Department of Economics
Dietrich School of Arts and Sciences

4508 Wesley Posvar Hall
(339) 227-5372
lep147@pitt.edu

February 3, 2025

To Whom it May Concern:

The manuscript “Got Beef with Beef? Evidence from a Large-Scale Carbon Labeling Experiment” was written with direct implementation support from HelloFresh. Lucy Page was supported by a National Science Foundation Graduate Research Fellowship throughout the duration of the projecct.

IRB approval for the project was obtained from the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects (Protocol # 2106000406). HelloFresh was given 30 days to review all data disclosures included in the paper.

Lucy Page declares that she has no relevant or material financial interests that relate to the research described in the paper.

Sincerely,

Lucy Page

Lucy Page

February 4, 2025

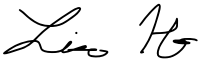
To Whom it May Concern:

The manuscript “Got Beef with Beef? Evidence from a Large-Scale Carbon Labeling Experiment” was written with direct implementation support from HelloFresh.

IRB approval for the project was obtained from the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects (Protocol # 2106000406). HelloFresh was given 30 days to review all data disclosures included in the paper.

Lisa Ho declares that she has no relevant or material financial interests that relate to the research described in the paper.

Sincerely,



Lisa Ho

Got Beef with Beef? Evidence from a Large-Scale Carbon Labeling Experiment*

Lisa Ho† and Lucy Page‡

February 6, 2025

Food systems account for approximately one-third of global greenhouse gas emissions, and carbon-footprint labeling is an increasingly common tool to shift consumers towards lower-carbon diets. In a randomized field experiment with over 200,000 customers at a major US food-services company, we find that carbon labels increase customer retention by 1.1% and company profits by 0.9%, despite reducing customers' carbon footprints by only 0.6%. These profit effects suggest that carbon labeling may remain a common sustainability tool, despite its small environmental benefits. Moreover, label targeting is crucial: labels may increase footprints among those who do not agree with their purpose.

*We are extremely grateful to Abhijit Banerjee, Esther Duflo, Ben Olken, and Frank Schilbach for their support and advice throughout this project. We also thank Hunt Allcott, David Atkin, Rohini Pande, Hannah Ruebeck, and participants in MIT Behavioral Lunch for many suggestions that improved this paper. We thank the entire team at HelloFresh who made this experiment possible, in particular Rémy Aldasoro and Jeff Yorzyk, whose constant dedication to supporting HelloFresh customers in making climate-friendly choices was pivotal to the implementation of carbon labels. This experiment was preregistered with the AEA RCT registry (AEARCTR-0009149) and received approval from MIT IRB (Protocol #2106000406). The conclusions expressed in this research project are solely the responsibility of the authors.

†Yale. Email: lisa.y.ho@yale.edu

‡University of Pittsburgh. Email: lep147@pitt.edu

1 Introduction

Anthropogenic climate change has already raised global average temperatures by about 1.1 °C from pre-industrial levels, and much larger increases are projected to occur without rapid cuts to greenhouse gas (GHG) emissions worldwide (Lindsey and Dahlman, 2024). Globally, food systems account for about one third of total GHG emissions, so limiting climate change will likely require major shifts in food consumption and production (Crippa et al., 2021; Ivanovich et al., 2023; Tubiello et al., 2022). Animal products—and beef in particular—account for a disproportionate share of food-system emissions: producing one kilogram of beef emits about 10 times as much as one kilogram of pork or chicken and about 20 times as much as one kilogram of tofu (Poore and Nemecek, 2018). Simple dietary shifts, especially away from beef consumption, could then substantially reduce emissions (Clark et al., 2020; Ripple et al., 2014; Li et al., 2024; Kozicka et al., 2023). Achieving these dietary shifts may be particularly important in the United States, which ranks among countries with both the highest per-capita GHG emissions (Jones et al., 2023; Ritchie et al., 2020) and per-capita meat consumption (Ritchie et al., 2019).

In the absence of systematic carbon pricing in the US, consumer-facing carbon labeling has gained traction as a potential lever to drive these demand shifts. Carbon-labeling schemes tag products with either numeric estimates of their associated emissions or coarser classifications of impact tiers, aiming to shift consumers' choices either by providing new information about products' environmental costs or by making salient their existing knowledge on these costs. The EPA has long ranked and labeled vehicles' GHG emissions, and the White House announced in 2024 that it would invest $160 million in carbon labeling of construction products (Lee, 2024). Alongside these public labeling schemes, it is increasingly common for individual companies to add carbon labels to their own products, including in food, beauty, and footwear.[1]

Even as carbon labels proliferate, it is unclear whether they offer a cost-effective tool to shift behavior, or are primarily a greenwashing tool for companies that may even divert attention from more impactful policy reforms (Chater and Loewenstein, 2023; Hagmann et al., 2019). How effective are labels in shifting consumer choices in a realistic US commercial setting?[2] Moreover, carbon labeling will likely remain voluntary in the US, with companies deciding both whether and how to present emissions-related information, so opportunities for profit may be the primary motivation for these labeling schemes. How does carbon labeling affect company profits?[3]

This paper reports results from the largest evaluation of carbon labeling to date, a randomized field experiment with over 200,000 US customers at HelloFresh, the largest meal-kit company both

[1] Food companies that have introduced carbon labels include Panera Bread, which identifies its climate-friendly options as "Cool Food" meals, and predominantly low-carbon food producers such as Oatly, Quorn, and JustSalad. This trend extends beyond the food industry: for example, L'Oreal introduced environmental labeling of its products in 2020, and ASICS began labeling its shoes with carbon footprints in 2023.

[2] In addition to a large literature on hypothetical choices, lab-based studies and small-scale experiments on real outcomes suggest that labels can shift consumers to more sustainable choices, reviewed in Taufique et al. (2022) and Potter et al. (2021). In larger evaluations, Tilling (2025) and Lohmann et al. (2022) find that carbon labels reduce diners' carbon footprints in university canteens in Germany and the United Kingdom, and Bilén (2024) uses a natural experiment to find that a bundle of pushes towards sustainable food purchases in a Swedish grocery store reduces customers' carbon footprints. None of these are in the US, which is characterized by higher per capita footprints and lower average support for climate action (Dechezleprêtre et al., 2024). While Bilén (2024) studies a commercial setting, the study does not isolate the impact of labels alone.

[3] Most related, Bilén (2024) uses a natural experiment and finds that a bundle of nudges towards sustainable consumption, including footprint labels, reduced customer visits to a Swedish grocery store, especially among those with high baseline carbon footprints.

in the US and globally.[4] Meal-kit services, which deliver pre-portioned ingredients for subscribers to cook at home, are a fast-growing form of food purchase in the US: between February 2021 and February 2022, the share of Americans who had tried a meal-kit service rose from 14% to 23% (Commisso, 2022). Meal-kit companies offer an ideal setting for food-based carbon labeling. Unlike supermarkets, meal-kit providers have information on the production of all foods they sell; unlike individual brands, they can label the full range of meals from which customers choose.

From May to July 2022, we worked with engineers and environmental scientists at HelloFresh to add carbon labels to the menus of randomly selected customers for eight weeks. Meals are categorized into one of three tiers based on the estimated carbon footprints of producing their ingredients: *"Climate Superstars"* with footprints under 2 kg CO_2/meal, *"Good"* meals with footprints of 2 to 7 kg CO_2/meal, and *"Fair"* meals with footprints over 7 kg CO_2/meal. Customers are randomized to see either control menus with no carbon labels or menus with one of two broad classes of labels: those that only highlight *Climate Superstar* meals and those that categorize all meals into carbon-footprint tiers.[5] We use administrative data to study the impacts of the labels on customer retention, company profits, and meal choices. We then combine this administrative data with a baseline survey of 5,592 customers to shed light on the mechanisms driving the labels' effects.

We find that the carbon-footprint labels had financial returns for HelloFresh, but yielded only small environmental benefits. Labels increase HelloFresh's customer retention by 1.1% ($p < 0.01$), translating into a 0.9% ($p = 0.09$) increase in profits per customer-week. We find no evidence that these effects dissipate over the course of the 8-week experimental period. At the same time, the labels only reduced customers' weekly carbon footprints by 0.6% ($p < 0.01$), equivalent on average to the emissions from charging a smartphone six times (EPA, 2024). These footprint reductions are driven by a 0.8% ($p = 0.08$) increase in the consumption of *Climate-Superstar* meals and a 1.5% ($p < 0.01$) reduction in consumption of *Fair* meals. Because all *Fair* meals contain beef, and vice versa, the labels reduce consumption of beef meals by 1.5%. These impacts vary over time, peaking at a 1.2% footprint reduction in week four before attenuating nearly to zero by the end of the 8-week experimental period. While the profit effects we estimate are in partial equilibrium and could dissipate as other companies introduce labels, they suggest that companies like HelloFresh have incentives to implement carbon-labeling schemes, despite their modest environmental benefits.

