WORKING PAPERS

N° 1690

November 2025

“Taxing and nudging to reduce carbon emissions: Results from an online shopping experiment”

Stefan Ambec, Henrik Andersson, Stéphane Cezera, Aysegul Kanay, Benjamin Ouvrard, Luca Panzone, Sebastian Simon

Toulouse School of Economics

Taxing and nudging to reduce carbon emissions: Results from an online shopping experiment[*]

Stefan Ambec[†], Henrik Andersson[‡§], Stéphane Cezera[¶], Aysegul Kanay[‖], Benjamin Ouvrard[**], Luca Panzone[††], Sebastian Simon[‡‡]

November 25, 2025

Abstract

What can be done to reduce the carbon footprint of consumption? To answer this, we conducted an online shopping experiment that tested the effects of two policy tools: a carbon tax (at two levels) and a behavioral nudge in the form of a traffic light-style label indicating a product's carbon footprint (green for low, orange for medium, and red for high). To disentangle the tax's substitution effect from its income effect, we held consumers' purchasing power constant. We find that the tax alone significantly reduces the carbon footprint per euro spent but not per basket purchased, implying that the reduction is driven purely by the income effect. The label alone makes consumers buy fewer red products and more green products, although without reducing significantly their carbon footprint. We do find some substitution effect and a significant reduction of the carbon footprint per basket only when the tax is high enough and combined with the label. Next, we perform a welfare analysis grounded on a theoretical framework that accommodates for several assumptions about consumer's preferences and motivations. We estimate the loss of consumer's surplus from nudging consumers with the label. We also estimate the consumers' valuation of a ton of CO2 avoided when they care about their climate impact.

Keywords : carbon tax, nudge, green label, carbon footprint, climate change, moral behavior.

JEL Codes : D12, D90, H23, Q58.

[*]The authors acknowledge the support of the Agence Nationale de la Recherche under grants ANR-2016-CE05-0018 and ANR-17-EURE-0010 (Investissements d'Avenir program). We have benefited from comments by participants of the TSE Yinghua He Empirical Micro Workshop, the CESifo Energy and Climate Conference 2024 in Munich, the EfD Annual Meeting 2024 in Nairobi, and the EAERE Conference 2024 in Leuven. We thank Ingela Alger for useful comments and suggestions. This paper is dedicated to the memory of Denis Hilton who initiated this research project.

[†]Toulouse School of Economics, INRAE, University of Toulouse Capitole, Toulouse, France. Email: stefan.ambec@tse-fr.eu. Phone: +33 5 61 12 85 16. Fax: +33 5 61 12 85 20.

[‡]Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France

[§]The Swedish National Road and Transport Research Institute (VTI), Lund, Sweden

[¶]Toulouse School of Economics, INRAE, University of Toulouse Capitole, Toulouse, France.

[‖]University Toulouse Jean Jaurès, Toulouse, France.

[**]Université Grenoble Alpes, INRAE, UMR GAEL, 38000 Grenoble, France.

[††]University of Newcastle upon Tyne, Newcastle upon Tyne, United Kingdom.

[‡‡]LIUM, Le Mans Université, 72085 Le Mans, Cedex 9, France.

1 Introduction

To effectively combat global warming, a drastic change in consumption habits is necessary. As food production is responsible for roughly one-quarter of the world's greenhouse gas emissions,[1] shifting to less carbon-intensive products can make a difference. For instance, eating chicken or salmon instead of beef reduces greenhouse gas emissions fourfold. Switching to a vegetarian diet further reduces emissions by half.[2] Clearly, the path to carbon neutrality requires the adoption of more sustainable consumption practices.

Economists recommend several policy instruments to effect this change. The most popular is pricing carbon through a carbon tax or tradable carbon emissions allowances. By pricing the carbon footprint of products, a carbon tax inflates their proportionally to their carbon intensity. Standard microeconomics predicts that consumers will substitute cheaper less carbon-intensive products (e.g. chicken or fish) for more expensive carbon-intensive ones (e.g. beef). Alternatively, nudges can guide consumers' behavior toward more sustainable consumption. Such instruments rely on a change of the choice architecture to push consumers to choose less carbon-intensive food products. For example, products with a lower carbon footprint can be made easily accessible in shops. They can be tagged with colored labels providing information on the carbon content of the product; for instance a green logo on the less carbon-intensive products and a red one on the more carbon-intensive ones. Both the carbon tax and the nudges aim at reducing the carbon footprint of food purchases, although they operate through different behavioral mechanisms.

How effective are carbon taxes and nudges in reducing the carbon footprint of food consumption? What is the role played by each policy instrument? How do they interact? Do they complement each other? To answer these questions, we ran an online supermarket experiment in which we varied the pricing and the framing of products, based on their carbon footprint. Subjects were endowed with a budget that they spent on purchasing food on an online shopping platform.

The subjects were divided into six groups, with each group assigned a distinct version of the online shopping platform. In some versions, prices were augmented by a tax applied to each product's carbon footprint. In addition to a benchmark without tax, we experimented with two levels of carbon tax, €80 and €250 per ton of CO_2. In other treatments, the subjects saw a "traffic light", that is, a label for each product, the color of which varied according to the product's carbon intensity (green, orange and red). Overall, this resulted in a 2×3 experimental design, depending on the tax level (no tax, €80 and €250 per ton of CO_2) and the presence or not of a traffic light.

Our experiment was designed to address two issues. The first was related to the informational content of environmental taxes. As the level of the tax associated with a product provided information about its carbon footprint, it might possibly nudge consumers, as might the traffic-light label.[3] To avoid such an effect, we did not display the tax content in the posted price. Instead, we modified the shopping platform by increasing each product's price after adding the carbon tax multiplied by the carbon footprint. Everything else remained unchanged, and consumers could not infer the carbon footprint of products from their price. In contrast, in the treatment with only the traffic

[1] Source: One World Data https://ourworldindata.org/food-ghg-emissions

[2] Based on the carbon footprint per 100 g we use in our experiment: 2100 g for beef, 600 g for chicken, 500 g for salmon and 240 g for avocado.

[3] Chetty et al. (2009) likewise show that making the tax salient in the posted price further reduces the consumption of the taxed products.

light, prices were unchanged compared to the control group. The traffic-light label provided information about the carbon footprint of products with a color code that nudged consumers to favor less carbon-intensive products ("green" products) instead of carbon-intensive ones ("orange" for medium-intensive ones and "red" for the most carbon-intensive ones). Our experimental design thus aimed to isolate the monetary impact of taxes from the nudge/informational impact of labels.[4]

Second, our experimental protocol aimed at disentangling the substitution effect from the income effect of taxing consumption. By inflating prices, a carbon tax reduces purchasing power, which lowers consumption and thus automatically reduces the overall carbon footprint of the purchased basket of food. However, this is obviously problematic for the political acceptability of carbon taxes because individuals will endorse taxes on the carbon content of food only if they themselves are not worse off.[5] Therefore, to avoid any income effect, we made the carbon tax budget-neutral by redistributing the money collected as a lump sum (described in section 2). We used a previous experiment run with a similar shopping platform to increase endowments by the total tax paid, on average, for the purchased basket. The purchasing power was thereby maintained insofar as subjects could still purchase the average basket of the control group when carbon was taxed.[6] Inflation was fully compensated for by multiplying the endowment with the inflation rate, based on the average basket. The substitution effect of the carbon tax was measured by the change of the shopping basket's carbon footprint, thus keeping the purchasing power constant. To measure the total effect of the carbon tax, we computed the carbon footprint *per euro spent*. This measure informed us of the variation of the food product's carbon footprint with a tax that kept the endowment unchanged. The total effect included both the *income* and the *substitution* effects.

Our results highlight the importance, with the carbon tax, of separating the substitution effect from the income effect. First, it turned out that all carbon footprint reductions driven by the carbon tax alone were primarily due to an income effect. We found that the carbon tax significantly reduced both the mean and the median carbon footprint per euro spent, but not per basket purchased, except with the high tax (€250) complemented by the traffic lights. Hence, the reduction of the carbon footprint per euro was mainly driven by an income effect.[7]

Second, we found evidence of a substitution effect but only when the tax was high and combined with the nudge. We did find that the label nudge made consumers switch from carbon-intensive products (i.e., those labeled "red") to less carbon-intensive ones (i.e., labeled "green"). However, despite this substitution, the nudge alone did not significantly impact the carbon footprint of the basket. It did so only when associated with the €250 tax, after adjusting for purchasing power. Overall, for the substitution effect to be effective in reducing the carbon footprint, the tax or the nudge alone was not enough.

[4] See Section 5 for a microeconomics formulation of these behavioral assumptions.

[5] On the acceptability of carbon pricing and the redistribution of tax revenues, see Douenne and Fabre (2022).

[6] To be precise, the purchasing power of the "average consumer" was unchanged insofar as they were still able to buy the same basket with the tax as in the control group (i.e., without any tax being implemented). However, consumers whose basket differed from the average one might lose or gain purchasing power. They would lose if their basket had a higher carbon footprint than the average one in the control treatment, and they would gain if it was lower.

[7] Furthermore, the tax should be high enough to obtain this reduction: we observed a significant impact if carbon was taxed at €250 per ton, but not at €80 per ton of CO_2 equivalent greenhouse gas. Nevertheless, we did find evidence of a reduction of the carbon footprint per euro spent (both in median and mean) for a lower tax rate of €80 per ton only when it was combined with the nudge.

Strong monetary incentives had to be combined with easy-to-read labels on the carbon content of products.

Our findings are consistent with the behavioral mechanisms described above. On the one hand, the tax is sufficient to guide consumers' choices toward less carbon-intensive products. Increasing prices proportionally to the carbon content of products made consumers buy less, but not more wisely. On the other hand, the label nudged consumers in the right direction: they did buy less "red" products and more "green" ones. However, switching products from "red" to "green" was not enough to reduce their carbon footprint.[8] Only when the nudge was combined with the high tax did the carbon footprint of food decrease significantly, after correcting for the inflation rate.

Finally, we investigated the impact of nudges on consumers' welfare. We considered two approaches. The first one was based on the assumption that consumers did not care about their environmental impact, yet that they were sensitive to the nudge and that they enjoyed some utility by buying "green" products and a disutility when purchasing "red" ones. The nudge modified their *decision utility* (i.e., the function they maximized when they shopped) but not their *experienced utility* (i.e., their welfare). This made them reduce carbon emissions at the cost of reducing their own welfare. We call them the *nudgeable* consumers. Using a behavioral economic framework, we identified the welfare loss expressed in nudgeable consumers' surplus when they were nudged with labels.

Our second approach assumed that consumers at least partially internalized their impact on the environment. This was a feature of *experienced utility* (or welfare). However, they had imperfect knowledge about the carbon footprint of products, which translated into some approximation about their environmental impact, based on the expected carbon footprint in their *decision utility* when they shopped. By providing information about the carbon footprint of their shopping basket, the label helped them to make better choices, and thereby to improve their own welfare. We thus identified the welfare gain for the so-called *moral* consumers, of being informed by labels.

Taking advantage of the price variation induced by the carbon tax, we estimated demand functions to calculate the welfare change for both the nudgeable and the moral consumers. In doing so, we were able to quantify some parameters of our behavioral economic model (e.g., the utility of buying a green product, the disutility of buying a red product, and the consumers' internalization of one ton of CO_2).

Related literature Our contribution to economics is threefold. First, we have furthered our knowledge on how taxes and nudges modify consumers' behavior. Second, we have enhanced our understanding of the interaction between public policies to mitigate climate change, specifically between market-based policy instruments and nudges. Third, we are contributing to the literature on the welfare impact of nudges.

Several studies have investigated how taxes and labels modify what people purchase in a laboratory setting (Muller et al., 2017, 2019; Crosetto et al., 2020, 2025; Pizzo et al., 2024), online (Panzone et al., 2018; Kanay et al., 2021; Kanay, 2021), in grocery stores (Chetty et al., 2009; Perino et al., 2014; Lanz et al., 2018; Dubois et al., 2021; Bilén, 2023; Barahona et al., 2023) and in cafeterias (Brunner et al., 2018; Lohmann et al., 2022; Handziuk and Lova, 2023). Most studies have found a significant impact of taxes and nudges on food purchases. In contrast, we provide evidence that the carbon tax and the

[8] This might be due to the three colors of the label, which leads to inter-category substitutions that are mitigated with intra-category substitutions: consumers purchase the more carbon-intensive products labeled as red and green.

nudge alone fail to reduce the carbon footprint of the shopping basket. We do find a significant reduction of the carbon footprint, but only when the tax is combined with the nudge.[9] This might be related to the fact that, unlike previous studies, we maintain consumers' purchasing power by increasing their income with an amount equivalent to the carbon tax. We thus measure only the substitution effect and abstract that from the income effect. Hence, the effectiveness of the tax in previous studies might be driven by the income effect. Consistently, Perino et al. (2014) and Lanz et al. (2018) found no impact on the carbon footprint of shopping baskets if a subsidy on less carbon-intensive products was implemented rather than a carbon tax.[10]

In terms of experimental design, the closest study to ours, Panzone et al. (2018), redistributes the revenue from the tax, as we do, to abstract it from the income effect. However, the redistribution is done after the experiment (once products have been purchased), whereas we assign the revenue from the tax before the experiment through a higher endowment (estimated based on a previous study). We do so because the fact of redistributing money after the experiment may lead some subjects to behave strategically to save money.[11]

Second, our study investigates how financial incentives and information interact in impacting individuals' behavior. One question addressed in economics is whether the two policy instruments are mutually substitutable (so that having the two together would be less effective than having only one of them) or complement (reinforce) each other. For instance, the motivation crowding-out effect of monetary compensation on moral duties, highlighted by Frey and Oberholzer-Gee (1997) and documented by Gneezi and Rustichini (2000)[12], suggests a substitution effect. This substitution effect can be explained by moral licensing: paying a tax exonerates one from feeling guilty for arriving late (or, in our experiments, for contributing to global warming).

In contrast, more recent experiments have found that nudges complement financial incentives. For instance, in random control trials of critical-peak pricing for electricity, sending price alerts (Jessoe and Rapson, 2014) or messages on the importance of energy saving during peak consumption (Ito et al., 2018) reduces further electricity consumption when the price increases.[13] Moreover, the type of information provided matters for the nudge to be a complement to financial incentives. Rodemeier (2023) has found that people tend to buy more subsidized light bulbs if the energy saved is expressed as a percentage, but fewer if it is expressed in monetary terms. In our experiment, the tax, and the

[9]Our result contrasts with Crosetto et al. (2025)'s finding that nutritional taxes and labels together do not perform better than the label alone in improving the basket's nutritional score.

[10]Kanay (2021) similarly found no impact on the carbon footprint of shopping baskets with a bonus-malus (or feebate) scheme, where the least carbon-intensive products are subsidized while the most carbon-intensive ones are taxed in an online shopping lab-experiment. Similarly, also with a bonus-malus, Handziuk and Lova (2023) find no impact on the carbon footprint of canteen meals.

[11]Another difference with Panzone et al. (2018), is that in their study subjects are informed that a carbon tax is implemented, whereas we do not inform them about the exact amount of tax they pay per food product. Informing consumers about the amount of tax paid can be seen as a nudge in itself. Thus, although Panzone et al. (2018) highlight the effectiveness of the carbon tax to reduce a food carbon footprint, they cannot clearly separate the effect of the tax from the effect of information provision about the amount of tax paid. Our approach therefore allows us to isolate the effect of the tax only.

