Transparency of Carbon-Neutral Labels: Evidence from a Choice Experiment*

Begüm Özdemir Oluk †‡

begum.ozdemiroluk@eawag.ch

begumozdemiroluk.github.io

June 5, 2025

*I express my gratitude to Stefano Carattini, Ivana Logar, and Sabrina Eisenbarth for their valuable guidance. I thank Fabian Dvorak, the participants of the Choice Experiment Clinics at the IVM Institute, VU Amsterdam, the 9th Workshop on Experimental Economics for the Environment, the Eawag Symposium 2024, the 1st ETH Domain Conference on Social Sciences on Environment, Technology and Sustainability, the University of St. Gallen Brown Bag Seminar Series, and the Eawag Environmental Social Sciences Colloquium for their helpful comments and suggestions. I acknowledge the support from Eawag: Swiss Federal Institute of Aquatic Science and Technology.

†SIAW-HSG and Department of Economics, University of St. Gallen, Switzerland

‡Eawag: Swiss Federal Institute of Aquatic Science and Technology Department of Environmental Social Sciences, Dübendorf, Switzerland

Abstract

This paper examines the effect of transparency in carbon-neutral labeling on consumer willingness to pay. Carbon-neutral labels indicate that a product's greenhouse gas emissions have been offset (compensated) outside the company and/or directly reduced within it. Although CO_2 offsets are generally viewed as less effective than CO_2 reductions, most labels on the market lack transparency regarding the proportion of CO_2 offset and CO_2 reduction. This study empirically investigates whether consumers are willing to pay for transparency on carbon-neutral labels and explores whether there are differences in consumers' valuation of CO_2 reductions versus CO_2 offsets. Using a discrete choice experiment survey among UK tea consumers, I compare willingness to pay for standard versus transparent carbon-neutral labels. The control group saw a standard carbon-neutral label, while the treatment groups saw transparent labels indicating different shares of CO_2 offsetting and CO_2 reduction. I find no evidence of consumers' willingness to pay for transparency on carbon-neutral labels or preference for CO_2 reductions over CO_2 offsets.

Keywords discrete choice experiment; stated preferences; environmental valuation; climate labeling; carbon neutrality; information asymmetry; environmental transparency

JEL codes C83, C90, D12, D80, D91, Q51, Q54

1 Introduction

As the negative effects of climate change become stronger, an increasing number of companies are making their products carbon-neutral. Carbon neutrality can be achieved by quantifying greenhouse gas (GHG) emissions of a product, service, or a company in terms of carbon dioxide equivalents (CO_2e), reducing these emissions within the company itself and/or by offsetting emissions externally. CO_2 offsetting involves compensating for GHG emissions by funding projects external to the company, such as reforestation. These offsets are traded in carbon markets, allowing companies to earn credits used to balance out their products' GHG emissions. In contrast, CO_2 reduction focuses on directly cutting down GHG emissions within the company through actions such as improving energy efficiency.

Carbon-neutral labels may reflect CO_2 offsets, CO_2 reductions, or a combination of both, but these activities are not equivalent, and how carbon neutrality is achieved matters. Both the IPCC (2022) and the Oxford Principles for Net Zero Aligned Carbon Offsetting (Axelsson et al., 2024) emphasize that organizations should prioritize direct CO_2 reductions over offsets. Recent debates in the literature and the media have increasingly questioned the effectiveness of CO_2 offsets (The Guardian, 2023b). These debates involve practical concerns that CO_2 offsets are less effective than direct reductions in achieving real atmospheric CO_2 decreases. Key issues include the credibility of offset projects, their additionality, and risks of double counting (Schneider et al., 2015; Becken and Mackey, 2017; Calel et al., 2021; Trencher et al., 2024).

Carbon-neutral labels generally lack information about the shares of CO_2 offset and CO_2 reduction behind the label. Recent developments highlight concerns about the quality of CO_2 offsets, including the European Parliament's proposed ban on unverified generic environmental claims such as "climate-neutral" (European Parlia-

ment, 2023), a lawsuit against Delta Air Lines over misleading carbon neutrality claims (The Guardian, 2023a), and the decision by two major carbon-neutral label certifiers, Carbon Trust and ClimatePartner, to withdraw their carbon-neutral labels (Carbon Trust, 2023; ClimatePartner, 2023). Given these developments, an important empirical question arises: Do consumers value transparency in carbon-neutral labels?

This pre-registered study[1] leverages a discrete choice experiment (DCE), a stated preference method, for eliciting consumers' preferences and willingness to pay (WTP) for carbon-neutral labels. An online survey is conducted among 1,339 tea consumers in the UK. The experimental design consists of one control group and two treatment arms. The participants were provided with a short, neutral description of the CO_2 offsetting and reduction activities that the label entails (with or without information on the exact share of each activity, depending on the experimental group). The control group saw the standard carbon-neutral label without any information about CO_2 offsetting and CO_2 reduction shares on the label. The treatment groups saw a transparent version of the carbon-neutral label, which includes information on the shares of CO_2 offsetting and CO_2 reduction. Specifically, the first treatment group saw a carbon-neutral label indicating a combination of 95% CO_2 offsetting and 5% CO_2 reduction, and the second treatment group saw a carbon-neutral label with an equal division, with 50% CO_2 offsetting and 50% CO_2 reduction.

The key strength of using a stated preference survey in this context is its ability to generate clean counterfactual comparisons between transparent and non-transparent carbon-neutral labels across experimental groups. DCE offers distinct advantages, such as capturing consumer trade-offs among various attributes and eliciting marginal

[1]Ozdemir Oluk, Begum. 2024. "Consumer Preferences for Transparent Carbon-Neutral Labels: A Choice Experiment." AEA RCT Registry. August 02. https://doi.org/10.1257/rct.12520-2.0

WTP (MWTP) (Hanley et al., 1998), while being less susceptible to biases like "yea-saying" compared to contingent valuation method (CVM) (Adamowicz et al., 1994). While stated preference methods may overstate WTP, such inflation, whether additive or proportional, is assumed to apply uniformly across groups. In both cases, the direction and relative magnitude of treatment effects remain informative. My design relies on this assumption, prioritizing internally valid comparisons across experimental arms rather than precise estimation of absolute WTP levels.

Due to pressures from investors and corporate boards, more and more companies make their products carbon-neutral (Kim and Lyon, 2011; Rogelj et al., 2021). While many studies find that consumers are willing to pay a premium for carbon-neutral labels (Gassler et al., 2015; Vecchio and Annunziata, 2015; Drichoutis et al., 2016; Birkenberg et al., 2021; Bek, 2022), a hedonic analysis by Carattini et al. (2024) suggests no significant difference from zero. In either case, examining how transparency of carbon-neutral labels affects consumers' demand remains essential, especially in light of recent developments about growing concerns about CO_2 offsets and increasing regulatory demands.

The main research questions are: (i) whether consumers value transparent carbon-neutral labels more than standard (non-transparent) ones, and (ii) whether they prefer CO_2 reductions over CO_2 offsets. Transparency of carbon-neutral labels, in terms of the shares of CO_2 offsetting and reduction, can influence consumer demand in multiple ways. On the one hand, it may increase consumers' trust in labels, as consumers are better able to infer sellers' incentives for offering carbon-neutral products (Darby and Karni, 1973). On the other hand, if the share of CO_2 offsetting is greater than CO_2 reduction, transparency may lower demand due to practical and ethical concerns about CO_2 offsetting (Carattini and Tavoni, 2016). Furthermore, because the market lacks standardized labeling, consumers may be unfamiliar with the additional infor-

mation provided. Consequently, confusion can arise, favoring lower-quality labels due to adverse selection (Akerlof, 1970; Brécard, 2017).

The main findings suggest that UK consumers are willing to pay a premium for carbon-neutral labels on tea products. However, I find no statistically significant difference in WTP between transparent and standard (non-transparent) carbon-neutral labels. On average, consumers are willing to pay more for the label with a higher share of CO_2 offsetting compared to the label indicating an equal share of CO_2 offsetting and reduction; however, this difference is not statistically significant. Exploratory analyses also examined the role of competing labels. The presence of organic and ethical trade labels reduced the estimated WTP for the carbon-neutral label, while the differences among experimental groups remained statistically significant.

Descriptive analyses explored potential mechanisms underlying these findings. Around half of the participants did not correctly understand the meaning of CO_2 offsetting, despite the explanations provided. Post-DCE questions revealed that, while the transparent label with an equal share of CO_2 offsetting and reduction was the most trusted, the label with a larger share of offsetting was the most confusing. This suggests that extra information may lower WTP when offsetting and reduction are equally important in achieving carbon neutrality. Interestingly, participants with greater concerns about CO_2 offsetting were associated with higher WTP for the label with a larger share of offsetting, indicating possible consumer confusion between the terms 'CO_2 offsetting' and 'CO_2 emissions'. Unlike in the two other experimental groups, for the group shown the label with a higher share of offsetting, trust in transparent labels was not associated with WTP. This suggests that transparency may not lead to increased consumer trust when the underlying concepts are misunderstood or when there are concerns about CO_2 offsetting.

Based on the findings of this paper, without reliable information indicating

whether offsets or reductions are more effective, consumers do not naturally pay more for transparency on carbon-neutral labels. However, regulating labels remains important[2], and aligns with the intentions of policymakers such as the EU, which plans to ban unverified generic environmental claims like "climate-neutral" (European Parliament, 2023). Therefore, the policy implications of this study suggest (i) the need for reliable third-party verification regarding companies' transparency, given the limited consumer capacity to understand detailed concepts such as CO_2 offsets and CO_2 reductions. Alternatively, (ii) exploring ways to simplify the information, while still communicating the different implications of CO_2 offsetting and reduction activities, may help improve consumer understanding and decision-making.

This study contributes to the following lines of the existing literature. Firstly, it adds to the literature focusing on consumers' valuation of climate labels (Drichoutis et al., 2016; Akaichi et al., 2017; Feucht and Zander, 2018; Grebitus et al., 2013; Onozaka and McFadden, 2011; Bek, 2022). While many stated preference studies and a few revealed preference studies have focused on consumers' WTP, to my knowledge, none of them focused on the effect of transparency in carbon-neutral labels. Secondly, this paper contributes to the extensive literature on the economics of CO_2 offsetting (Blasch and Farsi, 2014; Ziegler et al., 2012; Brouwer et al., 2008; MacKerron et al., 2009; Carattini and Tavoni, 2016; Chen et al., 2018). Similar to this study, a few others have examined how consumers value CO_2 offsets compared to CO_2 reductions (Bek, 2022; Roemer et al., 2023). However, these studies treated CO_2 offsets and reductions as different levels of the same attribute or as different attributes in choice experiments, meaning that each participant saw all versions of the labels. In contrast,

[2]Information asymmetry can distort market mechanisms, leading to adverse selection and the selection of lower-quality goods (Akerlof, 1970). Furthermore, a body of theoretical literature suggests that, under certain assumptions and in the presence of information asymmetry, competition can favor labels with lower environmental quality (Brécard, 2014; Heyes and Martin, 2017; Brécard, 2017; Heyes and Martin, 2018; Poret, 2019).

this study employs a between-subject design: participants were randomly assigned to one of three experimental groups. This type of design minimizes contrast effects, where respondents might otherwise compare transparent and non-transparent labels side by side make decisions based on relative salience rather than genuine valuation. It also reduces possible demand effects, such as respondents guessing the study's aim and making decisions accordingly. By ensuring that each participant only sees one label type, the design allows for a cleaner identification of the transparency effect, isolating how transparency affects WTP, without contamination from within-individual comparisons.

Furthermore, this study connects to several other lines of literature. First, it adds to the literature on how non-normative pro-social behaviors become widely accepted (Sparkman and Walton, 2017; Carattini et al., 2022; Mortensen et al., 2019; Kraft-Todd et al., 2018) by investigating whether increased transparency in environmental labeling can enhance the adoption of green products like carbon-neutral labels, which currently remain a niche market. Second, in a setting where the sellers have more information than the buyers, exploring whether consumers value transparent labels more compared to standard labels also adds to the literature on information asymmetry (Akerlof, 1970; Brounen and Kok, 2011; Brunnschweiler et al., 2021).

The remainder of the paper is outlined as follows: Section 2 provides a background and a review of the relevant literature. Section 3 details the methodology, including the survey and DCE design (3.1), data (3.2), and the empirical approach (3.3). Section 4 presents the results, which discusses the effect of transparency on WTP for carbon neutral labels (4.1), the underlying mechanisms of consumers' WTP (4.2), and competing labels (4.3). Section 5 concludes.

2 Background

This section introduces carbon-neutral labels and why they are interesting to study. Next, it examines factors that could derive the demand for carbon-neutral labels. Then, it highlights the knowledge gap regarding the demand for transparency on the label and discusses the study's implications for corporate social responsibility and policy-making.

Many companies have announced their commitment to reach carbon neutrality by 2050 or earlier, and global voluntary carbon markets are expected to grow 15-fold by 2030 (potentially reaching $50 billion) and expand by up to 100-fold by 2050 (McKinsey & Company, 2023). However, currently, most products have not yet been labeled carbon-neutral, and opting for carbon neutrality is still a niche behavior. Therefore, understanding the consumers' demand for carbon-neutral labels can help businesses and policymakers plan effective climate change mitigation.

A carbon-neutral certified product indicates that its lifecycle CO_2 emissions have been reduced, and the remaining emissions have been compensated through CO_2[3] offsetting (ClimatePartner, 2023). This means that a product can achieve carbon neutrality through 100% CO_2 offsetting, 100% CO_2 reduction[4], or a combination of both. CO_2 offsetting and reduction are different approaches. CO_2 offsetting projects involve initiatives outside of the production chain, such as reforestation (Climate Portal, 2023). Companies can buy and sell these offsets in carbon markets, allowing them to earn credits that balance (compensate) their products' CO_2 emissions. In contrast, CO_2 reduction focuses on directly lowering a company's emissions through

[3]In this paper, "CO_2" and "CO_2" are used interchangeably. The term CO_2 refers to the carbon dioxide equivalent (CO_2e), measuring the total CO_2 emissions of a product, expressed as an equivalent amount of CO_2.

[4]Although technically possible, achieving carbon neutrality through 100% CO_2 reduction alone is difficult due to residual emissions that are hard to eliminate entirely.

actions such as improving energy efficiency.

Given the link between carbon neutrality and CO_2 offsetting, the economics of CO_2 offsetting offers insights into the demand for carbon-neutral labels. The guilt associated with harming the environment (Kotchen, 2009) and the warm glow or satisfaction obtained from pro-environmental behavior (Andreoni, 1990; Kahneman and Knetsch, 1992) are factors that can positively drive demand.

Media skepticism and concerns about corporate greenwashing may negatively influence consumers' demand for carbon-neutral labels, which involve CO_2 offsets. There are growing reservations about CO_2 offsets, as opposed to direct CO_2 reductions. The practical concerns include the argument that CO_2 offsetting does not result in equivalent CO_2 reductions in the atmosphere (Becken and Mackey, 2017) due to issues of additionality (Hyams and Fawcett, 2013; Schneider and Kollmuss, 2015), credibility (Bumpus and Liverman, 2008; Hooper et al., 2008), and double-counting of emissions reductions (Schneider et al., 2015). Ethical concerns associated with CO_2 offsetting include issues such as moral licensing (Dorner, 2019) and putting a price on nature (Aldred, 2012).

While concerns about CO_2 offsets grow, carbon-neutral labels on products do not clearly indicate the proportion of CO_2 offsetting and CO_2 reduction on the label [5]. According to the European Commission, 53.3% of environmental claims were found to be vague, misleading, or unfounded, and based on the report by Changing Markets Foundation (2023), even the most carbon-intensive food products, such as beef, are labeled as carbon-neutral. Therefore, the empirical question of how people value these labels remains when the certifiers are more transparent regarding the environmental quality of the label.

[5]To my knowledge, Swiss Airlines is an exception, which recently started to offer CO_2 offsetting, or a combination of CO_2 offsetting and CO_2 reduction after flight purchases (Swiss International Air Lines, 2024)

This study examines how transparency in carbon-neutral labeling influences consumer demand. While consumers are willing to pay for carbon-neutral labels, the findings reveal that transparency about the shares of CO_2 offsetting and reduction alone does not lead to a higher WTP. This suggests that although transparency remains important for accountability and market regulation, it may not be sufficient to influence consumer behavior. These results may underscore the need for (i) reliable third-party verification to support transparency in carbon-neutral labeling without requiring consumers to understand these nuances, and (ii) simplified but informative communication strategies to convey the different environmental implications of CO_2 offsetting and reduction activities. These insights are relevant for firms and ongoing regulatory efforts to limit misleading environmental claims.

3 Methodology

3.1 Survey and choice experiment design

This section provides a detailed overview of the survey and experimental design. It begins by introducing the tea product focused in the survey and outlining the sample characteristics. Next, it presents the structure of the experimental design. It then describes the DCE in detail, including the attributes and levels, as well as the choice tasks presented to participants. Finally, it summarizes the additional survey questions that followed the DCE, which were designed to elicit participants' understanding, perceptions, and motivations related to climate-friendly products and carbon-neutral labels.

Participants are randomly assigned to one of the three experimental groups. All groups receive identical survey content and discrete choice experiment designs. The

only difference is in the type of carbon-neutral label, as shown in Table 1. The control group is shown a standard carbon-neutral label, stating "CO_2 neutral" only. The information on the shares of CO_2 offsetting and CO_2 reduction are not revealed to the participants [6]. While treatment groups are shown a "transparent" carbon-neutral label with additional text indicating the shares of CO_2 offsetting and CO_2 reduction. In treatment group 1, the carbon-neutral label indicates a 95% offset and a 5% reduction, while in treatment group 2, the carbon-neutral label indicates a 50% offset and a 50% reduction. Respondents in the treatment groups are informed about the specific offset and reduction percentages that make the life cycle of the tea product carbon neutral.

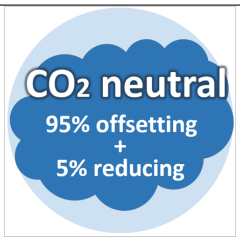
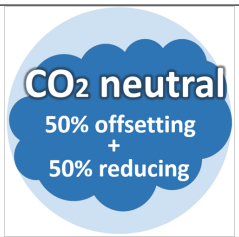
	Control	**Treatment 1**	**Treatment 2**
Label explanation	Yes	Yes	Yes
Carbon-neutral label			
CO_2 offset share	No information	%95	%50
CO_2 reduction share	No information	%5	%50
No. of participants	448	447	444

Table 1: Experimental design

This study focuses on tea consumers in the UK for several reasons. Firstly, it is common for tea products in the UK to have sustainability labels, including carbon-neutral labels, making tea a relevant product to study transparency in carbon-neutral

[6]Participants in the control group are provided the following information: "Product's greenhouse gas emissions, measured in carbon equivalent has been offset (compensated) by investing in activities outside of the company, such as tree planting projects, or reduced within the company in the last five years, such as through investments in cleaner production processes; or both offset and reduced."

labeling. Second, food systems significantly contribute to climate change, accounting for a third of global anthropogenic CO_2 emissions (Crippa et al., 2021). Although a box of tea itself may not be considered CO_2-intensive, its frequent consumption can still result in high emissions over time. Third, the UK ranks as the country with the third-highest per capita tea consumption globally (Statista, 2023).

In the survey, which is provided in Appendix A.1, respondents are asked to consider 80 teabags in their preferred form, equivalent to approximately 200 grams of tea. The pre-registered online survey[7] is pretested on 157 respondents, and the main survey included 1,339 tea drinkers in the UK.[8].

At the beginning of the survey, screening questions are asked about age, tea consumption and tea purchasing habits. Participants under the age of 18, or those who never consume or purchase tea, are screened out.

Attributes	**Attribute levels**
Carbon-neutral label	Carbon-neutral label, no label
Organic label	Organic label, no label
Ethical trade label	Ethical trade label, no label
Price	£0.90, £1.90, £2.90, £3.90, £4.90, £5.90, £6.90

Table 2: Choice attributes and levels

The DCE focuses on a 200-gram tea box with the following attributes: carbon-neutral label, organic label, ethical trade label, and price. Table 2 shows the attributes and attribute levels included in the DCE: three sustainability labels taking two levels each, the presence or absence of the label, and the price that takes seven levels ranging from £0.90 to £6.90 with £1 increments. The price levels in this survey

[7]Ozdemir Oluk, Begum. 2024. "Consumer Preferences for Transparent Carbon-Neutral Labels: A Choice Experiment." AEA RCT Registry. August 02. https://doi.org/10.1257/rct.12520-2.0

[8]Any deviations from the pre-registration, as well as exploratory and confirmatory hypotheses, are listed in the Appendix A.5.

were determined based on the average tea price in the UK, which is £2.46 for 250 grams of tea (Office for National Statistics, 2024c), as well as an online search for 100 tea products conducted on Sainsbury's, Tesco, Morrisons, Ocado, and Amazon UK during 2023.

Next, there are two choice attributes: organic and ethical trade labels, besides the carbon-neutral label. These are included because organic and fair trade labels are among the most common sustainability labels on tea products in the UK (based on an online search for 100 tea products in UK online grocery stores during 2023). Therefore, it makes the choice experiment more realistic and captures the trade-offs individuals would face in real decision-making scenarios. Organic and ethical trade do not vary across sub-samples, unlike carbon-neutral labels [9]. Participants are informed about the nature of these labels, including the carbon-neutral label, before the choice tasks. All five labels, three versions of carbon-neutral label, as well as organic, and ethical trade labels, are developed by the author and are hypothetical.

The type of tea blend (e.g., black, green, herbal) is not specified as a separate choice attribute because there is a strong preference for black tea in the UK. That is, 74% of people in the UK consume black tea, although they also drink other varieties such as green tea and herbal infusions at the same time (Tea and Infusions Association, 2022). Similarly, the packaging form (loose leaf or teabags) is not included as a separate attribute because 97.5% of tea sold in the UK is in teabag form (Tea and Herbal Association, 2024).

DCEs and CVMs are often criticized for being prone to hypothetical bias since participants are not required to make actual payments for their choices. To address this issue, the literature has employed various techniques such as cheap talk (Cummings

[9]The participants are informed that the organic label indicates products with only organic ingredients and no synthetic pesticides, while the ethical trade label ensures responsible labor practices and guarantees higher prices for exporters based on internationally recognized standards.

and Taylor, 1999), honesty priming (Howard et al., 2017), and oath scripts (de Magistris and Pascucci, 2014). In this study, I implement cheap talk, oath scripts, and a budget reminder to mitigate the limitations of stated preference methods. In the cheap talk script, I informed the respondents that survey participants are likely to overstate their WTP in hypothetical surveys and asked them to consider how they would feel about spending their money in a real situation. In the oath script, I asked participants to promise to provide honest responses by checking the box.