Judicious targeting is key to the labels' overall effects. The reduction in carbon emissions is driven by customers in the top 25 percentiles of pre-intervention beef consumption, for whom labels increase consumption of *Climate-Superstar* meals by 3.7% and reduce *Fair* meals by 1.9%. These high-beef customers also seem to most highly value the climate labels, with increases in retention and profit per customer-week of 4.2% and 3.9%, respectively. At the same time, carbon labels generate backlash among customers who disagree with their intent: those who state below-median beliefs that individuals or companies have a moral duty to combat climate change actually *increase* their carbon footprints in response to labels.[6] This pattern suggests caution in rolling out climate

[4] HelloFresh sold over 230 million meals and reported revenue of over 1.8 billion EUR in 2024 Q3 (HelloFresh, 2024). HelloFresh accounts for 75% of meal-kit sales in the US (Perri, 2024).

[5] Specifically, treatment customers were randomized across three labeling schemes. The "Superstars-Only" group saw menus on which only *Climate-Superstar* meals were labeled with the label title and an abstract globe symbol. Next, the "All-Tiers" group saw menus where all meals were labeled with their group name and a corresponding globe symbol. Finally, HelloFresh was interested in several user-interface designs, so another quarter of customers were randomized to an "All-Tiers–Letters" group, where footprint categories were shown with letters A, B, and C.

[6] Our findings for customers' differential responsiveness to carbon labels match Lohmann et al. (2022)'s findings that labels in UK university canteens had the largest effects on those with the highest baseline carbon footprints. In

labels in settings where some of the target audience do not support climate action. This backlash concern may be particularly important in the US, where the suggestion that Americans should alter their diets is politically contentious. For example, Republican lawmakers came out in 2021 against an (imagined) Democratic plan to reduce beef consumption with warnings to "stay out of my kitchen" (Washington Post Staff, 2021).[7] Altogether, the labels' treatment effects on both meal choices and retention are largest among customers who both had high baseline beef consumption and also believe that individuals and companies have a moral duty to address climate change.

Finally, we shed light on whether the labels primary operate by making salient customers' existing knowledge on foods' carbon footprints or by providing new information. Consistent with Tilling (2025) and Imai et al. (2022), our results suggest that the carbon labels primarily work via salience. In our full baseline sample, we find no gap in the labels' impacts on meal choices by baseline carbon-footprint knowledge. We also find that all of the label formats reduce consumption of *Fair* meals versus *Good* meals, including the scheme that only labels *Climate Superstars*; thus, customers seem to act on baseline knowledge that beef has a high carbon footprint. Finally, among the subset of customers with the largest meal-choice effects—high baseline beef consumers who believe in climate action—effects are fully concentrated among those who could correctly answer basic questions about the carbon footprints of different foods at baseline. The labels' retention effects are largest among the same subgroup, suggesting that these customers value the labels as salience nudges to help them reduce their environmental impacts.

2 Experimental Design

2.1 Meal classification and labeling

The experimental intervention added carbon-footprint labels to the HelloFresh menus from which customers choose meals each week. This menu includes an array of meal cards, each with a photo, name, and a series of labels characterizing the meal. (Appendix Figure A1 shows an example.) On average, each meal is labeled with 1.2 non-climate labels (e.g. low-calorie, vegetarian, spicy). From this main menu, customers can click on each meal to pull up a more detailed meal card including nutritional content and larger renditions of any labels.

Meal Classification: To develop the carbon-labeling scheme, HelloFresh first used the Agribalyse dataset (detailed in Appendix C.3) to estimate the carbon footprints of all menu offerings from their ingredients. Meals were then categorized into three tiers by their estimated footprints: (1) "*Climate Superstar*," with estimated footprints below 2 kg CO_2e, (2) "*Good*," with estimated footprints between 2 kg CO_2e and 7 kg CO_2e, and (3) "*Fair,*" with estimated footprints over 7 kg CO_2e.

Because different proteins vary widely in their associated GHG emissions, meals' carbon-footprint classifications are largely determined by their meat content (Figure 1). Vegetarian meals typically have lower carbon footprints than meals containing poultry, pork, and fish, while beef stands out

contrast, Bilén (2024) finds that while a bundle of pushes towards sustainable food purchases in a Swedish grocery store initially reduced carbon footprints both among those with above- and below-median baseline consumption, the effects differentially faded among consumers with high baseline footprints.

[7] During this controversy, Republican Representative Marjorie Taylor Greene called then-President Joe Biden "The Hamburglar," and the Republican Idaho governor Brad Little tweeted, "Idahoans also have beef with this agenda and for dinner!" (Bauder and Swenson, 2021).

as a disproportionate source of GHG emissions. All meals with a primary protein other than beef are classified as *Climate Superstar* or *Good*, while all beef meals have estimated carbon footprints well above the cutoff for the *Fair* category.

Label variations: We randomize the experimental sample across four arms: a control arm with no carbon-footprint information and three treatment arms with different carbon-labeling schemes. Table 1 depicts these labeling schemes. We test two broad classes of labels: those that only highlight climate-friendly meals and those that categorize all meals into carbon-footprint tiers. As in Panera Bread's "Cool Food" labels, companies may seek only to reward customers for good choices rather than risk upsetting customers with a taste for high-carbon meals. To test whether reward-only labeling schemes can affect meal choices, the *Superstars-Only* group saw menus on which only the lowest-footprint tier meals were labeled. In contrast, the *All-Tiers* group saw menus where all meals were labeled with their group name and a corresponding globe symbol. HelloFresh was interested in several user-interface designs, so another quarter of customers were randomized to an *All-Tiers–Letters* group, where footprint categories were denoted with letter-grades A, B, and C.[8] Appendix Figure A1 shows examples of the HelloFresh menu shown to the *All-Tiers* group. We find no differential effects by label type throughout the experiment, so we pool these treatment arms in the main text and present separate treatment effects by label variation in the appendix.

2.2 Experiment timing

HelloFresh assigned to the experimental sample any customers who visited the HelloFresh website in fiscal weeks 18, 19, or 20 of 2022. Labels were randomly added to the menu for weeks 21 through 28, first appearing on future menus midway through week 20. Customers select meals at least one week before delivery, so the addition of these labels could have affected meal choices for weeks 21 through 28. Customers can preview and choose meals from menus up to 5 weeks before delivery, however, so many customers would have chosen meals for early experimental weeks before seeing the labels. We define the pre-experimental period as weeks 8 through 20 and the experimental period as weeks 23 through 28 in our main analysis, but we show that results are robust to alternative definitions in Figures A3 and A5. For week-by-week treatment effects, see Figures A2 and A4.

2.3 Sample data and characteristics

2.3.1 Full experimental sample

A total of 234,511 customers were enrolled in the experiment. We observe administrative data on customers' meal choices from weeks 8 through 28, customers' weekly revenue and costs beginning in the week (18, 19, or 20) during which they were assigned to the experimental sample, and HelloFresh's estimates for the carbon footprint of each meal for weeks 17 through 28. While we do not observe customer identity or demographics, we do observe customer zipcodes; we merge these with zipcode-level data on voteshares in the 2020 presidential election and with demographics from

[8] In each of the labeling arms, menu cards also included short descriptions of what the climate labels meant. For example, the menu card for the Cherry Balsamic Bavette Steak stated, "This meal is rated **Fair** because it's among the least carbon-efficient options on this week's menu." Pre-existing, non-climate labels do not have explanations.

the 2020 American Community Survey.[9]

The HelloFresh customer population lives in areas that are somewhat whiter, more educated, and wealthier than the national population, so we may expect them to express above-average environmental concern (Angrist et al., 2024; Dechezleprêtre et al., 2024). If carbon labels are more effective among those with more concern about climate change, our estimated treatment effects—which are already small in magnitude—may be an upper bound for the impact of labels in the average US setting. On the other hand, if labels primarily work by providing information—which more educated and environmentally aware consumers may have at baseline—effects could be smaller in our setting than elsewhere. Treatment assignment is largely balanced on observable baseline characteristics. Appendix C.1 details the data, summary statistics, and balance for the full experimental sample.