[12]Gneezi and Rustichini (2000) report the results from a daycare experiment, showing that parents tend to arrive late more often to pick up their child when a fine is applied in case of late arrival. This suggests a substitution effect: financial incentives tend to be counter-productive in this case.

[13]Bollinger and Hartmann (2020) similarly show that equipping a household with an in-home display of consumption, rather than informing the customers about prices in real-time, further reduces electricity consumption after a price increase.

information provided on the carbon content through the three-color label are *strongly complementary* insofar as both are intended to achieve a significant carbon footprint reduction.

Our contribution to the literature on the welfare impact of nudges is twofold. First, on the theoretical foundation of welfare analysis, we distinguish between several approaches within a single framework. We consider the standard approach of nudges that modify *decision utility* with psychological reward or penalties to induce collectively superior choices (Farhi and Gabaix, 2020; Carlsson and Johansson-Stenman, 2024). Individuals reduce the climate impact of their behavior at the cost of making suboptimal choices from their own point of view. We contrast the standard approach of nudges to the moral approach in which consumers care about the morality of their behavior (Alger and Weibull, 2013; Ambec and De Donder, 2022; Herweg and Schmidt, 2022; Chan, 2024; Kaufmann et al., 2024). We characterize how the premise of the model (e.g., the psychological cost of buying a red product, or the reward from buying a green one, or the consumer's valuation of their own carbon emissions) determines the variation of welfare from nudging consumers with the two approaches. We identify how the variation of welfare can be measured with estimates of the demand function.

Second, to measure the welfare impact of nudges we employ a different empirical strategy than previous studies that have approximated it using treatment effects (List et al., 2023), a questionnaire to collect consumers' willingness to pay (Allcott and Kessler, 2019), or a method based on behavioral distortions (Allcott et al., 2025). We take advantage of price variations induced by the carbon tax to estimate demand functions,[14] and then to compute the consumer's surplus with and without a nudge. The welfare variation is derived directly from the consumer's surplus. Furthermore, the above-mentioned studies deal with one product and one bias (e.g. cigarettes and smoking, energy consumption and climate impact, vaccine intake and infection). In contrast, we consider two types of product, green and red, and two distortions or biases: under-consumption for green goods and over-consumption for red goods (compared to what is socially optimal). The dual dimension of the distortions makes the welfare analysis richer and more complex.

The paper is organized as follows. We describe the experiment in Section 2. In Section 3 we assess the impact of the taxes and the nudge (alone or combined) on the carbon footprint. In Section 4 we examine how the traffic-light nudge influences the choice of products, and we estimate a demand function and price-elasticities. In Section 5 we investigate the welfare impact of the nudge. We develop a theoretical analysis to measure welfare before performing a back-of-the-envelope estimation. Section 6 concludes the paper.

2 The experiment

We describe our experimental design which involves two parts: the online shopping platform and a survey (Section 2).

[14] We were inspired by Bollinger and Hartmann (2020) who also estimate demand functions to compute the consumer's surplus variation under some behavioral assumption. However, they evaluate not a nudge but rather real-time pricing that modifies consumers' demand in a very different way.

2.1 Design

Our experiment relied on a shopping platform called "green shop" that had previously been used for other lab experiments (see e.g., Kanay et al. 2021). The platform was adapted to perform online shopping with real consumers, with 480 food products in eight categories: fruits, vegetables, meat and fish, dairy, delicatessen, savoury, frozen products, and drinks (see Appendix A for an example). Each food product is presented with its name, a photo, its price and the price per kg.

Participants know that they can freely use their endowment to buy the goods they want, whatever their quantity (as long as the budget constraint is not met). They are also informed that money that is not used from their endowment is lost. In the control group, participants are endowed with €40.

2.2 Treatments

We randomly allocated each consumer to one of the six experimental conditions: a control group and five treatments with policy instruments. We considered one nudge (a traffic light) and two levels of carbon tax: €80 per ton of CO_2eq and €250 per ton of CO_2eq.[15] The nudge was a traffic light (i.e., a colored label) similar to the energy efficiency label on an electric appliance (Goeschl, 2019) or the nutritional label for food (Crosetto et al., 2016; Muller et al., 2019). The color varied with the carbon footprint of the product. We divided the set of products into three groups of roughly equal size, with a color assigned to each group: green, orange and red. The carbon footprint of a green product was between 6 and 137 grams of CO_2eq emissions per 100g (171 products were green); of an orange one between 140 and 456 grams of CO_2eq emissions per 100g (154 products); and of a red one between 480 and 2100 grams of CO_2eq emissions per 100g (155 products). The distribution of the carbon footprint per 100g in each category is presented in Appendix B.

To eliminate the income effect, we increased the endowment in the treatments with tax (tax only or tax and label) by the inflation rate computed on the average basket in the control treatment. Since we did not know the contents of the average basket before running the experiment, we used the one found in a previous experiment with the same platform (green shop) in Kanay et al. (2021). We computed the average carbon footprint in the control group in Kanay et al. (2021) and calculated the corresponding amount of tax that would be paid on it. After rounding upward this estimate, we came up with a total of €3 paid per basket with the €80 tax and €8 with the €250 tax.[16] Hence, participants with the €80 per ton of CO_2eq tax received an endowment of €43, while those with the €250 per ton of CO_2eq tax received an endowment of €48.[17]

[15] The former level corresponds to the average price of carbon in the European Union's Emission Trading Scheme (EU ETS) when the experiment was designed. The latter corresponds to the one adopted by the French administration following Quinet (2019).

[16] To be precise, in Kanay et al. (2021), the carbon footprint of subjects in the control group was 17.67 kg of CO_2 at the basket level on average, with an endowment of €25. Applying a €80 (€250) tax per ton of CO_2, this would lead to a tax bill of €1.41 (€4.42) on average. In our case, considering an endowment of €40, and considering that the basket's carbon footprint would increase proportionally to the endowment, this would lead to a tax of €2.26 (€7.07). We decided to consider conservative values and add €3 to the endowment with the €80 tax, and €8 with the €250 tax.

[17] Fortunately, those numbers, based on our previous experiment, turned out to be of the same magnitude in the current experiment: we obtained a carbon footprint of 29.79 kg of CO_2 per basket on average,

Since the money collected by taxing the average basket equalled the money obtained by taxing all baskets, divided by the number of baskets, compensating for the inflation rate was equivalent to redistributing as a lump sum the money collected through the carbon tax.

The six treatments involving the traffic-light (TL) label and two tax levels (Tax80 and Tax250) alone or combined are summarized in Table 1, with the number of observations in parentheses for each treatment.

Table 1: Experimental design (number of subjects in parentheses)

	No nudge	**Nudge**	Endowment
No tax	Control (109)	TL (92)	€40
Low tax	Tax80 (102)	Tax80+TL (97)	€43
High tax	Tax250 (140)	Tax250+TL (120)	€48

2.3 Survey

After the online shopping task, the participants were asked to answer a set of questions to better understand their motivations during the main task (see Appendix S1).

First, we measured the subjects' environmental preferences with the New Environmental Paradigm scale (Dunlap et al., 2000). This questionnaire consists of 15 statements that describe pro- and anti-environmental attitudes, and subjects have to indicate the extent to which they agree with each statement on a 5-likert scale ("Strongly disagree", "Disagree", "Neutral", "Agree" and "Strongly agree"). These environmental preferences are relevant for this exercise, as experimental studies have emphasized that the response to a nudge is influenced by the participant's environmental preferences (see e.g., Ouvrard and Stenger 2024).

We recoded the subjects' answers from 1 ("Strongly disagree") to 5 ("Strongly agree") for pro-environmental behaviors, and the other way round for anti-environmental behaviors. We then computed a total score per respondent. Considering the full sample, the Cronbach's alpha is 0.81, suggesting a high internal reliability of the NEP scale

Finally, we collected information on food habits (diet, etc.), meat consumption, and some additional attitudes and beliefs (e.g., political orientation). Meat consumption is generally associated with a high carbon footprint (see e.g., Bonnet et al. 2020). In the rest of the analysis, we identify those who eat more meat than the French average as those who consume butcher's meat, poultry or game other than cold meats at least four times a week.[18]

which implies a tax bill of €2.38 with the €80 tax, or €7.45 with the €250 tax. . Overall, subjects were able to buy the average basket of the control group when carbon was taxed.

[18] See https://www.credoc.fr/publications/les-nouvelles-generations-transforment-la-consommation-de-viande

2.4 Procedure

We pre-registered the design of this experiment (AEARCTR-0008676) and obtained the approval of TSE's ethics committee in December 2021. We then contracted with the survey company Enov to recruit the subjects.[19] To ensure that the delivery of the basket of products for the winning subjects was feasible, we selected subjects from the ten largest French cities. To improve data quality, subjects who reported not being (co)responsible for food shopping in their household were excluded from the study at the screening stage, and were thanked for their participation.[20]

For their participation in our experiment, participants received a €10 gift voucher. They also received an individual endowment (€40, €43 or €48 depending on the treatment), which they could only spend in our online grocery store. They were told that they had a 10% chance of winning the basket of products they had chosen, which ensured that they had incentives to choose the products they preferred. After being randomly selected, the winning subjects were contacted by the survey company Enov to receive their baskets.

The characteristic of the subjects are described in Table C.1 in Appendix C. Overall, our sample is well balanced between treatments as regards main demographics and socioeconomic characteristics (age, gender, income, number of children). However, we detect significant differences between treatments in the proportion of respondents who hold a university degree, and in their environmental score; both differences are significant only at the 10% level.

Due to the restrictions imposed as part of the data collection process, our sample is not representative of the French population: participants appear older, more educated, with a lower income and have fewer children than the French general population. As expected, the majority are women, as women are normally more likely to have responsibility for food provisioning. These differences may be explained by the fact that we conducted an online experiment, and restricted the geographical provenance of the participants. While these differences limit somewhat the external, but not internal, validity of this experiment, this is a common finding, as other recent experimental studies on food consumption also used samples that were not representative of the population of the country under consideration (see e.g., Enlund et al. 2023).

3 How effective are the taxes and the nudges in reducing a carbon footprint?

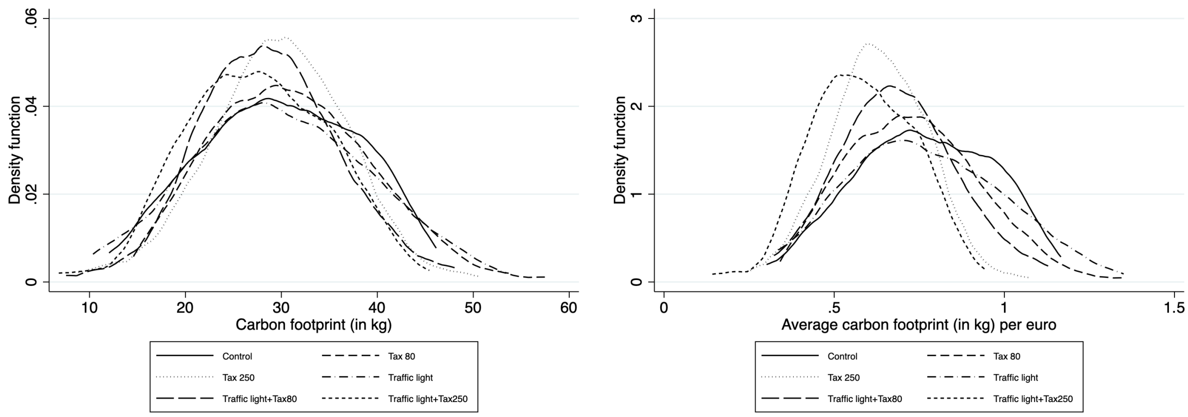
3.1 The distribution of a basket-level carbon footprint

In Figure 1, we graph the densities of the carbon footprint of subjects' groceries, measured per basket and per euro.

[19] See: `https://enov.fr`

[20] We introduced that question at the beginning of the study, and before the online shopping task. Subjects were not aware of that exclusion question when starting the survey.

Figure 1: Density of the carbon footprint per basket (left) and per euro (right)



The carbon footprint per basket (left-hand graph) exhibits a normally shaped density around an average of roughly 30kg of CO_2. The density of the nudge only (treatment TL) and of the low tax only (treatment Tax80) are very similar to the control treatment. However, the densities of the two combined nudge and tax treatments (Tax80+TL and Tax250+TL), as well as the high tax treatment (Tax250), seem to be less dispersed with a higher concentration around the mean. The density of the nudge and high tax treatments (Tax250 and TL) also appears to be shifted leftward compared to the control treatment, with a lower carbon footprint per basket. A Shapiro-Wilk normality test indicates that, considering the entire sample, the carbon footprint of the basket is normal (W = 0.998, *p-value* = 0.799). The log-transformed variable is, instead, not normally distributed (W = 0.963, *p-value* < 0.001).

The carbon footprint per euro (right-hand graph), on the other hand, shows an even more pronounced leftward shift for the combined nudge and tax treatments (Tax80+TL and Tax250+TL) and the high tax treatment (Tax250). A Shapiro-Wilk normality test indicates that, considering the entire sample, the carbon footprint of the basket is not normally distributed (W = 0.990, *p-value* < 0.001).

In Appendix S2 (see Table S2.1), we show with Kolmogorov tests that there are no significant differences between the treatments and the control group regarding the carbon footprint (in kg). However, we detect significant differences between the control group and the treatments Tax250, Tax80+TL and Tax250+TL in the average carbon footprint (in kg) per euro.

3.2 Nudge, tax, and basket-level carbon footprint

Descriptive analysis Table 2 presents the impact of each policy on the carbon footprint of the basket (left-hand side), as well as the carbon footprint per euro (right-hand side).

As previously argued, to assess whether this reduction is driven by a substitution or an income effect, we compare the carbon footprints per basket, keeping the purchasing power constant in the left-hand side of Table 2. With an endowment of €40 net of taxes, the mean and median carbon footprint are 29.79 kg and 29.67 kg respectively. We obtain a significantly lower carbon footprint only for the high tax and nudge policy Tax250+TL, with a mean of 27.58kg (instead of €29.79 kg without policy) and a median of 27.18 kg (instead of €29.67 kg without policy).

Table 2: Descriptive statistics for the carbon footprint per basket and per euro

	Carbon footprint per basket			**Carbon footprint per euro**		
	Median[1]	Mean[2]	St. dev.[3]	Median[1]	Mean[2]	St. dev.[3]
Control *N = 109*	29.67	29.79	8.11	0.76	0.76	0.20
Tax80 *N = 102*	30.32	31.06	8.20	0.72	0.73	0.19
Tax250 *N = 140*	30.03	29.71	7.10	0.63***	0.63***	0.14***
TL *N = 92*	29.56	29.95	9.14	0.75	0.76	0.23
TL+Tax80 *N = 97*	28.49	28.60	7.22	0.67**	0.69***	0.17
TL+Tax250 *N = 120*	27.18*	27.58**	7.34	0.57***	0.59***	0.15***

Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
[1] Two-sample Wilcoxon rank-sum test comparing each treatment to the control group.
[2] T-test comparing the means of each treatment vs. the mean of the control group.
[3] SD-test comparing the standard deviation of each treatment vs. the one for the control group.

In the case of the carbon footprint per euro, the median and the average carbon footprint per euro in the control group are 0.76kg of CO_2. We do not detect any significantly differences in the mean and the median of carbon footprint per euro for the nudge alone (TL) and the low tax alone (Tax80). It is only when the nudge is combined with the tax (Tax80+TL and Tax250+TL), or when the tax is high enough (Tax250), that we detect significant differences in both the mean and the median carbon footprint per euro. Hence, our experiment provides evidence of a reduction of the carbon footprint per euro with a mix of nudge and tax or with a high tax alone.