Ngene software is used to generate the DCE design, which consists of 16 different choice tasks blocked into 2^{10}. Figure 1 shows one of the choice cards shown to the control group. There are two tea products, and "none of the two" choice opinions. Restrictions are included to make choice cards more realistic. Alternatives with more labels had to be priced higher than alternatives with fewer labels. In addition to that, for the design of the main survey, additional restrictions are added for the lowest price level (£0.9) and the highest price level (£6.9) tea alternatives. If the price is 0.9 pounds, it should not have any labels; similarly, if it is £6.9, it must have all labels.. The estimated parameters from the pre-test were used to create the final CE design. Please refer to Table A.33 in Appendix A.2 for the respective choice design used for the survey.

The DCE is followed by questions that assess participants' choice certainty and identify protest responses. It further includes attribute non-attendance (ANA) and open-ended CVM questions. Additionally, the survey explores participants' attitudes toward climate change and sustainability labels using Likert-scale statements, which include trust in carbon-neutral labels, confusion about carbon-neutral labels, concerns about CO_2 offsetting, climate worry, warm glow (positive emotions from climate-friendly purchases), guilt (negative feelings when not making climate-friendly

[10]Therefore, each participant saw 8 choice cards.

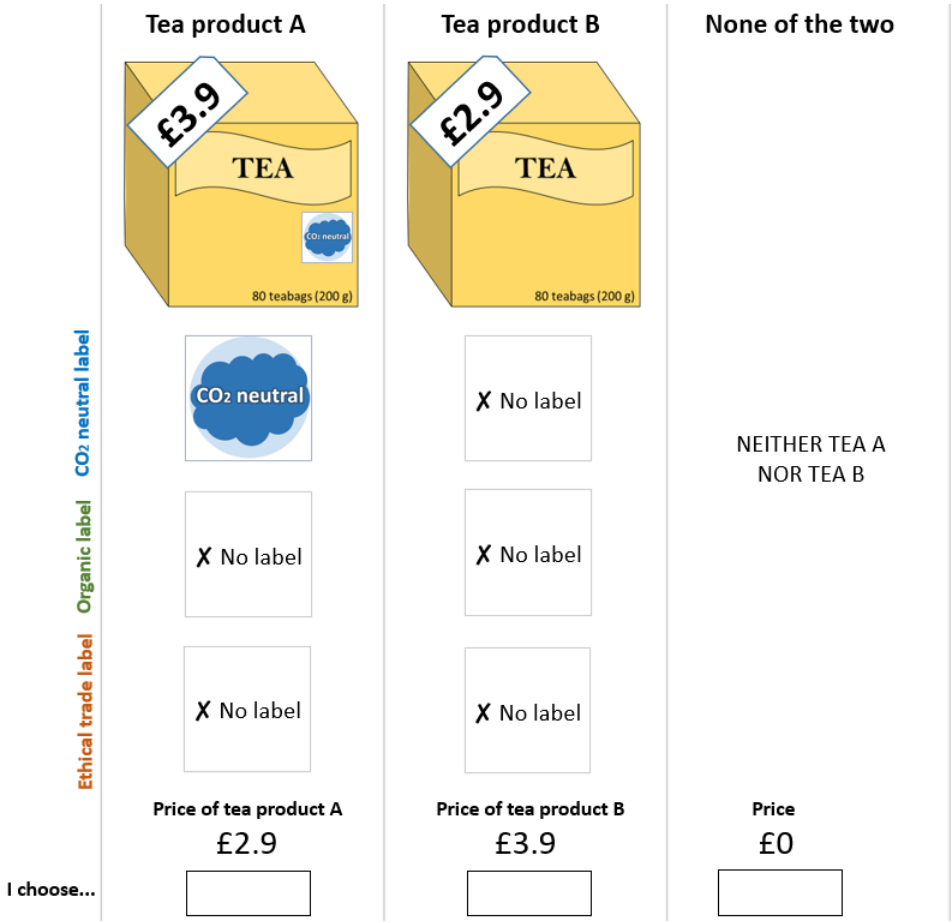


Figure 1: Example choice card presented to participants in the control group

Each participant was shown one of eight different cards. This example shows one of the cards presented to participants in the control group, block 2.

choices), social approval (perceived acceptance by others), the polluter pays principle (the opinion that producers should pay for climate mitigation), time restrictions (limited time for climate-friendly choices), and financial constraints (budget limitations preventing climate-friendly purchases). Then, participants are shown all three versions of carbon-neutral labels and asked which label they trust the most and find the most confusing. Finally, it collects information on participants' tea consumption habits and sociodemographic characteristics.

3.2 Data

The survey data were collected online in September 2024 in collaboration with a professional survey company. 54,219 respondents were initially invited to the survey. Of these, 2,366 started the survey, 580 were eliminated due to quota restrictions, and 225 were screened out or dropped out. Individuals who indicated that they never purchase or never drink tea, constituting 16% of the total sample, were disqualified from continuing the survey[11].

Consequently, a total of 1,339 individuals successfully completed the main survey[12]. However, the analyzed data consists of 1,321 individuals after excluding protest responses. Protest responses were identified as cases where respondents consistently chose the status quo option (no tea purchase), and when asked why, indicated reasons such as opposition to one or more labels, insufficient information about the products, a preference to spend money on other social or environmental initiatives, or disagreement with the question itself.

The sample mainly represents the UK population aged 18 and over, with some variations in general population statistics due to focusing on adult consumers older than 18 years old in this study. The median age of the sample is 46 years, whereas the national median age for adults is approximately 41 years (Office for National Statistics, 2022). Gender distribution is very close to the national profile, with 50% female participants in the sample compared to 51% in the population (Gov.uk, 2021b). 50% of the sample holding a post-secondary certificate (NQF Level 4) level or above, quite close to the national figure of 49% in the population (Gov.uk, 2021a). Additionally,

[11]Please refer to Table A.2 in Appendix A.1 for the comparison between tea drinkers/purchasers and non-tea drinkers/purchasers in terms of the three covariates (age, gender, and education) used as quota criteria to recruit a representative sample of the UK population before the screening questions.

[12]For the pre-test, 157 participants were recruited. The pre-test data is not included in the main survey data.

the median income of the sample[13] is £35,000, which is quite close to the national median of approximately £34,500 (Office for National Statistics, 2024a). Finally, the sample's employment rate of 65% is slightly below the national average of 75% (Office for National Statistics, 2024b). A summary of all detailed socioeconomic characteristics of tea consumers/drinkers is provided in Table A.1, and the balance of covariates is provided in Table A.7 in Appendix A.1.

Among tea consumers, approximately 78% drink tea daily, and 78% purchase tea at least once a month or more frequently. Detailed information on tea consumption and purchasing habits is provided in Table A.3 in Appendix A.1. Over 73% of respondents reported consuming black tea, while the remainder consume green tea, herbal tea, or other blends.

3.3 Empirical approach

To estimate consumers' WTP for carbon-neutral labels, I use a mixed logit (MXL) model in WTP space. The MXL model is advantageous over the multinomial logit (MNL) model because it accounts for unobserved preference heterogeneity across individuals and relaxes the independence of irrelevant alternatives (IIA) assumption of the MNL model.

According to the Random Utility Model (RUM) (Lancaster, 1966), consumers derive utility from the good's attributes. While making choices, they have to make tradeoffs between these attributes. The model assumes that consumers choose the option that provides the maximum expected utility (McFadden, 1973). Following Lancaster's random utility framework, the marginal utility of an individual derived from choosing an alternative is expressed as the sum of her marginal utility of at-

[13]Income was estimated by assigning the midpoint of each reported income bracket, with the lowest and highest categories approximated using a Pareto distribution.

tributes. The utility U_{inj} that individual i derives from choosing alternative j in choice situation n is:

The alternative is expressed as the sum of her marginal utility of attributes. The utility U_{inj} that individual i derives from choosing alternative j in choice situation n is:

$$U_{inj} = \alpha_i p_{inj} + \boldsymbol{\beta}_i^\top \mathbf{X}_{inj} + s_{in} + \varepsilon_{inj}, \tag{1}$$

where p_{inj} denotes the price (cost); $\alpha_i < 0$ is the individual-specific marginal disutility of price; $\boldsymbol{\beta}_i$ is the vector of individual-specific random coefficients for the non-price attributes in $\mathbf{X}_{inj}$; s_{in} is an individual- and choice situation-specific constant for the status quo alternative; and $\varepsilon_{inj} \sim$ i.i.d. Type I extreme value.

The MWTP for attribute level k for individual i is obtained by dividing the negative of the attribute level coefficient (-β_{ik}) by the price coefficient (α_{ik}), (Hensher et al., 2005). However, when both β_{ik} and α_i are random, this division can produce a heavy-tailed distribution with undefined moments, resulting in unstable estimates and wide confidence intervals (Hensher and Greene, 2003). obscure the problem by producing apparently finite moment estimates, even when the true underlying moments are infinite (Daly et al., 2012). To avoid these difficulties, Train and Weeks (2005) suggests estimating the model directly in WTP space. This involves treating the MWTPs, w_{ik}, as the random coefficients and reparameterize utility as (Train and Weeks, 2005):

$$U_{inj} = -\alpha_i \big(p_{inj} - \mathbf{w}_i^\top \mathbf{X}_{inj}\big) + s_{in} + \varepsilon_{inj}, \tag{2}$$

so that each w_{ik} can be interpreted directly as the MWTP for attribute level k, while s_{in} captures preference for the status quo alternative. The scale factor, $\alpha_i > 0$

is assumed to be positive.

The probability that individual i chooses the observed alternative j_{in} among J alternatives in choice situation n, under the WTP-space utility specification, is given by:

$$P_{in}(j = j_{in}) = \iiint \frac{\exp\left[-\alpha_i\big(p_{inj_{in}} - \mathbf{w}_i^\top \mathbf{X}_{inj_{in}}\big) + s_{in}\right]}{\sum_{j=1}^{J} \exp\left[-\alpha_i\big(p_{inj} - \mathbf{w}_i^\top \mathbf{X}_{inj}\big) + s_{in}\right]} \phi\big(\alpha_i, \mathbf{w}_i, s_i \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}\big)\, d\alpha_i\, d\mathbf{w}_i\, ds_i, \tag{3}$$

The likelihood for individual i, who faces L choice situations, is then:

$$L_i = \iiint \prod_{n=1}^{L} P_{in}(j = j_{in}) \cdot \phi(\alpha_i, \mathbf{w}_i, s_i \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})\, d\alpha_i\, d\mathbf{w}_i\, ds_i, \tag{4}$$

The log-likelihood is the sum over all individuals is given below:

$$\log L = \sum_{i=1}^{N} \log L_i. \tag{5}$$

By maximizing the simulated log-likelihood function in Equation 5, I estimate the MXL model. All parameters, including the price coefficient, the MWTP, and the status quo parameters, are modeled as random. I assume a log-normal distribution for the price coefficient[14] and normal distributions for the remaining parameters. I use 1,000 Halton draws to approximate the integral involved in the likelihood.[15] The starting values for the parameter means are taken from the MNL estimation, and the standard deviations are initialized at 0.5[16].

To compare the WTP for the carbon-neutral label among experimental groups,

[14]This is due to restricting the sign of the price coefficient.

[15]Following Train (2009), Halton draws are used to improve the precision of simulated maximum likelihood estimation compared to pseudo-random draws.

[16]Mariel et al. (2021) mentions that standard deviation starting values can be set to small positive numbers such as 0.5.

I use the approach introduced by Poe et al. (2005)[17] Although the Poe test is most relevant for WTP estimates derived from preference-space models, where the ratio of two random coefficients can result in analytically undefined distribution, I apply it to all specifications, including WTP-space estimations, to ensure full comparability between the main analysis (WTP space) and the robustness checks (which also include preference space estimations).

I conduct a comprehensive set of robustness checks that include both preference-space and WTP-space estimations. In the preference-space, I estimate an MNL model, an MXL model with random coefficients only for the label attributes, and an MXL model with all parameters specified as random. In the WTP-space, I estimate a MNL model, an MXL model with random coefficients only for the label attributes, and an MXL model with random coefficients for all attributes and with socioeconomic interactions, MXL model using 2,000 Halton draws instead of 1,000, and model with alternative-specific error components. I conduct a sensitivity analysis by varying the starting value of the standard deviation of the price coefficient to ensure that the estimation results are not sensitive to initial parameter values.

Additional robustness tests address sample restrictions. I re-estimate the model using the full sample including protest responses, a restricted sample excluding protest responses and survey speeders[18], and a more stringent sample excluding protest responses, survey speeders, and respondents who failed at least one attention or manipulation check.

To understand the underlying mechanisms of consumers' WTP, I interact the

[17]The Poe test is a simulation-based, non-parametric method that draws from the estimated distributions of two WTP measures. For each pair of draws, it records whether $\text{WTP}_{\text{group1}} < \text{WTP}_{\text{group2}}$. The two-sided p-value is calculated as $2 \times \min(P, , 1 - P)$, where P is the proportion of draws in which the first group's WTP is smaller than the second's.

[18]Survey speeders were identified as participants whose average survey completion time fell below the 5th percentile of the distribution.

WTP parameter for the carbon-neutral label with the variables listed in Table 3[19]. I then compare the resulting coefficients across experimental groups using Wald tests. Because the continuous variables are mean-centered, their coefficients represent the change in WTP associated with a one-unit increase; for dummy variables, they indicate a discrete shift compared to the reference group.

I conduct a series of robustness tests, including a preference-space estimation; the main WTP space estimation estimated without correcting for multiple hypothesis testing; a model with a reduced number of interaction terms; an alternative specification that includes an additional variable (warm glow) previously excluded due to high correlation with other variables; and a model that incorporates further interaction terms.

To explore competing sustainability labels, I include interaction terms between the WTP parameter for the carbon-neutral label and the organic and ethical-trade labels. This specification enables estimation of the WTP for the carbon-neutral label when no other label is present, when either organic or ethical-trade is present, and when both labels appear simultaneously, and it allows testing whether these differ across experimental groups in four different situations.

The importance of addressing multiple testing is well recognized in experimental economics (List et al., 2019). To account for multiple hypothesis testing, p-values are adjusted using the Holm-Bonferroni procedure (Holm, 1979). The p-values for the means of the random parameters, specifically for the WTP estimates of the carbon-neutral, organic, and ethical trade labels, as well as the status quo coefficient, are adjusted for four hypothesis tests[20]. Furthermore, estimations that include additional

[19]Since the warm glow variable is highly correlated with the guilt and social approval variables, it is excluded from the main estimation and included in a separate robustness test that omits the other two variables.

[20]In robustness analyses involving preference-space specifications, the p-value for the price coefficient is also adjusted for multiple testing, where it is interpreted as a structural utility parameter,

interaction terms with the WTP for the carbon-neutral label are also corrected for the number of tested hypotheses.

For the Poe tests comparing MWTP across experimental groups, p-values are adjusted for three pairwise comparisons. Wald tests comparing coefficients across bilateral comparisons of experimental groups are adjusted for the number of MWTP and status quo parameters. When additional interaction terms are included, the correction accounts for the total number of statistical tests.

but not in WTP-space estimations, where it functions as a scale parameter used to derive monetary values.

Variable	Description	Measurement
Only confused with transparent labels	Participants who are only confused with transparent labels but not with the standard labels	Binary variable (1 = yes, 0 = no)
Only trust in transparent labels	Participants who only trust transparent labels but not the standard labels	Binary variable (1 = yes, 0 = no)
Concern about CO_2 offsetting	Concern level about CO_2 offsetting	7-point Likert scale (mean-centered)
Confusion level	Confusion about carbon neutral labels	7-point Likert scale (mean-centered)
Trust level	Trust for carbon-neutral labels	7-point Likert scale (mean-centered)
Climate worry	Worry level about climate change	7-point Likert scale (mean-centered)
Warm glow	Positive emotions from climate-friendly purchases	7-point Likert scale (mean-centered)
Guilt	Negative feelings when not making climate-friendly choices	7-point Likert scale (mean-centered)
Social approval	Perceived acceptance by others for climate-friendly choices	7-point Likert scale (mean-centered)
Polluter pays	Belief that producers should pay for climate mitigation	7-point Likert scale (mean-centered)
Financial constraints	Limited financial resources for climate-friendly purchases	7-point Likert scale (mean-centered)
Time restrictions	Limited time for climate-friendly choices	7-point Likert scale (mean-centered)
Gender	Male respondents	Binary variable (1 = male, 0 = female, non-binary, or not disclosed)
Age	Age of the respondent	Continuous variable (mean-centered)
High education	Highest educational qualification: Bachelor's degree or higher	Binary variable (1 = bachelor's degree or higher, 0 = lower than bachelor's degree)
Employed	Employment status	Binary variable (1 = employed, 0 = unemployed, retired, full-time student, looking after family or home, other)
High income	Yearly household income equal to or greater than £50,000 after taxes	Binary variable (1 = high income, 0 = low income or not disclosed)
Not disclosed income	Participants' income information disclosure in the survey	Binary variable (1 = not disclosed, 0 = disclosed)

This table lists the covariates interacted with the carbon-neutral label to examine the factors associated with consumers' choices. For variables measured on 7-point Likert scales, higher values indicate stronger agreement, while lower values indicate stronger disagreement. Continuous variables are mean-centered.

Table 3: Description of interacted covariates

4 Results

4.1 Transparency and willingness to pay

This section presents the estimation results from the MXL model to estimate WTP for the carbon-neutral label. It examines differences in WTP across experimental groups to assess the effect of transparency and preference towards CO_2 reductions over offsets.

Table 4 presents the estimation results from the MXL model estimated in WTP space. The coefficients for the carbon-neutral label are positive and statistically significant at the 1% level across all groups, indicating a consistent positive WTP. The estimated WTP is £0.55 in the control group, £0.52 in treatment 1, and £0.32 in treatment 2. Relative to the average price used in the choice experiment (£3.90), these values correspond to 14%, 13%, and 8% of the price, respectively.

The Poe test results, reported in Table 5, show no statistically significant differences in WTP between the control and treatment groups, nor between the two treatment groups, after correcting for multiple hypothesis testing[21]. The difference between the control group and treatment 2 becomes significant at the 1% level only when no correction is applied. Therefore, there is no evidence of additional WTP for transparency in carbon-neutral labels. Similarly, no additional WTP is observed for CO_2 reductions compared to CO_2 offsets. If anything, consumers appear to prefer labels with a higher share of offset.

The comprehensive set of robustness checks, reported in Appendix A.3, assesses whether Poe test results vary across alternative model specifications and sample re-

[21]Holm-Bonferroni corrected Wald tests comparing coefficients across groups also yield the same results.

strictions. Across all estimations with DCE data, treatment 2 consistently yields the lowest WTP, while the control group and treatment 1 yield similar estimates.

In the preference-space models, Table A.10 presents an MNL specification without random parameters, while Tables A.11 and A.12 report MXL models with random label attributes and with all parameters random, respectively. The corresponding Poe test results in Tables A.14 show no statistically significant differences in WTP between groups.

In the WTP-space specifications, Tables A.16 and A.17 include MNL and MXL models, respectively. Next robustness checks vary the simulation precision and model specification. Table A.19 doubles the number of Halton draws to 2,000; Table A.20 introduces alternative-specific error components. Tables A.21, A.22, and A.23 estimate the model by changing sample restrictions, including or excluding protest responses, speeders, and participants who failed attention/manipulation checks. The Poe test results across Tables A.28, and A.29 show no significant differences in WTP across groups, except in two cases: estimation with twice as many Halton draws (Table A.19) and the inclusion of protest responses (Table A.21), both indicating marginal significance of 1% for treatment 2 versus control. No other group comparisons are significant.

Finally, a complementary robustness check using an open-ended CVM approach (Table A.30) based on 10,000 bootstrapped samples yields higher WTP in treatment 2 compared to the control and treatment 1, in contrast to the DCE results. However, none of the differences are statistically significant after applying the Holm adjustment, confirming the absence of statistically significant effects.

Table A.18 in Appendix A.3 explores the socioeconomic heterogeneity in the effect of transparency on consumers' WTP for carbon-neutral labels. The interaction terms indicate that both high education and high income are associated with higher WTP.

The association of high income is stronger in the control group and treatment 2, but less pronounced in treatment 1, which included a transparent label with a higher share of offsetting. In contrast, the association of high education is more prominent in the control group and treatment 1, while it is weaker in treatment 2, which displayed a transparent label with an equal share of offsetting and reduction.

The interaction with age is negative only in treatment 1, suggesting that older individuals may be more skeptical of offset-based labels. However, the effect size is relatively small: a one-year increase in age is associated with a decrease in WTP of approximately £0.02. No consistent patterns emerge for gender or employment status.

Bilateral Wald tests indicate that transparency may influence some of these associations. In particular, the difference in the high-income interaction between treatment 1 and treatment 2 is marginally significant, even after applying the Holm adjustment ($p = 0.091$).

The preference towards carbon reductions over offsets, is associated with higher income level. In treatment 2, individuals in the high-income group are willing to pay approximately £1.24 more for the carbon-neutral label compared to the low-income group. This estimate is £0.50 in treatment 1. This suggests that labels with larger reduction shares may increase the WTP of higher-income consumers more than intransparent labels or labels implying a larger share of offsetting would.

	Full sample	**Control**	**Treatment 1**	**Treatment 2**
MWTP estimates				
$\text{WTP}_{\text{Carbon neutral}}$	0.47***	0.55***	0.52***	0.32***
	(0.11)	(0.08)	(0.10)	(0.09)
$\text{WTP}_{\text{Organic}}$	1.10***	1.04***	1.15***	1.13***
	(0.07)	(0.09)	(0.11)	(0.10)
$\text{WTP}_{\text{Ethical trade}}$	1.17***	1.15***	1.14***	1.23***
	(0.07)	(0.11)	(0.12)	(0.11)
Means of the choice parameters				
μ_{Price}	-0.17***	-0.11	-0.18***	-0.21***
	(0.04)	(0.07)	(0.06)	(0.07)
$\mu_{\text{Status quo}}$	-4.26***	-4.73***	-3.97***	-4.28***
	(0.19)	(0.33)	(0.27)	(0.29)
Standard deviations				
$\sigma_{\text{Carbon neutral}}$	1.24***	1.31***	1.36***	1.25***
	(0.46)	(0.12)	(0.24)	(0.12)
σ_{Organic}	1.64***	1.41***	1.57***	1.69***
	(0.10)	(0.14)	(0.14)	(0.10)
$\sigma_{\text{Ethical trade}}$	1.48***	1.40***	1.43***	1.47***
	(0.12)	(0.19)	(0.15)	(0.15)
σ_{Price}	0.77***	0.78***	0.74***	0.82***
	(0.07)	(0.07)	(0.09)	(0.07)
$\sigma_{\text{Status quo}}$	2.27***	2.49***	2.12***	2.22***
	(0.13)	(0.24)	(0.18)	(0.19)
Log Likelihood	-8540.44	-2808.57	-2898.27	-2822.83
AIC	17100.88	5637.13	5816.54	5665.67
BIC	17173.54	5698.88	5878.24	5727.21
Pseudo-R^2	0.26	0.28	0.25	0.26
Number of observations	10568	3552	3536	3480
Number of participants	1321	444	442	435

Table 4: Mixed logit model, WTP space estimates, all parameters randomized

This table reports the results from a Mixed logit model estimated in WTP space. Robust standard errors are shown in parentheses. Holm-Bonferroni adjusted p-values (four tests for WTP estimates and status quo parameter) are used to determine statistical significance: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$.