2.3.2 Baseline survey sample

We pair administrative data for the full sample with detailed baseline data collected from a subsample of customers, recruited by email from the full experimental sample. Those who completed the survey were entered into a lottery for gift-card prizes. The survey elicited participants' beliefs about climate change, self-perceptions on traits like altruism and environmental consciousness, political affiliation, and baseline knowledge about the carbon footprints of different foods.[10]

7,259 customers completed the baseline survey, and 5,592 provided email addresses that merged with administrative data from HelloFresh. While the matched baseline sample comprises only 2.4% of the full experimental sample, it is similar to the full sample on most observable traits, with the exceptions that customers in the baseline sample are longer-standing customers of HelloFresh and have 3.2% lower baseline carbon footprints on average (see Appendix Table A6 for a full comparison). The baseline sample is broadly balanced across treatment arms, and we control for a range of traits in our baseline-sample analysis. Appendix C.2 summarizes the baseline sample in more detail.

2.4 Empirical strategy

Our main analysis uses the following simple regression model:

$$Y_{it} = \beta Label_i + \sum_{w=18}^{20} \gamma_{1,w} Y_{iw} + \sum_{w=18}^{20} \gamma_{2,w} \mathbb{1}(MiY_{iw}) + \delta_t + \Phi X_i + \epsilon_{it}$$

where Y_{it} is an outcome variable for customer i in week t, $Label_i$ indicates that customer i was assigned to see menus with carbon labels, and β is the coefficient of interest. We control for lagged outcomes in pre-experimental weeks 18, 19, and 20, $\{Y_{iw}\}_{w=18}^{20}$, and indicators that these values are missing, $\{\mathbb{1}(MiY_{iw})\}_{w=18}^{20}$.[11] We also control for week fixed effects (δ_t) and a vector of other customer controls (X_i). In our main specifications, X_i includes indicators for customers' meal plan with HelloFresh (e.g. vegetarian, chef's choice, premium), zipcode-level demographics, and

[9] We retrieve county-level vote shares in the 2020 presidential election from this public Github repository, with data scraped from Fox News, Politico, and the New York Times.

[10] The full baseline survey is presented at the end of the paper's appendices.

[11] For meal-choice outcomes, lagged controls may be missing because a customer did not order a meal box in that prior week. Lagged controls may also be missing because we observe the variables for a given customer only after they were allocated to the experimental sample in Weeks 18 through 20.

longevity at HelloFresh. In baseline-sample analysis, we also control for customers' educational attainment, gender, political affiliation, self-reported environmentalism, and the extent to which they report considering environmental factors in their food choices. Appendix C.4 details these controls, and we show robustness to excluding them in Appendix Figures A3 and A5.

3 Results

We present results in several steps. First, Sections 3.1 and 3.2 present the labeling scheme's main effects on customer meal choices and the company's financial outcomes, respectively. Section 3.3 then tests which subgroups of customers drive the treatment effects, and Section 3.4 explores the mechanisms underlying effects.

3.1 Main effects on meal choices

Adding carbon footprint labels to the HelloFresh menu induced small but significant shifts towards lower-carbon diets (Table 2, Panel A). Overall, the carbon labels reduced the total carbon footprint of meals a customer ordered each week by about 0.075 kg CO_2 ($p < 0.01$), or 0.6% of the control mean of 11.8 kg CO_2. The labels reduced the number of *Fair* meals ordered by about 0.009 (1.5%, $p < 0.01$) and increased the number of *Climate-Superstar* meals ordered by 0.007 (0.8%, $p = 0.08$). These shifts reflect substitution from beef meals to vegetarian, poultry, and pork meals.[12,13]

The per-customer impacts of the carbon labels are small, and they remain modest even when summed across HelloFresh customers. Our point estimates suggest that adding carbon labels to menus for all 234,511 customers in the experimental sample would reduce emissions by 17,588 kg CO_2e per week, equivalent to the GHG emissions from the weekly energy use of 114.4 homes (EPA, 2024).

Treatment effects over time. As expected, the treatment effects on meal choices are about zero in weeks 21 and 22, when many customers had made meal choices before labels were introduced (Appendix Figure A2). The treatment effects rise to a maximum in week 24 of a 1.2% drop in total carbon footprints before attenuating close to zero in weeks 27 and 28.

Robustness: Our estimates for the labels' effects on meal choices are largely robust to changes in control variables, though removing controls for past meal choices makes estimates imprecise and statistically indistinguishable from zero (Appendix Figure A3). Consistent with Appendix Figure A2, the labels' effects on meal choices are somewhat smaller when defining the treatment period as beginning in week 21.

3.2 Main effects on retention and profit

Regardless of labels' environmental impacts, whether companies choose to implement them likely depends on whether doing so increases profit. The labels could affect profit either by changing

[12] The label effects are quite similar across label variations (Appendix Table A3), a fact to which we return in Section 3.4.1 below.

[13] The labels' carbon-footprint effects are unlikely to be explained by the positive treatment effects on customer retention (Table 2, Panel B, described in Section 3.2). The customers driving the positive effects on retention had higher carbon footprints at baseline, and so we would expect the estimated treatment effect of climate labels on carbon footprints to underestimate the treatment effect holding customer composition fixed.

subscriber retention or by shifting customers towards more or less profitable meals, conditional on ordering a box in a given week. Our direct collaboration with HelloFresh allows us to directly test effects on back-end financial outcomes: customer retention, revenue net of customer discounts, total direct costs, and profits (Table 2, Panel B). The labels' effects on customer retention also allow us to speak to their effects on customer welfare. If the labels change meal choices but also induce shame among high-beef eaters (Butera et al., 2022), for example, these welfare costs could appear in label avoidance via customer drop-off.

We find that the carbon labels increase customer retention and HelloFresh's profit per customer overall (Table 2, Panel B). Adding carbon-footprint labels to menus makes customers about 0.3pp more likely to order a meal box in a given week, about 1.1% of the control mean of 28.1% ($p < 0.01$). The labels do not alter the average number of meals ordered per box, conditional on placing an order. On average, then, the carbon-footprint labels may increase customer welfare at HelloFresh. This result likely reflects that HelloFresh intentionally designed the labels to be neutral to positive—even beef meals are labeled as *Fair*—so the scheme was relatively unlikely to have negative emotional consequences.

The labels' positive retention effects translate into higher revenue, costs, and profit per customer. Setting these financial outcomes to zero for any week in which a customer did not order a box, the carbon labels increase revenue by €0.20 (1.0%; $p = 0.04$), direct costs by €0.13 (1.0%; $p = 0.04$), and profit by €0.07 (0.9%; $p = 0.09$) per customer-week.

In contrast, the labels' impacts on the composition of meal choices may reduce profits per box ordered. The direct costs to HelloFresh of providing each meal vary with its main protein (Appendix Table A4): vegetarian and pork meals are the cheapest to produce, followed by poultry, beef, and fish. HelloFresh more than recoups these additional costs with surcharges on high-cost meals, however, so beef and fish meals are ultimately more profitable than vegetarian or poultry. In line with these patterns, our point estimate suggests that the carbon labels somewhat reduce profit conditional on ordering a box, but this effect is statistically indistinguishable from zero.

Treatment effects over time. The labels' positive effects on retention and profit gradually rise over the first few weeks of the treatment, before remaining between 0.8% to 1.3% for weeks 25 through 28 (Appendix Figure A4). Unlike the labels' impacts on meal choices, their impacts on retention show no signs of dissipating by the end of the experimental period.

Robustness: Our estimates are broadly robust to controls, though they lose precision when we exclude controls for longevity fixed effects and lagged outcomes. As expected, the treatment estimates are slightly lower and become statistically insignificant when we define the experimental period as beginning in week 21 (Appendix Figure A5).

3.3 Backlash and heterogeneity in treatment effects

Larger meal-choice and retention effects among high-carbon customers: The carbon-footprint labels have larger impacts on both meal choices and customer retention among customers with high carbon footprints at baseline, approximated by the share of their pre-experimental meals that contained beef. Separately estimating treatment effects among those below the 25th percentile, between the 25th and 75th percentiles, and above the 75th percentile of baseline beef consumption (Figure 2), the labels only significantly decrease box-level carbon footprints and consumption of

Fair meals among customers with medium or high baseline beef consumption. Notably, the null treatment effects among those with low baseline beef consumption do not just arise from floor effects; this group consumed 0.25 *Fair* meals per week on average during the pre-experimental period, compared to means of 0.64 and 0.83 among the medium and high beef consumers, respectively.

Likewise, the carbon labels yield the largest retention increases among customers with high baseline beef consumption (Figure 3), making them 1pp (3.8%) more likely to order a meal box in a given week. In contrast, the labels had small and largely insignificant retention and profit effects among those with lower baseline beef consumption.

