.1.2 | *The Role of Range Knowledge*

Human judgments are generally context dependent (Haubensak 1992; Parducci 1965; Stewart et al. 2006), and environmental judgments are no exception (Skog and Sörqvist 2025; Tanner 2008). Energy labels are designed to provide people with different cues to the scale's context: (1) The letters can be matched with the beginning of the alphabet as a reference point, (2) the color-coding can be a familiar green-red "traffic light" spectrum, and (3) the range of the letter scale can be made explicit within the label (Figure 1). In the influential range-frequency theory (Parducci 1965, 1968), the range principle predicts that observers will judge a stimulus in relation to the endpoints of the stimulus range. In the case of energy labels, observers will judge a given energy class in relation to the highest and lowest available energy classes. This requires knowledge, as an observer must understand what the range is to make use of it in their judgments (cf. Skog and Sörqvist 2025). Because of this, we expected that judgments of environmental friendliness would be affected by different types of contextual information that reveal the scale's boundaries. We have already discussed how the choice of letters in the scales provided an important form of context that had a large impact on judgments (see above). Next, we focus on the effects of color-coding and information about the scale's range (also referred to as "range knowledge").

The results of Experiment 1 and 2, which used conventional letter scales (A to G, and A+++ to D), showed that the impact of range knowledge across energy classes was not significant when energy labels were color-coded (Experiment 1), but it was significant when energy labels were not color-coded (Experiment 2). This indicates that both color-coding and range knowledge can provide contextual information that helps participants discriminate energy classes. It also demonstrates that range knowledge is more important when energy labels are printed in black and white—that is, when other contextual cues are diminished. However, it must be noted that the absence of a cross-experiment three-way interaction for both scales (reported under Section 4.2.5) suggests that the effects of color-coding and range knowledge are weak or statistically unreliable for the conventional letter scales. The reason for this appears to be that the letters are a dominant stimulus cue that gains majority weight in the judgment formation process. When the letter cues are perceived as informative enough (such as when the range of energy classes align with the beginning of the alphabet), other cues are devalued.

Thus, our Experiment 3 abandoned the conventional letter scales in favor of an unintuitive scale with letters K to Q. Here, the contextual information provided by the letter scale was diminished since it is not anchored at an intuitive reference point (the beginning of the alphabet). When presented with the energy label "N," for example, without color-coding and without knowing the range of the scale, the observer's rating will be quite arbitrary and poorly organized. Unintuitive letters could therefore isolate the effects of color-coding and range knowledge. The results of Experiment 3 found that both color-coding and range knowledge were highly impactful for the participants' ability to discriminate energy classes. Both color-coding and range knowledge provided contextual information, which was more important than ever in Experiment 3, when the letter scale was unintuitive. Furthermore, range knowledge was a more effective cue than color-coding for the participants' ability to discriminate energy classes. All three experiments, taken together, provide evidence that range knowledge is a more important stimulus cue when energy labels are printed in black and white, where there is no available color-coding. Additionally, the results suggest that color-coding is the least important stimulus feature in energy labels, and it can be superfluous with range knowledge.

6.2 | Managerial and Applied Implications

From a managerial viewpoint, our results reveal how features of EU's energy labels shape consumers' product perception. The perceived environmental friendliness of products is a strong predictor of behavioral intentions, such as willingness to pay and product choice (e.g., Pichierri and Pino 2023; Sörqvist et al. 2015). Because of this, label design features that modulate perceived environmental friendliness shape these intentions. According to current EU legislation, energy labels printed in black and white may be used instead of their color-coded counterparts in visual advertisement. Our results suggest that this has marginal (if any) consequences for how consumers perceive products. Instead, perception is predominantly determined by the scales' categorical letters and range knowledge. Furthermore, the transition of replacing the old A+++ to D scale with the new A to G scale will have some predictable outcomes for advertising. Products will generally be perceived as less environmentally friendly overall (see also Stasiuk and Maison 2022). Our results suggest that this effect will be largest at the intermediate energy class (around "D" in the new A to G scale). And, we expect consumers to feel like there are larger differences in the perceived eco-value of products with the new scale, especially among the top-rated energy classes. In comparison, the "A+" classes are less distinct in the old scale—they are all seen as quite environmentally friendly.