Holm-Bonferroni adjusted Wald tests for WTP estimates (for pairwise comparisons across the three experimental groups) indicate no statistically significant differences in WTP coefficients for carbon-neutral, organic, and ethical trade labels.

	 Control	 Treatment 1	 Treatment 2
$WTP_{carbon\ neutral}$	0.55*** (0.08)	0.52*** (0.10)	0.32*** (0.09)
Holm-Bonferroni adjusted Poe test			
Control vs Treatment 1	0.40 (p = 0.793)		
Control vs Treatment 2	0.03 (p = 0.152)		
Treatment 1 vs Treatment 2		0.06 (p = 0.251)	
Unadjusted Poe test			
Control vs Treatment 1	0.40 (p = 0.793)		
Control vs Treatment 2	0.03* (p = 0.051)		
Treatment 1 vs Treatment 2		0.06 (p = 0.125)	

WTP estimates are reported with robust standard errors in parentheses.
Significance levels: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$.
Poe test statistics are reported with Holm-Bonferroni adjusted p-values (for three pairwise comparisons across experimental groups). Unadjusted p-values are also reported.

Table 5: WTP estimates for the carbon-neutral label across experimental groups and Poe test results

4.2 Mechanisms

This section examines the mechanisms that are associated with consumers' WTP for carbon-neutral labels. As shown in Table A.4, nearly half of the participants misunderstand the meaning of offsetting. Table 6 summarizes participants' trust in and confusion about transparent and standard carbon-neutral labels. After the choice experiment, respondents were asked to indicate which label they trusted most and which they found most confusing, selecting among the standard label, two transparent labels, and a none option. The transparent label with a 50/50 split between CO_2 reduction and offsetting was the most trusted (39.29%), while the version with 95% offsetting and 5% reduction was perceived as the most confusing (34.07%).

Estimation results presented in Table 7 explore how these perceptions, along with other psychological and contextual variables, interact with WTP for a carbon-neutral label.[22] Holm-adjusted Wald tests are conducted for pairwise comparisons of coefficients across experimental groups. Although no differences are statistically significant, variation in coefficient signs and levels across experimental groups provides informative patterns[23].

First, I examine the indicator for participants who are confused by transparent but not standard labels. The coefficient is positive in the control group and treatment group 1, and negative in treatment group 2, suggesting a potential negative association between confusion and WTP in that group, although the effect is not statistically significant.

Next, I examine the indicator for those who trust transparent but not standard

[22] All continuous covariates, including general trust, general confusion, climate worry, guilt, social approval, and financial constraints, have been mean-centered. These variables are measured on seven-point Likert scales; thus, a one-unit increase corresponds to an increase in WTP (in GBP). The binary indicators, for respondents who are 'only' (i) confused by or (ii) trust transparent labels (but not standard labels), reflect the average WTP difference relative to the remaining sample.

[23] Results not adjusted for multiple hypothesis testing are also included in the table notes.

labels. This variable shows one of the strongest associations: its coefficient is highly significant in the full sample, indicating an average WTP increase of £0.80 in the control group and £0.72 in treatment group 2. This association is insignificant in treatment group 1, possibly because participants in that group saw label dominated by offsetting, thereby weakening the influence of trust. Interestingly, trust in transparent labels increases WTP even in the control group, suggesting that trust do not play a role in WTP for transparency.

The variable capturing concern about carbon offsets is positive and significant in both the full sample and treatment group 1. This may reflect a misinterpretation: participants could confuse the term offsetting with emissions, particularly in treatment group 1, where offsetting was most emphasized.

I then turn to general trust and general confusion regarding carbon-neutral labels. Unlike the previous indicators, these variables reflect overall perceptions rather than relative evaluations between label types. Neither variable is statistically significant in any group, although general trust has a positive sign and general confusion a negative sign in the full sample, confirming the expectations.

Other covariates, such as climate worry and guilt, are not significantly associated with WTP. Social approval, however, is positive and significant in the full sample and the control group, suggesting that social image concerns may be more salient when the label is not transparent.

Financial constraints are negatively associated with WTP in the full sample even when socioeconomic controls are included, but this variable is insignificant within individual groups. Time constraints have no association with WTP.

The results are mostly robust; several robustness checks are presented in Appendix A.3, where different model specifications and additional variables are examined. Wald tests indicate that familiarity with carbon-neutral labels (i.e., respondents

who purchase carbon-neutral labeled products) is strongly associated with WTP (£ 1.57 increase), and differences among experimental groups are statistically significant. The association is greater when (i) the label transparent, (ii) implies higher share of reduction.

The mechanism analyses reveal important insights into how consumers interpret carbon-neutral labels. Although the confusion variable is not statistically significant, the high rate of misunderstanding, particularly regarding offsetting, suggests that limited comprehension may still influence preferences. Familiarity with carbon-neutral labels, measured by whether respondents have purchased such labeled products while grocery shopping, is consistently associated with higher WTP, indicating the importance of prior exposure in shaping valuation.

Trust in transparent labels is positively linked to WTP, but its effect does not vary with transparency, suggesting that trust reflects more general attitudes rather than direct responses to specific label content. The higher WTP for the label with a greater share of offsetting could also be explained by a larger order bias.

Overall, these findings imply that consumer responses are driven less by the exact information disclosed and more by their baseline understanding and familiarity with the concept of carbon neutrality.

		Full sample		Control		Treatment 1		Treatment 2	
		N	Share	N	Share	N	Share	N	Share
Trust for label 1		309	23.39%	109	24.55%	90	20.36%	110	25.29%
Trust for label 2		330	24.98%	114	25.68%	124	28.05%	92	21.15%
Trust for label 3		519	39.29%	179	40.32%	164	37.10%	176	40.46%
Trust (none)		264	19.98%	80	18.02%	97	21.95%	87	20.00%
Confusion about label 1		341	25.81%	111	25.00%	130	29.41%	100	22.99%
Confusion about label 2		450	34.07%	162	36.49%	124	28.05%	164	37.70%
Confusion about label 3		336	25.44%	106	23.87%	116	26.24%	114	26.21%
Confusion (none)		393	29.75%	131	29.50%	144	32.58%	118	27.13%

This table is based on participants' responses to the question about which label they trust the most and which they find the most confusing among the three label options (Label 1, Label 2, Label 3) and a 'none' option. This question was asked later in the survey, after the choice experiment. Multiple selections were allowed.
Label 1 represents the standard carbon-neutral label. Label 2 is the transparent carbon-neutral label with 95% CO_2 offsetting and 5% CO_2 reduction. Label 3 is the transparent carbon-neutral label with an equal split between CO_2 reduction and CO_2 offsetting (50%-50%).

Table 6: Confusion about and trust in standard and transparent carbon-neutral labels

	Full sample	**Control**	**Treatment 1**	**Treatment 2**
$WTP_{Carbon\ neutral}$	-0.26	-0.43	-0.25	-0.19
	(0.16)	(0.29)	(0.26)	(0.29)
Main interactions				
$WTP_{Carbon\ neutral}$ x Only confused with transparent labels	0.03	0.21	0.24	-0.32
	(0.15)	(0.20)	(0.19)	(0.24)
$WTP_{Carbon\ neutral}$ x Only trust in transparent labels	0.65***	0.80***	0.39	0.72**
	(0.14)	(0.23)	(0.20)	(0.23)
$WTP_{Carbon\ neutral}$ x Concern for CO_2 offsets	0.18***	0.08	0.30***	0.15
	(0.05)	(0.10)	(0.08)	(0.18)
$WTP_{Carbon\ neutral}$ x Trust level	0.09	0.1	0.12	-0.05
	(0.05)	(0.08)	(0.07)	(0.09)
$WTP_{Carbon\ neutral}$ x Confusion level	-0.03	0.01	-0.00	-0.16
	(0.06)	(0.06)	(0.06)	(0.08)
$WTP_{Carbon\ neutral}$ x Climate worry	0.06	0.01	-0.05	0.25
	(0.05)	(0.10)	(0.08)	(0.22)
$WTP_{Carbon\ neutral}$ x Guilt	0.12	0.15	0.11	0.13
	(0.05)	(0.07)	(0.08)	(0.13)
$WTP_{Carbon\ neutral}$ x Social approval	0.18***	0.29*	0.21	0.00
	(0.05)	(0.11)	(0.09)	(0.09)
$WTP_{Carbon\ neutral}$ x Polluter pays	-0.02	-0.01	-0.12	0.03
	(0.04)	(0.09)	(0.06)	(0.09)
$WTP_{Carbon\ neutral}$ x Financial constraints	-0.18***	-0.16	-0.15	-0.22
	(0.04)	(0.11)	(0.07)	(0.09)
$WTP_{Carbon\ neutral}$ x Time restrictions	-0.02	-0.08	-0.03	0.06
	(0.04)	(0.07)	(0.06)	(0.06)
Socioeconomic interactions	Yes	Yes	Yes	Yes
WTP for sustainability label attributes, price, status quo	Yes	Yes	Yes	Yes
Standard deviations of random parameters	Yes	Yes	Yes	Yes
Log Likelihood	-8371.60	-2746.65	-2835.84	-2756.99
AIC	16797.20	5547.30	5725.68	5567.97
BIC	16993.37	5714.04	5892.29	5734.15
Pseudo R^2	0.28	0.30	0.27	0.28
Number of observations	1056	3552	3536	3480
Number of participants	1321	444	442	435

This table shows the choice model output from the Mixed Logit model estimated in WTP space with interaction variables. Robust standard errors are reported in parentheses. All parameters, except for the interaction terms, are randomized. Holm-Bonferroni adjusted p-values (for 21 comparisons) and significance levels: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$. Bilateral Wald tests were conducted to assess the equivalence of interaction coefficients across experimental groups. Using coefficients reported in the table, adjusted Wald test results show that none of the above interactions are statistically significant after adjusting for multiple testing using Holm-Bonferroni method for 21 comparisons, which include all of the main interactions, socioeconomic variables, WTP parameters, and the status quo parameter. Unadjusted Wald test results show that the interaction with concern for carbon offsets is significant at $p = 0.079$ when comparing the control group and treatment group 1. The interaction with the binary variable indicating individuals confused only by transparent labels (but not by standard labels) is significant at $p = 0.065$ for the comparison between treatment group 1 and treatment group 2. For the comparison between the control group and treatment group 2, the interaction with the social approval variable is significant at $p = 0.038$, and the interaction with confusion about carbon-neutral labels is significant at $p = 0.095$. Wald test results for coefficients not reported in the table (e.g., socioeconomic variables) are also not included in the table notes.

Table 7: Mixed logit model, WTP space estimates with interaction variables

4.3 Competing labels

This section evaluates consumers' WTP for two further sustainability labels on tea, organic and ethical trade, and examines how their presence changes the WTP for transparent and standard versions of the carbon-neutral label.

Table 4 reports an average WTP of £0.47 for the carbon-neutral label, £1.10 for the organic label, and £1.17 for the ethical-trade label. Relative to the average product price of £3.90 in the choice experiment, these values correspond to approximately 12%, 28%, and 30% of the purchase price. Similar values are obtained in all robustness checks, including the preference-space results summarised in Table A.21 and the additional WTP-space estimates in Tables A.16-A.23.

To understand whether the organic and ethical-trade labels complement or compete with the carbon-neutral claim, interaction terms between the carbon-neutral label and each competing label are introduced. Figure 2 plots the estimated WTP for the carbon-neutral label under four scenarios. When no competing labels are present, the WTP equals £1.02 in the control group, £0.86 in treatment 1, and £0.83 in treatment 2. Adding only the ethical-trade label lowers the corresponding figures to £0.55, £0.44, and £0.30. When only the organic label is present, the estimates fall to £0.29, £0.40, and £0.11. Finally, when both labels are displayed simultaneously, the WTP for the carbon-neutral claim turns negative, reaching -£0.18 in the control group, -£0.02 in treatment 1, and -£0.43 in treatment 2.

The negative interaction coefficients in Table A.31 confirm competition among the labels: the carbon-neutral and organic label interaction equals –£0.73 in the control group, –£0.46 in Treatment 1, and –£0.73 in Treatment 2, while the carbon-neutral and ethical-trade label interaction is –£0.47, –£0.42, and –£0.54, respectively. While interaction effects are smaller for the label implying a higher offsetting share,

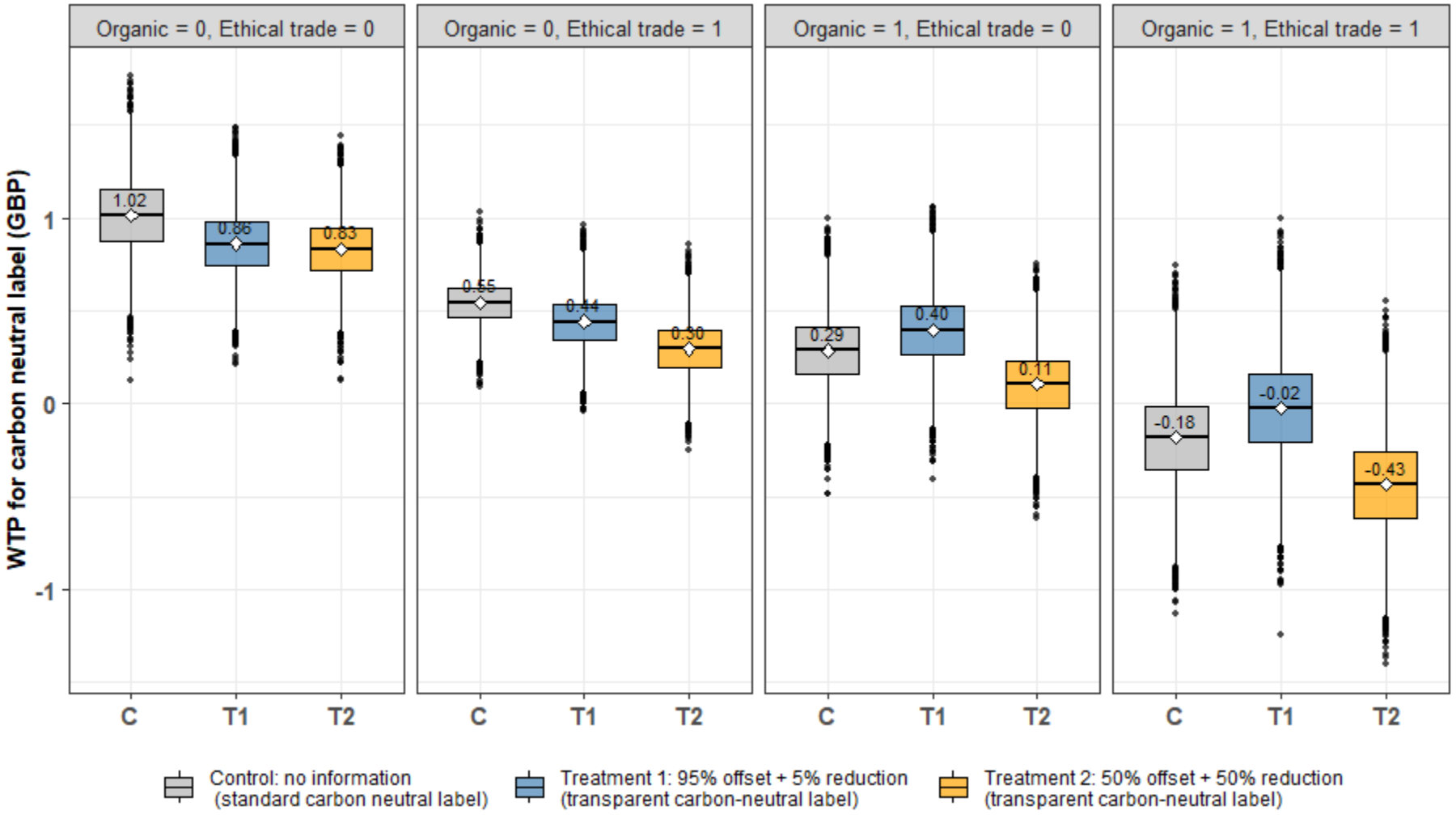


Figure 2: WTP for the carbon-neutral label in the presence or absence of other sustainability labels

This figure shows the WTP estimates derived from a mixed logit model with interaction effects between the carbon-neutral label with organic and ethical trade labels. It displays the WTP estimates for the carbon neutral label depending on whether the other labels are present (coded as 1) or not present (coded as 0). The corresponding choice model output is provided in Table A.31 in Appendix A.4.

the results indicate that the two competing labels substantially erode the premium otherwise paid for carbon neutrality.

Holm-adjusted Poe tests (Table A.32) show that, despite competition among labels, the differences among the experimental groups, standard carbon-neutral (control) versus transparent carbon-neutral labels (treatments 1 and 2), remain statistically insignificant.

To sum up, although consumers value each sustainability certification, the organic and ethical-trade labels reduce, and when combined, even reverse, the additional amount they are willing to pay for a carbon-neutral label, with the strongest reduction observed for the transparent 50% offset / 50% reduction label.

5 Conclusion

This study investigates whether transparency in carbon-neutral labels, in terms of disclosing the shares of CO_2 offsetting and CO_2 reduction, affects consumers' WTP. The motivation of this paper based on by growing criticism of CO_2 offsets and increasing regulatory scrutiny of environmental claims. While carbon-neutral labels are becoming more common, many lack information on how carbon neutrality is achieved. Given that offsets and reductions may significantly differ substantially in their environmental effectiveness, transparency may play a key role in shaping consumer WTP. This study provides empirical evidence on whether such transparency influences demand.

Using a pre-registered discrete choice experiment conducted with 1,339 UK tea consumers, this paper compares consumer preferences across a control group shown a standard carbon-neutral label and two treatment groups shown transparent labels with varying shares of CO_2 offsetting and reduction (95% offsetting + 5% reduction

vs. 50% offsetting + 50% reduction). The results reveal that while consumers are willing to pay a premium for carbon-neutral labels in general, there is no statistically significant difference in WTP between the standard label and either of the transparent versions. Moreover, no significant WTP difference is observed between the two transparency treatments. These findings suggest that transparency alone, without accompanying efforts to improve understanding, does not lead to higher WTP.

Exploratory analyses indicate that confusion and limited understanding may partly explain these results. Many participants misunderstood the concept of CO_2 offsetting despite receiving explanations. Interestingly, although the label with an equal share of CO_2 offsetting and reduction was most trusted overall, it did not generate the highest WTP. Participants with stronger concerns about CO_2 offsetting displayed higher WTP for the label with a larger offsetting share, suggesting potential misinterpretation of the term offsetting as the term emissions. This points to the complexity of consumer perceptions and highlights that transparency may not always increase trust or demand when underlying concepts are not well understood.

From a policy perspective, while transparency aligns with regulatory efforts to prevent greenwashing, I show that it may not be sufficient on its own to influence consumer behavior. Policymakers may consider supporting transparency mandates with simplified and accessible communication about the environmental implications of CO_2 offsetting versus reduction. Furthermore, given widespread information asymmetries between consumers and producers, reliable third-party verification may help ensure that transparency translates into meaningful consumer signals.

Taken together, the findings underscore the need for a an approach that balances the goals of not only transparency, but also clarity, and credibility. As green commitments by companies continue and regulations become stricter, understanding how consumers interpret and respond to carbon-neutral labels will remain essential for

designing effective climate claims in the marketplace.

References

Adamowicz, W., J. Louviere, and M. Williams (1994). Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management 26*(3), 271–292.

Akaichi, F., R. M. Nayga Jr, and L. L. Nalley (2017). Are there trade-offs in valuation with respect to greenhouse gas emissions, origin and food miles attributes? *European Review of Agricultural Economics 44*(1), 3–31.

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics 84*(3), 488–500.

Aldred, J. (2012). The ethics of emissions trading. *New Political Economy 17*(3), 339–360.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal 100*(401), 464–477.

Axelsson, K., A. Wagner, I. Johnstone, M. Allen, B. Caldecott, N. Eyre, S. Fankhauser, T. Hale, C. Hepburn, C. Hickey, R. Khosla, S. Lezak, E. Mitchell-Larson, Y. Malhi, N. Seddon, A. Smith, and S. M. Smith (2024). Oxford principles for net zero aligned carbon offsetting (revised 2024).

Becken, S. and B. Mackey (2017). What role for offsetting aviation greenhouse gas emissions in a deep-cut carbon world? *Journal of Air Transport Management 63*, 71–83.

Bek, D. (2022). Pricing sustainable shipping of coffee: Consumersâ preferences and willingness to pay for emission reductions and offsets. *Junior Management Science*.

Birkenberg, A., M. E. Narjes, B. Weinmann, and R. Birner (2021). The potential of carbon neutral labeling to engage coffee consumers in climate change mitigation. *Journal of Cleaner Production 278*, 123621.

Blasch, J. and M. Farsi (2014). Context effects and heterogeneity in voluntary carbon offsetting–a choice experiment in Switzerland. *Journal of Environmental Economics and Policy 3*(1), 1–24.

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

Brécard, D. (2017). Consumer misperception of eco-labels, green market structure and welfare. *Journal of Regulatory Economics 51*, 340–364.

Brounen, D. and N. Kok (2011). On the economics of energy labels in the housing market. *Journal of Environmental Economics and Management 62*(2), 166–179.

Brouwer, R. (2011). A mixed approach to payment certainty calibration in discrete choice welfare estimation. *Applied Economics 43*(17), 2129–2142.

Brouwer, R., L. Brander, and P. Van Beukering (2008). "a convenient truth": air travel passengers' willingness to pay to offset their co_2 emissions. *Climatic Change 90*, 299–313.

Brunnschweiler, C., I. Edjekumhene, and P. Lujala (2021). Does information matter? transparency and demand for accountability in ghana's natural resource revenue management. *Ecological Economics 181*, 106903.

Bumpus, A. G. and D. M. Liverman (2008). Accumulation by decarbonization and the governance of carbon offsets. *Economic Geography 84*(2), 127–155.

Calel, R., J. Colmer, A. Dechezleprêtre, and M. Glachant (2021). Do carbon offsets offset carbon?

Carattini, S. and J. Blasch (2024). Nudging when the descriptive norm is low: Evidence from a carbon offsetting field experiment. *Journal of Behavioral and Experimental Economics 110*, 102194.

Carattini, S., F. Dvorak, I. Logar, and B. Özdemir Oluk (2024). Demand for carbon-neutral products.

Carattini, S., K. Gillingham, X. Meng, and E. Yoeli (2022). Peer-to-peer solar and social rewards: Evidence from a field experiment. CESifo Working Paper 10173, Center for Economic Studies and ifo Institute (CESifo), Munich.

Carattini, S. and A. Tavoni (2016). How green are green economists? *CSLF Articles* (9).

Carbon Trust (2023). Carbon neutral verification. Accessed: 2024-12-19.

Changing Markets Foundation (2023). Feeding us greenwash. `http://changingmarkets.org/wp-content/uploads/2023/03/Feeding-Us-Greenwash-web.pdf`. Accessed: 2023-06-10.

Chen, N., Z.-H. Zhang, S. Huang, and L. Zheng (2018). Chinese consumer responses to carbon labeling: Evidence from experimental auctions. *Journal of Environmental Planning and Management 61*(13), 2319–2337.