But backlash in meal choices by climate beliefs: At the same time, the labels may generate backlash in carbon footprints among those who disagree with their purpose, elicited in a series of Likert-style scales in our baseline survey. In particular, the labels increase carbon footprints among customers with below-median agreement that companies and individuals have a moral duty to address climate change, while decreasing carbon footprints among those with at- or above-median beliefs on each dimension (Figure 2).[14] We find similar, though less precise, heterogeneity by correlated measures like self-perceived environmentalism and belief that climate change is human-caused. (See Appendix Table A2 for correlations and Appendix Figure A6 for heterogeneous treatment effects by these measures.) In contrast, there is no evidence of backlash among customers who believe that climate labels are simply ineffective.

These patterns suggest that the carbon-footprint labels may require careful targeting. Even among HelloFresh's customers, who are disproportionately drawn from wealthy and educated parts of the US, ideological non-congruence with the purpose of climate labels generates substantial backlash.

While customers who disagree with the labels' intent react to them by increasing their consumption of high-carbon meals, we find no evidence that this backlash extends to leaving HelloFresh. We estimate zero retention effects among customers with below-median climate morality beliefs in the full baseline sample (Figure 3).

Reconciling these patterns: The heterogeneous impacts of labels by beef consumption and climate beliefs raise a puzzle: customers with higher baseline beef consumption are less likely to believe that companies and individuals have duties to address climate (Appendix Figure A7), and yet on average the labels' impacts on meal choices go in opposite directions for customers who eat more beef on average and customers who disagree with the labels' moral underpinnings. Crossing these dimensions of heterogeneity, we find that one subgroup drives the labels' impacts on carbon footprints (Appendix Figure A8): customers who both had high baseline beef consumption and also more strongly believe that individuals and companies have a moral duty to address climate change. Notably, the labels' effects on retention are also largest among this customer subgroup (Appendix Figure A9). In the next section, we turn to understanding the mechanisms by which labels shaped this group's meal choices and retention.

[14] These heterogeneous effects by moral beliefs persist when collapsing the morality beliefs to an index and controlling simultaneously for interactions with baseline beef consumption and footprint knowledge (Appendix Table A1).

3.4 Mechanisms

3.4.1 Why do labels change meal choices?

The labels could change meal choices either by providing new information about the carbon footprints of different foods or by making customers' existing knowledge salient.

Several pieces of evidence suggest that the labels work at least in part through salience. First, all three of the labeling schemes have highly similar impacts on meal choices (Panel A, Appendix Table A3). The Climate Superstar labeling arm only provides information to distinguish the *Climate-Superstar* meals from all others, and yet it substantially reduces consumption of *Fair* meals relative to *Good* meals. This pattern suggests that the labels work in large part by activating—rather than adding to—customers' baseline knowledge about beef's large carbon footprint.

Moreover, we find no evidence that labels have larger effects among those without baseline carbon-footprint knowledge (Figure 2), as we would expect if the labels primarily work through information provision. We elicit baseline knowledge by asking baseline-survey respondents to rank six foods from that which generates the most to the least emissions per pound on average: beef, cheese, farmed shrimp, pork, chicken, and tomatoes, listed here in the correct order by median emissions.[15] We define baseline knowledge by whether participants correctly answer that beef and tomatoes have the highest and lowest carbon footprints, respectively; about 51% do so correctly. Only about 1% of the sample correctly produce the full scale, including farmed shrimp and cheese.

We find no gap in estimated effects on carbon footprint or number of *Fair* meals by baseline footprint knowledge. Moreover, among the subset of customers with the largest meal-choice effects—those with high baseline beef consumption and strong climate-morality beliefs—the treatment effects are driven entirely by those with pre-existing carbon-footprint knowledge (Appendix Figure A8). Thus, the climate labels seem to primarily work by providing a salience nudge to customers who believe they should be eating less beef due to its effect on the climate, rather than by correcting customers' misperceptions.

3.4.2 Why do customers value carbon labels?

Even if the labels largely change meal choices by reminding customers about their baseline footprint knowledge, customers may value the labels—and differentially remain at HelloFresh—for other reasons. First, customers may value labels because they derive warm-glow benefits from the labels' affirmation of their environmentally friendly choices. Second, customers may be more likely to continue their subscription if the labels increase their perception that HelloFresh is pro-social, a form of possible "greenwashing." Third, customers may value labels because they provide new information about the carbon footprints of different foods, even if this information does not change their meal choices. Finally, they may value the labels' reminders of their existing knowledge about foods' climate impacts.

Warm glow and retention effects by baseline footprints: As we note in Section 3.3, the labels have significantly larger retention effects among customers with higher baseline beef consumption

[15] Our estimates of the truth here come from data on greenhouse gas emissions per kilogram of food products estimated by Poore and Nemecek (2018) and compiled by Ritchie (2020). These estimates include emissions from the full production process: from land use change associated with food production to the production and disposal of packaging materials. Our measure of knowledge derives from simple rankings of emissions from proteins like beef and chicken and are robust to error in these estimates.

(Figure 3). This pattern suggests that customers do not primarily value carbon labels due to warm glow from a stamp of approval on low-carbon meals. If this were so, we would expect low-carbon customers, not high-beef customers, to drive retention increases.

Positive updating about HelloFresh: We find mixed, inconclusive evidence for whether the labels increase retention by changing perceptions of HelloFresh as a company. First, our point estimates for retention effects are higher among customers with greater beliefs in companies' moral imperative for climate action, who we might expect to more strongly value corporate pro-sociality (Figure 3). On the other hand, the retention and profit effects are driven by those with above-median baseline beliefs in HelloFresh's pro-sociality, whom we might expect to update these beliefs less (Appendix Figure A6, Panel B). Our ability to speak to this mechanism is limited in the absence of outcome data on customers' posterior perceptions of HelloFresh, and both heterogeneity patterns are imprecise.

Valuing reminders versus information: We find mixed heterogeneity in the labels' retention effects by customers' baseline knowledge of carbon footprints. In the full baseline sample, our point estimate for the impacts of carbon-footprint labels on customer retention are larger among those without carbon-footprint knowledge at baseline than those with knowledge (Figure 3). While this pattern is consistent with customers valuing new information about carbon footprints, we cannot reject that these coefficients are equal across groups.

On the other hand, baseline carbon-footprint ignorance only predicts higher retention effects among those with medium baseline beef consumption, while this pattern flips among the high-baseline-beef customers who drive the overall retention effects (Appendix Figure A9). Among high-beef customers with above-median climate-morality beliefs—those with the largest retention effects overall—the labels only increase retention and profit among those with baseline carbon-footprint knowledge (Appendix Figure A10). These results suggest that the labels' retention effects arise, at least in large part, because this group of customers values the labels as a nudge or reminder of their existing carbon-footprint knowledge.

4 Discussion and conclusion

Food systems contribute between 25-35% of total annual greenhouse gas emissions, and steep differences in the footprints of different foods mean that dietary shifts away from certain products, such as beef, could heavily reduce these emissions. We collaborate with HelloFresh on the first experimental evaluation of food-based carbon-footprint labels in a commercial US setting and the largest experimental evaluation of carbon labels in any setting to date.

Carbon-footprint labels increase customer retention by 1.1% and profit per customer-week by 0.9%, so carbon labeling may remain a common sustainability policy among companies like HelloFresh. However, the labels only reduce customers' carbon footprints by 0.6%. These effects are small relative to the changes necessary to reach global emissions targets. If widespread carbon labeling reduced Americans' per-capita total food emissions by 0.6%, this reduction would achieve only 1% of the required decrease in food-related emissions proposed by the EAT-Lancet commission to reach targets laid out in the Sustainable Development Goals and the Paris Agreement (Willett et al., 2019; Bassi et al., 2022). Moreover, these impacts fall somewhat over time. If the labels' impacts

on meal choices indeed fall to zero in the long run, commercial carbon labeling could ultimately serve as a form of greenwashing, regardless of companies' intentions.

Our results also suggest the need for careful label targeting. Even as the labels reduce customers' carbon footprints on average, they increase carbon footprints among those with weak beliefs in companies' and individuals' moral obligations to address climate change. This backlash suggests caution in proceeding—at least in the United States—with any form of mandatory carbon labeling. Our results suggest that some form of self-targeting could be effective: the subset of customers whose meal choices change most also show the largest retention increases at HelloFresh. A voluntary system that allows customers to opt out of seeing climate-related information could be an effective initial step.

Even as companies may have profit incentives to nudge customers towards greener products, consumer-facing carbon labels must be well-targeted and can only be one component in a broad, cross-sectoral response to climate change.