The results for the treatments including a traffic-light label are surprising, given the existing literature. As a robustness check, we have compared the average carbon footprint per basket across treatments with and without a TL label, regardless of the tax level. We find that, on average, the carbon footprint per basket is 30.13 kg (s.d. $= 7.75$) in treatments without a TL, compared to 28.60 kg (s.d. $= 7.92$) in treatments with a TL. This difference is statistically significant (two-sided t-test, $p = 0.013$).

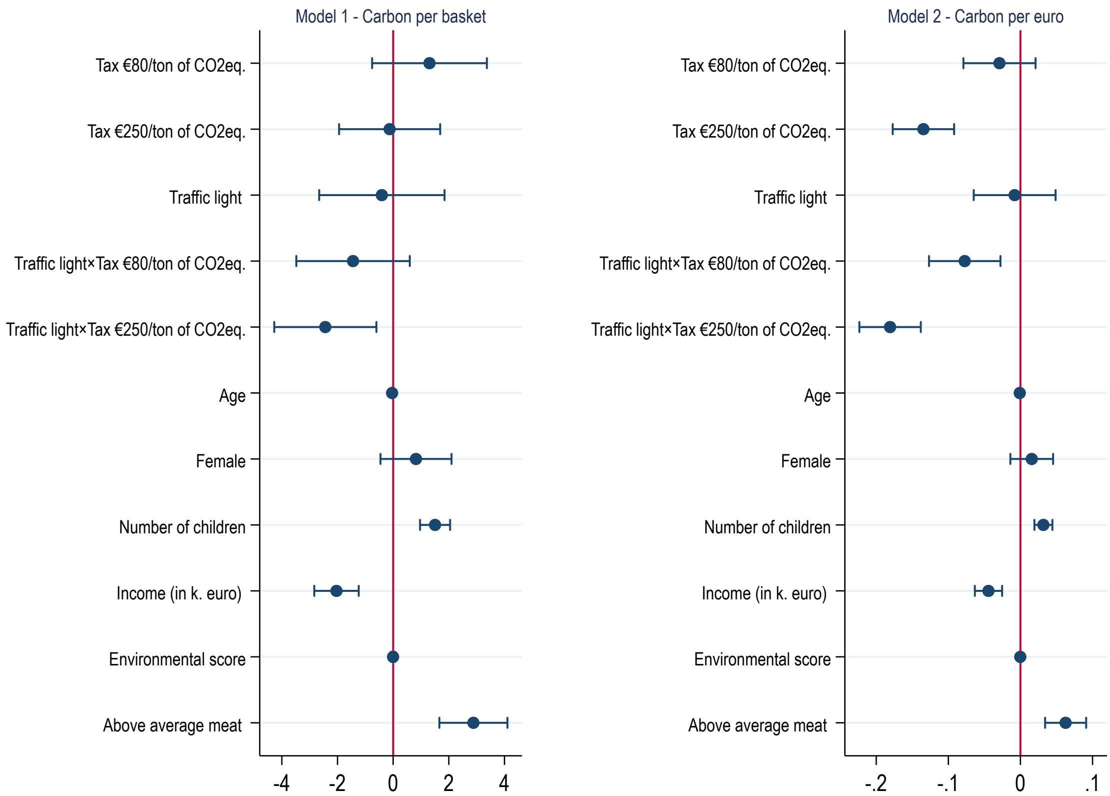
Overall, we find no evidence of a substitution effect for most policies, but some evidence of an effect with the high tax combined with the nudge. Therefore, the carbon footprint reduction induced by the tax alone, or the low tax and nudge, is driven exclusively by an income effect.[21]

Econometric analysis To further investigate the impact of each policy, we regress the carbon footprint per basket (in kg of CO_2eq) and the carbon footprint per euro spent (in kg of $CO_2eq/€$) over treatment groups using OLS. As mentioned in the introduction, the carbon footprint per basket measures the substitution effect of the tax, while the carbon footprint per euro spent measures the full effect, including the revenue effect. Our main focus is to identify treatment effects. We include the following variable: *Tax €80/ton of CO_2eq.*, *Tax €250/ton of CO_2eq.*, *Traffic light*, *Traffic light× Tax €80/ton of CO_2eq.*, *Traffic light× Tax €250/ton of CO_2eq.* These variables are dummy variables equal to one

[21] In Appendix S3, we report additional tests accounting for multiple hypothesis testing (List et al., 2019). While our results regarding carbon footprint per euro spent are robust, we no longer detect any significant differences regarding carbon footprint per basket.

if the subject was allocated to the corresponding treatment, while the control group was used as a reference dummy variable. We also control for subjects' age, gender, number of children, income (in k euro), environmental score, and whether they eat more meat than average. Results for the regression using the carbon footprint (in kg) per basket (model 1) and the carbon footprint (in kg) per euro spent (model 2) are reported in Figure 2.[22]

Figure 2: Effects of treatments on carbon footprint per basket (left panel) and per euro (right panel)



Note: This figure reports the effect of being exposed to the treatments on the carbon footprint per basket (left panel) and on the carbon footprint per euro spent (right panel). Treatment effects come from OLS regressions of the outcome variable (the carbon footprint per basket or the carbon footprint per euro spent). 95% confidence intervals are reported.

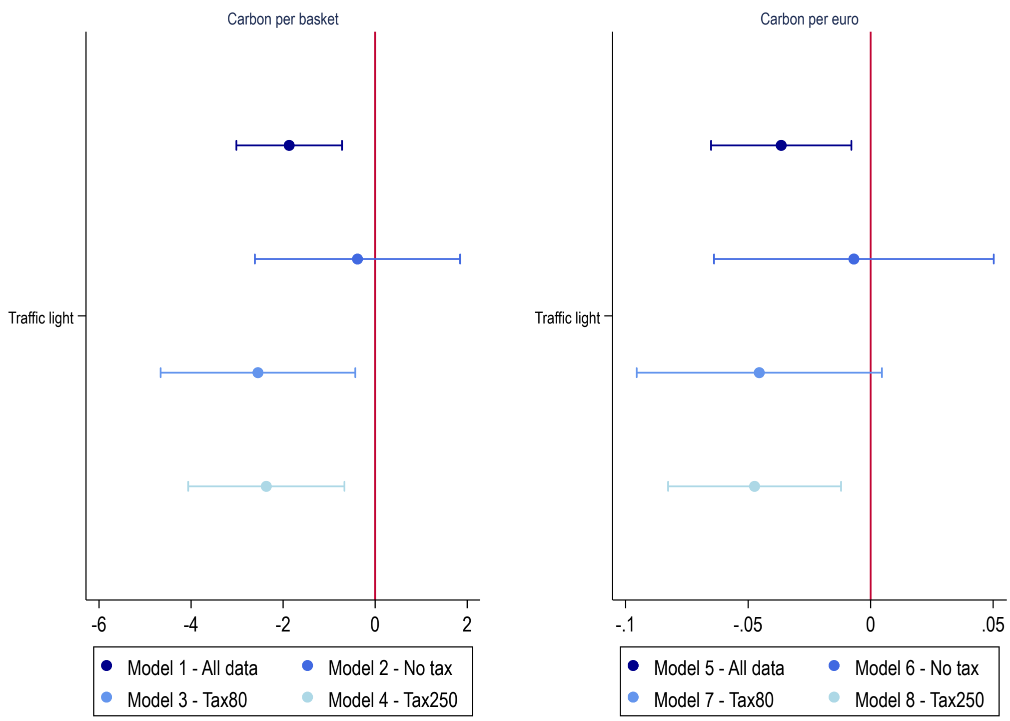
We confirm our previous findings: only the combination of the high tax with the nudge (variable *Traffic light× Tax €250/ton of* CO_2*eq.*) results in a significant decrease of the carbon footprint (in kg) at the basket level (negative and significant effect at the 1% level, model 1). Conversely, either the high tax alone (variable *Tax €250/ton of* CO_2*eq.*) or the combinations of the taxes and the nudge (variables *Traffic light× Tax €80/ton of* CO_2*eq.* and *Traffic light× Tax €250/ton of* CO_2*eq.*) results in a significant decrease of the carbon footprint (in kg) per euro spent (negative and significant coefficients at the 1% level, model 2). In both models, carbon emissions are lower for older consumers, and for households reporting a higher income; and higher in households with more children and with consumers who eat more meat than the French average.

[22]The full results are reported in Table D.1 in Appendix D.

3.3 Tax and nudge: substitute or complement?

To gain some insights into the complementarity between the nudge and the carbon tax, we run additional analyses to quantify the added effect of the nudge when there is a tax. We estimate eight different OLS regressions to explain the carbon footprint per basket and per euro, with a dummy for the nudge treatment (as well as the same control variables as before). First, we assess the global effect of the nudge, considering all data and abstracting from the tax treatments. We then assess the effect of the nudge in the absence of any tax, with the low tax (*Tax80*) and with the high tax (*Tax250*), by restricting the sample to the corresponding treatments in each regression: *Control+TL* in the absence of tax; *Tax80* and *TL+Tax80* for the low tax case; *Tax250*) and *TL+Tax250* for the high tax case. In all these models, the reference is the group in which no traffic light is implemented.[23] The results are reported in Figure 3.[24]

Figure 3: Added value of the traffic light with respect to the tax



Note: This figure reports the effect of being exposed to the traffic light on the carbon footprint per basket (left panel) and on the carbon footprint per euro spent (right panel), per tax level (no tax, *Tax80* and *Tax250*). Treatment effects come from OLS regressions of the outcome variable (the carbon footprint per basket and the carbon footprint per euro spent). Although not presented in these figures, the estimations controlled for age, gender, number of children, income, environmental score and meat consumption. 95% confidence intervals are reported.

[23] When assessing the global effect of the nudge considering all data, the reference group comprises the *Control*, *Tax80* and *Tax250* experimental conditions. When we assess the effect of the nudge in the absence of any tax, the reference group is the *Control* experimental condition. Finally, when assessing the effect of the nudge when being taxed, the reference group is the *Tax80* (respectively *Tax250*) experimental condition with the low (respectively high) tax.

[24] The full results are reported in Table D.2 in Appendix D.

The additional effect of the nudge is very similar on the carbon footprint per basket and the carbon footprint per euro. First, abstracting from any other treatment, having the nudge implemented has a negative and significant effect on the carbon footprint per basket and per euro. Second, we show that this effect comes from the data in complementarity with the tax (whether low or high). In the absence of a tax, there is no significant effect of the nudge, whereas we do detect a negative and significant effect of the nudge when a tax is applied (but at the 10% level only with the low tax for the carbon footprint per euro). These observations therefore confirm that, in our case, the nudge and the tax (and, in particular, the high one) are complementary to each other.

4 How taxes and nudges influence the choices of specific products?

4.1 Impact by product category

In this section we investigate how nudge and taxes influence the purchase of green, orange and red products.

In Table 3, we report the shares of green, orange and red products bought per treatment, as well as the results of median tests, t-tests and standard deviation tests. Our main observation is that the mean share of green (red) products bought significantly increases (decreases) only when the traffic light is implemented (t-test significant at the 1% level for the TL and TL+Tax250 treatments, and at the 10% level for the TL+Tax80 treatment). The mean share of orange products significantly decreases in the TL+Tax250 treatment only (t-test significant at the 5% level).

Table 3: Share of green, orange and red products bought per treatment

	Green products			Orange products			Red products		
	Median[1]	Mean[2]	St. dev.[3]	Median[1]	Mean[2]	St. dev.[3]	Median[1]	Mean[2]	St. dev.[3]
Control *N = 109*	0.364	0.350	0.142	0.333	0.338	0.119	0.313	0.312	0.130
Tax80 *N = 102*	0.363	0.361	0.149	0.312	0.325	0.120	0.297	0.314	0.134
Tax250 *N = 140*	0.350	0.350	0.137	0.333	0.339	0.114	0.289	0.311	0.133
TL *N = 92*	0.400	0.418***	0.162	0.326	0.333	0.127	0.239**	0.250***	0.131
TL+Tax80 *N = 97*	0.400	0.391*	0.154	0.333	0.333	0.122	0.273	0.276*	0.141
TL+Tax250 *N = 120*	0.412*	0.433***	0.181***	0.310	0.307**	0.114	0.229***	0.260***	0.157***

Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
[1] Two-sample Wilcoxon rank-sum test comparing each treatment to the control group.
[2] T-test comparing the means of each treatment vs. the mean of the control group.
[3] SD-test comparing the standard deviation of each treatment vs. the one for the control group.

Overall, these observations indicate that the nudge alone and the combination of the nudge and the high tax (Tax250+TL) result in significant substitution effects between food categories: subjects reduce their consumption of high-carbon foods (red category) and increase that of low-carbon foods (green category).

We then analyze the share of budget spent for each category of products at the basket level, starting with the green products. The densities between the different treatments are very similar, with a concentration around 40% of green products (stock-keeping unit, or SKU) in the basket, which account for around 27% of the budget in the absence of traffic-light labelling, and 30-33% in the presence of labelling. The expenditure share of orange products sits in the interval 31-32% across all groups, while the same share for red products is around 41% without a traffic-light label, declining to 35-37% in the presence of a label. A Kolmogorov-Smirnov test does not detect any significant difference in the distribution of budget share between the control group and the treatment groups (tests are provided in Appendix S2, Table S2.2). The only significant differences we find in the control group are for green and red products bought by consumers shopping with a label and a carbon tax of €250, and green products in the presence of traffic-light labelling alone (all significant at the 5% level).

4.2 Price-elasticity per category of product

To better assess how the price signal induced by the carbon tax impacts consumers' purchases, we investigate how they respond to a price change in each of the three product categories. We take advantage of price heterogeneity induced by the tax treatment to estimate price-elasticities. In Appendix E, we report the estimation of the demand function, relying on a linear Almost Ideal Demand System (AIDS) (Deaton and Muellbauer, 1980), used to compute the price-elasticities. Estimates of price-elasticities and their bootstrapped (200 replications) standard errors are reported in Table 4. These elasticities are estimated at the level of individual consumers, while Table S4.2 in Appendix E reports the average across the sample.

Table 4: Estimated price-elasticities in the experimental store

		Elasticity		
		Green	Orange	Red
Equation	Green	-1.0069***	0.1269***	-0.1200**
		(0.0622)	(0.0406)	(0.0545)
	Orange	0.1021***	-1.1025***	0.0003
		(0.0325)	(0.0336)	(0.0330)
	Red	-0.0845**	0.0003	-0.9158***
		(0.0389)	(0.0377)	(0.0500)

Bootstrapped standard errors (200 replications) are reported in parentheses.
Significance levels: $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

The results indicate that all categories are own-price elastic, with the green category showing the lowest sensitivity to own-price changes. At the same time, green and orange goods weakly substitute each other, with a cross-price elasticity being positive and significant, albeit small (0.10-0.12). Conversely, green and red products complement each other weakly, with a small but significant price elasticity (0.08-0.12). Red-orange cross-price elasticities are small and not significant.

These results indicate that consumers struggled to substitute products between categories, due to the different characteristics of food within each category – something that is reflected by the low elasticities. This limited ability to respond to price changes causes relatively small reductions in consumption with the price increase induced by the carbon tax. Furthermore, the complementarity between green and red goods indicates that consumers tend to purchase items from both categories jointly. It suggests that it is particularly hard to substitute green products for red products when carbon is taxed. Consumers rather pay the higher priced red products or replace them with lower carbon-intensive products within the red categories.