Climate Portal (2023). Carbon offsets. `https://climate.mit.edu/explainers/carbon-offsets`. Accessed: 2024-02-19.

ClimatePartner (2023). Carbon neutral. what does that actually mean? https://www.climatepartner.com/en/knowledge/glossary/carbon-neutral. Accessed: 2024-02-19.

Crippa, M., E. Solazzo, D. Guizzardi, F. Monforti-Ferrario, F. N. Tubiello, and A. Leip (2021). Food systems are responsible for a third of global anthropogenic ghg emissions. *Nature Food 2*(3), 198–209.

Cummings, R. G. and L. O. Taylor (1999). Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. *American Economic Review 89*(3), 649–665.

Daly, A., S. Hess, and K. Train (2012). Assuring finite moments for willingness to pay in random coefficient models. *Transportation 39*(1), 19–31.

Darby, M. R. and E. Karni (1973). Free competition and the optimal amount of fraud. *The Journal of Law and Economics 16*(1), 67–88.

de Magistris, T. and S. Pascucci (2014). The effect of the solemn oath script in hypothetical choice experiment survey: A pilot study. *Economics Letters 123*(2), 252–255.

Djimeu, E. W. and D.-G. Houndolo (2016). Power calculation for causal inference in social science: sample size and minimum detectable effect determination. *Journal of Development Effectiveness 8*(4), 508–527.

Dorner, Z. (2019). A behavioral rebound effect. *Journal of Environmental Economics and Management 98*, 102257.

Drichoutis, A. C., J. L. Lusk, and V. Pappa (2016). Elicitation formats and the WTA/WTP gap: A study of climate neutral foods. *Food Policy 61*, 141–155.

European Parliament (2023). EU to ban greenwashing and improve consumer information on product durability. https://www.europarl.europa.eu/news/en/press-room/20230918IPR05412/eu-to-ban-greenwashing-and-improve-consumer-information-on-product-durability. Accessed: 2023-11-06.

Feucht, Y. and K. Zander (2018). Consumers' preferences for carbon labels and the underlying reasoning. a mixed methods approach in 6 european countries. *Journal of Cleaner Production 178*, 740–748.

Gassler, B. et al. (2015). How green is your grüner? millennial wine consumers' preferences and willingness-to-pay for eco-labeled wine. *Jahrbuch der Österreichischen Gesellschaft für Agrarökonomie 24*, 131–140.

Gov.uk (2021a). Education and training statistics for the UK: 2021. https://explore-education-statistics.service.gov.uk/find-statistics/education-and-training-statistics-for-the-uk. Accessed on 3 November 2024.

Gov.uk (2021b). Male and female populations. https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/male-and-female-populations/latest/. Accessed on 3 November 2024.

Grebitus, C., B. Steiner, and M. Veeman (2013). Personal values and decision making: evidence from environmental footprint labeling in Canada. *American Journal of Agricultural Economics 95*(2), 397–403.

Gruchmann, T., G. Maugeri, and R. Wagner (2025). Do you feel guilty? a consumer-

centric perspective on green nudging in last-mile deliveries. *International Journal of Physical Distribution & Logistics Management 55*(5), 540–566.

Hanley, N., D. MacMillan, R. E. Wright, C. Bullock, I. Simpson, D. Parsisson, and B. Crabtree (1998). Contingent valuation versus choice experiments: estimating the benefits of environmentally sensitive areas in scotland. *Journal of agricultural economics 49*(1), 1–15.

Hensher, D. A. and W. H. Greene (2003). The mixed logit model: The state of practice. *Transportation 30*(2), 133–176.

Hensher, D. A., J. M. Rose, and W. H. Greene (2005). *Applied choice analysis: a primer*. Cambridge university press.

Heyes, A. and S. Martin (2017). Social labeling by competing NGOs: A model with multiple issues and entry. *Management Science 63*(6), 1800–1813.

Heyes, A. and S. Martin (2018). Inefficient NGO labels: Strategic proliferation and fragmentation in the market for certification. *Journal of Economics & Management Strategy 27*(2), 206–220.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics 6*(2), 65–70.

Hooper, P., B. Daley, H. Preston, and C. Thomas (2008). An assessment of the potential of carbon offset schemes to mitigate the climate change implications of future growth of UK aviation. Final OMEGA Project Report. Centre for Air Transport and the Environment, Manchester Metropolitan University.

Howard, G., B. E. Roe, E. C. Nisbet, and J. F. Martin (2017). Hypothetical bias mitigation techniques in choice experiments: Do cheap talk and honesty priming

effects fade with repeated choices? *Journal of the Association of Environmental and Resource Economists 4*(2), 543–573.

Hyams, K. and T. Fawcett (2013). The ethics of carbon offsetting. *Wiley Interdisciplinary Reviews: Climate Change 4*(2), 91–98.

International Initiative for Impact Evaluation (3ie) (2016). Power calculation for causal inference in social science: Sample size and minimum detectable effect determination [excel spreadsheet]. `https://www.3ieimpact.org/evidence-hub/publications/working-papers/power-calculation-causal-inference-social-science-sample`.

IPCC (2022). *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, NY, USA: Cambridge University Press.

Kahneman, D. and J. L. Knetsch (1992). Valuing public goods: the purchase of moral satisfaction. *Journal of environmental economics and management 22*(1), 57–70.

Kim, E.-H. and T. Lyon (2011). When does institutional investor activism increase shareholder value?: The Carbon Disclosure Project. *The B.E. Journal of Economic Analysis & Policy 11*(1).

Kotchen, M. J. (2009). Voluntary provision of public goods for bads: A theory of environmental offsets. *The Economic Journal 119*(537), 883–899.

Kraft-Todd, G. T., B. Bollinger, K. Gillingham, S. Lamp, and D. G. Rand (2018). Credibility-enhancing displays promote the provision of non-normative public goods. *Nature 563*, 245–248.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of political economy 74*(2), 132–157.

List, J. A., A. M. Shaikh, and Y. Xu (2019). Multiple hypothesis testing in experimental economics. *National Bureau of Economic Research Working Paper No. 25848*.

MacKerron, G. J., C. Egerton, C. Gaskell, A. Parpia, and S. Mourato (2009). Willingness to pay for carbon offset certification and co-benefits among (high-) flying young adults in the uk. *Energy policy 37*(4), 1372–1381.

Mariel, P., D. Hoyos, J. Meyerhoff, M. Czajkowski, T. Dekker, K. Glenk, J. B. Jacobsen, U. Liebe, S. B. Olsen, J. Sagebiel, et al. (2021). *Environmental valuation with discrete choice experiments: Guidance on design, implementation and data analysis*. Springer Nature.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. Academic Press New York, New York, NY, USA.

McKinsey & Company (2023). A blueprint for scaling voluntary carbon markets to meet the climate challenge. `https://www.mckinsey.com/capabilities/sustainability/our-insights/a-blueprint-for-scaling-voluntary-carbon-markets-to-meet-the-climate-challenge`. Accessed: March 2023.

Mortensen, C. R., R. Neel, R. B. Cialdini, C. M. Jaeger, R. P. Jacobson, and M. M. Ringel (2019). Trending norms: A lever for encouraging behaviors performed by the minority. *Social Psychological and Personality Science 10*(2), 201–210.

Office for National Statistics (2022). UK population estimates, mid-2021. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2021. Accessed on 3 November 2024.

Office for National Statistics (2024a). Average household income, UK: financial year ending 2023. https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/bulletins/householddisposableincomeandinequality/financialyearending2023. Accessed on 3 November 2024.

Office for National Statistics (2024b). Employment in the UK: October 2024. https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/employmentintheuk/october2024. Accessed on 3 November 2024.

Office for National Statistics (2024c, February). RPI: Ave price - Tea bags, per 250g. https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/cznq/mm23. Release date: 14 February 2024.

Onozaka, Y. and D. T. McFadden (2011). Does local labeling complement or compete with other sustainable labels? A conjoint analysis of direct and joint values for fresh produce claim. *American Journal of Agricultural Economics 93*(3), 693–706.

Poe, G. L., K. L. Giraud, and J. B. Loomis (2005). Computational methods for measuring the difference of empirical distributions. *American Journal of Agricultural Economics 87*(2), 353–365.

Poret, S. (2019). Label wars: Competition among NGOs as sustainability standard setters. *Journal of Economic Behavior & Organization 160*, 1–18.

Roemer, N., G. C. Souza, C. Tröster, and G. Voigt (2023). Offset or reduce: How should firms implement carbon footprint reduction initiatives? *Production and Operations Management 32*(9), 2940–2955.

Rogelj, J., O. Geden, A. Cowie, and A. Reisinger (2021). Net-zero emissions targets are vague: Three ways to fix. *Nature 591*(7850).

Schneider, L. and A. Kollmuss (2015). Perverse effects of carbon markets on hfc-23 and sf6 abatement projects in russia. *Nature Climate Change 5*(12), 1061–1063.

Schneider, L., A. Kollmuss, and M. Lazarus (2015). Addressing the risk of double counting emission reductions under the UNFCCC. *Climatic Change 131*(4), 473–486.

Sparkman, G. and G. M. Walton (2017). Dynamic norms promote sustainable behavior, even if it is counternormative. *Psychological Science 28*(11), 1663–1674.

Statista (2023). Global per capita tea consumption by country. Accessed: 2024-02-19.

Swait, J. and J. Louviere (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of marketing research 30*(3), 305–314.

Swiss International Air Lines (2024). Sustainable choices. Accessed: 2024-10-09.

Tea and Herbal Association (2024). Tea facts. `https://www.tea.co.uk/tea-facts`. Accessed: 2024-02-19.

Tea and Infusions Association (2022). UK Tea Census Report 2022. Accessed: 2024-02-19.

The Guardian (2023a). Delta air lines lawsuit carbon neutrality. https://www.theguardian.com/environment/2023/may/30/delta-air-lines-lawsuit-carbon-neutrality-aoe. Accessed: 2023-11-06.

The Guardian (2023b, January). Revealed: more than 90% of rainforest carbon offsets by biggest provider are worthless, analysis shows. Accessed: 2024-11-01.

Theotokis, A. and E. Manganari (2015). The impact of choice architecture on sustainable consumer behavior: The role of guilt. *Journal of Business Ethics 131*, 423–437.

Train, K. and M. Weeks (2005). Discrete choice models in preference space and willingness-to-pay space. In R. Scarpa and A. Alberini (Eds.), *Applications of Simulation Methods in Environmental and Resource Economics*, Volume 6 of *The Economics of Non-Market Goods and Resources*, pp. 1–17. Springer, Dordrecht.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press.

Trencher, G., S. Nick, J. Carlson, and M. Johnson (2024). Demand for low-quality offsets by major companies undermines climate integrity of the voluntary carbon market. *Nature Communications 15*(1), 6863.

Vecchio, R. and A. Annunziata (2015). Willingness-to-pay for sustainability-labelled chocolate: an experimental auction approach. *Journal of Cleaner Production 86*, 335–342.

Ziegler, A., J. Schwarzkopf, and V. H. Hoffmann (2012). Stated versus revealed knowledge: Determinants of offsetting CO_2 emissions from fuel consumption in vehicle use. *Energy Policy 40*, 422–431.

Appendix

A Survey Data

A.1 Descriptive Statistics

This section presents the descriptive statistics of the main survey data. Table A.1 includes the age, gender, education, employment, and income of the participants, while Table A.2 shows the socio-demographic comparison between tea drinkers/purchasers and never tea drinkers/purchasers. Table A.3 presents participants' tea consumption habits, including their frequency of tea drinking, purchasing habits, and preferences for tea blends. Table A.4 shows the survey clarity, attention checks, and completion time, while Table A.5 presents the levels of certainty, attribute non-attendance, consequentiality, and protest responses of participants. Table A.6 presents the level of agreement with various statements using a Likert scale. Finally, Table A.7 shows covariate balance.

	Full sample		Control		Treatment 1		Treatment 2	
	N	Share	N	Share	N	Share	N	Share
Sample size	1321		444		442		435	
Age								
18 - 34 years	410	31.04%	135	30.41%	134	30.32%	141	32.41%
35 - 54 years	469	35.50%	157	35.36%	158	35.75%	154	35.40%
55+ years	442	33.46%	152	34.23%	150	33.94%	140	32.18%
Gender								
Male	645	48.83%	227	51.13%	206	46.61%	212	48.74%
Female	666	50.42%	214	48.20%	231	52.26%	221	50.80%
Non-binary	9	0.68%	3	0.68%	5	1.13%	1	0.23%
I prefer not to say	1	0.08%	-	-	-	-	1	0.23%
Education								
Primary school	3	0.23%	2	0.45%	1	0.23%	-	-
Secondary school: High school or equivalent	359	27.18%	130	29.28%	124	28.05%	105	24.14%
Post-secondary vocational training (up to 1 year)	55	4.16%	16	3.60%	27	6.11%	12	2.76%
Post-secondary vocational training (2 and more years)	153	11.58%	44	9.91%	55	12.44%	54	12.41%
Post-secondary academic below-degree level qualification (up to 1 year)	80	6.06%	31	6.98%	23	5.20%	26	5.98%
Post-secondary academic below-degree level qualification (2 and more years)	171	12.94%	49	11.04%	50	11.31%	72	16.55%
Bachelors or equivalent first degree qualification (e.g., BA, BSc, BEng)	311	23.54%	111	25.00%	95	21.49%	105	24.14%
Masters or equivalent higher degree level qualification (e.g., MA, MSc, MBA)	140	10.60%	49	11.04%	47	10.63%	44	10.11%
PhD or equivalent doctoral level qualification (e.g., PhD)	39	2.95%	8	1.80%	19	4.30%	12	2.76%
None of above	10	0.76%	4	0.90%	1	0.23%	5	1.15%

This figure displays the number of participants (N) and their share.

Table A.1: Summary statistics of sociodemographics

	Full sample		Control		Treatment 1		Treatment 2	
	N	Share	N	Share	N	Share	N	Share
Sample size	1321		444		442		435	
Employment								
Full or part time employment	760	57.53%	248	55.86%	257	58.14%	255	58.62%
Self-employed	95	7.19%	32	7.21%	29	6.56%	34	7.82%
Unemployed	82	6.21%	24	5.41%	28	6.33%	30	6.90%
Retired	263	19.91%	90	20.27%	87	19.68%	86	19.77%
Looking after family or home	77	5.83%	30	6.76%	26	5.88%	21	4.83%
Full-time student	32	2.42%	14	3.15%	11	2.49%	7	1.61%
None of above	12	0.91%	6	1.35%	4	0.90%	2	0.46%
Annual household income after taxes								
Under £10,000	75	5.68%	23	5.18%	22	4.98%	30	6.90%
£10,000 - £19,999	177	13.40%	50	11.26%	62	14.03%	65	14.94%
£20,000 - £29,999	250	18.93%	85	19.14%	88	19.91%	77	17.70%
£30,000 - £39,999	174	13.17%	71	15.99%	48	10.86%	55	12.64%
£40,000 - £49,999	136	10.30%	43	9.68%	53	11.99%	40	9.20%
£50,000 - £59,999	103	7.80%	32	7.21%	34	7.69%	37	8.51%
£60,000 - £69,999	68	5.15%	24	5.41%	16	3.62%	28	6.44%
£70,000 - £79,999	49	3.71%	15	3.38%	18	4.07%	16	3.68%
£80,000 - £89,999	43	3.26%	10	2.25%	21	4.75%	12	2.76%
£90,000 - £99,999	31	2.35%	12	2.70%	9	2.04%	10	2.30%
£100,000 - £129,999	70	5.30%	22	4.95%	27	6.11%	21	4.83%
£130,000 or more	49	3.71%	16	3.60%	15	3.39%	18	4.14%
No answer	96	7.27%	41	9.23%	29	6.56%	26	5.98%
Environmental organization membership								
Member	149	11.28%	46	10.36%	55	12.44%	48	11.03%

This figure displays the number of participants (N) and their share.

Table A.1: Summary statistics of sociodemographics

	Non-tea consumers (N = 222)	**Tea consumers** (N = 1,339)	**SMD**
Descriptive statistics			
Age	Mean: 50.54 (SD: 16.78)	Mean: 47.54 (SD: 16.52)	0.181
Male, n (%)	111 (50.0%)	686 (51.2%)	0.024
Higher than post-secondary education (up to 1 year), n (%)	82 (36.9%)	664 (49.6%)	-0.254

This table displays descriptive statistics and standardized mean differences (SMDs) between tea consumers and non-tea consumers across age, gender, and education characteristics. Note that the tea consumer sample includes protest responses.

Table A.2: Comparison of tea consumers and non-tea consumers

	Full sample		Control		Treatment 1		Treatment 2	
	N	Mean	N	Mean	N	Mean	N	Mean
Sample size	1321		444		442		435	
Tea drinking frequency								
Daily	1027	77.74%	338	76.13%	343	77.60%	346	79.54%
Once a week	138	10.45%	52	11.71%	50	11.31%	36	8.28%
Once every two weeks	60	4.54%	18	4.05%	21	4.75%	21	4.83%
Once a month	35	2.65%	10	2.25%	10	2.26%	15	3.45%
Several times a year	61	4.62%	26	5.86%	18	4.07%	17	3.91%
Tea purchase frequency								
Once a week	252	19.08%	94	21.17%	83	18.78%	75	17.24%
Once every two weeks	271	20.51%	80	18.02%	98	22.17%	93	21.38%
Once a month	505	38.23%	173	38.96%	160	36.20%	172	39.54%
Several times a year	253	19.15%	83	18.69%	85	19.23%	85	19.54%
Once a year	40	3.03%	14	3.15%	16	3.62%	10	2.30%
Preferred tea blend								
Black tea	955	72.29%	317	71.40%	324	73.30%	314	72.18%
Green tea	212	16.05%	75	16.89%	62	14.03%	75	17.24%
Herbal tea	105	7.95%	36	8.11%	39	8.82%	30	6.90%
Other	49	3.71%	16	3.60%	17	3.85%	16	3.68%
Regular grocery shopping								
Carbon-neutral label	168	12.72%	54	12.16%	61	13.80%	53	12.18%
Organic label	321	24.30%	109	24.55%	111	25.11%	101	23.22%
Fair trade label	485	36.71%	169	38.06%	161	36.43%	155	35.63%
No sustainability label	249	18.85%	80	18.02%	81	18.33%	88	20.23%
Other label	26	1.97%	15	3.38%	4	0.90%	7	1.61%
No knowledge about label	391	29.60%	127	28.60%	131	29.64%	133	30.57%

This figure displays the number of participants (N), along with either their share for binary variables or their mean for continuous variables.

Table A.3: Summary statistics: tea consumption

	Full		Control		Treatment 1		Treatment 2	
	N	Mean	N	Mean	N	Mean	N	Mean
Sample size	1321		444		442		435	
Survey completion time								
Choice experiment time (min.)	1321	3.09	444	4.69	442	1.86	435	2.70
Full survey time (min.)	1321	14.61	444	16.14	442	11.41	435	16.31
Device								
Smartphone	741	56.09%	239	53.83%	260	58.82%	242	55.63%
Tablet	42	3.18%	12	2.70%	13	2.94%	17	3.91%
Desktop	538	40.73%	193	43.47%	169	38.24%	176	40.46%
Attention and manipulation								
The number of labels								
One label	65	4.92%	28	6.31%	20	4.52%	17	3.91%
Two labels	334	25.28%	103	23.20%	119	26.92%	112	25.75%
Three labels (correct resp.)	821	62.15%	281	63.29%	266	60.18%	274	62.99%
Not remember.	101	7.65%	32	7.21%	37	8.37%	32	7.36%
Percentage info. on label								
Yes (correct for samples 2 & 3)	567	42.41%	95	21.30%	242	54.14%	230	51.80%
No (correct for sample 1)	273	20.42%	163	36.55%	51	11.41%	59	13.29%
I do not remember.	497	37.17%	188	42.15%	154	34.45%	155	34.91%
Definition of CO_2 offsetting								
Correct resp.	650	49.21%	223	50.23%	225	50.90%	202	46.44%
False resp. (CO_2 reduction def.)	475	35.96%	156	35.14%	149	33.71%	170	39.08%
Not remember.	196	14.84%	65	14.64%	68	15.38%	63	14.48%
Survey clarity								
Clear instructions								
No.	9	0.68%	-	-	5	1.13%	4	0.92%
Yes.	1312	99.32%	444	100.00%	437	98.87%	431	99.08%
Confusion with survey								
No.	1232	93.26%	422	95.05%	415	93.89%	395	90.80%
Yes.	89	6.74%	22	4.95%	27	6.11%	40	9.20%

This figure displays the number of participants (N), along with either their share of the total sample for binary variables or the mean for continuous variables.

Table A.4: Summary statistics: survey clarity, and attention

	Full sample		Control		Treatment 1		Treatment 2	
	N	Share	N	Share	N	Share	N	Share (%)
Certainty level (0-10)								
0-2 (very uncertain)	10	0.76%	3	0.68%	5	1.13%	2	0.46%
3-5	119	9.01%	39	8.78%	42	9.50%	38	8.74%
6-8	627	47.46%	214	48.20%	198	44.80%	215	49.43%
9-10 (very certain)	565	42.77%	188	42.34%	197	44.57%	180	41.38%
Attribute non-attendance								
Not considered: CN label	440	33.31%	134	30.18%	154	34.84%	152	34.94%
Not considered: Organic label	450	34.07%	128	28.83%	154	34.84%	168	38.62%
Not considered: Ethical trade l.	315	23.85%	97	21.85%	108	24.43%	110	25.29%
Not considered: Price	266	20.14%	103	23.20%	73	16.52%	90	20.69%
Considered all attributes	400	30.28%	144	32.43%	138	31.22%	118	27.13%
Consequentiality								
Yes (policy and price impact).	706	53.44%	229	51.58%	253	57.24%	224	51.49%
No (no impact).	615	46.56%	215	48.42%	189	42.76%	211	48.51%
Consistent status-quo response								
Total protest responses	18	1.34%	4	0.89%	5	1.12%	9	2.03%
The products were too expensive.	10	0.75%	4	0.89%	3	0.67%	3	0.68%
I oppose one or more of the labels. [P]	3	0.22%	1	0.22%	-	-	2	0.45%
Insufficient information was provided about the labels or the products. [P]	7	0.52%	2	0.45%	2	0.45%	3	0.68%
I prefer to spend money on other social and environmental responsibility projects.	3	0.22%	1	0.22%	2	0.45%	-	-
I disagree with the way the choice question was asked. [P]	2	0.15%	-	-	-	-	2	0.45%
Other reason* [P]	6	0.45%	1	0.22%	3	0.67%	2	0.45%

This table displays the number of participants (N) and their share. Protest responses indicated by [P]. *This category includes protest responses based on various open-ended reasons. In addition to pre-specified categories, other responses were classified as protests if participants indicated that they consume only one brand, do not drink tea, or never purchase loose-leaf tea (possibly due to misinterpreting the question). One respondent mentioned that the options were expensive; however, since a pre-specified category already covered this reason, the response was coded accordingly and not classified as a protest.