References

Angrist, Noam, Kevin Winseck, Harry Anthony Patrinos, and Joshua Graff Zivin, "Human capital and climate change," *Review of Economics and Statistics*, 2024, pp. 1–28. 2.3.1, C.1

Bassi, Clare, Rachael Maysels, and Rob Anex, "Declining greenhouse gas emissions in the US diet (2003–2018): Drivers and demographic trends," *Journal of Cleaner Production*, 2022, *351*, 131465. 4

Bauder, David and Ali Swenson, "The Hamburglar? How a story about meat limits fell apart," *The Associated Press*, 2021. 7

Bilén, David, "Do carbon labels cause consumers to reduce their emissions? Evidence from a large scale natural experiment," *Working Paper*, 2024. 2, 3, 6

Butera, Luigi, Robert Metcalfe, William Morrison, and Dmitry Taubinsky, "Measuring the Welfare Effects of Shame and Pride," *American Economic Review*, January 2022, *112* (1), 122–68. 3.2

Chater, Nick and George Loewenstein, "The i-frame and the s-frame: How focusing on individual-level solutions has led behavioral public policy astray," *Behavioral and Brain Sciences*, 2023, *46.* 1

Clark, Michael A, Nina GG Domingo, Kimberly Colgan, Sumil K Thakrar, David Tilman, John Lynch, Inês L Azevedo, and Jason D Hill, "Global Food System Emissions Could Preclude Achieving the 1.5 and 2 C Climate Change Targets," *Science*, 2020, *370* (6517), 705–708. 1

Commisso, Danielle, "Meal Kit Delivery Continues to Thrive In the Stay-at-Home Economy," Technical Report, Civic Science 2 2022. 1

Crippa, Monica, Efisio Solazzo, Diego Guizzardi, Fabio Monforti-Ferrario, Francesco N Tubiello, and AJNF Leip, "Food Systems are Responsible for a Third of Global Anthropogenic GHG Emissions," *Nature Food*, 2021, *2* (3), 198–209. 1

Dechezleprêtre, Antoine, Adrien Fabre, Tobias Kruse, Bluebery Planterose, Ana Sanchez Chico, and Stefanie Stantcheva, "Fighting climate change: International attitudes toward climate policies," *NBER Working Paper #30265*, 2024. 2, 2.3.1, C.1

EPA, "Greenhouse Gas Equivalencies Calculator," 2024. https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator. 1, 3.1

Hagmann, David, Emily H. Ho, and George Loewenstein, "Nudging out support for a carbon tax," *Nature Climate Change*, 05 2019, *9*, 484–489. 1

HelloFresh, "Q3 2024: HelloFresh Group achieved a year-on-year constant currency revenue growth of 1.9% and an adjusted EBITDA of €72 million," Quarterly Financial Report, HelloFresh 10 2024. Corporate News, Q3 2024 Financial Results. 4

Imai, Taisuke, Davide D Pace, Peter Schwardmann, and Joël J van der Weele, "Correcting Consumer Misperceptions About CO 2 Emissions," *CESifo working paper 10138*, 2022. 1

Ivanovich, Catherine C., Tianyi Sun, Doria R. Gordon, and Ilissa B. Ocko, "Future warming from global food consumption," *Nature Climate Change*, 2023, *13*, 297–302. 1

Jones, Matthew W, Glen P Peters, Thomas Gasser, Robbie M Andrew, Clemens Schwingshackl, Johannes Gütschow, Richard A Houghton, Pierre Friedlingstein, Julia Pongratz, and Corinne Le Quéré, "National contributions to climate change due to historical emissions of carbon dioxide, methane, and nitrous oxide since 1850," *Scientific Data*, 2023, *10* (1), 155. 1

Kozicka, M., P. Havlík, H. Valin et al., "Feeding climate and biodiversity goals with novel plant-based meat and milk alternatives," *Nature Communications*, 2023, *14*, 5316. 1

Lee, Stephen, "White House Awards Help Construction Sector Make Carbon Labels," July 2024. 1

Li, Yanxian, Pan He, Yuli Shan, Yu Li, Ye Hang, Shuai Shao, Franco Ruzzenenti, and Klaus Hubacek, "Reducing climate change impacts from the global food system through diet shifts," *Nature Climate Change*, 08 2024, *14*, 943–953. 1

Lindsey, Rebecca and Luann Dahlman, "Climate Change: Global Temperature," Jan 2024. 1

Lohmann, Paul M, Elisabeth Gsottbauer, Anya Doherty, and Andreas Kontoleon, "Do carbon footprint labels promote climatarian diets? Evidence from a large-scale field experiment," *Journal of Environmental Economics and Management*, 2022, *114*, 102693. 2, 6

Perri, Janine, "Competition heats up among meal kit companies like Blue Apron and HelloFresh," Technical Report, Bloomberg Second Measure 1 2024. 4

Poore, Joseph and Thomas Nemecek, "Reducing Food's Environmental Impacts Through Producers and Consumers," *Science*, 2018, *360* (6392), 987–992. 1, C.4.3

Potter, Christina, Anastasios Bastounis, Jamie Hartmann-Boyce, Cristina Stewart, Kerstin Frie, Kate Tudor, Filippo Bianchi, Emma Cartwright, Brian Cook, Mike Rayner et al., "The effects of environmental sustainability labels on selection, purchase, and consumption of food and drink products: a systematic review," *Environment and behavior*, 2021, *53* (8), 891–925. 2

Ripple, William J, Pete Smith, Helmut Haberl, Stephen A Montzka, Clive McAlpine, and Douglas H Boucher, "Ruminants, Climate Change and Climate Policy," *Nature Climate Change*, 2014, *4* (1), 2–5. 1

Ritchie, Hannah, Pablo Rosado, and Max Roser, "Meat and Dairy Production," *Our World in Data*, 2019. https://ourworldindata.org/meat-production. 1

_ , _ , and _ , "Greenhouse gas emissions," *Our World in Data*, 2020. https://ourworldindata.org/greenhouse-gas-emissions. 1

Taufique, Khan MR, Kristian S Nielsen, Thomas Dietz, Rachael Shwom, Paul C Stern, and Michael P Vandenbergh, "Revisiting the promise of carbon labelling," *Nature Climate Change*, 2022, *12* (2), 132–140. 2

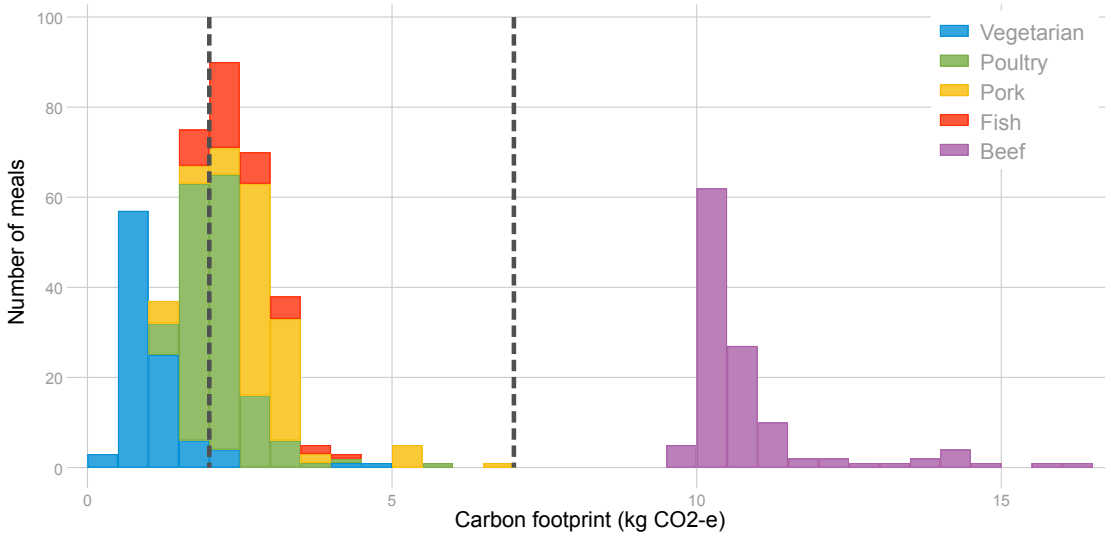
Tilling, Anna Schulze, "Changing consumption behavior with carbon labels: Causal evidence on behavioral channels and effectiveness," *Working Paper*, 2025. 2, 1

Tubiello, Francesco N., Kevin Karl, Alessandro Flammini, Johannes Gütschow, Griffiths Obli-Laryea, Giulia Conchedda, Xueyao Pan, Sally Yue Qi, Hörn Halldórudóttir Heiarsdóttir, Nathan Wanner, Roberta Quadrelli, Leonardo Rocha Souza, Philippe Benoit, Matthew Hayek, David Sandalow, Erik Mencos Contreras, Cynthia Rosenzweig, Jose Rosero Moncayo, Piero Conforti, and Maximo Torero, "Pre- and post-production processes increasingly dominate greenhouse gas emissions from agri-food systems," *Earth System Science Data*, 2022, *14*, 1795–1809. 1