Showing the range with letters and an arrow on the energy labels as an external cue has smaller yet important effects on perception, particularly when other external cues to scale range is absent (e.g., when energy labels are shown in black and white) and when consumers cannot use internal cues (such as alphabetical knowledge) to form an intuitive understanding of scale range. Providing reference points in advertisement will modulate how consumers contextualize a product and shape its perceived value, especially when other contextual cues are sparse. This conclusion goes beyond the use of ecodesign in advertisement. We argue that product evaluation will be strongly influenced by scale range knowledge with any type of eco-label, and more so when the eco-label's scale is ambiguous or unintuitive. Beyond eco-labels, we would expect range knowledge to be very impactful when quantitative information (numbers) is presented to consumers (e.g., annual kilowatt-hours of energy consumption, or carbon dioxide emissions; Skog and Sörqvist 2025).

From a policy point of view, the present study supports the recent (and ongoing) efforts made by the EU to rescale energy labels to A to G, while phasing out the A+++ to D scale. Our behavioral results show that "A+" classes are less discriminable, which in turn can have negative implications for how consumers form beliefs about environmental impact (see also Beck and Toulouse 2023; Faure et al. 2021; Heinzle and Wüstenhagen 2012).

Our results also add to the ongoing discussion of how labeling systems should be designed to reduce complexity, ambiguity and confusion (see e.g., Boe-Lillegraven and Demmers 2025; Donato and Adıgüzel 2024; Gosselt et al. 2019; Sejas-Portillo et al. 2025; Torma and Thøgersen 2024; Thøgersen et al. 2024). The EU issued a recent notice that information about the scale's range should be included in energy labels. The novel results of the present study bolster this directive. We have demonstrated that range knowledge helps participants to better understand the context of a given energy label. We note that range knowledge is even more important when color-coding is not available. The directive to include information about the scale's range also aligns with leading psychological theories on how judgments depend on contextual information. The EU's 2024 directive to include explicit information about the scale's range in energy labels is supported by range-frequency theory (Parducci 1965).

6.3 | Limitations and Future Studies

The study has several limitations that should be addressed. One concerns the generalizability of the results beyond EU. While energy labeling is used in many countries across the world (see e.g., Andika et al. 2025; Asif et al. 2023; Basiru et al. 2024; Davis and Metcalf 2016; Durmus Senyapar et al. 2024; Elmor et al. 2024; Newell and Siikamäki 2014; Wang et al. 2019), the design varies. Our conclusions that specifically concern effects with Latin letters might not generalize to other letter systems. Arguably, the results should be generalizable to any culture that uses letter systems with alphabetical ordering. We suggest that ordering can produce an intuitive understanding that the first entry represents a "high point," and that this might be a general mechanism. But whether the present effect of letters generalizes beyond the Latin alphabet requires evaluation in future studies.

Another limitation concerns the ecological validity of the study. While the stimulus material used in the present study resembles how white goods are advertised in online web shops, future research should test if our results generalize to more ecologically valid contexts. This can be done, for example, by conducting studies of simulated shopping or field experiments. In future studies, it is also important to directly measure behavioral intentions (e.g., willingness to pay) or purchase decisions in choice settings to solidify the results' relevance in marketing.

A further set of limitations concerns control variables and measures. Participants' familiarity with the energy labeling scales were not controlled for. Similarly, individual differences in environmental concern, energy literacy or education were not controlled. The random assignment of participants to between-subject conditions should in theory even out such individual differences and not compromise comparisons between conditions. Furthermore, many of the key effects (e.g., scale differences) found in Experiment 1 replicated in Experiment 2, with an independent sample of participants. Still, not controlling individual differences in environmental attitudes, familiarity, and literacy could have added unknown noise to the measurements. Another issue concerning measurement is the single-item measure of perceived environmental friendliness. Single-item measures can often be appropriate for measuring non-complex constructs (Allen et al. 2022) but it is often recommended to use muti-item measures to create an index measure of the psychological construct of interest. Perceived environmental friendliness is arguably a non-complex construct that aligns with a single good-to-bad continuum. A single-item measure might therefore be valid here. Indeed, single-item measures (Andersson et al. 2024; Gorissen and Weijters 2016) and multi-item measures (Kim and Schuldt 2018) of perceived environmental impact tend to yield similar results. Future studies should still consider using multi-item indices to increase construct validity.