Table A.5: Summary statistics: certainty, attribute non-attendance, consequentiality, and protest responses

Statement	SD	MD	SltD	N	SltA	MA	SA
I worry about climate change	5.98	5.45	7.57	13.02	28.08	22.79	17.11
Limited financial resources prevent me from buying climate-friendly products	5.83	4.69	8.55	16.73	23.54	21.88	18.77
Lack of time prevents me from buying climate-friendly products	16.35	13.02	19.53	24.30	12.94	8.71	5.15
My positive emotions increase when I choose climate-friendly products	8.02	6.43	9.31	31.87	22.71	13.40	8.25
I feel guilty when I buy conventional products	14.84	11.36	18.85	21.88	16.96	8.71	7.42
Most people approve of my choice of climate-friendly products	5.53	3.71	8.10	47.16	15.97	12.34	7.19
Producers are responsible for climate change mitigation costs	2.35	3.63	8.48	23.54	23.92	21.88	16.20
I trust carbon-neutral labels	5.90	6.06	12.64	31.26	25.44	13.55	5.15
I am confused about carbon-neutral labels	5.00	8.25	14.84	26.87	29.37	10.14	5.53
I am concerned about carbon offsets	5.07	4.69	13.63	30.81	25.28	11.73	8.78
Carbon offsetting reduces carbon emissions	5.60	4.92	12.49	31.11	27.93	12.41	5.53
Carbon offsetting allows producers to continue polluting	1.89	2.95	8.02	30.20	30.96	15.37	10.60
Carbon offsetting is a misleading sense of relief	2.57	1.67	6.28	27.18	32.10	19.15	11.05
Carbon offsetting is a form of greenwashing	1.74	1.97	6.43	33.61	28.16	15.82	12.26

Values indicate the percentage of respondents who selected each agreement level. SD = Strongly Disagree, MD = Mostly Disagree, SltD = Slightly Disagree, N = Neutral, SltA = Slightly Agree, MA = Mostly Agree, SA = Strongly Agree.

Table A.6: Agreement to the list of statements - full sample

Covariate	**Control (Mean)**	**Treatment 1 (Mean)**	**Treatment 2 (Mean)**	**C - T1 (SMD)**	**C - T2 (SMD)**	**T1 - T2 (SMD)**
Only confused with transparent labels	0.45	0.38	0.50	0.15	-0.09	-0.24
Only trust in transparent labels	0.57	0.58	0.55	-0.01	0.05	0.06
Confusion level	4.25	4.16	4.18	0.06	0.05	-0.01
Trust level	4.28	4.21	4.28	0.05	0.00	-0.05
Concern for CO_2 offsets	4.45	4.32	4.34	0.09	0.07	-0.01
Climate worry	4.96	4.79	4.91	0.10	0.03	-0.08
Warm glow	4.43	4.20	4.21	0.15	0.14	-0.00
Guilt	3.80	3.67	3.64	0.08	0.09	0.02
Social approval	4.44	4.29	4.17	0.11	0.20	0.09
Polluter pays	4.97	4.86	4.97	0.08	0.00	-0.07
Financial constraints	4.97	4.74	4.94	0.14	0.02	-0.12
Time restrictions	3.62	3.36	3.55	0.16	0.04	-0.11
Gender (male)	0.51	0.47	0.49	0.09	0.05	-0.04
Age	47.72	47.88	46.73	-0.01	0.06	0.07
High education	0.38	0.36	0.37	0.03	0.02	-0.01
Employed	0.63	0.65	0.66	-0.03	-0.07	-0.04
High income	0.33	0.34	0.35	-0.03	-0.05	-0.02
Not disclosed income	0.09	0.07	0.06	0.10	0.12	0.02

This table continues on the next page, where table notes are provided.

Table A.7: Covariate balance

Covariate	Control (Mean)	Treatment 1 (Mean)	Treatment 2 (Mean)	C - T1 (SMD)	C - T2 (SMD)	T1 - T2 (SMD)
Effectiveness of offsetting	4.37	4.18	4.35	0.14	0.01	-0.12
Putting a price on nature	4.79	4.73	4.70	0.05	0.07	0.02
Moral licensing	4.91	4.87	4.81	0.03	0.07	0.04
Offsetting as greenwashing	4.88	4.71	4.84	0.13	0.03	-0.10
Survey time	16.14	11.41	16.31	0.11	-0.00	-0.10
Familiarity with carbon neutral label	0.12	0.14	0.12	-0.05	-0.00	0.05
Manipulation checker	0.50	0.51	0.46	-0.01	0.08	0.09
Attention checker (number of labels)	0.63	0.60	0.63	0.06	0.01	-0.06
Attention checker (percentage info.)	0.37	0.55	0.52	-0.36	-0.32	0.05
Not consider carbon neutral label (ANA)	0.30	0.35	0.35	-0.10	-0.10	-0.00
Consequentiality	0.52	0.57	0.51	-0.11	0.00	0.12
Member of environmental organization	0.90	0.88	0.89	0.07	0.02	-0.04
Response certainty	8.03	8.00	8.05	0.02	-0.01	-0.03

SMD refers to the standardized mean difference between experimental groups. c refers to the control group; t1 refers to treatment 1; t2 refers to treatment 2. The first part of the table includes key variables used to test mechanisms, including: only confused with transparent labels (participants confused by at least one transparent label but not the standard label), only trust in transparent labels (participants who trust at least one transparent label but not the standard label), confusion level (degree of confusion about carbon-neutral labels), trust level (trust in carbon-neutral labels), concern for CO_2 offsets, climate worry (concern about climate change), warm glow (positive emotions from climate-friendly purchases), guilt (negative feelings when making non-climate-friendly choices), social approval (perceived social acceptance for climate-friendly behavior), polluter pays (belief that producers should bear climate mitigation costs), financial constraints (perceived financial constraints to climate-friendly purchases), and time restrictions (perceived lack of time to make climate-friendly choices). Among these, all except the two binary variables (only confused and only trust in transparent labels) are measured using 7-point Likert scales. Demographic covariates include gender, age, high education (post-secondary or higher), employment status, high income (annual household income above £50,000), and undisclosed income. The part of the table reports additional covariates included as robustness checks. Among them, the following are measured using 7-point Likert scales: perceived effectiveness of offsetting (belief that offsetting reduces emissions), putting a price on nature (perception that offsetting is about putting monetary value to nature), moral licensing (belief that offsetting allows companies to pollute without real effort to reduce), and offsetting as greenwashing (belief that offsetting is a form of greenwashing). Other robustness covariates include: survey duration, familiarity with carbon-neutral labels, manipulation checker, attention checkers (number of labels and percentage information), attribute non-attendance (ignoring the carbon-neutral label), consequentiality perception (whether participants believe their responses affect tea prices or label policies), environmental organization membership, and response certainty.

Table A.7: Covariate balance

A.2 Choice data

This section presents the choice experiment data. Table A.8 summarizes the overall choice design, including the different blocks, choice scenarios (cards), and respective attribute levels. Table A.9 details the specific choices available to participants within each block, presenting the number of alternatives (Alt A, Alt B, Alt SQ) and their respective distribution percentages for each choice card.

Block	Choice card	Price AltA	Price AltB	CN AltA	CN AltB	Organic AltA	Organic AltB	ET AltA	ET AltB
1	1	3.9	4.9	No	Yes	Yes	No	No	Yes
1	2	2.9	6.9	No	Yes	No	Yes	No	Yes
1	3	2.9	3.9	No	No	Yes	No	No	Yes
1	4	4.9	4.9	No	No	No	Yes	Yes	No
1	5	0.9	4.9	No	Yes	No	Yes	No	Yes
1	6	6.9	0.9	Yes	No	Yes	No	Yes	No
1	7	2.9	0.9	Yes	No	Yes	No	No	No
1	8	4.9	2.9	Yes	Yes	Yes	No	No	Yes
2	1	1.9	5.9	Yes	No	No	Yes	No	Yes
2	2	1.9	4.9	No	Yes	Yes	No	Yes	Yes
2	3	0.9	1.9	No	Yes	No	Yes	No	Yes
2	4	1.9	1.9	No	No	Yes	No	No	Yes
2	5	6.9	3.9	Yes	No	Yes	No	Yes	Yes
2	6	5.9	1.9	Yes	No	Yes	Yes	No	Yes
2	7	3.9	2.9	Yes	No	No	No	No	No
2	8	5.9	5.9	No	Yes	Yes	No	No	No

CN = Carbon-neutral label; ET = Ethical Trade label. Values in the table represent attribute levels shown on each choice card for Alternatives A and B. For the status quo option, all attribute levels take a value of 0 (i.e., the no-purchase option). Price values are in GBP. There were two blocks, shown in the first column.

Table A.8: Choice design

Block	Choice card	N Alt A	N Alt B	N Alt SQ	Total	AltA	AltB	AltSQ
1	1	319	189	168	676	47.19%	27.96%	24.85%
1	2	363	166	147	676	53.70%	24.56%	21.75%
1	3	403	172	101	676	59.62%	25.44%	14.94%
1	4	228	203	245	676	33.73%	30.03%	36.24%
1	5	309	252	115	676	45.71%	37.28%	17.01%
1	6	174	333	169	676	25.74%	49.26%	25.00%
1	7	337	284	55	676	49.85%	42.01%	8.14%
1	8	110	449	117	676	16.27%	66.42%	17.31%
2	1	433	155	75	663	65.31%	23.38%	11.31%
2	2	506	103	54	663	76.32%	15.54%	8.14%
2	3	204	420	39	663	30.77%	63.35%	5.88%
2	4	258	342	63	663	38.91%	51.58%	9.50%
2	5	123	313	227	663	18.55%	47.21%	34.24%
2	6	77	537	49	663	11.61%	81.00%	7.39%
2	7	149	296	218	663	22.47%	44.65%	32.88%
2	8	186	147	330	663	28.05%	22.17%	49.77%

This table reports the number of times each alternative was chosen in each choice card, as well as the share of respondents who selected each option. N Alt A, N Alt B, and N Alt SQ represent the number of times Alternatives A, B, and the status quo were chosen, respectively. Alt A, Alt B, and Alt SQ indicate their corresponding shares. Total refers to the total number of respondents per choice card. These results are based on the sample prior to the exclusion of protest responses.

Table A.9: Choices: detailed information

A.3 Robustness tests

This section presents robustness tests of the main findings. Preference-space estimations are reported in Section A.3.1, and begins with the MNL model estimates in Table A.10. Next, a MXL model where label attributes treated as random (Table A.11) and then with all coefficients, including price and the status-quo, treated as random (Table A.12). To check the mechanisms underlying consumer WTP, the model with interaction terms is estimated (Table A.13). Corresponding Poe statistics for WTP estimates are collected in Table A.14, and WTP point estimates derived from these preference-space models are collected in Table A.15.

Section A.3.2 repeats the same estimations directly in WTP space. It starts with a MNL model (Table A.16), proceeds to MXL (Table A.17), doubles the simulation precision with 2,000 Halton draws (Table A.19), and adds alternative-specific error components (Table A.20). Socio-economic interactions are added to the main MXL model in Table A.18; the main model without adjusting p-values for multiple hypothesis testing (Table A.24), decrease the interaction set for parsimony (Table A.25), introduce the warm-glow variable (Table A.26), and finally incorporate additional interaction variables (Table A.27). Holm-adjusted Poe tests for the the core WTP-space specifications are provided in Table A.28.

Robustness to alternative samples is examined by estimating (i) the raw data of 1,339 responses without excluding protest responses (Table A.21), (ii) a dataset that excludes survey speeders (Table A.22), and (iii) a strictly-screened sample that also removes every participant who failed any manipulation or attention check (Table A.23). Associated Poe comparisons for these three samples appear in Table A.29. The results of the open-ended contingent-valuation question with 10,000 bootstrapped WTP samples and Holm-corrected Poe tests are summarized in Table A.30.

Finally, Section A.4 explores how the presence or absence of organic and ethical-trade labels changes the valuation of the carbon-neutral label: interaction estimates are presented in Table A.31, and the implied scenario-specific WTPs with their Poe comparisons are laid out in Table A.32.

A.3.1 Preference space estimations

	Full sample	Control	Treatment 1	Treatment 2
Means of the choice parameters				
$\mu_{\text{Carbon neutral}}$	0.27***	0.31***	0.31***	0.20***
	(0.03)	(0.05)	(0.05)	(0.05)
μ_{Organic}	0.63***	0.61***	0.65***	0.63***
	(0.03)	(0.06)	(0.06)	(0.06)
$\mu_{\text{Ethical trade}}$	0.63***	0.64***	0.62***	0.62***
	(0.03)	(0.06)	(0.06)	(0.06)
μ_{Price}	-0.48***	-0.50***	-0.48***	-0.47***
	(0.02)	(0.03)	(0.03)	(0.03)
$\mu_{\text{Status quo}}$	-1.84***	-1.95***	-1.77***	-1.81***
	(0.06)	(0.11)	(0.10)	(0.10)
Log Likelihood	-9986.11	-3312.74	-3367.50	-3302.05
AIC	19982.23	6635.47	6745.01	6614.09
BIC	20018.56	6666.35	6775.86	6644.87
Pseudo R^2	0.14	0.15	0.13	0.14
Number of observations	10568	3552	3536	3480
Number of participants	1321	444	442	435

This table reports the results from a multinomial logit model estimated in preference space. Robust standard errors are shown in parentheses.

P-values (Holm adjusted for 5 comparisons) and are used to determine statistical significance: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$.

Table A.10: Multinomial logit model, preference space estimates

	Full sample	Control	Treatment 1	Treatment 2
Means of the choice parameters				
$\mu_{\text{Carbon neutral}}$	0.30***	0.37***	0.37***	0.18***
	(0.04)	(0.07)	(0.08)	(0.07)
μ_{Organic}	0.76***	0.74***	0.80***	0.74***
	(0.05)	(0.08)	(0.08)	(0.08)
$\mu_{\text{Ethical trade}}$	0.81***	0.84***	0.80***	0.79***
	(0.05)	(0.08)	(0.08)	(0.08)
μ_{Price}	-0.67***	-0.69***	-0.67***	-0.65***
	(0.02)	(0.04)	(0.04)	(0.04)
$\mu_{\text{Status quo}}$	-2.42***	-2.54***	-2.35***	-2.38***
	(0.08)	(0.14)	(0.13)	(0.14)
Standard deviations				
$\sigma_{\text{Carbon neutral}}$	1.18***	1.18***	1.27***	1.05***
	(0.06)	(0.10)	(0.11)	(0.12)
σ_{Organic}	1.22***	1.18***	1.24***	1.25***
	(0.06)	(0.10)	(0.09)	(0.10)
$\sigma_{\text{Ethical tade}}$	1.13***	1.18***	1.11***	1.13***
	(0.06)	(0.11)	(0.10)	(0.10)
Log Likelihood	-9218.19	-3063.44	-3098.03	-3051.75
AIC	18452.38	6142.88	6212.06	6119.50
BIC	18510.50	6192.28	6261.42	6168.74
Pseudo R^2	0.21	0.21	0.20	0.20
Number of observations	10568	3552	3536	3480
Number of participants	1321	444	442	435

This table reports the results from a mixed logit model estimated in preference space. Robust standard errors are shown in parentheses.
P-values (Holm adjusted for 5 comparisons) and are used to determine statistical significance: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$.

Table A.11: Mixed logit model, preference space estimates, label parameters randomized

	Full sample	Control	Treatment 1	Treatment 2
Means of the choice parameters				
$\mu_{\text{Carbon neutral}}$	0.46***	0.59***	0.50***	0.34***
	(0.05)	(0.08)	(0.08)	(0.08)
μ_{Organic}	0.97***	0.99***	0.97***	0.96***
	(0.05)	(0.09)	(0.09)	(0.09)
$\mu_{\text{Ethical trade}}$	1.01***	1.10***	0.97***	1.01***
	(0.05)	(0.10)	(0.09)	(0.09)
μ_{Price}	-0.23***	-0.13*	-0.25***	-0.30***
	(0.05)	(0.08)	(0.08)	(0.09)
$\mu_{\text{Status quo}}$	-5.19***	-5.93***	-4.66***	-5.41***
	(0.25)	(0.56)	(0.37)	(0.49)
Standard deviations				
$\sigma_{\text{Carbon neutral}}$	0.83***	0.88***	0.88***	0.68***
	(0.08)	(0.14)	(0.12)	(0.14)
σ_{Organic}	1.00***	1.00***	0.99***	1.04***
	(0.07)	(0.11)	(0.11)	(0.12)
$\sigma_{\text{Ethical trade}}$	0.84***	0.93***	0.89***	0.78***
	(0.08)	(0.14)	(0.14)	(0.13)
σ_{Price}	0.92***	0.93***	0.86***	0.98***
	(0.04)	(0.08)	(0.06)	(0.07)
$\sigma_{\text{Status quo}}$	3.24***	3.78***	2.92***	3.37***
	(0.21)	(0.41)	(0.28)	(0.40)
Log Likelihood	-8248.47	-2697.14	-2811.60	-2726.17
AIC	16516.94	5414.28	5643.19	5472.34
BIC	16589.60	5476.03	5704.90	5533.89
Pseudo R^2	0.29	0.31	0.28	0.29
Number of observations	10568	3552	3536	3480
Number of participants	1321	444	442	435

This table reports the results from a mixed logit model estimated in preference space. Robust standard errors are shown in parentheses.
P-values (Holm adjusted for 5 comparisons) and are used to determine statistical significance: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$.

Table A.12: Mixed logit model, preference space estimates, all parameters randomized

	Full sample	Control	Treatment 1	Treatment 2
$\mu_{\text{Carbon neutral}}$	0.00	-0.03	0.01	-0.01
	(0.12)	(0.21)	(0.21)	(0.23)
Main interactions				
$\mu_{\text{Carbon neutral}}$ * Only confused with transparent labels	-0.19	-0.13	-0.08	-0.31
	(0.09)	(0.17)	(0.17)	(0.15)
$\mu_{\text{Carbon neutral}}$ * Only trust in transparent labels	0.40***	0.51**	0.19	0.60***
	(0.10)	(0.17)	(0.18)	(0.17)
$\mu_{\text{Carbon neutral}}$ * Concern for CO_2 offsets	0.07	-0.06	0.18*	0.09
	(0.04)	(0.07)	(0.06)	(0.07)
$\mu_{\text{Carbon neutral}}$ * Trust level	0.04	0.10	0.07	-0.08
	(0.04)	(0.06)	(0.06)	(0.07)
$\mu_{\text{Carbon neutral}}$ * Confusion level	-0.05	0.00	-0.05	-0.16*
	(0.03)	(0.06)	(0.06)	(0.06)
$\mu_{\text{Carbon neutral}}$ * Climate worry	0.12**	0.13	0.02	0.25***
	(0.03)	(0.06)	(0.06)	(0.06)
$\mu_{\text{Carbon neutral}}$ * Guilt	0.07	0.11	0.04	0.07
	(0.04)	(0.06)	(0.07)	(0.06)
$\mu_{\text{Carbon neutral}}$ * Social approval	0.13**	0.18	0.16	0.01
	(0.04)	(0.08)	(0.07)	(0.07)
$\mu_{\text{Carbon neutral}}$ * Polluter pays	-0.04	-0.02	-0.09	0.03
	(0.03)	(0.06)	(0.05)	(0.06)
$\mu_{\text{Carbon neutral}}$ * Financial constraints	-0.08*	-0.10	-0.07	-0.13
	(0.03)	(0.06)	(0.05)	(0.05)
$\mu_{\text{Carbon neutral}}$ * Time restrictions	-0.08	-0.10	-0.09	-0.05
	(0.03)	(0.06)	(0.06)	(0.05)
Socioeconomic interactions	Yes	Yes	Yes	Yes
Means of the sustainability label attributes, price, status quo	Yes	Yes	Yes	Yes
Standard deviations of random parameters	Yes	Yes	Yes	Yes
Log Likelihood	-8031.02	-2614.74	-2739.11	-2640.57
AIC	16116.05	5283.47	5532.23	5335.15
BIC	16312.22	5450.20	5698.84	5501.33
Pseudo R^2	0.31	0.33	0.29	0.31
Number of Observations	10568	3552	3536	3480
Number of Participants	1321	444	442	435

P-values (Holm adjusted for 5 comparisons) and are used to determine statistical significance: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$.
All parameters, except for the interaction terms, are randomized. Holm-Bonferroni adjusted p-values (for 11 comparisons) and significance levels: ***$p \leq 0.01$, **$0.01 < p \leq 0.05$, *$0.05 < p \leq 0.1$. Bilateral Wald tests were conducted to assess the equivalence of interaction coefficients across experimental groups. Holm-Bonferroni adjusted p-values (for 21 Wald test comparisons), covering all main interactions, socioeconomic variables, WTP parameters, and the status quo parameter, were used to account for multiple testing. Two interaction effects remained statistically significant: concern for carbon offsets (control vs. treatment 1, $p = 0.033$) and climate worry (treatment 1 vs. treatment 2, $p = 0.020$). Note that Wald test results for coefficients not reported in the table (e.g., socioeconomic variables) are also not included in the table notes.

Table A.13: Mixed logit model, preference space estimates with interaction variables

WTP estimates derived from MNL model in preference space (Table A.10)				
	Full sample	Control	Treatment 1	Treatment 2
$WTP_{\text{Carbon neutral}}$	0.56*** (0.06)	0.61*** (0.11)	0.66*** (0.11)	0.42*** (0.10)
Poe test: Control vs Treatment 1		0.61 (p = 0.78)		
Poe test: Control vs Treatment 2		0.09 (p = 0.37)		
Poe test: Treatment 1 vs Treatment 2			0.06 (p = 0.34)	
WTP estimates derived from MXL model in preference space, label attributes as random parameters (Table A.11)				
	Full sample	Control	Treatment 1	Treatment 2
$WTP_{\text{carbon neutral}}$	0.45*** (0.06)	0.54*** (0.10)	0.56*** (0.11)	0.27*** (0.10)
Poe test: Control vs Treatment 1		0.53 (p = 0.93)		
Poe test: Control vs Treatment 2		0.03 (p = 0.16)		
Poe test: Treatment 1 vs Treatment 2			0.03 (p = 0.16)	
WTP estimates derived from MXL model in preference space, , all parameters random (Table A.12)				
	Full sample	Control	Treatment 1	Treatment 2
$WTP_{\text{carbon neutral}}$	0.58*** (0.06)	0.67*** (0.10)	0.65*** (0.11)	0.45*** (0.11)
Poe test: Control vs Treatment 1		0.16 (p = 0.62)		
Poe test: Control vs Treatment 2			0.07 (p = 0.40)	
Poe test: Treatment 1 vs Treatment 2				0.16 (p = 0.62)

WTP estimates are reported with standard errors in parentheses.
Significance levels: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$
Poe test statistics are provided with p-values in parentheses, which are Holm-Bonferroni adjusted for three comparisons.