Washington Post Staff, "Republicans seize on Biden's climate plan to inaccurately claim it will limit meat consumption," *The Washington Post*, 2021. Accessed: 2024-10-16. 1

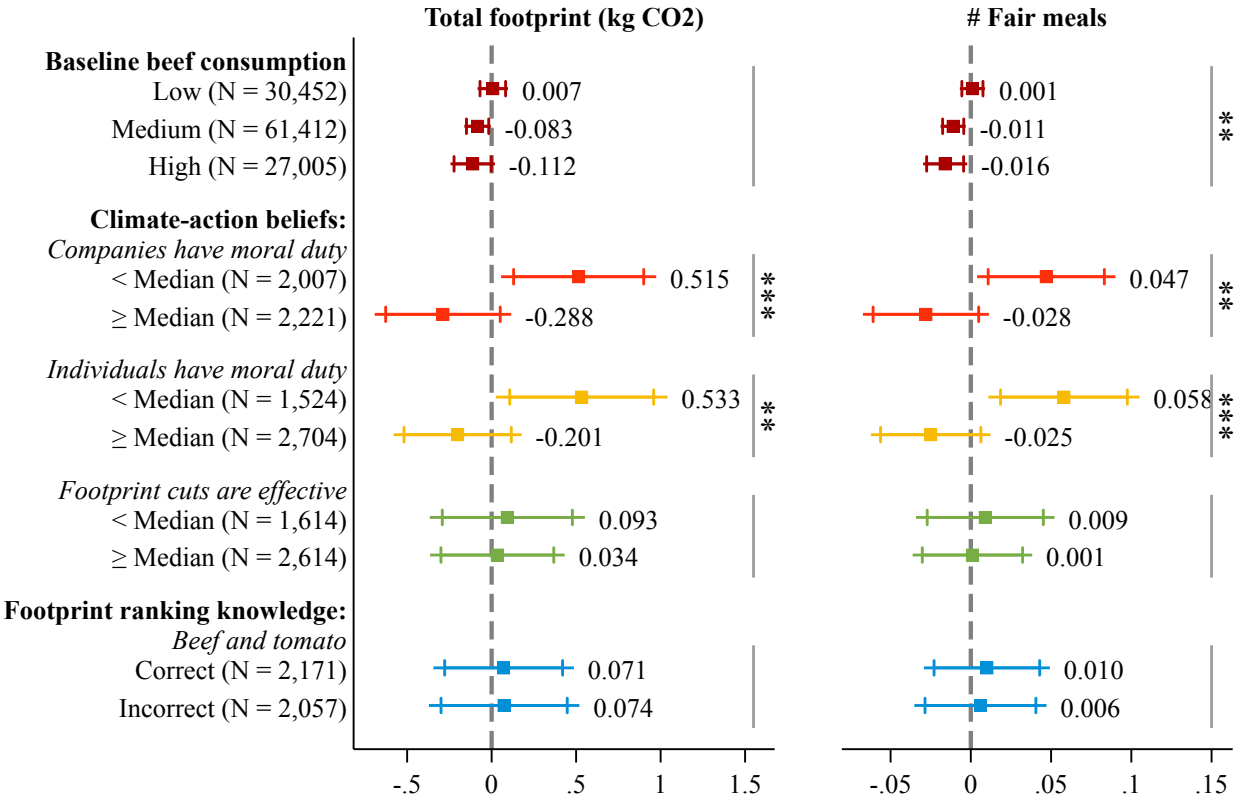
Willett, Walter, Johan Rockström, Brent Loken, Marco Springmann, Tim Lang, Sonja Vermeulen, Tara Garnett, David Tilman, Fabrice DeClerck, Amanda Wood et al., "Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems," *The lancet*, 2019, *393* (10170), 447–492. 4

Figure 1: The distributions of meals' carbon footprints



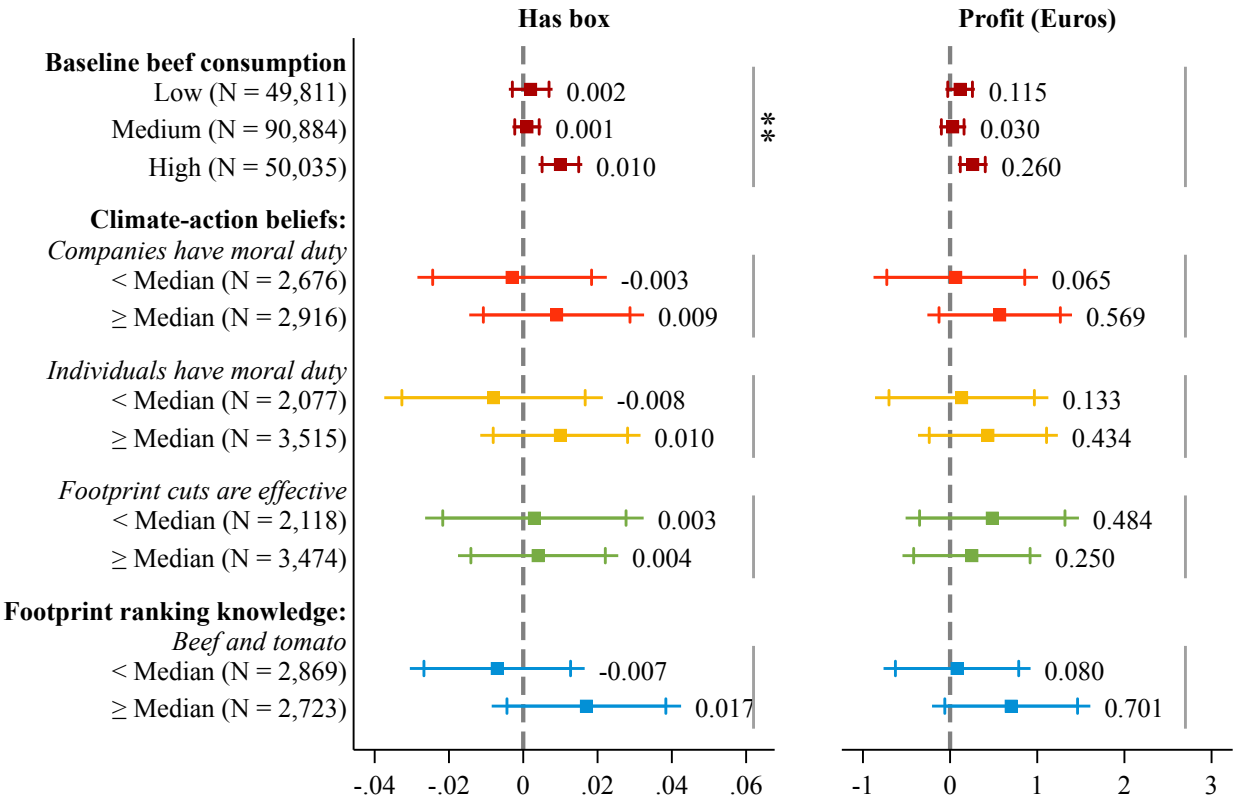
Note: This figure plots the distribution of carbon footprints, measured in CO_2-equivalents, for the 505 meals that HelloFresh offered from fiscal week 21 through week 28, separating the distributions for meals with different proteins. When a meal had multiple proteins, we categorize it as containing the higher-emissions protein (Poore and Nemecek 2018). We ignore minor protein additions such as a sprinkle of bacon or a slice of prosciutto. The two black dashed vertical lines denote the thresholds separating *Fair* vs *Good* meals (2 kg CO_2e) and *Good* vs *Superstar* meals (7 kg CO_2e). HelloFresh calculated these carbon footprints using the Agribalyse dataset, which estimates products' carbon footprint using Life Cycle Assessments (LCA) and is widely cited, including by the EU and the UN.

Figure 2: Heterogeneous effects on meal choices by customer traits



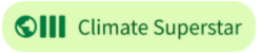
Note: This figure plots heterogeneous impacts of climate labels on meal choices by customers' baseline traits. Outcomes in the left and right panels are the total carbon footprint of ordered meal boxes and number of fair meals ordered, respectively. We first test heterogeneity by baseline beef consumption during the pre-experimental period (fiscal weeks 8 through 20). We define low, medium, and high baseline beef consumers as those whose share of pre-period meals containing beef falls below the 25th percentile, between the 25th and 75th percentile, or above the 75th percentile across customers, respectively. Customers who did not order at least one meal during the pre-treatment period are excluded from these regressions. Here and for all other dimensions of heterogeneity shown, we show the number of customers in each heterogeneity category in parentheses beside each label. Next, we test heterogeneity by three climate-action beliefs measured in the baseline survey: (i) whether *companies* have a moral duty to address climate change, (ii) whether *individuals* have a moral duty to address climate change, and (iii) whether individual efforts to reduce personal carbon footprints can effectively address climate change. For each, we separately estimate treatment effects among those whose beliefs on these dimensions are below versus at or above those of the median baseline survey participant. Finally, we test heterogeneity by baseline carbon-footprint knowledge, measured by whether customers correctly select that beef and tomatoes have the highest and lowest carbon footprints in a list of foods. Appendix C.4 details how we define each of the dimensions of heterogeneity. For each dimension, we separately estimate treatment effects via Equation 2.4 in each subgroup. Regressions are otherwise identical to those in Table 2. Spiked and capped error bars show 90% and 95% confidence intervals for these estimates, respectively, with standard errors clustered by customer. We test that coefficients are equal across subgroups, indicating that we can reject equality at the 10%, 5%, or 1% levels with *, **, and ***, respectively. When there are three subgroups, the null hypothesis is that all three coefficients are equal.