These estimations are consistent with our theoretical predictions regarding the impact of the nudge alone, detailed in the next Section: it significantly decreases the consumption of red products and increases the consumption of green ones (coefficients significant at the 1% level). As for the impact of the tax, we find that it is mostly driven by a reduction in consumption (as indicated by own-price elasticities), and a more modest substitution from orange to green products, with no substitution away from red products. This result explains why we find a significant effect only when the tax is high (€250), as only a sufficiently high carbon tax is effective in achieving visible changes in carbon emissions in the presence of inelastic substitution.

Regarding individual characteristics, the only robust effects we find are that females operate the same substitution highlighted above between red and green food products (significant at the 1% level), while those who eat more meat than average operate the opposite substitution in their basket (significant at the 1% level).

5 Welfare analysis

5.1 A behavioral economic framework to measure welfare

We now rely on consumer theory to ground our welfare analysis. We consider the three types of product: green g, orange o and red r. Let $u(\boldsymbol{x})$ denote the utility from consuming the basket $\boldsymbol{x} = (x_g, x_o, x_r)$, where x_i is the quantity of product (of type) i for $i = g, o, r$. Utility is expressed in euros. The function u has the properties of being twice continuously differentiable, increasing and concave. Prices are denoted p_i for $i = g, o, r$. As usual in behavioral economics (see e.g., Allcott and Kessler 2019), we distinguish between *decision utility* and *experienced utility*. The decision utility guides the shopping choices, while the experienced utility measures the consumer's welfare.

Under the materially selfish "homo economicus" approach, the decision and experienced utility coincide. Consumers make choices under full information about products and do not care about the climate impact of their consumption choices. Hence, nudging them with labels should not have an impact on their consumption choices or their welfare. The utility $u(\boldsymbol{x})$ determines both the consumers' choices and their welfare. Denoting the budget to spend as m, a consumer maximizes $u(\boldsymbol{x})+m-\sum_i p_i x_i$ with or without labels,[25], which leads them to equalize marginal utility to prices:

$$\frac{\partial u(\boldsymbol{x})}{\partial x_i} = p_i, \tag{1}$$

for $i = g, o, r$. The purchased basket satisfying (1), which we denote $\boldsymbol{x}^e$, leads to a welfare

[25]Note that in our experiment the budget not spent is lost to consumers so that they do not value the money per se. A more consistent modeling assumption would be that consumers maximize their utility $u(\boldsymbol{x})$ under the constraint $\sum_i p_i x_i \leq m$. That would lead to similar first-order conditions (and then willingness to pay, or inverse demand function, for each type of product), *modulo* a Lagrangian multiplier associated with the budget constraint. The same would apply to the nudgeable and moral consumers defined below. Our simpler formulation avoids carrying out Lagrangian multipliers throughout our analysis without impacting the welfare analysis.

of $u(\boldsymbol{x}^e)$.

From (1), we can then define the inverse demand function, or willingness to pay (WTP), for a product of type i by:

$$P_i^e(\boldsymbol{x}) = \frac{\partial u(\boldsymbol{x})}{\partial x_i}. \tag{2}$$

5.2 Nudgeable consumers

Our first behavioral approach is inspired by previous studies which assume that nudges modify decision utility, through some behavioral bias. In this respect, it departs from experienced utility (Allcott and Taubinsky, 2015; Allcott and Kessler, 2019; Farhi and Gabaix, 2020; List et al., 2023). As a result, decisions change in a way which is welfare-improving for society but not individually. In our experiment, *nudgeable consumers* react to the label by buying more green products and fewer red products. Let us assume in their decision utility when they shop, but not (necessarily) in their experienced utility, nudgeable consumers assign a benefit $b > 0$ to purchasing the product labeled as green, and a cost $c > 0$ to buying the product labeled as red.

The parameters b and c quantify the extent to which people can be nudged, as defined in Farhi and Gabaix (2020), under several behavioral motivations. For instance, consumers might follow simple heuristics or rule of thumb when they do their shopping (Thaler and Sunstein, 2008). One of such rule is to use the label color to select between 480 products, and thus favor green products and avoid red ones. Alternatively, consumers might enjoy a warm-glow utility from buying products labeled as green (Ambec and De Donder, 2022). They might feel guilt or shame for not doing so and instead buying products labeled as red. Their motivation is not necessarily the carbon footprint *per se*, but rather the feeling that green products are "good" and red product are "bad".[26]

Nudgeable consumers maximize their decision utility $u(\boldsymbol{x}) + bx_g - cx_r + m - \sum_i p_i x_i$ with respect to $\boldsymbol{x}$ when they shop in the store with labeled products, which leads to the following first-order conditions:

$$\begin{aligned} \frac{\partial u(\boldsymbol{x})}{\partial x_g} + b &= p_g, \\ \frac{\partial u(\boldsymbol{x})}{\partial x_r} - c &= p_r. \end{aligned} \tag{3}$$

Nudgeable consumers equalize their marginal utility plus b to the price of green products, and their marginal utility minus c to the price of red products.

The nudgeable consumers' WTP for the green and red products when nudged with labels are derived from (3) as follows:

$$P_g^n(\boldsymbol{x}) = \frac{\partial u(\boldsymbol{x})}{\partial x_g} + b, \tag{4}$$

$$P_r^n(\boldsymbol{x}) = \frac{\partial u(\boldsymbol{x})}{\partial x_g} - c. \tag{5}$$

[26] Orange products are assumed to be "neutral".

Using (2), we can write $P_g^n(\boldsymbol{x}) = P_g^e(\boldsymbol{x}) + b$ and $P_r^n(\boldsymbol{x}) = P_r^e(\boldsymbol{x}) - c$. The label shifts the WTP for green products upward by b, and the WTP of red products downward by c. The magnitude of this shift is determined by the behavioral wedge b and c for green and red products respectively. It is illustrated in Figures 4 and 5 below with a linear demand function per product.[27]

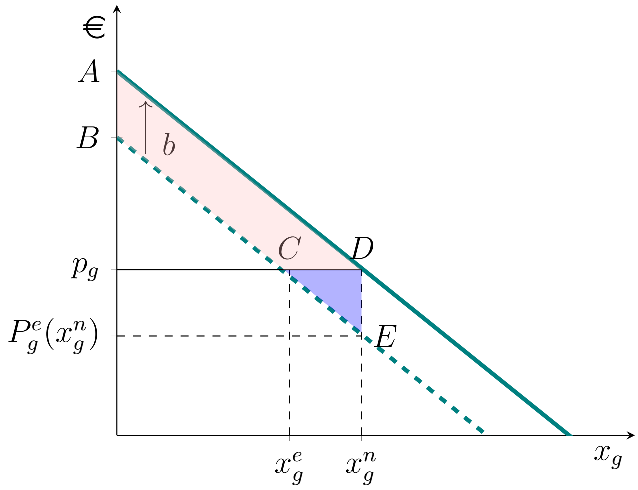


Figure 4: Welfare change for a nudgeable consumer when a product is labeled as green.

The green dashed line is the WTP for an unlabeled green product $P_g^e(\boldsymbol{x})$; it is the inverse demand function without the nudge. The solid green line is the WTP for the product labeled as green $P_g^n(\boldsymbol{x})$,; it is the inverse demand function with the nudge.

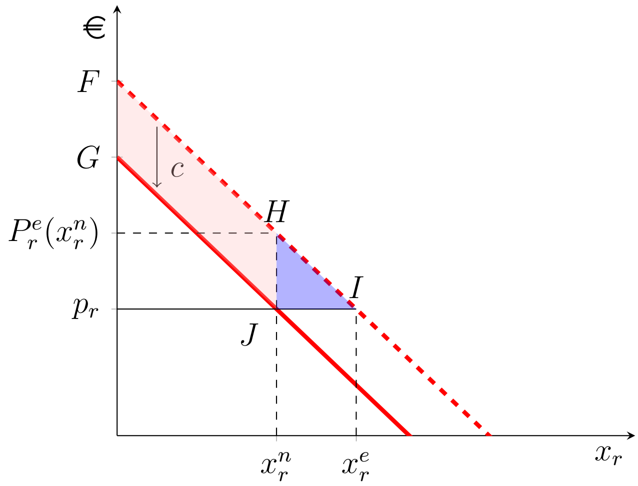


Figure 5: Welfare loss for a nudgeable consumers when a product is labeled as red.

The red dashed line is the WTP for an unlabeled red product $P_g^e(\boldsymbol{x})$. It is the inverse demand function without the nudge. The plain red line is the WTP for the product labeled as green $P_g^n(\boldsymbol{x})$, the inverse demand function with the nudge.

[27]We abstract from the cross-demand elasticity among products.

As usual, welfare is measured with the consumer's surplus. We need to select the utility that measures welfare, i.e., *experienced utility*, which might differ from the *decision utility*. To avoid taking a stand on the experienced utility, we consider two alternative assumptions. First, following Harsanyi (1995), Goodin (1986) and Diamond (2006), we "launder" preferences by removing b and c from the utility to measure welfare. The experienced utility is $u(\boldsymbol{x})$. It differs from the decision utility by the parameters b and c. The consumption basket $\boldsymbol{x}^n$ maximizes the decision utility, but not the experienced utility. Hence, the nudge leads to an allocative inefficiency: the optimal basket for the consumer (but not for society) is $\boldsymbol{x}^e$, not $\boldsymbol{x}^n$. Comparing (3) with (1) shows that b and c can be interpreted as the "behavioral wedge" defined in Farhi and Gabaix (2020), that is, the gap between marginal experienced utility and the price. It reduces their welfare by shifting their consumption choice away from the one that maximize their own welfare. Denoting the shopping basket maximizing the decision utility of a nudgeable consumer as $\boldsymbol{x}^n$, we obtain $u(\boldsymbol{x}^n) < u(\boldsymbol{x}^e)$. Green products are over-consumed, and red products are under-consumed: $x_g^n > x_g^e$ and $x_r^n < x_r^n$.

We call the welfare impact *allocative* since the nudge leads to suboptimal consumption without changing the intrinsic value of products for consumers (the experienced utility). It is graphed by the blue triangles labeled CDE in Figure 4 and HIJ in Figure 5. For green products, all units exceeding x_g^e are bought at a price higher than the consumers' WTP for those units according to their experienced utility. The blue CDE area sums up the difference between the price p_g and the "real" WTP for those units $P_g^e(x_g)$. Symmetrically, for red products, all units in-between x_r^n and x_r^e are not purchased despite being valued more than the price p_r with the experienced utility. The blue HIJ area sums up the difference between the price p_r and the "real" WTP for those units $P_r^e(x_r)$.

Second, drawing on other studies such as Allcott and Kessler (2019), one can choose not to launder preferences by including b and c in the experienced utility.[28] The nudge changes both the decision and the experienced utility to $u(\boldsymbol{x}) + bx_g - cx_r$. The consumption basket x^n is individually optimal. The nudge changes preferences and thus WTP. The marginal valuation of green products is increased by b. It leads to a welfare gain captured by the area $ABCD$ in Figure 4, the pink parallelogram. In contrast, the marginal valuation of red products is reduced by c. It leads to a welfare loss captured by the area $FGJI$ in Figure 5. We call the welfare impact *valuation* because the nudge changes the valuation of products (that is, the experienced utility).

5.3 Moral consumers

Our second behavioral approach is based on the assumption that consumers care about their impact on the climate[29], and that the label provides them with useful information about this impact. As consumers who have an imperfect knowledge about the carbon footprint of products are then likely to shop differently once they see the labels, they make better choices and their welfare is improved. Contrary to the nudgeable consumers, they enjoy a welfare gain from being nudged.

The *moral consumers* are internalizing part, if not all, of the climate impact of their consumption.[30] They maximize the social impact of their action, given the available

[28] Allcott and Taubinsky (2015) refers to this welfare shift c as a "moral tax" of the nudge.

[29] Empirical evidence of such behavior is reported in Rodemeier (2023), Kaufmann et al. (2024).

[30] A motive for doing so is being guided by Emmanuel Kant's categorical imperative: "*Act only according to that maxim whereby you can at the same time will that it should become a universal law*", see

information on products.

Being imperfectly informed about the carbon footprint of products, moral consumers evaluate it based on their expectation. Without the label, they assign to any product the average carbon footprint in the shop, that is, that of orange products. We denote it as o. With the label, they know that the green products have a carbon footprint of $g < o$ on average while red products have one of $r > o$ ton of CO_2eq per kilo of product. Let δ denote the value that a moral consumer assigns to a ton of carbon dioxide embedded in their shopping basket.[31] Without the label, moral consumers maximize their decision utility $u(\boldsymbol{x}) - \delta o \sum_i x_i + m - \sum_i p_i x_i$ when they do their grocery shopping. The first-order conditions yield:

$$\frac{\partial u(\boldsymbol{x})}{\partial x_i} = p_i + o\delta, \tag{6}$$

for the three types of products $i = g, o, r$. Moral consumers equalize their marginal utility of consumption to the price of products and the perceived climate impact of their consumption measured in euros. With the superscript "0" referring a treatment without label, moral consumers' WTP for a product of type i is thus:

$$P_i^0(\boldsymbol{x}) = \frac{\partial u(\boldsymbol{x})}{\partial x_i} - o\delta. \tag{7}$$

With the label, moral consumers maximize $u(\boldsymbol{x}) - \delta[g x_r + o x_o + r x_r] + m - \sum_i p_i x_i$.[32] The first-order conditions of the maximization program yields:

$$\begin{aligned} \frac{\partial u(\boldsymbol{x})}{\partial x_g} &= p_g + g\delta, \\ \frac{\partial u(\boldsymbol{x})}{\partial x_o} &= p_o + o\delta, \\ \frac{\partial u(\boldsymbol{x})}{\partial x_r} &= p_r + r\delta. \end{aligned} \tag{8}$$

Moral consumers equalize the marginal utility from consuming the good to its price and the marginal climate impact of they internalize. The shopping basket $\boldsymbol{x}^n$ solution to the maximization problem with nudge is optimal for moral consumers.

Comparing (7) and (8) shows that the "behavioral wedge" from nudging moral consumers is the unobserved marginal damage: $(o - g)\delta$ for green products and $(r - o)\delta$ for red products. Like for nudgeable consumers, the label makes moral consumers buy more green products and fewer red products. As long as $g < o < r$, we have $x_g^m > x_g^0$ and $x_r^m < x_r^0$ where superscript 'm' refers to moral consumers and a treatment with nudge. However, the motivation and the magnitude of the wedge are not related to some psychological reward or cost but rather to differences in the carbon footprint.

From (8), we can derive the following moral consumer's WTP for the three types of

Laffont (1975) and Alger and Weibull (2013).

[31] Under the internalization of their climate impact, δ is the social cost of carbon. Under partial internalization, δ is lower than the social cost of carbon.

[32] Although the categorization of products does not give the exact carbon footprint, for simplicity we ignore the welfare loss due to intra-categorical differences in the carbon footprint. This assumption does not qualitatively alter our results.

products with a label:

$$
\begin{aligned}
P_g^m(\boldsymbol{x}) &= \frac{\partial u(\boldsymbol{x})}{\partial x_g} - g\delta, \\
P_o^m(\boldsymbol{x}) &= \frac{\partial u(\boldsymbol{x})}{\partial x_o} - o\delta = P_o^0(\boldsymbol{x}), \\
P_r^m(\boldsymbol{x}) &= \frac{\partial u(\boldsymbol{x})}{\partial x_r} - r\delta.
\end{aligned} \tag{9}
$$

Combining (7) with (8), we conclude that $P_g^m(\boldsymbol{x}) - P_g^0(\boldsymbol{x}) = (g - o)\delta$ and $P_g^m(\boldsymbol{x}) - P_g^0(\boldsymbol{x}) = -(r - o)\delta$. The label shifts the WTP for green products upward by $(g - o)\delta$ and the WTP of red products downward by $(r - o)\delta$.