Table A.14: Poe test results, preference space estimates

WTP estimates derived from MNL model in preference space (Table A.10)				
	Full sample	Control	Treatment 1	Treatment 2
$WTP_{\text{carbon neutral}}$	0.56*** (0.06)	0.61*** (0.10)	0.65*** (0.11)	0.42*** (0.10)
WTP_{organic}	1.31*** (0.06)	1.22*** (0.11)	1.37*** (0.11)	1.33*** (0.12)
$WTP_{\text{ethical trade}}$	1.29*** (0.07)	1.28*** (0.11)	1.29*** (0.12)	1.31*** (0.12)
WTP estimates derived from MXL in preference space, label attributes as random parameters (Table A.11)				
	Full sample	Control	Treatment 1	Treatment 2
$WTP_{\text{Carbon neutral}}$	0.45*** (0.06)	0.54*** (0.10)	0.56*** (0.11)	0.27*** (0.10)
WTP_{Organic}	1.13*** (0.06)	1.08*** (0.11)	1.19*** (0.11)	1.14*** (0.12)
$WTP_{\text{Ethical trade}}$	1.20*** (0.06)	1.22*** (0.11)	1.19*** (0.11)	1.21*** (0.11)
WTP estimates derived from MXL model in preference space, all attributes and status quo as random p. (Table A.12)				
	Full sample	Control	Treatment 1	Treatment 2
$WTP_{\text{Carbon neutral}}$	0.58*** (0.06)	0.67*** (0.10)	0.65*** (0.11)	0.45*** (0.11)
WTP_{Organic}	1.21*** (0.07)	1.13*** (0.10)	1.25*** (0.12)	1.29*** (0.13)
$WTP_{\text{Ethical trade}}$	1.27*** (0.07)	1.26*** (0.11)	1.25*** (0.12)	1.35*** (0.13)
Number of participants	1321	444	442	435

This table shows the WTP estimated based on choice model outputs shown in Tables A.10, A.11, A.12.
Standard errors are reported in parentheses.
For the model shown in Tables A.10 and A.11, the Delta method is used to derive the standard errors, while for the model shown in Table A.12, the Krinsky-Robb simulation with 10,000 draws is used.
P-values (Holm adjusted for 3 comparisons) and are used to determine statistical significance: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$.

Table A.15: WTP estimates

A.3.2 WTP space estimations

	Full sample	Control	Treatment 1	Treatment 2
Means of the choice parameters				
$\text{WTP}_{\text{Carbon neutral}}$	0.56***	0.61***	0.65***	0.42***
	(0.06)	(0.10)	(0.11)	(0.10)
$\text{WTP}_{\text{Organic}}$	1.31***	1.22***	1.37***	1.33***
	(0.06)	(0.11)	(0.11)	(0.12)
$\text{WTP}_{\text{Ethical trade}}$	1.29***	1.28***	1.29***	1.31***
	(0.07)	(0.11)	(0.12)	(0.12)
Means of the choice parameters				
μ_{Price}	-0.48***	-0.50***	-0.48***	-0.47***
	(0.02)	(0.03)	(0.03)	(0.03)
$\mu_{\text{Status quo}}$	-1.84***	-1.95***	-1.77***	-1.81***
	(0.06)	(0.11)	(0.10)	(0.10)
Log Likelihood	-9986.11	-3312.74	-3367.50	-3302.05
AIC	19982.23	6635.47	6745.01	6614.09
BIC	20018.56	6666.35	6775.86	6644.87
Pseudo R^2	0.14	0.15	0.13	0.14
Number of observations	10568	3552	3536	3480
Number of participants	1321	444	442	435

This table shows the choice model output from the multinomial logit model estimated in WTP-space.
Robust standard errors are reported in parentheses.
P-values (Holm adjusted for 4 comparisons) and are used to determine statistical significance: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$.

Table A.16: Multinomial logit model, WTP-space estimates

	Full sample	Control	Treatment 1	Treatment 2
MWTP estimates				
$\text{WTP}_{\text{Carbon neutral}}$	0.45***	0.54***	0.56***	0.27***
	(0.06)	(0.10)	(0.11)	(0.10)
$\text{WTP}_{\text{Organic}}$	1.13***	1.08***	1.19***	1.14***
	(0.06)	(0.11)	(0.11)	(0.12)
$\text{WTP}_{\text{Ethical trade}}$	1.20***	1.22***	1.19***	1.21***
	(0.06)	(0.11)	(0.11)	(0.11)
Means of the choice parameters				
μ_{Price}	-0.67***	-0.69***	-0.67***	-0.65***
	(0.02)	(0.04)	(0.04)	(0.04)
$\mu_{\text{Status quo}}$	-2.42***	-2.54***	-2.35***	-2.38***
	(0.08)	(0.14)	(0.13)	(0.14)
Standard deviations				
$\sigma_{\text{Carbon neutral}}$	1.75***	1.72***	1.89***	1.61***
	(0.10)	(0.17)	(0.18)	(0.19)
σ_{Organic}	1.81***	1.70***	1.84***	1.92***
	(0.09)	(0.16)	(0.16)	(0.17)
$\sigma_{\text{Ethical trade}}$	1.69***	1.71***	1.65***	1.73***
	(0.09)	(0.16)	(0.16)	(0.16)
Log Likelihood	-9218.19	-3063.44	-3098.03	-3051.75
AIC	18452.38	6142.88	6212.06	6119.50
BIC	18510.50	6192.28	6261.42	6168.74
Pseudo R^2	0.21	0.21	0.20	0.20
Number of observations	10568	3552	3536	3480
Number of participants	1321	444	442	435

This table shows the choice model output from the MXL model estimated in WTP-space.
Robust standard errors are reported in parentheses.
P-values (Holm adjusted for 4 comparisons) and are used to determine statistical significance: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$.

Table A.17: Mixed logit model, WTP space estimates

	Full sample	Control	Treatment 1	Treatment 2
Socioeconomic interactions				
$WTP_{Carbon\ neutral}$ x Male	0.01	-0.09	0.41	-0.15
	(0.12)	(0.23)	(0.21)	(0.23)
$WTP_{Carbon\ neutral}$ x Age	-0.01	-0.00	-0.02**	-0.01
	(0.00)	(0.01)	(0.01)	(0.01)
$WTP_{Carbon\ neutral}$ x High education	0.55***	0.63*	0.76**	0.40
	(0.13)	(0.25)	(0.28)	(0.27)
$WTP_{Carbon\ neutral}$ x Employed	-0.09	-0.06	0.01	-0.08
	(0.13)	(0.27)	(0.25)	(0.26)
$WTP_{Carbon\ neutral}$ x High income	0.75***	0.72**	0.50	1.24***
	(0.13)	(0.25)	(0.24)	(0.25)
$WTP_{Carbon\ neutral}$ x Not disclosed income	-0.29	-0.68*	0.16	-0.47
	(0.16)	(0.27)	(0.40)	(0.79)
MWTP estimates				
$WTP_{Carbon\ neutral}$	0.13	0.29	-0.09	-0.03
	(0.12)	(0.20)	(0.22)	(0.24)
$WTP_{Organic}$	1.07***	1.05***	1.16***	1.10***
	(0.06)	(0.09)	(0.10)	(0.12)
$WTP_{Ethical\ trade}$	1.16***	1.19***	1.16***	1.12***
	(0.06)	(0.10)	(0.12)	(0.11)
Means of the choice parameters				
$\mu_{Status\ quo}$	-4.23***	-4.63***	-3.93***	-4.22***
	(0.17)	(0.31)	(0.26)	(0.29)
μ_{Price}	-0.17***	-0.11*	-0.18***	-0.21***
	(0.04)	(0.07)	(0.06)	(0.07)
Random coefficients included	Yes	Yes	Yes	Yes
Log Likelihood	-8480.40	-2790.12	-2869.27	-2800.80
AIC	16992.80	5612.25	5770.55	5633.59
BIC	17109.05	5711.05	5869.28	5732.07
Pseudo R^2	0.27	0.28	0.26	0.27
Number of observations	10568	3552	3536	3480
Number of participants	1321	444	442	435

This table shows the choice model output from the MXL model estimated in WTP-space.
Robust standard errors are reported in parentheses.
P-values (Holm adjusted for 10 comparisons) and are used to determine statistical significance: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$.
The equivalence of interaction coefficients across experimental groups was tested using bilateral Wald tests. For the comparison between treatment group 1 and treatment group 2, the interaction with *high income* remains marginally significant after applying a Holm-Bonferroni correction for six comparisons ($p = 0.091$).

Table A.18: Mixed logit model, WTP space estimates

	Full sample	Control	Treatment 1	Treatment 2
MWTP estimates				
$WTP_{Carbon\ neutral}$	0.45***	0.53***	0.54***	0.25***
	(0.06)	(0.09)	(0.10)	(0.08)
$WTP_{Organic}$	1.09***	1.03***	1.16***	1.11***
	(0.05)	(0.09)	(0.10)	(0.14)
$WTP_{Ethical\ trade}$	1.16***	1.15***	1.17***	1.17***
	(0.06)	(0.09)	(0.12)	(0.13)
Means of the choice parameters				
μ_{Price}	-0.17***	-0.11	-0.20***	-0.21***
	(0.04)	(0.07)	(0.07)	(0.08)
$\mu_{Status\ quo}$	-4.36***	-4.82***	-3.96***	-4.36***
	(0.18)	(0.36)	(0.26)	(0.33)
$\sigma_{Carbon\ neutral}$	1.17***	1.13***	1.44***	0.98***
	(0.22)	(0.26)	(0.22)	(0.09)
Standard deviations				
$\sigma_{Organic}$	1.58***	1.47***	1.58***	1.67***
	(0.08)	(0.14)	(0.13)	(0.27)
$\sigma_{Ethical\ trade}$	1.47***	1.44***	1.45***	1.49***
	(0.09)	(0.13)	(0.15)	(0.21)
σ_{Price}	0.82***	0.81***	0.72***	0.85***
	(0.06)	(0.08)	(0.07)	(0.07)
$\sigma_{Status\ quo}$	2.30***	2.61***	2.08***	2.32***
	(0.13)	(0.25)	(0.20)	(0.24)
Log Likelihood	-8536.84	-2808.48	-2898.52	-2823.37
AIC	17093.68	5636.96	5817.04	5666.73
BIC	17166.34	5698.71	5878.75	5728.28
Pseudo R^2	0.26	0.28	0.25	0.26
Number of Observations	10568	3552	3536	3480
Number of Participants	1321	444	442	435

This table reports the results from a MXL model estimated in WTP space.
Robust standard errors are shown in parentheses.
P-values (Holm adjusted for 4 comparisons) and are used to determine statistical significance: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$.
2,000 Halton draws are used.

Table A.19: Mixed logit model, WTP space estimates, 2000 Halton draws

	Full sample	Control	Treatment 1	Treatment 2
MWTP estimates				
$\text{WTP}_{\text{Carbon neutral}}$	0.47***	0.55***	0.52***	0.32***
	(0.11)	(0.08)	(0.10)	(0.08)
$\text{WTP}_{\text{Organic}}$	1.10***	1.04***	1.15***	1.14***
	(0.08)	(0.09)	(0.11)	(0.10)
$\text{WTP}_{\text{Ethical trade}}$	1.17***	1.16***	1.14***	1.23***
	(0.07)	(0.11)	(0.12)	(0.11)
Means of the choice parameters				
μ_{Price}	-0.17***	-0.11	-0.18***	-0.21***
	(0.04)	(0.07)	(0.07)	(0.07)
$\mu_{\text{Status quo}}$	-4.26***	-4.73***	-3.97***	-4.29***
	(0.19)	(0.33)	(0.27)	(0.29)
Standard deviations				
$\sigma_{\text{Carbon neutral}}$	1.24***	1.31***	1.37***	1.25***
	(0.47)	(0.12)	(0.25)	(0.12)
σ_{Organic}	1.64***	1.41***	1.57***	1.68***
	(0.10)	(0.14)	(0.14)	(0.11)
$\sigma_{\text{Ethical trade}}$	1.48***	1.40***	1.43***	1.47***
	(0.12)	(0.19)	(0.15)	(0.15)
σ_{Price}	0.77***	0.78***	0.74***	0.82***
	(0.07)	(0.07)	(0.09)	(0.08)
$\sigma_{\text{Status quo}}$	2.27***	2.49***	2.13***	2.22***
	(0.13)	(0.24)	(0.19)	(0.19)
Alternative specific error components				
$\sigma_{\text{Alternative 1}}$	0.02	-0.03	-0.12	0.13
	(0.05)	(0.05)	(0.15)	(0.14)
$\sigma_{\text{Alternative 2}}$	-0.00	-0.05	-0.11	0.06
	(0.10)	(0.06)	(0.18)	(0.18)
Log Likelihood	-8540.42	-2808.50	-2898.05	-2822.61
AIC	17104.84	5641.01	5820.10	5669.21
BIC	17192.03	5715.11	5894.15	5743.07
Pseudo R^2	0.26	0.28	0.25	0.26
Number of observations	10568	3552	3536	3480
Number of participants	1321	444	442	435

This table reports the results from the mixed logit model estimated in WTP space.
Robust standard errors are shown in parentheses.
P-values (Holm adjusted for 4 comparisons) and are used to determine statistical significance: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$.

Table A.20: Mixed logit model, WTP space estimates, alternative specific error components

	Full sample	Control	Treatment 1	Treatment 2
MWTP estimates				
$\text{WTP}_{\text{Carbon neutral}}$	0.44***	0.53***	0.48***	0.24***
	(0.06)	(0.09)	(0.09)	(0.08)
$\text{WTP}_{\text{Organic}}$	1.07***	1.06***	1.04***	1.13***
	(0.06)	(0.10)	(0.09)	(0.10)
$\text{WTP}_{\text{Ethical trade}}$	1.13***	1.16***	1.09***	1.15***
	(0.06)	(0.11)	(0.10)	(0.09)
Means of the choice parameters				
μ_{Price}	-0.15***	-0.11	-0.18**	-0.17**
	(0.04)	(0.07)	(0.07)	(0.08)
$\mu_{\text{Status quo}}$	-4.42***	-4.78***	-4.03***	-4.56***
	(0.19)	(0.33)	(0.26)	(0.38)
Standard deviations				
$\sigma_{\text{Carbon neutral}}$	1.21***	1.25***	1.50***	-0.74***
	(0.16)	(0.19)	(0.25)	(0.21)
σ_{Organic}	1.57***	1.50***	1.54***	1.71***
	(0.09)	(0.16)	(0.08)	(0.09)
$\sigma_{\text{Ethical trade}}$	1.36***	1.41***	1.38***	1.55***
	(0.09)	(0.16)	(0.17)	(0.09)
σ_{Price}	0.85***	0.80***	0.81***	0.97***
	(0.05)	(0.08)	(0.08)	(0.11)
$\sigma_{\text{Status quo}}$	2.49***	2.64***	2.22***	2.74***
	(0.14)	(0.24)	(0.22)	(0.26)
Log Likelihood	-8609.12	-2827.09	-2918.09	-2855.94
AIC	17238.24	5674.18	5856.17	5731.87
BIC	17311.03	5736.02	5917.99	5793.62
Pseudo R^2	0.27	0.28	0.26	0.27
Number of observations	10712	3584	3576	3552
Number of participants	1339	448	447	444

This table reports the results from a MXL model estimated in WTP space.
Robust standard errors are shown in parentheses.
P-values (Holm adjusted for 4 comparisons) and are used to determine statistical significance: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$..

Table A.21: Mixed logit model, WTP space estimates, the sample including protest responses

	Full sample	Control	Treatment 1	Treatment 2
MWTP estimates				
$WTP_{Carbon\ neutral}$	0.47***	0.56***	0.51***	0.31***
	(0.05)	(0.09)	(0.13)	(0.11)
$WTP_{Organic}$	1.10***	1.10***	1.12***	1.12***
	(0.07)	(0.11)	(0.11)	(0.12)
$WTP_{Ethical\ trade}$	1.14***	1.15***	1.09***	1.15***
	(0.07)	(0.11)	(0.11)	(0.11)
Means of the choice parameters				
μ_{Price}	-0.11***	-0.04	-0.15**	-0.14*
	(0.04)	(0.07)	(0.07)	(0.08)
$\mu_{Status\ quo}$	-4.30***	-4.73***	-3.96***	-4.30***
	(0.17)	(0.37)	(0.29)	(0.33)
Standard deviations				
$\sigma_{Carbon\ neutral}$	1.24***	1.18***	1.42***	1.17**
	(0.12)	(0.22)	(0.48)	(0.47)
$\sigma_{Organic}$	1.57***	1.40***	1.59***	1.64***
	(0.09)	(0.14)	(0.18)	(0.18)
$\sigma_{Ethical\ trade}$	1.44***	1.39***	1.31***	1.33***
	(0.12)	(0.16)	(0.23)	(0.18)
σ_{Price}	0.76***	0.72***	0.70***	0.80***
	(0.05)	(0.10)	(0.10)	(0.08)
$\sigma_{Status\ quo}$	2.28***	2.62***	2.05***	2.30***
	(0.13)	(0.25)	(0.23)	(0.24)
Log Likelihood	-8609.12	-2827.09	-2918.09	-2855.94
AIC	17238.24	5674.18	5856.17	5731.87
BIC	17311.03	5736.02	5917.99	5793.62
Pseudo R^2	0.27	0.28	0.26	0.27
Number of observations	10712	3584	3576	3552
Number of participants	1339	448	447	444

This table reports the results from a MXL model estimated in WTP space.
Robust standard errors are shown in parentheses.
P-values (Holm adjusted for 4 comparisons) and are used to determine statistical significance: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$.
Protest responses and survey speeders are excluded.

Table A.22: Mixed logit model, WTP space estimates. all parameters randomized

	Full sample	Control	Treatment 1	Treatment 2
MWTP estimates				
$WTP_{Carbon\ neutral}$	0.87***	0.97***	1.05***	0.67***
	(0.13)	(0.27)	(0.24)	(0.20)
$WTP_{Organic}$	1.14***	1.39***	0.92***	1.24***
	(0.14)	(0.29)	(0.22)	(0.26)
$WTP_{Ethical\ trade}$	1.17***	1.68***	0.75***	1.21***
	(0.14)	(0.34)	(0.23)	(0.24)
Means of the choice parameters				
μ_{Price}	0.05	0.22*	-0.02	-0.00
	(0.08)	(0.13)	(0.12)	(0.16)
$\mu_{Status\ quo}$	-4.19***	-4.65***	-3.80***	-4.74***
	(0.40)	(0.92)	(0.56)	(0.87)
Standard deviations				
$\sigma_{Carbon\ neutral}$	1.59***	1.46***	1.56***	1.84***
	(0.22)	(0.38)	(0.31)	(0.42)
$\sigma_{Organic}$	1.65***	1.76***	1.26***	1.68***
	(0.17)	(0.31)	(0.33)	(0.18)
$\sigma_{Ethical\ trade}$	1.32***	1.74***	1.16***	0.95*
	(0.16)	(0.46)	(0.35)	(0.52)
σ_{Price}	0.53***	0.31	0.38**	0.81***
	(0.11)	(0.20)	(0.18)	(0.27)
$\sigma_{Status\ quo}$	2.17***	2.57***	1.79***	2.45***
	(0.31)	(0.59)	(0.45)	(0.50)
Log Likelihood	-1423.90	-365.32	-524.64	-520.53
AIC	2867.79	750.64	1069.27	1061.07
BIC	2922.84	792.70	1114.01	1106.17
Pseudo R^2	0.29	0.33	0.26	0.29
Number of observations	1816	496	648	672
Number of participants	227	62	81	84

This table reports the results from a MXL model estimated in WTP space.
Robust standard errors are shown in parentheses.
P-values (Holm adjusted for 4 comparisons) and are used to determine statistical significance: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$.
Protest responses, survey speeders and participants who fail at least one attention or manipulation tests are excluded.

Table A.23: Mixed logit model, WTP space estimates, excluding protest responses, survey speeders, and attention/manipulation test failers

	Full sample	**Control**	**Treatment 1**	**Treatment 2**
$WTP_{\text{Carbon neutral}}$	-0.26	-0.43	-0.25	-0.19
	(0.16)	(0.29)	(0.26)	(0.29)
Main interactions				
$WTP_{\text{Carbon neutral}}$ x Only confused with transparent labels	0.03	0.21	0.24	-0.32
	(0.15)	(0.20)	(0.19)	(0.24)
$WTP_{\text{Carbon neutral}}$ x Only trust in transparent labels	0.65***	0.80***	0.39*	0.72***
	(0.14)	(0.23)	(0.20)	(0.23)
$WTP_{\text{Carbon neutral}}$ x Concern for CO_2 offsets	0.18***	0.08	0.30***	0.15
	(0.05)	(0.10)	(0.08)	(0.18)
$WTP_{\text{Carbon neutral}}$ x Trust level	0.09**	0.1	0.12**	-0.05
	(0.05)	(0.08)	(0.07)	(0.09)
$WTP_{\text{Carbon neutral}}$ x Confusion level	-0.03	0.01	-0.00	-0.16**
	(0.06)	(0.06)	(0.06)	(0.08)
$WTP_{\text{Carbon neutral}}$ x Climate worry	0.06	0.01	-0.05	0.25
	(0.05)	(0.10)	(0.08)	(0.22)
$WTP_{\text{Carbon neutral}}$ x Guilt	0.12***	0.15***	0.11	0.13
	(0.05)	(0.07)	(0.08)	(0.13)
$WTP_{\text{Carbon neutral}}$ x Social approval	0.18***	0.29***	0.21**	0.00
	(0.05)	(0.11)	(0.09)	(0.09)
$WTP_{\text{Carbon neutral}}$ x Polluter pays	-0.02	-0.01	-0.12*	0.03
	(0.04)	(0.09)	(0.06)	(0.09)
$WTP_{\text{Carbon neutral}}$ x Financial constraints	-0.18***	-0.16	-0.15**	-0.22**
	(0.04)	(0.11)	(0.07)	(0.09)
$WTP_{\text{Carbon neutral}}$ x Time restrictions	-0.02	-0.08	-0.03	0.06
	(0.04)	(0.07)	(0.06)	(0.06)
Socioeconomic interactions	Yes	Yes	Yes	Yes
WTP for other sustainability label attributes, price, status quo	Yes	Yes	Yes	Yes
Standard deviations of random parameters	Yes	Yes	Yes	Yes
Log Likelihood	-8371.60	-2746.65	-2835.84	-2756.99
AIC	16797.20	5547.30	5725.68	5567.97
BIC	16993.37	5714.04	5892.29	5734.15
Pseudo R^2	0.28	0.30	0.27	0.28
Number of Observations	1056	3552	3536	3480
Number of Participants	1321	444	442	435

This table shows the choice model output from the MXL model estimated in WTP-space.
Robust standard errors are reported in parentheses.
P-values and significance levels: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$

Table A.24: Mixed logit model, WTP space estimates, with interaction variables (unadjusted for multiple hypothesis testing)

	Full sample	Control	Treatment 1	Treatment 2
Interactions				
$WTP_{\text{Carbon neutral}}$ x Only confused with transparent labels	-0.01	0.24	0.21	-0.35
	(0.11)	(0.20)	(0.21)	(0.17)
$WTP_{\text{Carbon neutral}}$ x Only trust in transparent labels	0.77***	1.04***	0.57*	0.77***
	(0.12)	(0.21)	(0.21)	(0.19)
$WTP_{\text{Carbon neutral}}$ x Concern for CO_2 offsets	0.31***	0.25***	0.35***	0.30***
	(0.04)	(0.07)	(0.07)	(0.07)
MWTP estimates				
$WTP_{\text{Carbon neutral}}$	-0.37*	-0.55	-0.45	-0.25
	(0.14)	(0.29)	(0.26)	(0.25)
WTP_{Organic}	1.09***	1.03***	1.17***	1.10***
	(0.06)	(0.10)	(0.11)	(0.12)
$WTP_{\text{Ethical trade}}$	1.17***	1.17***	1.19***	1.16***
	(0.06)	(0.13)	(0.11)	(0.10)
Means of the choice parameters				
μ_{Price}	-0.18***	-0.12*	-0.18***	-0.21***
	(0.04)	(0.07)	(0.06)	(0.08)
$\mu_{\text{Status quo}}$	-4.20***	-4.63***	-3.90***	-4.22***
	(0.16)	(0.32)	(0.26)	(0.29)
Socioeconomic interactions	Yes	Yes	Yes	Yes
Standard deviations of random parameters	Yes	Yes	Yes	Yes
Log Likelihood	-8413.84	-2768.77	-2848.03	-2777.06
AIC	16865.68	5575.54	5734.06	5592.12
BIC	17003.73	5692.87	5851.30	5709.06
Pseudo R^2	0.28	0.29	0.27	0.27
Number of Observations	10568	3552	3536	3480
Number of Participants	1321	444	442	435

This table shows the choice model output from the MXL model estimated in WTP-space. (with main variables
Robust standard errors are reported in parentheses.
P-values (Holm adjusted for 13 comparisons) and significance levels: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$
Bilateral Wald tests were conducted to assess the equivalence of interaction coefficients across experimental groups. Holm-Bonferroni adjusted p-values, covering all main interactions, socioeconomic variables, WTP parameters, and the status quo parameter, were used to account for multiple hypothesis testing. For the comparison between treatment group 1 and treatment group 2, the interaction with the binary indicator for respondents confused only by transparent labels is significant (p=0.077). In the comparison between the control group and treatment group 2, the same interaction is also significant (p=0.084).