Finally, the results of the present study inform future studies on label development. These studies can consider the relative weight consumers assign to categorical letters, color-coding, and explicit range cues, to design new labels that optimize consumers' ability to appropriately contextualize energy classes. This extends beyond energy labeling and can inform development of new labels in other fields such as eco-labeling (carbon footprint) and health labeling (calorific content and nutritional value).

6.4 | Concluding Remarks

The present study has demonstrated that consumers assign different weight to different contextual label cues: the most weight to categorical letters, intermediate weight to explicit range cues, and least weight to color-coding. However, the relative importance of these cues depends on the availability of the other cues. One key takeaway is that consumers need sufficient contextual information to understand energy labels. To ensure that consumers can reliably perceive the meaning of energy labels, we favor a multicue approach where the eco-design of energy labels should contain: (1) A letter scale that begins with the alphabet, (2) a color-coded scale, and (3) information about the scale's range. This provides multiple contextual cues, which can lead to a better understanding of energy labels that will in turn maximize the ability to tell energy classes apart. These insights are important in any situation where eco-labeling is used in advertisement, and to future design and development of eco-labels. Consumers often consider the environmental impact of products when making purchasing decisions. A better understanding of energy labels can lead to more informed decisions which can support pro-environmental consumer behavior.

Acknowledgments

Sörqvist was supported by Riksbankens Jubileumsfond (P23-0067) and Skog was supported by Kempe foundations (JCSMK23-0179).

Ethics Statement

The study was conducted in accordance with the Declaration of Helsinki and conformed to the guidelines of the Swedish Research Ethics Authority (Dnr 2024-05795-01).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data for all three experiments are publicly available via the Open Science Framework (https://doi.org/10.17605/OSF.IO/QXBER).

References

Akroush, M. N., M. I. Zuriekat, H. I. Al Jabali, and N. A. Asfour. 2019. "Determinants of Purchasing Intentions of Energy-Efficient Products: The Roles of Energy Awareness and Perceived Benefits." *International Journal of Energy Sector Management* 13, no. 1: 128–148. https://doi.org/10.1108/IJESM-05-2018-0009.

Allen, M. S., D. Iliescu, and S. Greiff. 2022. "Single Item Measures in Psychological Science: A Call to Action." *European Journal of Psychological Assessment* 38: 1–5. https://doi.org/10.1027/1015-5759/a000699.

Andersson, H., M. Holmgren, P. Sörqvist, et al. 2024. "The Negative Footprint Illusion Is Exacerbated by the Numerosity of Environment-Friendly Additions: Unveiling the Underpinning Mechanisms." *Journal of Cognitive Psychology* 36: 295–307. https://doi.org/10.1080/20445911.2024.2313568.

Andika, A., M. Maslikhah, N. K. T. Tampubolon, D. Wahyudi, T. Dwiwinarno, and M. A. K. Amirullah. 2025. "Exploring Consumer Transition to Energy-Efficient Household Appliances in Indonesia: The Mediating Role of Curiosity in a Push-Pull-Mooring Framework." *International Journal of Energy Economics and Policy* 15, no. 2: 9–22. https://doi.org/10.32479/ijeep.17744.

Andor, M. A., A. Gerster, and S. Sommer. 2020. "Consumer Inattention, Heuristic Thinking and the Role of Energy Labels." *Energy Journal* 41, no. 1: 83–112. https://doi.org/10.5547/01956574.41.1.mand.

Asif, M. H., T. Zhongfu, M. Irfan, B. Ahmad, and M. Ali. 2023. "Assessing Eco-Label Knowledge and Sustainable Consumption Behavior in Energy Sector of Pakistan: An Environmental Sustainability Paradigm." *Environmental Science and Pollution Research* 30, no. 14: 41319–41332. https://doi.org/10.1007/s11356-023-25262-8.