The variation of WTP and demand with the label is graphed in Figures 6 and 7 for green and red products respectively.[33]

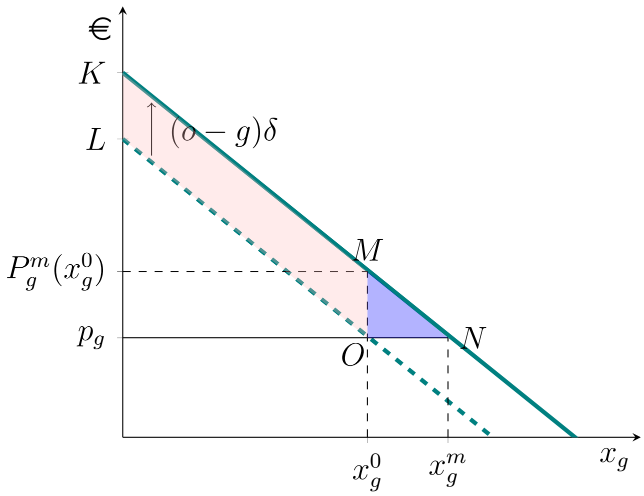


Figure 6: Welfare gain for a moral consumer when informed that a product is green.

The dashed green line is the WTP for green products when they are not labeled as green $P_g^0(\boldsymbol{x})$ (the inverse demand function without the nudge). The solid green line is the WTP for green products labeled as green $P_g^m(\boldsymbol{x})$ for moral consumers (the inverse demand function with nudge). The nudge moves the inverse demand function for green products from the dashed line to the plain line. The same logic applies for red products in Figure 7 below.

[33] Note that we abstract for the for cross-demand elasticity among products.

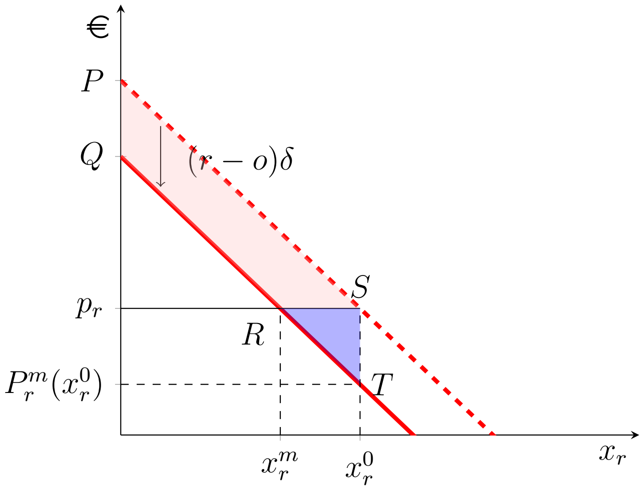


Figure 7: Welfare change for a moral consumer from being informed that a product is red.

As above, we distinguish between allocative and valuation welfare change based on how experienced utility is defined, and whether it changes with the nudge. The *allocative* welfare change compares welfare with and without a label with the same metric: the experienced utility $u(\boldsymbol{x}) - \delta[gx_r + ox_o + rx_r]$. It is the welfare gain from consuming $\boldsymbol{x}^m$ instead of $\boldsymbol{x}^0$ expressed in a consumer's surplus with the demand functions with nudge $P_g^m(\boldsymbol{x})$ and $P_r^m(\boldsymbol{x})$. The allocative welfare changes are the blue triangles MNO for green products and RST for red products in Figures 6 and 7 respectively. For green products, all units between x_g^0 and x_g^m have a higher WTP than the price, and should therefore be bought. The welfare gain from buying those units when the green label is seen is the difference between the WTP $P_g^m(\boldsymbol{x})$ and the price p_g in Figure 6. Inversely, for red products, all units between x_r^m and x_r^0 have a lower WTP than the price, and should theefore not be purchased. The welfare gain from not purchasing those units thanks to the red label is the difference between the price p_r and the WTP $P_r^m(\boldsymbol{x})$ in Figure 7.

The *valuation* welfare change assumes that the experienced utility is modified by revealing the true carbon footprint on all units consumed with the label.[34] It is graphed by the sum of the pink parallelograms and blue triangles, the area $KNOL$ and $PSTQ$ in Figures 6 and 7 respectively. The welfare change is a gain for green products because the carbon-footprint is lower than expected. In contrast, the welfare change is a loss for red products because moral consumers learn that the carbon footprint of red products is higher than expected.

Our behavioral assumptions are summarized in Tables 5 and below.

Table 5: Experienced utility by consumer's type and by measure of welfare change.

Type of consumer	Allocative welfare change	Valuation welfare change
Nudgeable	$u(\boldsymbol{x})$ with and without nudge	$u(\boldsymbol{x})$ without nudge $u(\boldsymbol{x}) + bx_g - cx_r$ with nudge
Moral	$u(\boldsymbol{x}) - \delta\left[gx_g + ox_o + rx_r\right]$ with and without nudge	$u(\boldsymbol{x}) - \delta o \sum_i x_i$ without nudge $u(\boldsymbol{x}) - \delta\left[gx_g + ox_o + rx_r\right]$ with nudge

[34]The experienced utility is $u(\boldsymbol{x}) - \delta o \sum_i x_i$ without label and $u(\boldsymbol{x}) - \delta[gx_r + ox_o + rx_r]$ with the label.

For nudgeable consumers, the allocative welfare change is a loss for both types of product (green or red) because their decision utility differs from their experienced utility. By contrast, their valuation welfare change is a gain with the green products only because they enjoy extra utility b in their experienced utility with nudge. These consumers' valuation welfare change is a loss with the red products because they incur a disutility c for consuming red products in their experienced utility with the nudge.

For moral consumers, the allocative welfare change is a gain for both types of products (green or red) because their decision utility coincides with their experienced utility. By contrast, their valuation welfare change is a gain for the green products because moral consumers value the lower-than-expected carbon-footprint on green products in their experienced utility. It is a loss for the red products because they internalize the higher than expected carbon-footprint of red products in their experienced utility.

5.4 Welfare estimation

Now that we have a theoretically grounded measure of consumers' welfare, we perform a back-of-the-envelope estimation of how the traffic-light label impacts it, closely following the graphs presented in the previous section.

First, we estimate the WTPs reported Figures 4-7: what the WTP for the basket purchased in the shop with the label, $\boldsymbol{x}^n$, would be without a label for nudgeable consumers; and what the WTP for the basket purchased in the shop without the label, $\boldsymbol{x}^e$, would be if it had a label for moral consumers. To simplify the analysis, we mirror the graphs in Section 5 by assuming a linear system of demand equations in the form

$$Q_l = \alpha_l + \beta_l \mathbb{1}_{TL} + \gamma_g p_g + \gamma_o p_o + \gamma_r p_r + \delta_l D + u_l, \tag{10}$$

where $\mathbb{1}_{TL}$ is a dummy for the nudge (traffic-light label) treatment, D is a vector of control variables (age, gender, number of children, environmental score and a dummy for meat consumption), and u_l is the error term. Using (10), we compute the inverse demand curves (see Appendix S5 for the details). Table F.1 in Appendix F presents the coefficient estimates for the linear model, which are consistent with the AIDS estimate in Table E.1.

5.4.1 Nudgeable Consumers

Once we have obtained the inverse demand functions, we set Q_l equal to the average quantity purchased by all participants where a nudge is present, to obtain an estimate of $P_g^e(\boldsymbol{x}^n)$ and $P_r^e(\boldsymbol{x}^n)$ setting $\mathbb{1}_{TL} = 1$. We then set Q_l equal to the average quantity purchased by all participants without a nudge setting $\mathbb{1}_{TL} = 0$, to obtain the same estimate for p_g and p_r, which are the WTP for the quantity purchased without nudge. We compute $P_l^e(\boldsymbol{x}^n)$, p_g and p_r for the different carbon tax levels, multiply it with by the variation of quantity $x_l^n - x_l^e$ and divide it by 2 to obtain the area of the blue triangle in Figures 4 or 5, i.e., the allocative welfare change on product $l = r, g$.

Using a similar approach we can also estimate the valuation welfare change. To do so, we use the same inverse demand function to determine the price corresponding to $x_l = 0$, that is, point A and B in Figure 4, and points F and G in Figure 5. Once these WTPs are estimated, for green products we estimate welfare as the difference between the triangles ADp_g and BCp_g, in Figure 4. For red products, welfare is instead estimated more straightforwardly as the the difference in the area of the triangles FIp_r and GJp_r in Figure 5.

5.4.2 Moral consumers

For moral consumers, we use a similar approach as for nudgeable consumers. In the inverse demand function, we set Q_l equal to the average quantity purchased by all participants where a nudge is absent, to obtain an estimate of $P_g^m(\boldsymbol{x}^e)$ and $P_r^m(\boldsymbol{x}^e)$, setting $\mathbb{1}_{TL} = 0$. We then set Q_l equal to the average quantity purchased in all participants in treatments with a nudge to obtain the same estimate for p_g and p_r setting $\mathbb{1}_{TL} = 1$. We compute $P_l^e(\boldsymbol{x}^n)$, p_g and p_r for the different carbon tax levels, multiply it with by the variation of quantity $x_l^n - x_l^e$ and divide by 2 to obtain the area of the blue triangle in Figures 6 and 7, i.e., the allocative welfare change on product $l = r, g$.

As before, the valuation welfare change is estimated by first using the inverse demand function to determine the price corresponding to $x_l = 0$, that is, point K and L in Figure 6, and points P and Q in Figure 7. Once these prices are estimated, for green products we estimate valuation welfare directly as the difference between the triangles KNp_g and LOp_g in 6. For red products, valuation welfare is instead estimated as the the difference in the area of the triangles PSp_r and QRp_r in Figure 7, plus the blue triangle RST in Figure 7 , the latter being the allocative welfare estimate explained above.

5.4.3 Welfare estimates

In Table Table F.2, Appendix F, we report in the estimation of the WTP and prices used in the estimation of welfare. The observed average prices for each category in treatment with or without nudges can be found in Table F.3.

In Table 6 below, we present our estimates of the welfare change in absolute value (the areas highlighted in Figures 4 to 7). Following our previous theoretical analysis, we focus on green and red products, and do not present the results for orange products, for which we do not have specific predictions. Moreover, we do not present the results for the *Tax80* treatments because we observe a decrease in the purchase of green products with the label (and thus a decreasing WTP for green products with the label), which is not consistent with our theoretical analysis. However, for full transparency, we report all the welfare estimates, including those for orange products and *Tax80* treatments, in Table S5.1 in Appendix S5.

As expected, the valuation welfare changes are always greater in absolute value than the allocative welfare changes. Since the tax reinforces the impact of the label, the size of the welfare change is always higher in absolute value in the high-tax treatment for both the allocative and the valuation welfare (both the variation of WTP and the quantities purchased are higher). However, the valuation welfare for the high-tax seems too high given the value of the shopping basket.

The above impact on consumers' surplus should be contrasted with the climate impact of the nudge. Since the nudge is effective in reducing significantly carbon footprint per basket only if combined with the high tax (see Table 2), we focus on the treatment with the nudge and the €250 tax (*TL+Tax250*) that we compare with the control treatment (*Control*). We obtain a reduction of the carbon footprint per basket, of 29.79-27.58=2,21 kg (see Table 2) on average, which, with a social cost of carbon (SCC) of €250 per

Table 6: Welfare change estimates for nudgeable and moral consumers (in €).

Consumer type	Welfare type	Product type	No tax		Tax250	
			Value	S.E.	Value	S.E.
Nudgeable	Allocative	Green	0.9739	1.5912	1.7573	1.5416
		Red	0.0707	0.2289	-1.0062*	0.5312
	Valuation	Green	11.5406	7.4886	12.4396*	6.7982
		Red	3.1733**	1.5553	5.4657***	1.2667
Moral	Allocative	Green	0.9739	1.5912	1.7573	1.5416
		Red	0.0707	0.2289	1.0062*	0.5312
	Valuation	Green	11.5406	7.4886	12.4396*	6.7982
		Red	3.1733**	1.5553	5.4657***	1.2667
Observations			201		260	

S.E. refer to bootstrapped standard errors (1,000 replications).
Significance levels: $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

ton, is valued at €0.55. With higher SSC such as Bilal and Känzig (2024)'s estimate of \$1367(€1173) per ton of CO_2, the value goes up to about €2.6.[35]

To obtain the full welfare impact of the label and high-tax regulation, we combine those estimates with the consumer's surplus welfare impact focusing on the allocative welfare change. For nudgeable consumers, we obtain the net welfare impact of the label. The consumer's surplus welfare losses adds up to €2.7635 (1.7573 + 1.0062) which, net of the climate change mitigation valued at €0.55 or €2.6, depending on the SSC, leads to a welfare loss of €2.2135 or €0.1635 respectively. The welfare loss is about 4.7% and 0.3% of the €47 budget respectively, about 5 cents or less than 1 cent per euro.

For moral consumers, the allocative welfare change shows how much they internalize their climate impact once they have been better informed by the label about the carbon footprint. It reveals how much they value one ton of CO_2 emitted. Moral consumers value at €2.7635 (1.7573+1.0062) the 0.00221 tons of CO_2 avoided per basket when they know the carbon-footprint of the products. Hence, each ton of CO_2 avoided is valued at €2.7635/0.00221= €1250.

6 Conclusion

Limiting global warming requires the implementation of public policies aimed at curbing greenhouse gas emissions. Carbon should therefore be priced, and citizens should be informed about the carbon footprint of products, economic activities, investment strategies, and transportation choices. Both types of policy impact market supply and demand.

Our study focused on the demand side by running an online shopping experiment with a representative sample of products. We provide evidence that a carbon tax alone is not sufficient to reduce carbon emissions if the consumers' purchasing power is maintained. An hypothetical average consumer would be likely to buy the same shopping basket that they would have bought without a carbon tax. By recycling the revenue from pricing carbon to the consumers, we make the carbon tax budget-neutral.

As expected, the carbon footprint per euro spent is lower with a carbon tax because

[35] It is a relatively small share of the €40 budget, about 1.4% and 13% with a social cost of carbon of €250 and \$1367 (€1173) respectively.

consumers can buy fewer products overall. However, the carbon footprint per basket purchased is not significantly different in treatments with a tax alone. It is only when the tax is complemented with a label on the carbon footprint of each product that we find a significant reduction of the carbon footprint per basket. Our results suggest a strong complementarity between the two policy instruments. The label on the carbon footprint of products provides consumers with useful information that helps them to make better choices for the climate. The price signal is not sufficient to change consumers' habits toward the reduction of carbon emissions.

Climate policies such as taxing or labeling goods based on their carbon footprint should be assessed not only on their impact on carbon emissions, but also on the consumers' welfare. The carbon footprint labels modify consumers' behavior and, potentially, their preferences. Based on behavioral economic theory, we posit a method to evaluate the consumer's welfare change when they are nudged by the label. This method relies on an estimation of the demand functions with and without the nudge, which allows to measure the variations of consumers' surplus. The method encompasses various motives and consequences of the behavioral response nudges.