Table A.25: Mixed logit model, WTP space estimates (fewer interactions)

	Full sample	**Control**	**Treatment 1**	**Treatment 2**
$WTP_{\text{Carbon neutral}}$	-0.27	-0.47	-0.17	-0.14
	(0.15)	(0.25)	(0.26)	(0.25)
Main interactions				
$WTP_{\text{Carbon neutral}}$ x Only confused with transparent labels	0.02	0.21	0.15	-0.29
	(0.11)	(0.18)	(0.19)	(0.20)
$WTP_{\text{Carbon neutral}}$ x Only trust in transparent labels	0.66***	0.76***	0.33	0.70**
	(0.12)	(0.19)	(0.20)	(0.22)
$WTP_{\text{Carbon neutral}}$ x Trust level	0.08	0.14	0.07	-0.03
	(0.05)	(0.07)	(0.07)	(0.09)
$WTP_{\text{Carbon neutral}}$ x Confusion level	-0.05	0.00	0.02	-0.16
	(0.04)	(0.07)	(0.06)	(0.08)
$WTP_{\text{Carbon neutral}}$ x Concern for CO_2 offsets	0.21***	0.09	0.28***	0.16
	(0.05)	(0.08)	(0.08)	(0.08)
$WTP_{\text{Carbon neutral}}$ x Climate worry	0.05	0.05	-0.07	0.25**
	(0.04)	(0.07)	(0.08)	(0.08)
$WTP_{\text{Carbon neutral}}$ x Warm glow	0.25***	0.31***	0.34***	0.14
	(0.06)	(0.08)	(0.10)	(0.09)
$WTP_{\text{Carbon neutral}}$ x Polluter pays	-0.02	0.00	-0.11	0.02
	(0.04)	(0.07)	(0.06)	(0.07)
$WTP_{\text{Carbon neutral}}$ x Financial constraints	-0.18***	-0.12	-0.16	-0.24***
	(0.04)	(0.08)	(0.07)	(0.06)
$WTP_{\text{Carbon neutral}}$ x Time restrictions	-0.00	-0.05	-0.04	0.06
	(0.04)	(0.06)	(0.06)	(0.06)
Socioeconomic interactions	Yes	Yes	Yes	Yes
WTP for other sustainability labels, price, status quo	Yes	Yes	Yes	Yes
Standard deviations of random parameters	Yes	Yes	Yes	Yes
Log Likelihood	-8372.29	-2750.23	-2834.09	-2757.73
AIC	16796.58	5552.46	5720.18	5567.45
BIC	16985.49	5713.01	5880.62	5727.48
Pseudo R^2	0.28	0.30	0.27	0.28
Number of Observations	10568.00	3552.00	3536.00	3480.00
Number of Participants	1321	444	442	435

This table shows the choice model output from the MXL model estimated in WTP-space. (with main variables
Robust standard errors are reported in parentheses.
P-values (Holm-adjusted for 20 comparisons) and significance levels: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$

Table A.26: Mixed logit model, WTP space estimates with interaction variables (warm glow variable added)

	Full sample	Control	Treatment 1	Treatment 2
$WTP_{\text{Carbon neutral}}$	0.51	0.36	1.05	0.09
	(0.26)	(0.45)	(0.51)	(0.46)
Additional interactions				
$WTP_{\text{Carbon neutral}}$ x Effectiveness of offsetting	0.04	-0.02	0.16	-0.00
	(0.04)	(0.07)	(0.07)	(0.07)
$WTP_{\text{Carbon neutral}}$ x Putting a price on nature	-0.02	-0.02	0.06	0.00
	(0.05)	(0.09)	(0.09)	(0.07)
$WTP_{\text{Carbon neutral}}$ x Moral licensing	0.01	0.06	-0.05	0.01
	(0.06)	(0.10)	(0.10)	(0.11)
$WTP_{\text{Carbon neutral}}$ x Offsetting as greenwashing	-0.07	-0.05	-0.17	-0.00
	(0.05)	(0.08)	(0.09)	(0.10)
$WTP_{\text{Carbon neutral}}$ x Survey time	-0.00	-0.00	0.00	-0.00
	(0.00)	(0.00)	(0.01)	(0.00)
$WTP_{\text{Carbon neutral}}$ x Familiarity with carbon neutral label	1.57***	1.14**	1.29**	2.71***
	(0.27)	(0.34)	(0.39)	(0.55)
$WTP_{\text{Carbon neutral}}$ x Manipulation checker	0.16	-0.09	0.17	0.32
	(0.10)	(0.16)	(0.18)	(0.17)
$WTP_{\text{Carbon neutral}}$ x Attention checker (number of labels)	0.28	0.34	0.14	0.37
	(0.11)	(0.19)	(0.20)	(0.18)
$WTP_{\text{Carbon neutral}}$ x Label attention checker (info. on labels)	0.11	0.18	0.10	0.03
	(0.10)	(0.17)	(0.19)	(0.17)
$WTP_{\text{Carbon neutral}}$ x Not consider carbon neutral label (ANA)	-0.54***	-0.24	-0.64**	-0.63**
	(0.11)	(0.17)	(0.20)	(0.18)
$WTP_{\text{Carbon neutral}}$ x Perceived consequentiality	-0.25	-0.11	-0.44	-0.20
	(0.10)	(0.17)	(0.19)	(0.18)
$WTP_{\text{Carbon neutral}}$ x Environmental organization membership	-0.76***	-0.97	-0.97	-0.30
	(0.20)	(0.35)	(0.45)	(0.34)
$WTP_{\text{Carbon neutral}}$ x Response certainty	-0.03	0.00	0.01	-0.14
	(0.03)	(0.05)	(0.06)	(0.06)
Main interactions	Yes	Yes	Yes	Yes
Socioeconomic interactions	Yes	Yes	Yes	Yes
WTP for other sustainability labels, price, status quo	Yes	Yes	Yes	Yes
Standard deviations of random parameters	Yes	Yes	Yes	Yes
Log Likelihood	-8297.24	-2733.19	-2803.33	-2712.95
AIC	16674.49	5546.38	5686.66	5505.89
BIC	16965.11	5793.39	5933.49	5752.08
Pseudo-R^2	0.29	0.30	0.28	0.29
Number of Observations	10568	3552	3536	3480
Number of Participants	1321	444	442	435

This table shows the choice model output from the MXL model estimated in WTP-space.
Robust standard errors are reported in parentheses.
P-values (Holm adjusted for 32 comparisons) and significance levels: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$
All parameters, except for the interaction terms, are randomized. Holm-Bonferroni adjusted p-values and significance levels: ***$p \leq 0.01$, **$0.01 < p \leq 0.05$, *$0.05 < p \leq 0.1$. Bilateral Wald tests were conducted to assess the equivalence of interaction coefficients across experimental groups. Holm-Bonferroni adjusted p-values, covering all main interactions, socioeconomic variables, WTP parameters, and the status quo parameter, were used to account for multiple testing. Adjusted Wald test results indicate that the binary variable for participants participants who purchase carbon-neutral labeled products during grocery shopping (as captured by the label familiarity variable), the difference between the control group and treatment 1 is significant at the 5% level ($p = 0.043$), and the difference between treatment 1 and treatment 2 is marginally significant at the 10% level ($p = 0.080$).

Table A.27: Mixed logit model, WTP space estimates, additional interactions

WTP estimates: MNL model, WTP space (Table A.16)				
	Full sample	Control	Treatment 1	Treatment 2
$WTP_{carbon\ neutral}$	0.56*** (0.06)	0.61*** (0.10)	0.65*** (0.11)	0.41*** (0.10)
Poe test: Control vs Treatment 1		0.61 (p = 0.78)		
Poe test: Control vs Treatment 2		0.09 (p = 0.35)		
Poe test: Treatment 1 vs Treatment 2			0.06 (p = 0.34)	
WTP estimates: MXL model, WTP space, label attributes as random parameters (Table A.17)				
	Full sample	Control	Treatment 1	Treatment 2
$WTP_{carbon\ neutral}$	0.45*** (0.06)	0.54*** (0.10)	0.56*** (0.11)	0.27*** (0.10)
Poe test: Control vs Treatment 1		0.54 (p = 0.925)		
Poe test: Control vs Treatment 2		0.03 (p = 0.163)		
Poe test: Treatment 1 vs Treatment 2			0.03 (p = 0.163)	
WTP estimates: MXL model, WTP space, all parameters as random, using twice as many Halton draws (Table A.19)				
	Full sample	Control	Treatment 1	Treatment 2
$WTP_{carbon\ neutral}$	0.45*** (0.06)	0.53*** (0.10)	0.54*** (0.10)	0.25*** (0.08)
Poe test: Control vs Treatment 1		0.54 (p = 0.916)		
Poe test: Control vs Treatment 2		0.01* (p = 0.064)		
Poe test: Treatment 1 vs Treatment 2			0.01* (p = 0.064)	
WTP estimates: MXL model, WTP space, all parameters as random, error components for alternatives (Table A.20)				
	Full sample	Control	Treatment 1	Treatment 2
$WTP_{carbon\ neutral}$	0.47*** (0.12)	0.55*** (0.08)	0.52*** (0.10)	0.32*** (0.08)
Poe test: Control vs Treatment 1		0.41 (p = 0.811)		
Poe test: Control vs Treatment 2		0.03 (p = 0.168)		
Poe test: Treatment 1 vs Treatment 2			0.06 (p = 0.255)	

WTP estimates are reported with standard errors in parentheses.
Significance levels: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$
Poe test statistics are provided with p-values in parentheses, which are Holm-Bonferroni adjusted for three comparisons.

Table A.28: Poe test results, estimations with various model specifications estimated in WTP space

WTP estimates: MXL model, WTP space, including protest responses (Table A.21)				
	Full sample	Control	Treatment 1	Treatment 2
$WTP_{carbon\ neutral}$	0.44*** (0.06)	0.53*** (0.09)	0.48*** (0.09)	0.24*** (0.08)
Poe test: Control vs Treatment 1		0.35 (p = 0.705)		
Poe test: Control vs Treatment 2		0.01* (p = 0.054)		
Poe test: Treatment 1 vs Treatment 2			0.03 (p = 0.105)	
WTP estimates: MXL model, WTP space, excluding protest responses and survey speeders (Table A.22)				
	Full sample	Control	Treatment 1	Treatment 2
$WTP_{carbon\ neutral}$	0.47*** (0.05)	0.56*** (0.09)	0.51*** (0.13)	0.32*** (0.11)
Poe test: Control vs Treatment 1		0.38 (p = 0.756)		
Poe test: Control vs Treatment 2		0.04 (p = 0.262)		
Poe test: Treatment 1 vs Treatment 2			0.13 (p = 0.526)	
WTP estimates: MXL model, WTP space, excluding protest responses, attention/manipulation checker failers (Table A.23)				
	Full sample	Control	Treatment 1	Treatment 2
$WTP_{carbon\ neutral}$	0.87*** (0.13)	0.96*** (0.27)	1.04*** (0.24)	0.67*** (0.20)
Poe test: Control vs Treatment 1		0.59 (p = 0.821)		
Poe test: Control vs Treatment 2		0.19 (p = 0.747)		
Poe test: Treatment 1 vs Treatment 2			0.11 (p = 0.673)	

WTP estimates are reported with standard errors in parentheses.
Significance levels: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$
Poe test statistics are provided with p-values in parentheses, which are Holm-Bonferroni adjusted for three comparisons.

Table A.29: Poe test results, estimations with different sample restrictions

A.3.3 Open-ended contingent valuation method MWTP estimates

	 Control	 Treatment 1	 Treatment 2
$WTP_{carbon\ neutral}$	4.33*** (0.16)	4.45*** (0.18)	4.82 (0.16)
Holm-Bonferroni adjusted Poe test			
Control vs Treatment 1	0.70 (p=0.60)		
Control vs Treatment 2	0.98 (p=0.10)		
Treatment 1 vs Treatment 2		0.93 (p=0.26)	
Unadjusted Poe test			
Control vs Treatment 1	0.70 (p=0.60)		
Control vs Treatment 2	0.99** (p=0.03)		
Treatment 1 vs Treatment 2		0.94 (p=0.13)	

WTP estimates are based on 10,000 bootstraped samples.
Significance levels: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$.
Poe test statistics are reported with Holm-Bonferroni adjusted p-values (adjusted for three pairwise comparisons).

Table A.30: Poe test results, contingent valuation method

A.4 Competing labels

	Full sample	Control	Treatment 1	Treatment 2
MWTP estimates				
$\text{WTP}_{\text{Carbon neutral}}$	0.91***	1.01***	0.86***	0.84***
	(0.10)	(0.20)	(0.17)	(0.17)
$\text{WTP}_{\text{Organic}}$	1.38***	1.35***	1.36***	1.45***
	(0.08)	(0.15)	(0.15)	(0.19)
$\text{WTP}_{\text{Ethical trade}}$	1.43***	1.45***	1.38***	1.51***
	(0.09)	(0.18)	(0.21)	(0.20)
MWTP interactions				
$\text{WTP}_{\text{Carbon neutral}}$ x $\text{WTP}_{\text{Organic}}$	-0.65***	-0.73**	-0.46	-0.73***
	(0.13)	(0.29)	(0.25)	(0.25)
$\text{WTP}_{\text{Carbon neutral}}$ x $\text{WTP}_{\text{Ethical trade}}$	-0.52***	-0.47**	-0.42	-0.54***
	(0.12)	(0.21)	(0.23)	(0.18)
Means of the choice parameters				
μ_{Price}	-0.18***	-0.12	-0.19***	-0.22***
	(0.04)	(0.07)	(0.07)	(0.08)
$\mu_{\text{Status quo}}$	-4.12***	-4.51***	-3.82***	-4.13***
	(0.17)	(0.33)	(0.28)	(0.31)
Standard deviations				
$\sigma_{\text{Carbon neutral}}$	1.10***	1.32***	1.35***	1.20***
	(0.14)	(0.19)	(0.25)	(0.20)
σ_{Organic}	1.71***	1.53***	1.62***	1.80***
	(0.06)	(0.16)	(0.13)	(0.21)
$\sigma_{\text{Ethical trade}}$	1.55***	1.53***	1.52***	1.56***
	(0.06)	(0.20)	(0.18)	(0.24)
σ_{Price}	0.89***	0.86***	0.78***	0.91***
	(0.06)	(0.09)	(0.10)	(0.10)
$\sigma_{\text{Status quo}}$	2.25***	2.46***	2.10***	2.26***
	(0.13)	(0.24)	(0.18)	(0.20)
Log Likelihood	-8517.55	-2800.77	-2893.98	-2813.97
AIC	17059.10	5625.53	5811.97	5651.95
BIC	17146.29	5699.64	5886.01	5725.80
Pseudo R^2	0.27	0.28	0.26	0.26
Number of observations	10568.00	3552.00	3536.00	3480.00
Number of participants	1321	444	442	435

This table presents the choice model results from the MXL model estimated in WTP space.
Robust standard errors are reported in parentheses.
Holm-Bonferroni adjusted p-values (for 6 comparisons) and significance levels: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$

Table A.31: Mixed logit model, WTP space estimates, sustainability label interactions

WTP estimates: MNL model, WTP space, Org=0, ET=0 (Table A.31)				
	Full sample	Control	Treatment 1	Treatment 2
WTP_{CN}(Org=0,ET=0)	0.91*** (0.10)	1.02*** (0.20)	0.86*** (0.17)	0.83*** (0.17)
Poe test: Control vs Treatment 1		0.29 (p = 0.570)		
Poe test: Control vs Treatment 2		0.25 (p = 0.494)		
Poe test: Treatment 1 vs Treatment 2			0.45 (p = 0.907)	
WTP estimates: MNL model, WTP space, Org=0, ET=1 (Table A.31)				
	Full sample	Control	Treatment 1	Treatment 2
WTP_{CN}(Org=0,ET=1)	0.40*** (0.08)	0.55*** (0.12)	0.44*** (0.14)	0.30 (0.15)
Poe test: Control vs Treatment 1		0.29 (p = 0.581)		
Poe test: Control vs Treatment 2		0.09 (p = 0.188)		
Poe test: Treatment 1 vs Treatment 2			0.24 (p = 0.476)	
WTP estimates: MNL model, WTP space, Org=1, ET=0 (Table A.31)				
	Full sample	Control	Treatment 1	Treatment 2
WTP_{CN}(Org=1,ET=0)	0.26** (0.10)	0.29 (0.19)	0.40* (0.19)	0.11 (0.18)
Poe test: Control vs Treatment 1		0.66 (p = 0.685)		
Poe test: Control vs Treatment 2		0.25 (p = 0.493)		
Poe test: Treatment 1 vs Treatment 2			0.14 (p = 0.274)	
WTP estimates: MNL model, WTP space, Org=1, ET=1 (Table A.31)				
	Full sample	Control	Treatment 1	Treatment 2
WTP_{CN}(Org=1,ET=1)	-0.25* (0.14)	-0.18 (0.26)	-0.02 (0.28)	-0.43 (0.26)
Poe test: Control vs Treatment 1		0.66 (p = 0.672)		
Poe test: Control vs Treatment 2		0.25 (p = 0.495)		
Poe test: Treatment 1 vs Treatment 2			0.14 (p = 0.282)	

WTP estimates are reported with standard errors in parentheses.
Significance levels: *** $p \leq 0.01$, ** $0.01 < p \leq 0.05$, * $0.05 < p \leq 0.1$
Poe test statistics are provided with p-values in parentheses, which are Holm-Bonferroni adjusted for three comparisons.
CN stands for carbon neutral, Org stands for Organic and ET for Ethical Trade. A value of 0 indicates that the label is present on the choice card, while 1 indicates that it is absent.

Table A.32: Poe test results for competing label scenarios

A.5 Confirmatory and exploratory analyses

This section outlines the deviations from the pre-analysis plan and clarifies the distinction between confirmatory and exploratory analyses.

The study is pre-registered on the AEA RCT Registry (ID: 12520-2.0), with the main objective of estimating the effect of transparency regarding CO_2 offsetting and reduction on consumers' WTP for a carbon-neutral label. The following deviations from the pre-registration occurred:

First, although the pre-registration stated that 1,200 participants would be recruited and payment was made for 1,200 responses, the survey company provided data for a total of 1,339 participants.

Second, while the original plan indicated the use of a preference-space model, the final main analysis uses an MXL model estimated in WTP space. This deviation was motivated by the methodological advantages of WTP-space models, such as direct estimation of MWTP values and better interpretability. All preference-space model results are reported in the robustness checks.

Third, although the variable 'warm glow' variable was originally planned to be included in the estimation exploring underlying mechanisms, it was excluded due to its high correlation with guilt and social approval variables. Its role is examined separately in the robustness checks.

Fourth, to detect differences among experimental groups, both the Poe et al. (2005) and Swait and Louviere (1993) tests were pre-registered. However, only the Poe test was used in the analysis, as the main model was estimated using an MXL specification that accounts for individual heterogeneity. The Swait and Louviere (1993) test, by contrast, requires an MNL model.

Fifth, since the primary goal of the paper is not to estimate WTP for a carbon-

neutral label per se, but rather to assess the effect of transparency, the translation of WTP into CO_2 reduction or offset equivalents, and its comparison with the social cost of carbon, is not discussed. This decision responds to criticisms that the study focuses on a single product, which may not be representative of all products and services, and relies on a stated preference method that may be subject to hypothetical bias.

Fourth, although the pre-registration did not mention multiple hypothesis correction, all p-values for WTP estimates and hypothesis tests are corrected using the Holm-Bonferroni procedure. This correction has become a standard practice in experimental economics, and strengthens the reliability of analyses. Results without Holm-Bonferroni correction are reported in the robustness checks.

Finally, while the exploration of the role of different sustainability labels was pre-registered, the interaction effects between sustainability labels are considered exploratory analyses.

All deviations and additional analyses are transparently reported, and robustness tests confirm that the main findings remain consistent across alternative specifications and samples.

B Survey

This section includes the survey instrument, as well as the choice experiment design and power analysis.

In the survey instrument, the statements capturing respondents' concerns about CO_2 offsetting are designed based on the ethical and practical considerations discussed in (Carattini and Blasch, 2024). The protest-response screeners and the statement assessing beliefs about the polluter-pays principle are based on Brouwer (2011). Finally, the items measuring social approval and guilt are adapted from the questionnaire used by Theotokis and Manganari (2015) and Gruchmann et al. (2025).

A.1 Survey instrument

This survey is part of a research project conducted by Eawag, the Swiss Federal Institute of Aquatic Science and Technology, and the University of St. Gallen in Switzerland. The project studies the role of sustainability labels on people's preferences for tea products. By participating in this survey, you are contributing to improving our understanding of this topic. Your participation is voluntary, your responses are anonymous and will be kept confidential. You can opt out anytime without providing a reason, in which case your responses will not be stored. You will receive a payment from the survey company after submitting a completed survey. The survey is expected to take 10 minutes.

While you can return to previous pages in the survey, you will not be able to change your answers. We would like to ask you to read each question carefully. Please refrain from seeking information online or from any other source while answering the survey. There is no right or wrong answer, we are simply interested in knowing your opinion.