Figure 3: Heterogeneous effects on retention and profit by customer traits



Note: This figure shows the heterogeneous treatment effects of climate labels on customer retention and profits by baseline customer traits. The outcome variables in the left and right panels are an indicator for whether the customer ordered a box in a given week and profit that HelloFresh made from a customer in a given week (in Euros), respectively. This figure is constructed using the same baseline traits as in Figure 3 and takes the same structure as that figure. Spiked and capped error bars show 90% and 95% confidence intervals for these estimates, respectively, with standard errors clustered by customer. We test that coefficients are equal across subgroups, indicating that we can reject equality at the 10%, 5%, or 1% levels with *, **, and ***, respectively. When there are three subgroups, the null hypothesis is that all three coefficients are equal.

Table 1: Randomized label variations

Treatment group	N	Labels on main menu	Labels on menu cards
Control	58,645	None	None
Superstars Only	59,040		
All Tiers	58,385		
All Tiers–Letters	58,441		

Note: This table shows the carbon-label variants included in the experiment. The second column presents the number of customers assigned to see menus with each label variation. The third column shows the carbon-footprint labels shown on the main HelloFresh menu for *Climate Superstar*, *Good*, and *Fair* meals, respectively from left to right. Note that only *Climate-Superstar* meals were labeled with the tag shown for the Superstars-Only group. The fourth column shows the expanded labels visible on detailed menu cards that customers could click to from the main menu. Appendix Figure A1 shows screenshots of how the All-Tiers labels appeared on sample menus, including explanations of the carbon-footprint categories.

Table 2: Main treatment effects of climate labels

Panel A. Meal-choice outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Footprint	# Meals by footprint tier			# Meals by protein				
	(kg CO2)	Superstar	Good	Fair	Veggie	Poultry	Pork	Fish	Beef
Any label	-0.075***	0.007*	0.002	-0.009***	0.005*	0.003	0.003	-0.001	-0.009***
	(0.028)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.002)	(0.003)
Cont. mean	11.775	0.856	1.635	0.589	0.439	1.070	0.851	0.167	0.589
# Customers	131,017	131,017	131,017	131,017	131,017	131,017	131,017	131,017	131,017
# Cust-weeks	397,514	397,514	397,514	397,514	397,514	397,514	397,514	397,514	397,514

Panel B. Retention and profit outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
		# Meals	Revenue	Cost	Profit (Euros)	
	Has box	/ box	(Euros)	(Euros)	All	If box
Any label	0.003***	0.001	0.204**	0.129**	0.072*	-0.071
	(0.001)	(0.004)	(0.100)	(0.062)	(0.043)	(0.076)
Cont. mean	0.281	3.078	20.628	12.720	7.975	28.384
# Customers	234,511	131,049	234,511	234,511	234,511	131,049
# Cust-weeks	1,407,066	397,565	1,407,066	1,407,066	1,407,066	397,565

Note: This table reports the treatment effects of the climate labels on customer meal choices (Panel A), customer retention, and profit per customer (Panel B). We estimate the impacts of an indicator that a customer was assigned to any of the three climate-label variations (Table 1), as in Equation 2.4. Appendix Table A3 presents results of parallel regressions in which we separately estimate the treatment effects of each label variation. The sample in Panel A comprises all customer-weeks in which customers ordered a meal box from HelloFresh. The outcomes in Panel A are the total estimated carbon-footprint of each customers' box (column 1), the number of meals ordered in each carbon-footprint category (columns 2-4), and the number of meals ordered that contain each protein type (column 5-9). The outcomes in Panel B are an indicator for whether customers ordered a meal box in a given week (column 1), the number of meals they ordered conditional on ordering a box (column 2), HelloFresh's revenue (column 3), costs (column 4), and profit (column 5) per customer in each week (setting these fields to zero if customers did not order a box), and profit per customer conditional on ordering a box (column 6). The samples in columns 1 and 3 to 5 of Panel B include observations for each customer in each experimental week, while the samples for Panel B columns 2 and 6 only includes customer-weeks in which customers ordered a meal box. Appendix C.3 details the construction of all outcomes. Regressions include controls for lagged outcomes in weeks 18, 19, and 20 (as well as indicators that those values are missing), fixed effects for customers' meal plan with HelloFresh (e.g. vegetarian or chef's choice), zipcode-level characteristics (political leanings, education, and racial mix), customer longevity at HelloFresh, and week fixed effects. Appendix C.4 details these controls. Standard errors are clustered at the customer level. We indicate statistical significance at the 10%, 5%, and 1% levels by *, **, and ***, respectively.