The consumers that we call "nudgeable", they do not care about their climate impact, but still respond to the nudge by departing from their preferred food choices. The nudgeable consumer's loss of surplus is measured by the valuation of the "green" products that they over-consume and the "red" products that they under-consume. Our estimate of this loss of welfare is about €3.85, which is way above the reduction of the carbon footprint valued at €0.55 with a carbon price and a social cost of carbon of €250. The consumers that we call "moral" care about the climate impact of their behavior. Without information on the carbon footprint of products, they over-consume "red" products and under-consume "green" products. The label provides them with useful information to consume better. The increase of welfare from consuming a shopping basket more adapted to their preferences is about €3.85.[36] It reveals an internalization of a social cost of carbon of about €1742 per ton of CO2 avoided.

Obviously, our estimates are limited by our sample size. More precise and robust estimates could be obtained by scaling up the experiment. Furthermore, the method we use can be used to evaluate the impact on welfare of any non-monetary policies to address any environmental externalities and moral values on food consumption, such as contamination with pesticides, water use and fair trade. We plan to pursue this research agenda in the future.

[36]Note that we assumed linear demand in our evaluation, which implies that the allocative welfare change of nudgeable and moral consumers is the same by construction.

References

Alger, I. and J. W. Weibull (2013). Homo moralis - preference evolution under incomplete information and assortative matching. *Econometrica 81*(6), 2269–2301.

Allcott, H., D. Cohen, W. Morrison, and D. Taubinsky (2025). When do"nudges" increase welfare? *American Economic Review 115*, 1555–96.

Allcott, H. and J. B. Kessler (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics 11*(1), 236–276.

Allcott, H. and D. Taubinsky (2015). Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market. *American Economic Review 105*(8), 29501–2538.

Ambec, S. and P. De Donder (2022). Environmental policy with green consumerism. *Journal of Environmental Economics and Management 111*, 102584.

Barahona, N., C. Otero, and S. Oter (2023). Equilibrium effects of food labeling policies. *Econometrica 91*(3), 839–868.

Bilal, A. and D. Känzig (2024). The macroeconomic impact of climate change: Global vs. local temperature. NBER Working paper 32450.

Bilén, D. (2023). Do carbon labels cause consumers to reduce their emissions? evidence from a large-scale natural experiment. *Working Paper, University of Gothenburg*.

Bollinger, B. K. and W. R. Hartmann (2020). Information versus automation and implications for dynamic pricing. *Management Science 66*(1), 290–314.

Bonnet, C., Z. Bouamra-Mechemache, V. Réquillart, and N. Treich (2020). Regulating meat consumption to improve health, the environment and animal welfare. *Food Policy 97*, 101847.

Brunner, F., V. Kurz, D. Bryngelsson, and F. Hedenus (2018). Carbon label at a university restaurant–label implementation and evaluation. *Ecological economics 146*, 658–667.

Carlsson, F. and O. Johansson-Stenman (2024). Optimal prosocial nudging. Working paper, University of Gothenburg.

Chan, N. W. (2024). Pigouvian Policies under Behavioral Motives. *Journal of the Association of Environmental and Resource Economists 11*(1), 97–135.

Chetty, R., A. Looney, and K. Kroft (2009). Salience and taxation: Theory and evidence. *American Economic Review 99*(4), 1145–77.

Crosetto, P., A. Lacroix, L. Muller, and B. Ruffieux (2020). Nutritional and economic impact of five alternative front-of-pack nutritional labels: experimental evidence. *European Review of Agricultural Economics 47*(2), 785–818.

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

Crosetto, P., L. Muller, and B. Ruffieux (2025). Label or taxes: Why not both? testing nutritional mixed policies in the lab. *Journal of Economic Behavior & Organization 229*, 106825.

Deaton, A. and J. Muellbauer (1980). An almost ideal demand system. *The American economic review 70*(3), 312–326.

Diamond, P. (2006). Optimal tax treatment of private contributions for public goods with and without warm glow preferences. *Journal of Public Economics 90*, 897–019.

Douenne, T. and A. Fabre (2022). Yellow vests, pessimistic beliefs, and carbon tax aversion. *American Economic Journal: Economic Policy 14*(1), 81–110.

Dubois, P., P. Albuquerque, O. Allais, C. Bonnet, P. Bertail, P. Combris, S. Lahlou, N. Rigal, and d. P. C. Bernard Ruffieux a (2021). Effects of front-of-pack labels on the nutritional quality of supermarket food purchases: evidence from a large-scale randomized controlled trial. *Journal of the Academy of Marketing Science 49*, 119—138.

Dunlap, R. E., K. D. Van Liere, A. G. Mertig, and R. E. Jones (2000). New trends in measuring environmental attitudes: measuring endorsement of the new ecological paradigm: a revised nep scale. *Journal of social issues 56*(3), 425–442.

Enlund, J., D. Andersson, and F. Carlsson (2023). Individual carbon footprint reduction: Evidence from pro-environmental users of a carbon calculator. *Environmental and Resource Economics*, 1–35.

Farhi, E. and X. Gabaix (2020). Optimal taxation with behavioral agents. *American Economic Review 110*(1), 298–336.

Filippini, M. (1995). Swiss residential demand for electricity by time-of-use. *Resource and Energy Economics 17*(3), 281–290.

Frey, B. S. and F. Oberholzer-Gee (1997). The cost of price incentives: An empirical analysis of motivation crowding- out. *The American Economic Review 87*(4), 746–755.

Gneezi, U. and A. Rustichini (2000). A fine is a price. *The Journal of Legal Studies 29*(1), 1–17.

Goeschl, T. (2019). Cold case: The forensic economics of energy efficiency labels for domestic refrigeration appliances. *Energy Economics 84*, 104468.

Goodin, R. E. (1986). Laundering preferences. In *Jon Elster and Aanund Hylland, eds.,Foundations of Social Choice Theory.* Cambridge: Cambridge University Press.

Handziuk, Y. and S. Lova (2023). Carbon information, pricing, and bans: Evidence from a field experiment. *Working Paper, HEC Paris*.

Harsanyi, J. C. (1995). A theory of social values an a rule utilitarian theory of morality. *Social Choice and Welfare 12*, 319–344.

Herweg, F. and K. M. Schmidt (2022, 08). How to Regulate Carbon Emissions with Climate-Conscious Consumers. *The Economic Journal 132*(648), 2992–3019.

Ito, K., T. Ida, and M. Tanaka (2018). Moral suasion and economic incentives: Field experimental evidence from energy demand. *American Economic Journal: Economic Policy 10*(1), 240–267.

Jessoe, K. and D. Rapson (2014). Knowledge is (less) power: Experimental evidence from residential energy use. *American Economic Review 104*(4), 1417–1438.

Kanay, A. (2021). *Encourager la consommation durable : évaluer l'efficacité des instruments économiques et psychologiques dans le contexte des courses en ligne*. Ph. D. thesis, Université Toulouse II - Jean Jaures, France.

Kanay, A., D. Hilton, L. Charalambides, J.-B. Corrégé, E. Inaudi, L. Waroquier, and S. Cezera (2021). Making the carbon basket count: Goal setting promotes sustainable consumption in a simulated online supermarket. *Journal of Economic Psychology 83*, 102348.

Kaufmann, M., P. Andre, and B. Köszegi (2024). Understanding Markets with Socially Responsible Consumers. *The Quarterly Journal of Economics 139*(3), 1989–2035.

Laffont, J.-J. (1975). Macroeconomic constraints, economic efficiency and ethics: An introduction to kantian economics. *Economica 42*(168), 430–437.

Lanz, B., J.-D. Wurlod, L. Panzone, and T. Swanson (2018). The behavioral effect of pigovian regulation: Evidence from a field experiment. *Journal of Environmental Economics and Management 87*, 190–205.

List, J., G. Sun, and S. Roy (2023). Judging nudging: Understanding the welfare effects of nudges versus taxes information nudges, subsidies, and crowding out of attention: Field evidence from energy efficiency investments. Working paper, University of Chicago.

List, J. A., A. M. Shaikh, and Y. Xu (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics 22*, 773–793.

Lohmann, P. M., E. Gsottbauer, A. Doherty, and A. Kontoleon (2022). Do carbon footprint labels promote climatarian diets? evidence from a large-scale field experiment. *Journal of Environmental Economics and Management 114*, 102693.

Muller, L., A. Lacroix, J. L. Lusk, and B. Ruffieux (2017). Distributional impacts of fat taxes and thin subsidies. *The Economic Journal 127*(604), 2066—-2092.

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

Ouvrard, B. and A. Stenger (2024). Informational nudges and environmental preferences in a public goods game in networks. *Environmental Modeling & Assessment*, 1–22.

Panzone, L. A., A. Ulph, D. J. Zizzo, D. Hilton, and A. Clear (2018). The impact of environmental recall and carbon taxation on the carbon footprint of supermarket shopping. *Journal of Environmental Economics and Management*.

Perino, G., L. A. Panzone, and T. Swanson (2014). Motivation crowding in real consumption decisions: Who is messing with my groceries? *Economic Inquiry 52*(2), 592–607.

Pizzo, A., C. Gravert, J. M. Bauer, and L. Reich (2024). Carbon taxes crowd out climate concerns: Experimental evidence from sustainable consumer choices. CEBI Working paper, University of Copenhagen.

Quinet, A. (2019). La valeur de l'action pour le climat. Report, France Stratégie.

Rodemeier, M. (2023). Willingness to pay for carbon mitigation: Field evidence from the market for carbon offsets. *IZA Discussion Paper No. 15939*.

Thaler, R. and C. Sunstein (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness.* Yale University Press.

A Screenshots

Figure A.1: Screenshot of the online shopping platform in treatments without label

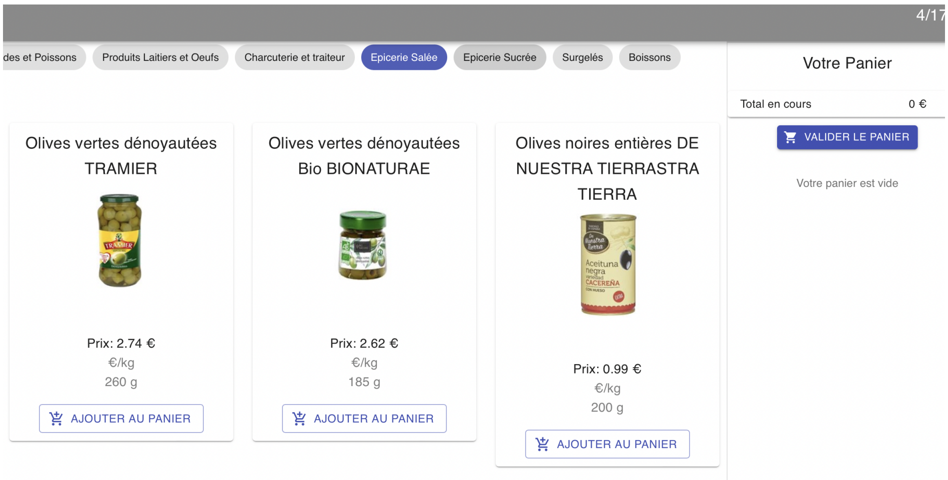


Figure A.2: Screenshot of the online shopping platform in treatments with labels (case of green products)

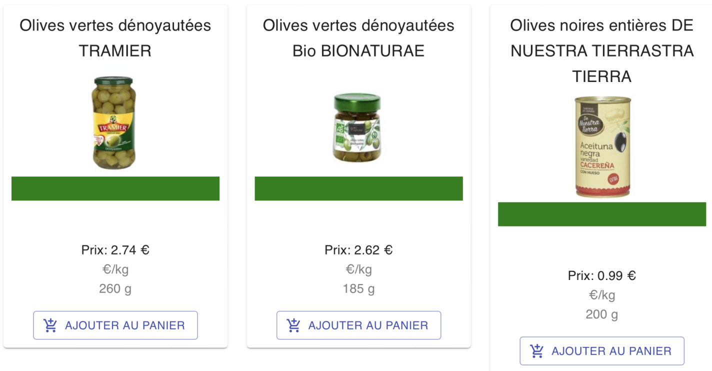
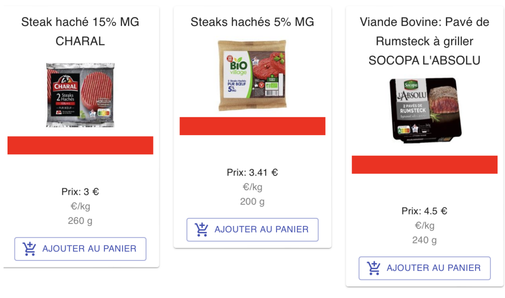


Figure A.3: Screenshot of the online shopping platform in treatment with label (case of red products)



B Categories in the traffic-light label

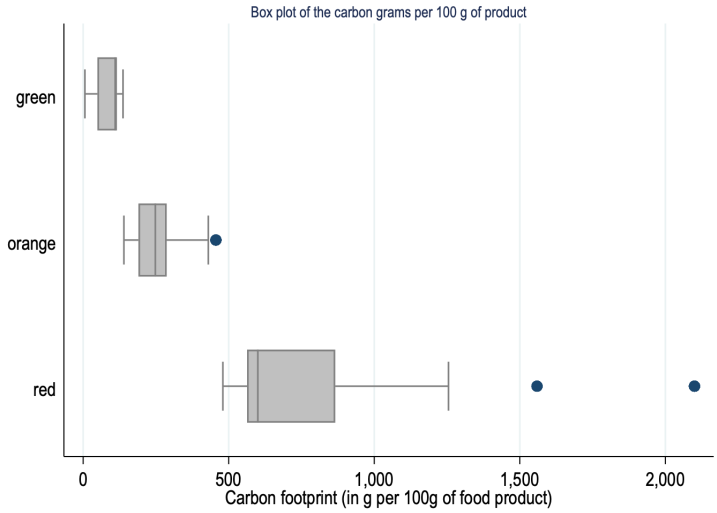


Figure B.1: Carbon footprint of the three product categories: green, orange, red.

C Descriptive statistics

In Table C.1, we present the subjects' socio-demographic characteristics (column *All*), and compare them between the experimental conditions.

Table C.1: Demographic characteristics of the sample

	All	France[a]	Control	Tax 80	Tax 250	Traffic lights	TL/ Tax 80	TL/ Tax 250	Kruskal -Wallis
Age	45.03	42.4	44.71	44.80	46.53	44.98	46.48	42.64	0.345
	(14.05)		(14.39)	(13.94)	(14.47)	(14.45)	(14.64)	(12.37)	
Female	0.735	0.516	0.697	0.716	0.793	0.761	0.711	0.717	0.522
	(0.442)		(0.462)	(0.453)	(0.407)	(0.429)	(0.455)	(0.453)	
Higher education	0.591	0.415	0.633	0.569	0.536	0.500	0.608	0.692	0.050
	(0.492)		(0.484)	(0.498)	(0.501)	(0.503)	(0.491)	(0.464)	
Income (in k. euro)	1.514	2.424	1.517	1.597	1.513	1.329	1.484	1.604	0.168
	(0.740)		(0.715)	(0.690)	(0.716)	(0.734)	(0.713)	(0.840)	
Number of children	0.927	1.80	0.853	0.882	0.914	1.054	0.866	1.00	0.911
	(1.097)		(1.008)	(1.037)	(1.049)	(1.235)	(1.086)	(1.181)	
Env. score	57.72	N.A	58.51	57.23	57.45	56.50	57.25	59.02	0.090
	(7.649)		(7.663)	(6.955)	(7.451)	(7.285)	(8.599)	(7.790)	
Above average meat	0.324	N.A	0.321	0.294	0.343	0.315	0.392	0.283	0.607
	(0.468)		(0.469)	(0.458)	(0.476)	(0.467)	(0.491)	(0.453)	
Observations	660		109	102	140	92	97	120	

[a]: See `https://www.insee.fr/fr/accueil`

First, when considering globally the socio-demographic characteristics of subjects overall (column *All*), our sample appears to differ slightly from the general French population. Participants in our study tend to be older, include a higher proportion of women, have lower levels of higher education, are poorer, and have fewer children on average. Such differences are not really surprising when dealing with online panels. In addition, in our case we rely on a specific sample since we consider only individuals who are responsible of shopping for food which, *de facto*, increases the likelihood of differing from the general French population.