Basiru, I., Y. Xu, V. E. Arkorful, B. K. Lugu, and A. H. Ibrahim. 2024. "Energy Efficiency Labels and Urban Residents' Intention to Purchase Energy-Efficient Household Appliances: An Empirical Study Based on the Theory of Planned Behavior." *Clean Technologies and Environmental Policy* 27: 1–16. https://doi.org/10.1007/s10098-024-03070-z.

Baca, G., and N. Reshidi. 2025. "Green Branding and Consumer Behavior, Unveiling the Impact of Environmental Marketing Strategies on Purchase Decisions." *Business Strategy and the Environment* 34: 3701–3713. https://doi.org/10.1002/bse.4172.

Beck, M., and N. Ö. Toulouse. 2023. "Assessing the Impact of Energy Labels on Attitude and Behavioral Intention: An Empirical Investigation." *Journal of Cleaner Production* 415: 137751. https://doi.org/10.1016/j.jclepro.2023.137751.

Bengart, P., and B. Vogt. 2023. "Effects and Interactions of Labels' Color Scheme and the Individual Difference Variable Lay Rationalism on Pro-Environmental Choices." *Journal of Environmental Psychology* 87: 101998. https://doi.org/10.1016/j.jenvp.2023.101998.

Birnbaum, M. H., and J. Rouvere. 2023. "Contextual Effects in Salary Satisfaction." *Judgment and Decision Making* 18: e31. https://doi.org/10.1017/jdm.2023.26.

Boe-Lillegraven, S., and J. Demmers. 2025. "Leveling Up on Labels? Consumer Preferences for Firm-Level Eco-Labels as Substitutes for or Complements to Product-Level Eco-Labels." *Corporate Social Responsibility and Environmental Management* 32: 1920–1944. https://doi.org/10.1002/csr.3051.

Cox, D. F. 1962. "The Measurement of Information Value: A Study in Consumer Decision-Making." In *Emerging Concepts in Marketing*, edited by W. S. Decker, 413–421. American Marketing Association.

Davis, L. W., and G. E. Metcalf. 2016. "Does Better Information Lead to Better Choices? Evidence From Energy-Efficiency Labels." *Journal of the Association of Environmental and Resource Economists* 3, no. 3: 589–625. https://doi.org/10.1086/686252.

Donato, C., and F. Adıgüzel. 2024. "The Effects of Visual Design on Eco-Labels Evaluations: Guidelines for Effective Green Advertising." *Journal of Marketing Theory and Practice* 33: 271–288. https://doi.org/10.1080/10696679.2024.2307889.

Durmus Senyapar, H. N., B. Duzgun, and F. E. Boran. 2024. "Energy Labels and Consumer Attitudes: A Study Among University Staff." *Sustainability* 16, no. 5: 1754. https://doi.org/10.3390/su16051754.

Easterbrook, J. A. 1959. "The Effect of Emotion on Cue Utilization and the Organization of Behavior." *Psychological Review* 66: 183–201. https://doi.org/10.1037/h0047707.

Elmor, L., G. A. Ramos, Y. Vieites, B. Andretti, and E. B. Andrade. 2024. "Environmental Sustainability Considerations (or Lack Thereof) in Consumer Decision Making." *International Journal of Research in Marketing* 42: 1203–1228. https://doi.org/10.1016/j.ijresmar.2024.08.003.

Epley, N., and T. Gilovich. 2006. "The Anchoring-and-Adjustment Heuristic: Why the Adjustments Are Insufficient." *Psychological Science* 17: 311–318. https://doi.org/10.1111/j.1467-9280.2006.01704.x.

Faul, F., E. Erdfelder, A.-G. Lang, and A. Buchner. 2007. "G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences." *Behavior Research Methods* 39: 175–191. https://doi.org/10.3758/bf03193146.

Faure, C., M. C. Guetlein, and J. Schleich. 2021. "Effects of Rescaling the EU Energy Label on Household Preferences for Top-Rated Appliances." *Energy Policy* 156: 112439. https://doi.org/10.1016/j.enpol.2021.112439.

Gigerenzer, G., and D. G. Goldstein. 1996. "Reasoning the Fast and Frugal Way: Models of Bounded Rationality." *Psychological Review* 103: 650–669. https://doi.org/10.1037/0033-295X.103.4.650.