Please indicate your consent to participate in this study by selecting the appropriate option.

☐ I consent to participate in this study.

☐ I do not consent to participate in this study.

1 What is your year of birth?

[4 digit number]

2 What is your gender?

- Male.
- Female.
- Non-binary.
- I prefer not to say.

3. What is the highest level of education you have completed?

- Primary school
- Secondary school: High school or equivalent
- Post-secondary vocational training (2 and more years)
- Post-secondary vocational training (up to 1 year)
- Post-secondary academic below-degree level qualification (2 and more years)
- Post-secondary academic below-degree level qualification (up to 1 year)
- Bachelors or equivalent first degree qualification (e.g., BA, BSc, BEng)
- Masters or equivalent higher degree level qualification (e.g., MA, MSc, MBA)
- PhD or equivalent doctoral level qualification (e.g., PhD)
- None of above

4 How frequently do you purchase tea?

- Once a week
- Once every two weeks
- Once a month
- Several times a year
- Once a year
- Never

5 How frequently do you drink tea?

- Daily
- Once a week
- Once every two weeks
- Once a month
- Several times a year
- Never

If you are participating in the survey on a smartphone, please turn it to a **horizontal** orientation for the best display of the upcoming survey section.

Imagine that a box of tea containing **80 teabags** (roughly 200 grams) as shown in this figure is your preferred tea type-be it black, green, or herbal tea-and in your favored packaging type, either teabags or loose leaf. In the following questions, you will be asked to choose between various versions of this tea, differentiated by the presence or absence of carbon-neutral, organic, and ethical trade labels, and price. Note that the tea product and the labels are hypothetical and designed for the purpose of this survey.

This specific **carbon-neutral (CO_2 neutral) label** indicates that the product's greenhouse gas emissions, measured in carbon equivalents, have been:

- **offset** (compensated) by investing in activities **outside of the company** such as tree planting projects; **or**
- **reduced within the company** in the last five years, such as through investments in cleaner production processes; **or**
- both **offset** and **reduced**.

This specific **organic label** indicates that the product contains **only organic ingredients** and no synthetic pesticides.

This specific **ethical trade label** indicates that the product is produced following **responsible labor practices**, which guarantee higher prices for exporters based on internationally recognized standards.

If you are participating in the survey on a smartphone, please turn it to a **vertical orientation** for the best display of the remainder part of survey.

In the next questions, you will be asked to choose between two tea products with certain labels and a given price and 'none of the two' option. Note that the questions are hypothetical, i.e. you are not required to pay for your choices at any point in the survey. All you have to do is to indicate your most preferred option as if you were choosing between such products in a supermarket.

We would like to inform you that people are likely to overstate their willingness to pay for a product in a survey and would not pay the stated amount in real life. Please consider how you would feel spending your money on such products in a real-life situation, and answer accordingly.

Please also consider your budget limitations. Depending on the amount you choose to spend on tea, you will have less money available for other products.

Please check the box if you agree with the statement below.

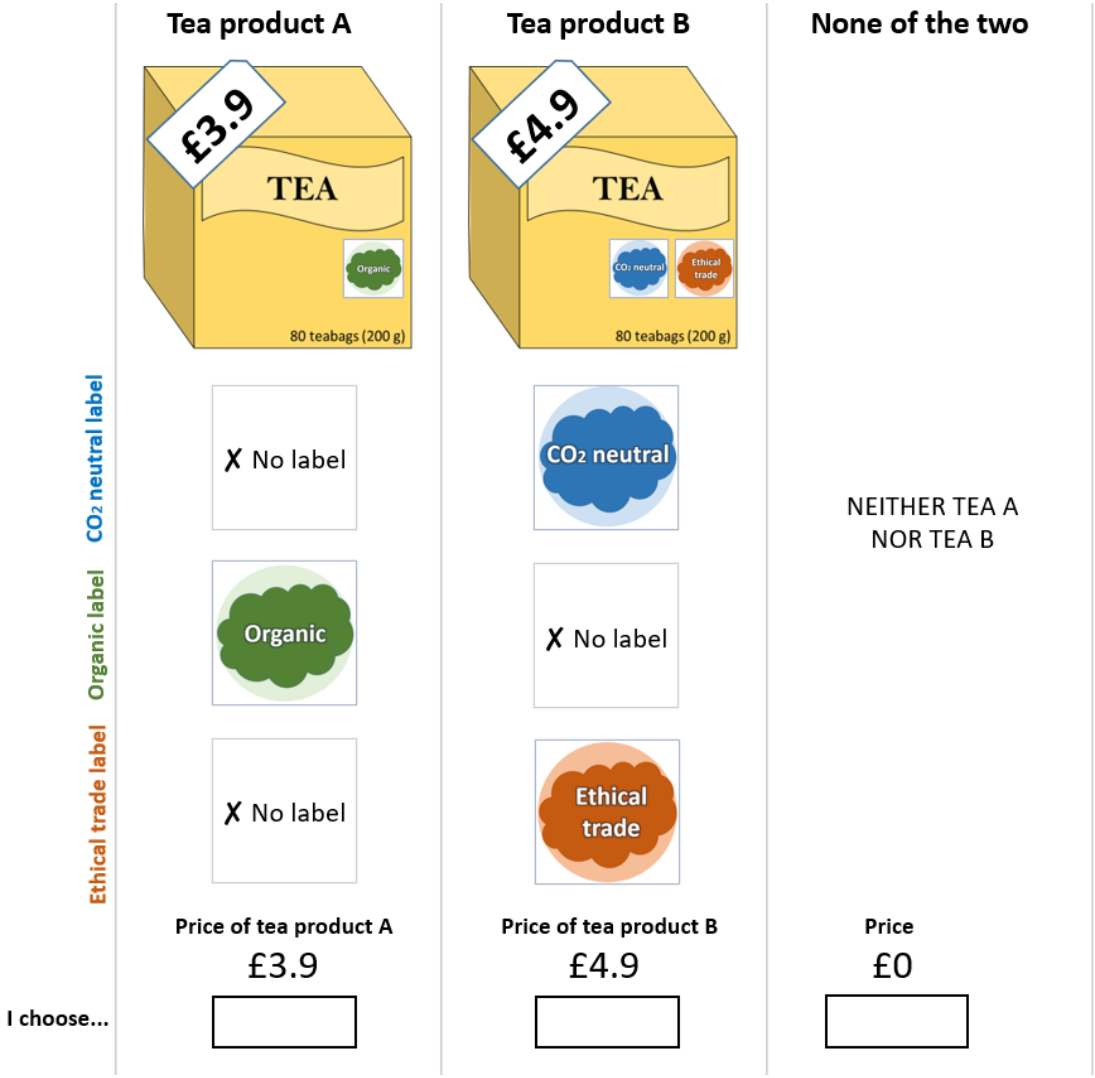
I **understand** the importance of providing truthful answers as if I was making a decision in a real-life setting and promise to provide honest and accurate responses to the questions that follow.

☐

6 Imagine a box containing 80 teabags or 200 grams of loose tea. In the following questions, please indicate which option you prefer the most.

If you are participating in the survey on a smartphone, please keep it in a vertical orientation for the best display.

CHOICE CARD 1/8



CHOICE CARD 2/8

	Tea product A	Tea product B	None of the two
			
CO₂ neutral label	✗ No label	CO₂ neutral	NEITHER TEA A NOR TEA B
Organic label	✗ No label	Organic	
Ethical trade label	✗ No label	Ethical trade	
	Price of tea product A £2.9	Price of tea product B £6.9	Price £0
I choose...	☐	☐	☐

CHOICE CARD 3/8

	Tea product A	Tea product B	None of the two
			
CO_2 neutral label	✗ No label	✗ No label	NEITHER TEA A NOR TEA B
Organic label	Organic	✗ No label	
Ethical trade label	✗ No label	Ethical trade	
	Price of tea product A £2.9	Price of tea product B £3.9	Price £0
I choose...	☐	☐	☐

CHOICE CARD 4/8

	Tea product A	Tea product B	None of the two
			
CO_2 neutral label	✗ No label	✗ No label	NEITHER TEA A NOR TEA B
Organic label	✗ No label	Organic	
Ethical trade label	Ethical trade	✗ No label	
	Price of tea product A £4.9	Price of tea product B £4.9	Price £0
I choose...	☐	☐	☐

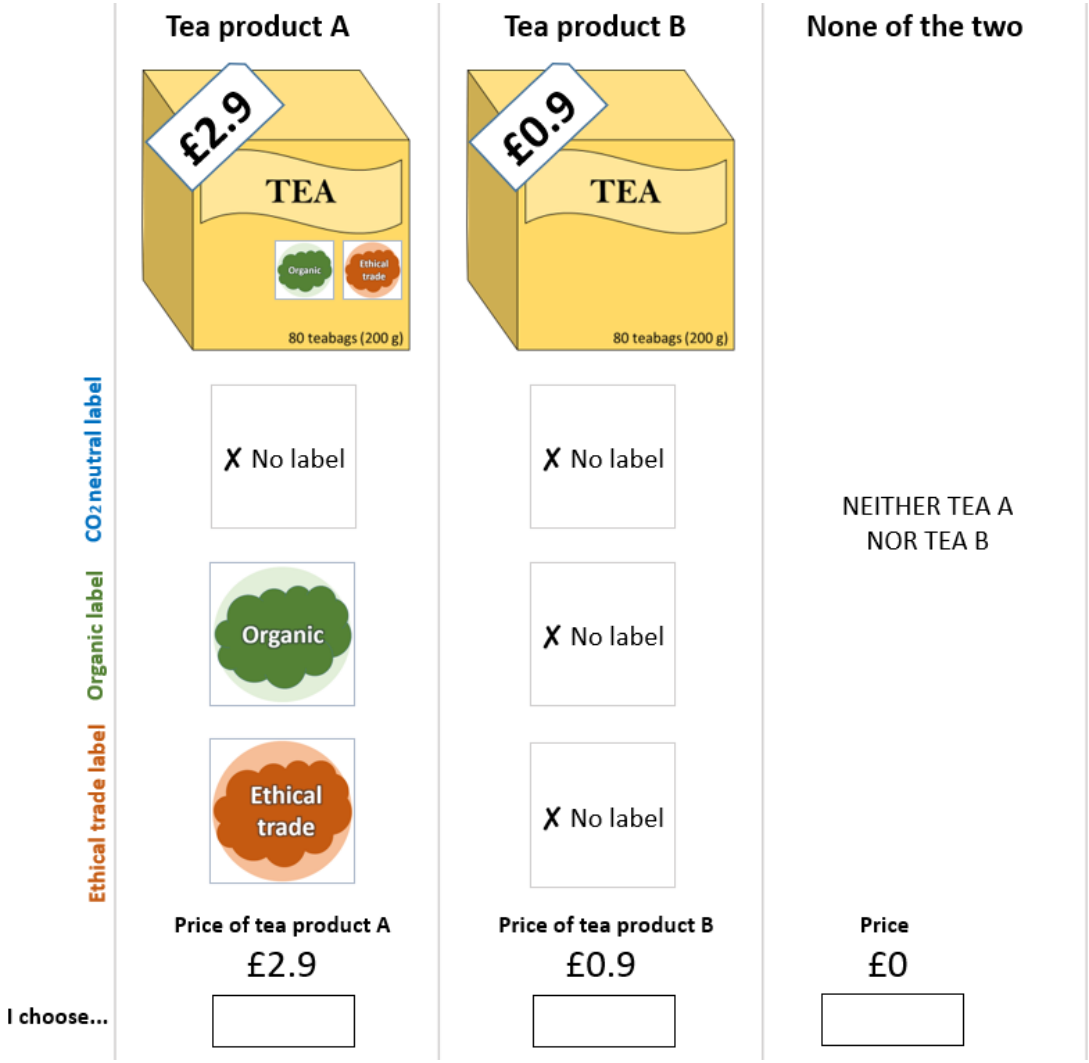
CHOICE CARD 5/8

	Tea product A	Tea product B	None of the two
			
CO_2 neutral label	✗ No label	CO_2 neutral	NEITHER TEA A NOR TEA B
Organic label	✗ No label	Organic	
Ethical trade label	✗ No label	Ethical trade	
	Price of tea product A £0.9	Price of tea product B £4.9	Price £0
I choose...			

CHOICE CARD 6/8

	Tea product A	Tea product B	None of the two
			
CO_2 neutral label	CO_2 neutral	✗ No label	NEITHER TEA A NOR TEA B
Organic label	Organic	✗ No label	
Ethical trade label	Ethical trade	✗ No label	
	Price of tea product A: £6.9	Price of tea product B: £0.9	Price: £0
I choose...	☐	☐	☐

CHOICE CARD 7/8



CHOICE CARD 8/8

	Tea product A	Tea product B	None of the two
			
CO₂ neutral label	CO₂ neutral	CO₂ neutral	NEITHER TEA A NOR TEA B
Organic label	Organic	✗ No label	
Ethical trade label	✗ No label	Ethical trade	
	Price of tea product A £4.9	Price of tea product B £2.9	Price £0
I choose...	☐	☐	☐

7 How certain are you about your choices? Please use the slider below to indicate your level of certainty.

Very uncertain (0) ——slider——(10) Very certain

8. Could you please indicate the main reason why you always chose the "neither tea A nor tea B" option? (This question will only be shown to the participants who always chose the "none of the two" option.)

- The products were too expensive.
- I oppose one or more of the labels.
- Insufficient information was provided about the labels or the products.
- I prefer to spend money on other social and environmental responsibility projects.
- I disagree with the way the choice question was asked.
- Other reason, please specify: ...

9. Can you shortly describe how you made your choices?

... (Open-ended)

10. Which tea characteristics did you **not consider** when making your choices? You can choose one, more than one, or none of the characteristics.

- Carbon neutral label
- Organic label
- Ethical trade label
- Price

- None of above (I considered all tea characteristics).

11. What was your main reason for not considering this tea characteristic/these tea characteristics? (This question will only be shown to those who have not chosen the "None of above" option in the previous question .)

...

12. What is the maximum amount you would be willing to pay for **the carbon-neutral label** (shown on the previous choice cards) **in addition to the cost of the tea product?**

£...

13. Which of the following can be considered **carbon offsetting**?

- **Compensating emissions** by investing in activities outside of the company, such as tree planting projects.
- **Reducing emissions** within the company, such as through investments in cleaner production processes.
- I do not remember.

14. How many different **tea product labels** appear on the choice cards?

- 1 label
- 2 label
- 3 labels
- I do not remember.

15. In this survey, did carbon-neutral label include any percentages (%) of carbon offsetting and carbon reduction?

- Yes.
- No.
- I do not remember.

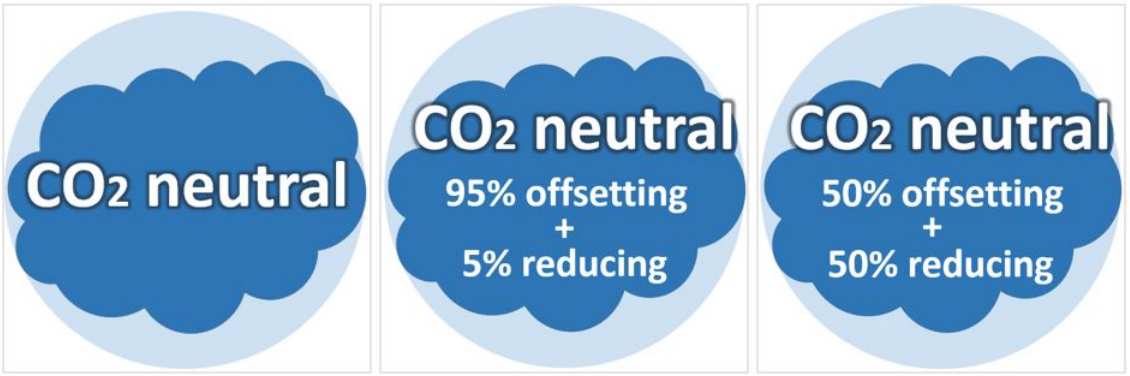
16. Please indicate to which degree you agree or disagree with the following statements.

Statement	Strongly Disagree	Mostly Disagree	Slightly Disagree	Neutral	Slightly Agree	Mostly Agree	Strongly Agree
I worry about climate change.							
Limited financial resources prevent me from buying climate-friendly products instead of the conventional ones.							
Lack of time prevents me from buying climate-friendly products instead of conventional ones.							
My positive emotions increase when I choose climate-friendly products over conventional ones.							
I feel guilty when I buy conventional products instead of climate-friendly ones.							
Most people who are important to me approve of my choice of climate-friendly products over the conventional ones.							
Producers, not consumers, are responsible for covering climate change mitigation costs.							

17. Remember that carbon offsetting involves compensating emissions by investing in projects outside of the company, such as tree planting projects, while carbon reductions occur within the company, such as investments in cleaner production processes. Emission offsetting, reduction, or a combination of both ensures that the entire life-cycle of a product is carbon-neutral.

Please indicate to which degree you agree or disagree with the following statements.

Statement	Strongly Disagree	Mostly Disagree	Slightly Disagree	Neutral	Slightly Agree	Mostly Agree	Strongly Agree
I trust carbon neutral labels.							
I am confused about carbon-neutral labels.							
I am concerned about carbon offsetting.							



Label 1 Label 2 Label 3

Please look at the different types of hypothetical labels shown above carefully, and answer the questions below. You may choose one or more label options, or none.

18. Which label(s) do you trust the most?

- Label 1
- Label 2
- Label 3
- None

19. Which label(s) do you find the most confusing?

- Label 1
- Label 2
- Label 3
- None

20. Please indicate to which degree you agree or disagree with the following statements.

Statement	Strongly Disagree	Mostly Disagree	Slightly Disagree	Neutral	Slightly Agree	Mostly Agree	Strongly Agree
Carbon offsetting effectively reduces carbon emissions.							
Carbon offsetting puts a price tag on emissions, thereby allowing producers to continue polluting.							
Carbon offsetting generates a misleading sense of relief, without encouraging further efforts to reduce emissions.							
Carbon offsetting is a form of greenwashing.							

Please now consider your usual **real-life** grocery shopping.

21. Which type of tea do you consume the most? Please select only one.

- Black tea (e.g., English Breakfast, Early Grey tea)
- Green tea
- Herbal tea
- Other, please specify: ...

22. How much do you typically pay for tea?

- £...
- I do not know.

23. Which tea quantity do you usually buy?

- I usually buy ... **grams** of tea.
- I usually buy ... **teabags**.
- I do not know.

24. Please indicate which sustainability labels the grocery products you buy have.

- Carbon neutral label
- Organic label
- Fair trade label
- No sustainability label
- I do not know
- Other, please specify: ...

Finally, a few questions about yourself.

25. What is your approximate annual household income after taxes?

- Under £10,000
- £10,000 - £19,999
- £20,000 - £29,999
- £30,000 - £39,999
- £40,000 - £49,999
- £50,000 - £59,999

- £60,000 - £69,999
- £70,000 - £79,999
- £80,000 - £89,999
- £90,000 - £99,999
- £100,000 - £129,999
- £130,000 or more

26. What is your current employment status?

- Full or part time employment
- Self-employed
- Unemployed
- Retired
- Looking after family or home
- Full-time student
- None of above

27. Are you a member of any environmental organization?

- Yes
- No

Now, we want to learn about your opinions and experience with this survey.

28. Do you think that your responses in this survey will influence tea product labeling or pricing policies?

- Yes
- No

29. Is there anything about the tea choices that you find confusing or unclear?

- Yes. Please specify what was confusing or unclear: ...
- No.

30.Were the instructions clear to you?

- Yes
- No. Please specify what was unclear: ...

31. Do you have any further comments or feedback about the survey?

... [open-ended]

Thank you for your participation in this survey. Your responses are very valuable to us and contribute to the research project by Eawag, the Swiss Federal Institute of Aquatic Science and Technology, and the University of St.Gallen, Switzerland.

If you have any questions or concerns about the survey please contact Begüm Özdemir Oluk (begum.ozdemiroluk@eawag.ch).

Please click "continue" to submit the survey and receive your payment.

A.2 Choice design

This section includes detailed information on the choice design created for the main survey on software Ngene using the MNL model. Table A.33 shows the details of the attributes and combinations for each choice situation, respectively.

Block	Choice Situation	Price AltA (£)	CN AltA	Org AltA	ET AltA	Price AltB (£)	CN AltB	Org AltB	ET AltB
1	1	3.9	No	Yes	No	4.9	Yes	No	Yes
1	2	2.9	No	No	No	6.9	Yes	Yes	Yes
1	3	2.9	No	Yes	No	3.9	No	No	Yes
1	4	4.9	No	No	Yes	4.9	No	Yes	No
1	5	0.9	No	No	No	4.9	Yes	Yes	Yes
1	6	6.9	Yes	Yes	Yes	0.9	No	No	No
1	7	2.9	Yes	Yes	No	0.9	No	No	No
1	8	4.9	Yes	Yes	No	2.9	Yes	No	Yes
2	1	1.9	Yes	No	No	5.9	No	Yes	Yes
2	2	1.9	No	Yes	Yes	4.9	Yes	No	Yes
2	3	0.9	No	No	No	1.9	Yes	Yes	Yes
2	4	1.9	No	Yes	No	1.9	No	No	Yes
2	5	6.9	Yes	Yes	Yes	3.9	No	No	Yes
2	6	5.9	Yes	Yes	No	1.9	No	Yes	Yes
2	7	3.9	Yes	No	No	2.9	No	No	No
2	8	5.9	No	Yes	No	5.9	Yes	No	No

AltA refers to tea alternative A (on the left of the choice card), and AltB refers to tea alternative B (on the right of the choice card).
D error = 0.287, A error = 0.447, B estimate = 73.188, S estimate = 29.530.

Table A.33: Main survey choice design

C Power Analysis

In this section, I discuss the ex-ante power calculations for my main analysis to understand the consumers' WTP differences between samples. This analysis involves three bilateral comparisons, each with subsamples of 400 participants. The objective of the power calculation is to determine the minimum detectable effect size (MDEs) for WTP differences between subsamples, using the formula from Djimeu and Houndolo (2016) and the power calculation sheet from International Initiative for Impact Evaluation (3ie) (2016):

$$\text{MDEs} = \frac{(t_1 + t_2) \times \text{sd}(y)}{\sqrt{p \times (1-p) \times n}}$$

where sd(y) is the pooled total standard deviation of the estimated effect on the outcome variable, p is the proportion of the study that is randomly assigned to the treatment group, n is the sample size, t1 is the t-value corresponding to the significance level (0.05) of the test, t2 is the t-value corresponding to the power of the design (0.80), and MDEs is the minimum detectable effect size. I assume sd(y) follows a truncated normal distribution, and the mean and standard deviation of the distribution are based on both the literature (Bek, 2022; Carattini et al., 2024) and the my expectations regarding the differences between the subsamples.

I find that the MDEs is 0.18 pounds for the difference between sample 3 and samples 1 and 2, which corresponds to 18% of the standard deviation. Furthermore, the MDEs for the difference between samples 1 and 2 is 0.36 pounds, which also corresponds to 18% of the standard deviation.