Turning to the comparison between treatments, subjects' characteristics are, overall, well balanced regarding age, gender, income, and the number of children. We, however, detect significant differences between treatments in the proportion of respondents who hold a university degree, and in their environmental score (though at the 10% level only in both cases).

D Regression results for treatment effects

In this Appendix we presents the detailed results of the estimations reported in the main body of the text.

Table D.1: OLS estimations to explain carbon footprint (total and per euro spent)

	Carbon per basket	Carbon per euro spent
	(1)	(2)
Tax €80/ton of CO_2eq.	1.308	-0.0290
	(1.051)	(0.025)
Tax €250/ton of CO_2eq.	-0.126	-0.135***
	(0.925)	(0.022)
Traffic light	-0.406	-0.00803
	(1.147)	(0.029)
Traffic light×Tax €80/ton of CO_2eq.	-1.441	-0.0773***
	(1.038)	(0.025)
Traffic light×Tax €250/ton of CO_2eq.	-2.437***	-0.181***
	(0.935)	(0.022)
Age	-0.0367*	-0.000930**
	(0.020)	(0.000)
Female	0.822	0.0157
	(0.650)	(0.015)
Number of children	1.507***	0.0318***
	(0.275)	(0.006)
Income (in k. euro)	-2.036***	-0.0443***
	(0.407)	(0.010)
Environmental score	-0.00177	-0.000183
	(0.040)	(0.001)
Above average meat	2.884***	0.0627***
	(0.623)	(0.014)
Constant	31.83***	0.823***
	(2.822)	(0.065)
Obs.	633	633
R^2	0.157	0.236

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.2: OLS estimations to explain the added value of the traffic light

	Carbon footprint per basket				**Carbon footprint per euro**			
	All data	No tax	Low tax	High tax	All data	No tax	Low tax	High tax
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TL	-1.867***	-0.383	-2.546**	-2.365***	-0.0365**	-0.00684	-0.0454*	-0.0474***
	(0.585)	(1.131)	(1.072)	(0.862)	(0.015)	(0.029)	(0.025)	(0.018)
Age	-0.0345*	-0.0479	-0.00579	-0.0616**	-0.000830*	-0.00171*	-0.000117	-0.00113*
	(0.020)	(0.041)	(0.034)	(0.028)	(0.000)	(0.001)	(0.001)	(0.001)
Female	0.808	2.771**	0.253	-0.656	0.00939	0.0452	0.0146	-0.0135
	(0.655)	(1.258)	(1.149)	(0.972)	(0.016)	(0.032)	(0.027)	(0.020)
Children	1.514***	1.706***	1.815***	1.279***	0.0318***	0.0356***	0.0384***	0.0271***
	(0.271)	(0.490)	(0.558)	(0.397)	(0.007)	(0.012)	(0.013)	(0.008)
Income	-2.100***	-2.420***	-1.521*	-2.272***	-0.0511***	-0.0471**	-0.0436**	-0.0467***
	(0.410)	(0.701)	(0.830)	(0.570)	(0.010)	(0.019)	(0.020)	(0.012)
Env. score	-0.0105	0.155**	-0.0467	-0.0925	-0.000664	0.00240	-0.00108	-0.00163
	(0.040)	(0.078)	(0.072)	(0.057)	(0.001)	(0.002)	(0.002)	(0.001)
Above average meat	2.869***	5.440***	1.318	1.825*	0.0622***	0.127***	0.0316	0.0332*
	(0.623)	(1.137)	(1.132)	(0.951)	(0.016)	(0.029)	(0.027)	(0.020)
Constant	32.68***	21.40***	34.09***	40.17***	0.800***	0.667***	0.812***	0.822***
	(2.690)	(5.397)	(4.669)	(3.819)	(0.065)	(0.139)	(0.104)	(0.080)
Obs.	633	193	193	247	633	193	193	247
R^2	0.148	0.231	0.108	0.195	0.120	0.181	0.0965	0.182

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

E Demand estimation to compute the price-elasticity

We model demand using a linear Almost Ideal Demand System (AIDS) (Deaton and Muellbauer, 1980), taking advantage of the tax treatments to obtain price heterogeneity. In this model, where we omit the subscript i of consumers for ease of reporting, we define demand w_l as the expenditure share within a category l for $l = g, o, r$, and model the three-equation system:

$$w_l = \alpha_l + \gamma_{gl} ln(p_g) + \gamma_{ol} ln(p_o) + \gamma_{rl} ln(p_r) + \beta_l \mathbb{1}_T + \pi_l D + \epsilon_l, \tag{11}$$

where p_k are prices for each type of product $k = g, o, r$ (for green, orange and red), $\mathbb{1}_T$ is a dummy for the traffic-light label treatment ($\mathbb{1}_T = 0$ in treatments without label and $\mathbb{1}_T = 1$ in treatments with label), D is a vector of control variables (age, gender, number of children, environmental score and a dummy for meat consumption), while ϵ_l is the error term.

Compared to the standard AIDS model, in this specification we omit the expenditure term because participants were given the same budget expressed in terms of purchasing power. A full estimation including the (endogenous) expenditure term, as well as an adjustment of price on purchasing power using the Laspeyres index, can be found in Appendix S4. The model is estimated using a nonlinear seemingly unrelated regression to account for the simultaneity of the three demand functions, and to align with the model with expenditures in Appendix S4. In the estimation of the demand system, we did not omit nor correct for the presence of zeros, because of their rare occurrence: out of 660 participants, we have only 28 zeros, with 9 consumers buying no green products, 3 buying no orange products, and 16 buying no red products. Furthermore, a series of Pearson χ^2 tests find no association between treatment and no expenditures in a category.[37]

Results of the AIDS are reported in Table E.1.[38]

As expected, prices are an important determinant of the demand for all products. The share of budget devoted to a category of products is negatively impacted by the price increase of the same category of good for green and orange products. In contrast, the impact is positive for red products, suggesting that the income effect dominates the substitution effect for these products. The cross-substitution between products is negative and significant except for the red-orange products. Importantly, the traffic-light label shifts the demand function (as expected by the behavioral economic theory introduced in Section 5). The coefficient of the label dummy is positive for green goods and negative for red goods. It implies that when the consumers see the label, the share of green goods increases and the share of red goods decreases for the same prices.

Regarding the control variables (individual characteristics), age increases the demand for green products and reduces slightly the consumption of orange products; female respondents allocate a higher share of expenditures to green products, whilst consuming fewer red products; and the number of children in a household reduces the expenditure share of green products, to the advantage of orange ones. As expected, high consumption of meat is associated with a lower demand for green products, and a higher demand for

[37] Green category: $\chi^2(5) = 3.98$, *p-value* $= 0.552$; Orange category: $\chi^2(5) = 3.40$, *p-value* $= 0.639$; Red category: $\chi^2(5) = 7.84$, *p-value* $= 0.165$.

[38] Note that the estimation of the model requires the use of three restrictions that ensure that the parameters align with economic theory: *homogeneity*, so that $\Sigma(\gamma_{ln}) = 0$; *symmetry*, so that $\gamma_{ln} = \gamma_{nl}$; and *adding-up*, where $\Sigma(\alpha_l) = 1$, $\Sigma(\gamma_l) = 0$.

Table E.1: Demand system estimation using an AIDS approach

	Green	Orange	Red
Price - Green goods	-0.0014	0.0256***	-0.0242***
	(0.0090)	(0.0064)	(0.0081)
Price - Orange goods	0.0256***	-0.0257***	0.0001
	(0.0064)	(0.0082)	(0.0076)
Price - Red goods	-0.0242***	0.0001	0.0241**
	(0.0081)	(0.0076)	(0.0110)
Traffic-light label (TL)	0.0582***	-0.0042	-0.0540***
	(0.0113)	(0.0104)	(0.0130)
Age	0.0011***	-0.0007*	-0.0004
	(0.0004)	(0.0004)	(0.0005)
Female	0.0568***	0.0027	-0.0595***
	(0.0128)	(0.0118)	(0.0147)
Number of children	-0.0095*	0.0154***	-0.0059
	(0.0053)	(0.0048)	(0.0060)
Env. score	0.0001	-0.0001	0.0000
	(0.0006)	(0.0006)	(0.0007)
Above average meat	-0.0329***	-0.0057	0.0386***
	(0.0122)	(0.0111)	(0.0140)
Intercept	0.2201***	0.3573***	0.4226***
	(0.0449)	(0.0413)	(0.0480)
Observations	660	660	660
RMSE	0.1452	0.1329	0.1671
R^2	0.0936	0.0392	0.0684
Romano-Wolf *p-value* for TL	0.0040	0.7092	0.0040

Standard error in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

red products. Finally, attitudes (the environmental score) have no impact on demand.

Following Filippini (1995), from the parameters of Table E.1 we estimate own-price elasticities as $\eta_l = -1 + \gamma_l/w_l$, and cross-price elasticities as $\eta_{ln} = \gamma_{ln}/w_l$ in Table 4.[39] These elasticities are estimated at the level of individual consumers.

[39]Note that these formulas differ slightly from Deaton and Muellbauer (1980) and Filippini (1995) because we do not consider expenditure in the regression, as all subjects are assigned the same budget adjusted for purchasing power. For this reason, elasticities should purely capture substitution, without income effects.

F Welfare estimation

F.1 Inversion of the demand curves

After estimating (10), we invert it to obtain the WTP function $p_l = f(Q_l)$ to estimate counterfactual prices as follows. Firstly, we derive the inverted form of the linear demand function (10) for each of the three prices to obtain the three equations:

$$p_g = \frac{Q_g - \alpha_g - \beta_g \mathbb{1}_{TL} - \gamma_o p_o - \gamma_r p_r - \delta_g D - u_g}{\gamma_g} \tag{12}$$

$$p_o = \frac{Q_o - \alpha_o - \beta_l \mathbb{1}_{TL} - \gamma_g p_g - \gamma_r p_r - \delta_o D - u_o}{\gamma_o} \tag{13}$$

$$p_r = \frac{Q_r - \alpha_r - \beta_r \mathbb{1}_{TL} - \gamma_g p_g - \gamma_o p_o - \delta_r D - u_r}{\gamma_r} \tag{14}$$

To obtain the inverse demand equation, we isolate the quantity Q_l for $l = g, , o, r$ to obtain:

$$p_g = \frac{1}{\gamma_g} Q_g - \frac{1}{\gamma_g} \beta_g \mathbb{1}_{TL} - \frac{1}{\gamma_g} \left(\alpha_g + \gamma_o p_o + \gamma_r p_r + \delta_g D + u_g\right) \tag{15}$$

$$p_o = \frac{1}{\gamma_o} Q_o - \frac{1}{\gamma_o} \beta_o \mathbb{1}_{TL} - \frac{1}{\gamma_o} \left(\alpha_o + \gamma_g p_g + \gamma_r p_r + \delta_o D + u_o\right) \tag{16}$$

$$p_r = \frac{1}{\gamma_r} Q_r - \frac{1}{\gamma_r} \beta_r \mathbb{1}_{TL} - \frac{1}{\gamma_r} \left(\alpha_r + \gamma_g p_g + \gamma_o p_o + \delta_r D + u_r\right) \tag{17}$$

where the last term of the right-hand side of each equation corresponds to the individual-level intercept of a demand equation. If we define the slope coefficient $b_l = \frac{1}{\gamma_l}$ for $l = g, , o, r$, and define the intercept term (individual-specific) as:

$$A_g = -\frac{1}{\gamma_g} \left(\alpha_g + \gamma_o p_o + \gamma_r p_r + \delta_g D + u_g\right) \tag{18}$$

$$A_o = -\frac{1}{\gamma_o} \left(\alpha_o + \gamma_g p_g + \gamma_r p_r + \delta_o D + u_o\right) \tag{19}$$

$$A_r = -\frac{1}{\gamma_r} \left(\alpha_r + \gamma_g p_g + \gamma_o p_o + \delta_r D + u_r\right) \tag{20}$$

then each inverted equation simplifies to:

$$p_g = A_g - \frac{\beta_g}{\gamma_g} \mathbb{1}_{TL} + b_g Q_g \tag{21}$$

$$p_o = A_o - \frac{\beta_o}{\gamma_o} \mathbb{1}_{TL} + b_o Q_o \tag{22}$$

$$p_r = A_r - \frac{\beta_r}{\gamma_r} \mathbb{1}_{TL} + b_r Q_r \tag{23}$$

where the terms $\frac{\beta_l}{\gamma_l}$ apply only to participants in a group with traffic light labels, but not to those who do not see the label.

Finally, since A_j is defined at the individual level l, we can obtain an average intercept over all individuals:

$$\bar{A}_l = \frac{1}{N} \sum_{i=1}^{N} A_{il}, \quad \text{for } l \in \{g, o, r\}$$

This gives us average inverse demand curves with the form:

$$\begin{aligned} p_g &= \bar{A}_g - \frac{\beta_g}{\gamma_g} \mathbb{1}_{TL} + b_g Q_g \\ p_o &= \bar{A}_o - \frac{\beta_o}{\gamma_o} \mathbb{1}_{TL} + b_o Q_o \\ p_r &= \bar{A}_r - \frac{\beta_r}{\gamma_r} \mathbb{1}_{TL} + b_r Q_r \end{aligned} \tag{24}$$

F.2 Estimation of the demand function

Table F.1: Linear demand models, by tax and product type

	No Tax			Tax80			Tax250		
	Green	Orange	Red	Green	Orange	Red	Green	Orange	Red
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
P green	-1762.5300***	-8.1116	-12.9853	-1475.3966***	135.1268	-59.1838	-554.0475***	-47.2595	14.7329
	(283.9249)	(140.5744)	(36.1960)	(224.0425)	(148.2126)	(35.0789)	(105.7528)	(64.8070)	(15.4847)
P orange	67.4256	-701.2090***	-31.6449	-33.7607	-725.7813***	21.9232	87.9542*	-146.1749***	2.4078
	(208.9735)	(103.4652)	(26.6409)	(124.6108)	(82.4347)	(19.5106)	(41.7533)	(25.5870)	(6.1136)
P red	-114.8436	-6.8303	-69.1477***	-64.3727	-0.5952	-81.6444***	-190.3167**	-31.1781	-61.1767***
	(104.2628)	(51.6217)	(13.2919)	(82.1030)	(54.3142)	(12.8551)	(68.5084)	(41.9830)	(10.0312)
TL	1019.5855	-410.5038	-132.2631	-225.8253	117.1407	-125.4056	2342.8352***	-471.1120	-319.7985***
	(811.1301)	(401.5996)	(103.4065)	(625.2671)	(413.6377)	(97.8995)	(618.1829)	(378.8322)	(90.5163)
Age	-0.1589	-19.2602	-0.1235	3.9814	-17.7893	2.0076	26.5146	-10.7280	-5.9419
	(28.3233)	(14.0232)	(3.6108)	(21.8962)	(14.4852)	(3.4283)	(22.9825)	(14.0840)	(3.3652)
Female	1227.6741	492.5208	-57.4523	1722.9300*	305.8836	-269.2673*	540.5329	820.0465	-194.1503
	(909.5115)	(450.3094)	(115.9486)	(704.8704)	(466.2983)	(110.3632)	(724.5684)	(444.0269)	(106.0936)
Has children	-297.6660	294.2594	33.3276	97.7106	149.8581	61.2683	-565.9561*	908.7949***	-86.8057*
	(375.5013)	(185.9149)	(47.8706)	(304.0478)	(201.1391)	(47.6055)	(280.1681)	(171.6914)	(41.0231)
Env. score	-34.5444	26.8534	4.7949	-51.5525	-8.8248	0.6820	-14.5429	21.8990	-5.4311
	(54.6828)	(27.0741)	(6.9712)	(39.8639)	(26.3715)	(6.2416)	(39.8830)	(24.4409)	(5.8398)
Above average meat	251.7366	874.9148	146.8620	-176.5547	368.3573	21.8126	-1242.0151	330.1541	319.7734***
	(901.3413)	(446.2642)	(114.9071)	(667.1162)	(441.3225)	(104.4519)	(651.0571)	(398.9780)	(95.3299)
Constant	13291.6129**	5876.1340**	2142.9567***	12882.0126***	8131.5550***	2592.0151***	9865.3968***	3863.4768*	3105.0694***
	(4049.0375)	(2004.7241)	(516.1896)	(2802.5503)	(1853.9925)	(438.8018)	(2882.0762)	(1766.1816)	(422.0028)
Observations	201.00	201.00	201.00	199.00	199.00	199.00	260.00	260.00	260.00
R^2	0.21	0.31	0.20	0.25	0.31	0.22	0.19	0.24	0.23
Adjusted R^2	0.18	0.28	0.16	0.22	0.28	0.19	0.16	0.22	0.21
F-statistic	5.77	9.45	5.24	7.07	9.47	6.07	6.44	9.01	8.53
Romano-Wolf p for TL	0.4622	0.4622	0.4622	0.9363	0.9363	0.5299	0.0080	0.1793	0.0120