Gorissen, K., and B. Weijters. 2016. "The Negative Footprint Illusion: Perceptual Bias in Sustainable Food Consumption." *Journal of Environmental Psychology* 45: 50–65. https://doi.org/10.1016/j.jenvp.2015.11.009.

Gosselt, J. F., T. van Rompay, and L. Haske. 2019. "Won't Get Fooled Again: The Effects of Internal and External CSR Eco-Labeling." *Journal of Business Ethics* 155: 413–424. https://doi.org/10.1007/s10551-017-3512-8.

Grankvist, G., U. Dahlstrand, and A. Biel. 2004. "The Impact of Environmental Labelling on Consumer Preference: Negative vs. Positive Labels." *Journal of Consumer Policy* 27: 213–230. https://doi.org/10.1023/B:COPO.0000028167.54739.94.

Haubensak, G. 1992. "The Consistency Model: A Process Model for Absolute Judgments." *Journal of Experimental Psychology: Human Perception and Performance* 18, no. 1: 303–309. https://doi.org/10.1037/0096-1523.18.1.303.

Heinzle, S. L., and R. Wüstenhagen. 2012. "Dynamic Adjustment of Eco-Labeling Schemes and Consumer Choice—The Revision of the EU Energy Label as a Missed Opportunity?" *Business Strategy and the Environment* 21, no. 1: 60–70. https://doi.org/10.1002/bse.722.

Holenweger, G., S. Stöckli, and A. Brügger. 2023. "Carbon Footprint Labels Involving Traffic Lights Foster Sustainable Food Choices." *Food Quality and Preference* 106: 104813. https://doi.org/10.1016/j.foodqual.2023.104813.

Kim, B., and J. P. Schuldt. 2018. "Judging the Environmental Impact of Green Consumption: Evidence of Quantity Insensitivity." *Journal of Environmental Psychology* 60: 122–127. https://doi.org/10.1016/j.jenvp.2018.10.005.

Lurie, N. H., and C. H. Mason. 2007. "Visual Representation: Implications for Decision Making." *Journal of Marketing* 71: 160–177. https://doi.org/10.1509/jmkg.71.1.160.

Majer, J. M., H. A. Henscher, P. Reuber, D. Fischer-Kreer, and D. Fischer. 2022. "The Effects of Visual Sustainability Labels on Consumer Perception and Behavior: A Systematic Review of the Empirical Literature." *Sustainable Production and Consumption* 33: 1–14. https://doi.org/10.1016/j.spc.2022.06.012.

Mei, D., Z. Zhu, J. Li, H. Wu, J. Li, and Q. Xiao. 2025. "Comparing the Effectiveness of Green Attributes in Food Consumption: A Cue Utilization Theory Perspective." *British Food Journal*: 1–16. https://doi.org/10.1108/BFJ-12-2024-1251.

Meyers-Levy, J., and D. Maheswaran. 1991. "Exploring Differences in Males' and Females' Processing Strategies." *Journal of Consumer Research* 18, no. 1: 63–70. https://doi.org/10.1086/209241.

Meyers-Levy, J., and B. Sternthal. 1991. "Gender Differences in the Use of Message Cues and Judgments." *Journal of Marketing Research* 28, no. 1: 84–96. https://doi.org/10.1177/002224379102800107.

Newell, R. G., and J. Siikamäki. 2014. "Nudging Energy Efficiency Behavior: The Role of Information Labels." *Journal of the Association of Environmental and Resource Economists* 1, no. 4: 555–598. https://doi.org/10.1086/679281.

Pancer, E., L. McShane, and T. J. Noseworthy. 2017. "Isolated Environmental Cues and Product Efficacy Penalties: The Color Green and Eco-Labels." *Journal of Business Ethics* 143: 159–177.

Parducci, A. 1965. "Category Judgments: A Range-Frequency Model." *Psychological Review* 72: 407–418. https://doi.org/10.1037/h0022602.

Parducci, A. 1968. "The Relativism of Absolute Judgments." *Scientific American* 219: 84–90. https://doi.org/10.1038/scientificamerican1268-84.

Peirce, J., J. R. Gray, S. Simpson, et al. 2019. "PsychoPy2: Experiments in Behavior Made Easy." *Behavior Research Methods* 51: 195–203. https://doi.org/10.3758/s13428-018-01193-y.

Pichierri, M., and G. Pino. 2023. "Less Saturated, More Eco-Friendly: Color Saturation and Consumer Perception of Product Sustainability." *Psychology & Marketing* 40: 1830–1849.

Sammer, K., and R. Wüstenhagen. 2006. "The Influence of Eco-Labelling on Consumer Behaviour–Results of a Discrete Choice Analysis for Washing Machines." *Business Strategy and the Environment* 15, no. 3: 185–199. https://doi.org/10.1002/bse.522.

Samson, A., and B. G. Voyer. 2012. "Two Minds, Three Ways: Dual System and Dual Process Models in Consumer Psychology." *AMS Review* 2: 48–71. https://doi.org/10.1007/s13162-012-0030-9.

Schuldt, J. P. 2013. "Does Green Mean Healthy? Nutrition Label Color Affects Perceptions of Healthfulness." *Health Communication* 28: 814–821. https://doi.org/10.1080/10410236.2012.725270.

Sejas-Portillo, R., M. Moro, and T. Stowasser. 2025. "The Simpler the Better? Threshold Effects of Energy Labels on Property Prices and Energy Efficiency Investments." *American Economic Journal: Applied Economics* 17: 41–89. https://doi.org/10.1257/app.20220329.

Seo, J. Y., and D. L. Scammon. 2017. "Do Green Packages Lead to Misperceptions? The Influence of Package Colors on Consumers' Perceptions of Brands With Environmental Claims." *Marketing Letters* 28: 357–369. https://doi.org/10.1007/s11002-017-9420-y.

Shah, A. K., and D. M. Oppenheimer. 2007. "Easy Does It: The Role of Fluency in Cue Weighting." *Judgment and Decision Making* 2: 371–379. https://doi.org/10.1017/S1930297500000516.

Shankar, M. U., C. A. Levitan, J. Prescott, and C. Spence. 2009. "The Influence of Color and Label Information on Flavor Perception." *Chemosensory Perception* 2: 53–58. https://doi.org/10.1007/s12078-009-9046-4.

Silva, A. R. A., A. S. Bioto, P. Efraim, and G. C. Queiroz. 2017. "Impact of Sustainability Labeling in the Perception of Sensory Quality and Purchase Intention of Chocolate Consumers." *Journal of Cleaner Production* 141: 11–21. https://doi.org/10.1016/j.jclepro.2016.09.024.

Skog, E., I. Sandlund, and P. Sörqvist. 2025. "Cognitive Bias in Perceived Concern With Rainfall: Implications for Climate Adaptation." *Journal of Environmental Psychology* 106: 102737. https://doi.org/10.1016/j.jenvp.2025.102737.

Skog, E., and P. Sörqvist. 2025. "Context Foreknowledge Can Make Emissions Seem More Environmentally Friendly or Harmful: Evidence From Distribution-Density Effects in Human Judgment." *Journal of Applied Research in Memory and Cognition* 14, no. 2: 288–299. https://doi.org/10.1037/mac0000213.

Skourtos, M., D. Damigos, C. Tourkolias, and A. Kontogianni. 2021. "Efficient Energy Labelling: The Impact of Information Content and Style on Product Choice." *Energy Efficiency* 14, no. 6: 58. https://doi.org/10.1007/s12053-021-09950-3.

Sloman, S. A. 2002. "Two Systems of Reasoning." In *Heuristics and Biases: The Psychology of Intuitive Judgment*, edited by T. Gilovich, D. Griffin, and D. Kahneman, 379–396. Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.024.

Slovic, P. 1966. "Cue-Consistency and Cue-Utilization in Judgment." *American Journal of Psychology* 79: 427–434. https://doi.org/10.2307/1420883.

Slovic, P., M. L. Finucane, E. Peters, and D. G. MacGregor. 2007. "The Affect Heuristic." *European Journal of Operational Research* 177: 1333–1352. https://doi.org/10.1016/j.ejor.2005.04.006.