Notes: Standard errors in parentheses. Statistical significance is as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

F.3 Estimation of the WTP and prices

Table F.2: Estimated WTP and prices by product type and tax level

Consumer	Price	No tax		Tax80		Tax250	
		Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Nudgeable	$\hat{p}_g$	1.7942	0.5715	1.7079	0.5716	2.6962	0.4566
	$P^e_g(x^n_g)$	0.2877	1.1555	2.1515	0.6109	0.6726	0.9285
	$\hat{p}_o$	4.2480	1.2100	3.6337	1.3958	4.2521	1.1073
	$P^e_o(x^n_o)$	5.1345	1.8896	4.8236	1.7220	5.8439	1.8268
	$\hat{p}_r$	12.4787	0.8614	11.7431	0.9184	10.7221	0.7806
	$P^e_r(x^n_r)$	13.9155	1.4327	13.8087	1.4157	16.1418	1.2470
Moral	$P^m_g(x^0_g)$	1.6891	0.5971	1.6028	0.6006	2.5911	0.4479
	$\hat{p}_g$	0.1827	1.1984	2.0464	0.6095	0.5675	0.9578
	$P^m_o(x^0_o)$	4.4014	1.2652	3.7871	1.4258	4.4055	1.1949
	$\hat{p}_o$	5.2879	1.8623	4.9770	1.7293	5.9973	1.8041
	$P^m_r(x^0_r)$	12.5183	0.8803	11.7827	0.9191	10.7618	0.8102
	$\hat{p}_r$	13.9552	1.4334	13.8484	1.4123	16.1815	1.2500
Observations				199		260	

Note: S.E. refers to bootstrapped standard errors (1,000 replications). Statistical significance is as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table F.3: Average observed prices by product type and tax level

		No TL		**TL**	
		Mean	SD	Mean	SD
P green	No tax	€2.01	€1.78	€1.92	€1.03
	Tax80	€2.09	€1.73	€2.15	€1.05
	Tax250	€2.26	€0.99	€2.66	€4.04
P orange	No tax	€3.41	€2.47	€3.14	€1.58
	Tax80	€3.37	€1.79	€3.91	€3.24
	Tax250	€5.10	€9.57	€4.44	€2.75
P red	No tax	€10.88	€3.85	€10.42	€4.48
	Tax80	€11.28	€3.16	€11.65	€4.56
	Tax250	€12.27	€3.65	€13.23	€5.28

Online Appendix

Taxing and nudging to reduce carbon emissions: Results from an online shopping experiment

Contents

S1 New Environmental Paradigm scale

We implement the following survey from Dunlap et al. (2000).

Please indicate your level of agreement for each of the following statements, between "Strongly agree" and "Strongly disagree".

1. We are approaching the limit of the number of people the Earth can support.
2. Humans have the right to modify the natural environment to suit their needs.
3. When humans interfere with nature it often produces disastrous consequences.
4. Human ingenuity will ensure that we do not make the Earth unlivable.
5. Humans are seriously abusing the environment.
6. The Earth has plenty of natural resources if we just learn how to develop them.
7. Plants and animals have as much right as humans to exist.
8. The balance of nature is strong enough to cope with the impacts of modern industrial nations.
9. Despite our special abilities, humans are still subject to the laws of nature.
10. The so-called "ecological crisis" facing humankind has been greatly exaggerated.
11. The Earth is like a spaceship with very limited room and resources.
12. Humans were meant to rule over the rest of nature.
13. The balance of nature is very delicate and easily upset.
14. Humans will eventually learn enough about how nature works to be able to control it.
15. If things continue on their present course, we will soon experience a major ecological catastrophe.

S2 Additional analyses

Table S2.1: *P-values* of the two-sample Kolmogorov–Smirnov tests for equality to compare the distributions of the carbon footprint (in kg) and of the average carbon footprint (in kg) per euro between treatments

Treatments	Carbon footprint (in kg)	Average carbon footprint (in kg) per euro
Tax80	0.749	0.256
Tax250	0.335	$<$0.01
TL	0.955	0.821
Tax80+TL	0.185	0.017
Tax250+TL	0.117	$<$0.01

We compare the treatments to the control group.

Table S2.2: *P-values* of the two-sample Kolmogorov–Smirnov tests for equality to compare the distributions of the budget share of green, orange and red products between treatments

Treatments	Green products	Orange products	Red products
Tax80	0.1245	0.1023	0.0754
Tax250	0.0662	0.0638	0.0773
TL	0.2091**	0.1182	0.1714
Tax80+TL	0.1506	0.0910	0.1401
Tax250+TL	0.1833**	0.0987	0.1823**

We compare the treatments to the control group.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

S3 Multiple hypothesis testing

Table S3.1: Multiple Hypothesis Testing Robustness Check - Carbon footprint per basket and carbon per euro

Comparison	Difference in means	Multiplicity-adjusted $p-value$
Carbon footprint per basket		
Control vs. Tax80	1.264	0.659
Control vs. Tax250	0.084	0.927
Control vs. TL	0.155	0.997
Control vs. TL + Tax80	1.194	0.663
Control vs. TL + Tax250	2.217	0.126
Carbon footprint per euro		
Control vs. Tax80	0.029	0.590
Control vs. Tax250	0.131***	0.0003
Control vs. TL	0.003	0.991
Control vs. TL + Tax80	0.073**	0.033
Control vs. TL + Tax250	0.176***	0.0003

Notes: In this Table we report the results of tests to compare means using the *mhtexp* command in Stata (List et al., 2019). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

S4 Alternative demand estimation

In this section, we estimate the same demand for products in the three carbon categories (green, orange and red) using an Quadratic Almost Ideal Demand System (QUAIDS) (Banks, Blundell, Lewbel, 1997). In the model, where we omit the subscript i of consumers for ease of reporting, we define demand w_l as the expenditure share within a category, and model the three-equation system:

$$w_l = \alpha_l + \gamma_{gl} ln(p_g) + \gamma_{ol} ln(p_o) + \gamma_{rl} ln(p_r) + \beta_l \; ln\left(\frac{m}{a(p)}\right) + \frac{\lambda_l}{b(p)} \; ln\left(\frac{m}{a(p)}\right)^2 + \pi_l D + \epsilon_l$$

where p are prices for each type of product $l = g, o, r$, the term m refers to total expenditures, and ϵ_g is the error term. The terms $a(p)$ and $b(p)$ correspond to

$$ln\,(a(p)) = \alpha_0 + \sum_l \alpha_l ln(p_l) + \frac{\lambda_1}{2} \sum_l \sum_n \gamma_{ln} ln(p_l) ln(p_n)$$

and

$$b(p) = \prod_l p_l^{\beta_l}$$

The model requires the use of three restrictions: **homogeneity**, so that $\Sigma(\gamma_{ln}) = 0$; **symmetry**, so that $\gamma_{ln} = \gamma_{nl}$; and **adding-up**, where $\Sigma(\alpha_l) = 1$, $\Sigma(\beta_l) = 0$, $\Sigma(\gamma_l) = 0$, $\Sigma(\lambda_l) = 0$. Finally, D is a vector of personal characteristics of the respondent. Prior to the estimation of the QUAIDS, we corrected prices to ensure all groups had the same purchasing parity using the Laspeyres index $I_L = \frac{\Sigma_j (q_{j0} p_{jt})}{\Sigma_j (q_{j0} p_{j0})}$, where the suffix 0 refers to the control group, and t refers to the tax treatments. We estimated a Laspeyres index for each of the three carbon categories separately. This adjustment was required because consumers in the tax groups would face a different price schedule than those in non-tax groups, so that any change in behaviour could not be unequivocally be ascribed to the nudge itself. We therefore estimated the costs of the same bundle of goods in the control group at the three different tax levels (€0, €80, €250), obtaining prices in "control-equivalent" euros to ensure that consumers had the same purchasing power in all price regimes. Finally, in our analyses we did not omit zeros, nor correct for the presence of zeros, because of their rare occurrence: out of 660 participants, we have only 27 zeros, with 9 consumers buying no green products, 3 buying no orange products, and 16 buying no red products. A series of Pearson χ^2 tests show no association between treatment and no expenditures in a category (green category: $\chi^2(5) = 3.98$, $p = 0.552$; orange category: $\chi^2(5) = 3.40$, $p = 0.639$; red category: $\chi^2(5) = 7.84$, $p = 0.165$). Results of the QUAIDS in Table S4.1 indicate that, as expected, prices are an important determinant of the demand for products in all categories. At the same time, the quadratic expenditure term is significant for orange and red products-an indication that these demand functions are of rank third. A likelihood ratio test indicates that a quadratic AIDS fits the data better than the linear AIDS, with $\chi^2(2) = 15.28$, p = 0.0005.

In terms of personal characteristics, age increases the demand for green products, while a high consumption of meat is associated with a lower demand for green products; female respondents allocate a lower expenditure share to both red and orange products; while consumers with high environmental score purchase fewer red and green products. Importantly, the label increases the expenditure share of both green and orange products (at 10% level of significance), with a non-significant reduction in the demand for red products. Robustness checks using unadjusted prices (that is, not corrected with a Laspeyres index) show no notable difference in the estimated coefficients (data are available on request).

Table S4.1: Demand system estimation using a Quadratic AIDS approach

	Green products	Orange products	Red products
Intercept	0.3969***	0.2574***	0.3457***
	(0.0385)	(0.0269)	(0.0335)
Price - Green products	0.0533***	0.0167**	-0.0700***
	(0.0144)	(0.0085)	(0.0117)
Price - Orange products	0.0167**	0.0061	-0.0228**
	(0.0085)	(0.0124)	(0.0115)
Price - Red products	-0.0700***	-0.0228**	0.0928***
	(0.0117)	(0.0115)	(0.0150)
Expenditures	0.0428	0.0394	-0.0822*
	(0.0347)	(0.0298)	(0.0428)
Expenditures2	-0.0362	0.1484***	-0.1122**
	(0.0395)	(0.0391)	(0.0499)
Age	0.0296**	0.0033	-0.0004
	(0.0128)	(0.0064)	(0.0033)
Female	0.0297	-0.3679*	-0.4018***
	(0.4251)	(0.1936)	(0.1170)
Number of children	-0.2752	0.1382	-0.0376
	(0.1878)	(0.0885)	(0.0448)
Env. score	-0.0550***	-0.0033	-0.0092*
	(0.0135)	(0.0044)	(0.0052)
Above average meat	-1.2332***	-0.1657	0.1064
	(0.4550)	(0.2231)	(0.1033)
Traffic Light	1.4464***	0.3561*	-0.1231
	(0.3824)	(0.1880)	(0.0953)
Observations	660	660	660
R^2	0.1212	0.0560	0.1052

From the parameters of the equations above, we estimate elasticities based on Banks, Blundell, and Lewbel (1997). Estimates and their standard errors, obtained using the Delta method, are reported in Table S4.2. These elasticities are estimated at the level of the individual consumers, and Table S4.2 reports the average across the sample. Results indicate that all categories are own-price inelastic, with the red category showing the lowest sensitivity to own-price changes. Moreover, all compensated price elasticities are positive and inelastic (always below 0.45), an indication that the three categories weakly substitute each other. Finally, the green and orange categories are expenditure elastic, while the red category is inelastic and less responsive to changes in total expenditures. These results provide an explanation of the limited impact of taxation on carbon emissions observed above: within a color category, there is limited responsiveness to changes in own-price, as consumers show an inelastic response to the increase of prices, and struggle to substitute products across categories due to the different characteristics of the food within each category. This limited ability to respond to price changes leads to relatively small reductions in consumption, causing consumers to internalize the tax to some extent, rather than changing their consumption.

Table S4.2 below reports the average elasticities across the sample.

Table S4.2: Estimated elasticities in the experimental store

		Elasticity Price Green	Orange	Red	Expenditure
Equation	Green	-0.7115***	0.3662***	0.2126***	1.1148***
		(0.0393)	(0.0239)	(0.0277)	(0.0714)
	Orange	0.4204***	-0.7259***	0.2994***	1.1455***
		(0.0275)	(0.0316)	(0.0264)	(0.0591)
	Red	0.2912***	0.3597***	-0.5121***	0.8401***
		(0.0377)	(0.0313)	(0.039)	(0.0633)

Standard errors, estimated using the Delta methods, are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

S5 Welfare analysis with all product categories

Table S5.1: WTP and welfare estimates by consumer type and tax level

Consumer type	Welfare type	Product type	No tax		Tax80		Tax250	
			Value	S.E.	Value	S.E.	Value	S.E.
Nudgeable	Allocative	Green	0.9739	1.5912	0.0844	0.5112	1.7573	1.5416
		Orange	0.0867	0.9356	0.1562	0.9043	0.2795	0.8663
		Red	0.0707	0.2289	0.1462	0.2863	1.0062*	0.5312
	Valuation	Green	11.5406	7.4886	3.2205	4.6833	12.4396*	6.7982
		Orange	5.9152	6.9290	6.7366	7.3046	7.3549	6.2875
		Red	3.1733**	1.5553	3.6533**	1.5358	5.4657***	1.2667
Moral	Allocative	Green	-0.9739	1.5912	-0.0844	0.5112	-1.7573	1.5416
		Orange	-0.0867	0.9356	-0.1562	0.9043	-0.2795	0.8663
		Red	-0.0707	0.2289	-0.1462	0.2863	-1.0062*	0.5312
	Valuation	Green	11.5406	7.4886	3.2205	4.6833	12.4396*	6.7982
		Orange	5.9152	6.9290	6.7366	7.3046	7.3549	6.2875
		Red	3.1733**	1.5553	3.6533**	1.5358	5.4657***	1.2667
Observations					199		260	

S.E. refer to bootstrapped standard errors (1,000 replications). Statistical significance is as follows: * = 10%; ** = 5%; *** = 1%