Sörqvist, P., A. Haga, L. Langeborg, et al. 2015. "The Green Halo: Mechanisms and Limits of the Eco-Label Effect." *Food Quality and Preference* 43: 1–9. https://doi.org/10.1016/j.foodqual.2015.02.001.

Sörqvist, P., D. Hedblom, M. Holmgren, et al. 2013. "Who Needs Cream and Sugar When There Is Eco-Labeling? Taste and Willingness to Pay for "Eco-Friendly" Coffee." *PLoS One* 8, no. 12: e80719. https://doi.org/10.1371/journal.pone.0080719.

Stadelmann, M., and R. Schubert. 2018. "How Do Different Designs of Energy Labels Influence Purchases of Household Appliances? A Field Study in Switzerland." *Ecological Economics* 144: 112–123. https://doi.org/10.1016/j.ecolecon.2017.07.031.

Stasiuk, K., and D. Maison. 2022. "The Influence of New and Old Energy Labels on Consumer Judgements and Decisions about Household Appliances." *Energies* 15, no. 4: 1260. https://doi.org/10.3390/en15041260.

Stewart, N., N. Chater, and G. D. A. Brown. 2006. "Decision by Sampling." *Cognitive Psychology* 53, no. 1: 1–26. https://doi.org/10.1016/j.cogpsych.2005.10.003.

Sucapane, D., C. Roux, and K. Sobol. 2021. "Exploring How Product Descriptors and Packaging Colors Impact Consumers' Perceptions of Plant-Based Meat Alternative Products." *Appetite* 167: 105590. https://doi.org/10.1016/j.appet.2021.105590.

Sullivan, G. L., and K. J. Burger. 1987. "An Investigation of the Determinants of Cue Utilization." *Psychology & Marketing* 4: 63–74. https://doi.org/10.1002/mar.4220040107.

Tanner, C. 2008. "Context Effects in Environmental Judgments: Assimilation and Contrast Effects in Separate and Joint Evaluation Modes." *Journal of Applied Social Psychology* 38, no. 11: 2759–2786. https://doi.org/10.1111/j.1559-1816.2008.00413.x.

Thøgersen, J. 2002. "Promoting Green Consumer Behavior With Eco-Labels." In *New Tools for Environmental Protection: Education, Information, and Voluntary Measures*, edited by T. Dietz and P. C. Stern, 83–104. National Academy Press.

Thøgersen, J., F. J. Dessart, G. Marandola, and S. L. Hille. 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, no. 7: 6795–6813. https://doi.org/10.1002/bse.3838.

Thøgersen, J., and K. S. Nielsen. 2016. "A Better Carbon Footprint Label." *Journal of Cleaner Production* 125: 86–94. https://doi.org/10.1016/j.jclepro.2016.03.098.

Torma, G., and J. Thøgersen. 2024. "Can a Meta Sustainability Label Facilitate More Sustainable Consumer Choices?" *Business Strategy and the Environment* 33: 283–306. https://doi.org/10.1002/bse.3488.

Tversky, A., and D. Kahneman. 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185: 1124–1131. https://doi.org/10.1126/science.185.4157.1124.

Wang, Z., Q. Sun, B. Wang, and B. Zhang. 2019. "Purchasing Intentions of Chinese Consumers on Energy-Efficient Appliances: Is the Energy Efficiency Label Effective?" *Journal of Cleaner Production* 238: 117896. https://doi.org/10.1016/j.jclepro.2019.117896.

Wason, P. C., and J. S. B. T. Evans. 1974. "Dual Processes in Reasoning?" *Cognition* 3: 141–154. https://doi.org/10.1016/0010-0277(74)90017-1.

Zhong, K., Z. Yan, H. Sun, and R. Huang. 2025. "Overcoming the Challenges of "Green Enough": Exploring the Dark Side of Incremental Green Options." *Psychology & Marketing* 42: 892–906. https://doi.org/10.1002/mar.22155.

Appendix A

Oven Images Used as Stimuli

FIGURE A1 | The seven oven images used as stimuli in all three experiments. Ovens were randomly paired with energy labels.