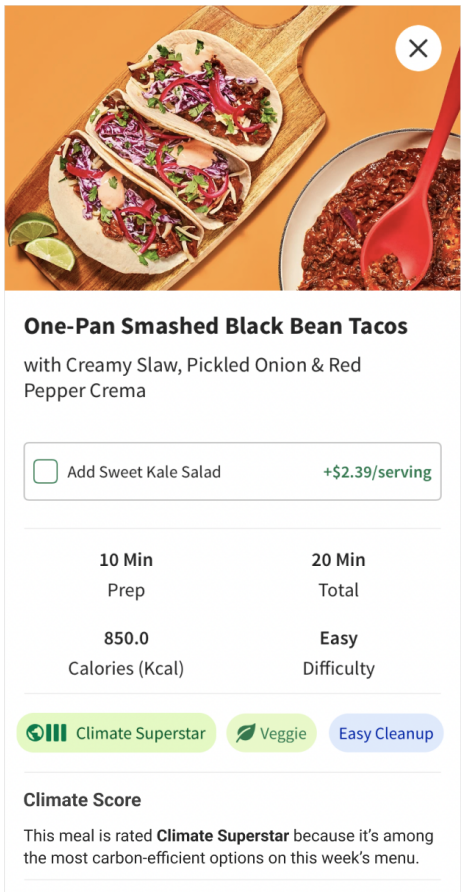
A Additional figures

Figure A1: HelloFresh menu except with All-Tiers labels

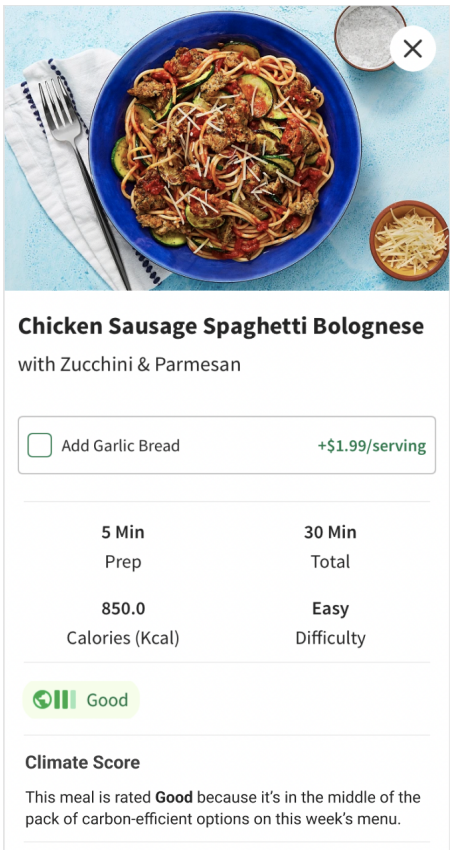
(a) Excerpt from a sample menu



(b) Menu card example 1

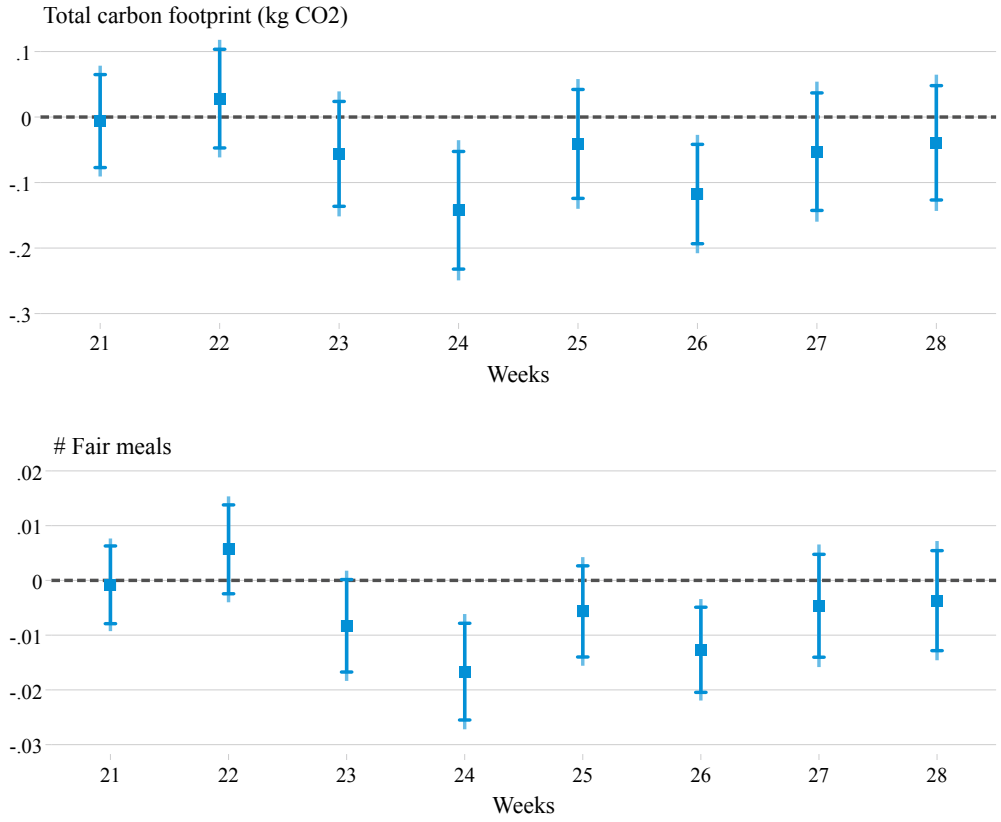


(c) Menu card example 2



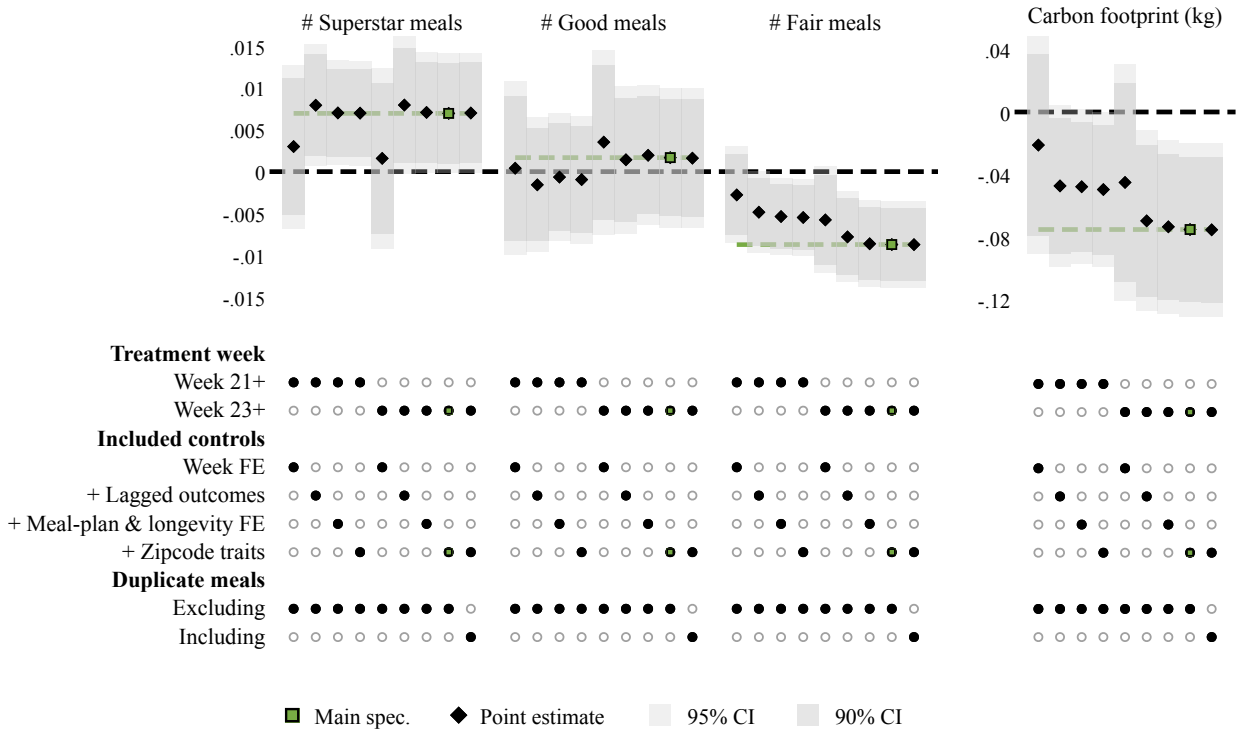
Note: This figure shows excerpts of sample HelloFresh menus, including the carbon labels from the All-Tiers variant (Table 1). Subfigure (a) shows an excerpt from a sample main menu, which includes photo entries for each of about 25 recipes per week. These small placards include recipe tags including carbon-footprint labels and tags for being e.g. vegetarian, spicy, or easy-cleanup. Customers can click on each of the recipe placards to pull up a more detailed recipe card, with two examples shown in Subfigures (b) and (c). These include more detailed versions of the carbon-footprint labels, a short explanation of what these labels mean, other recipe tags, information on prep and cook times, nutrition information, and a detailed recipe description. (These screenshots cut off the detailed nutrition information and recipe description that are provided.)

Figure A2: Treatment effects on meal choices over time



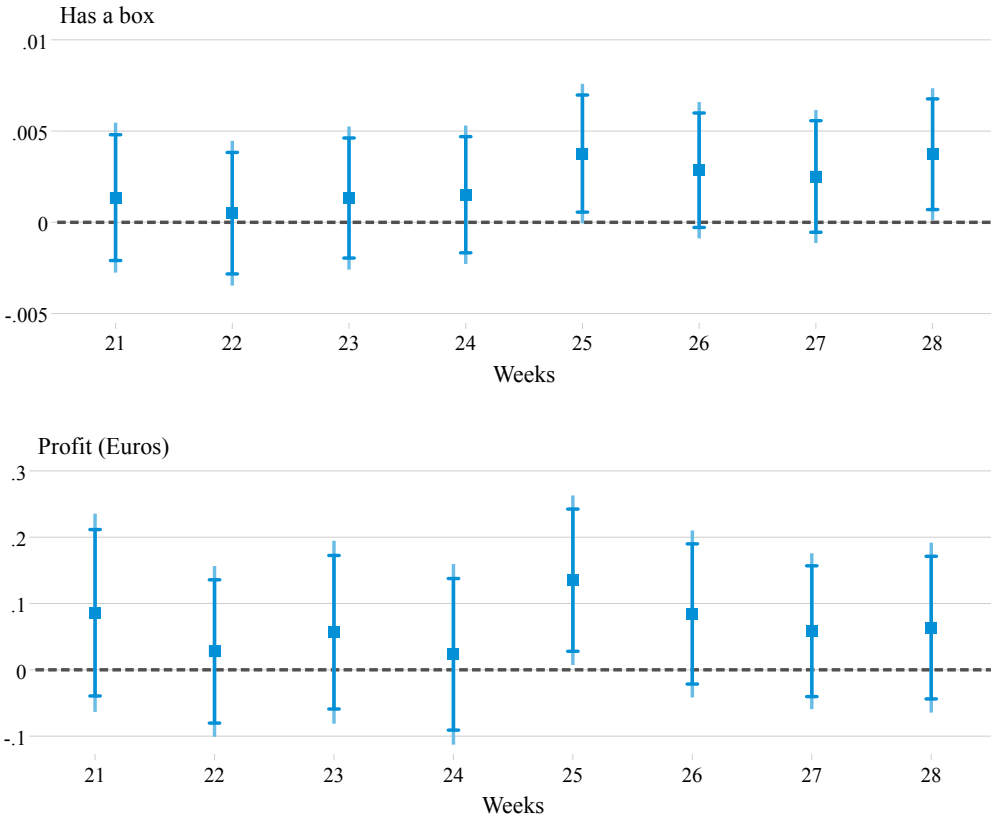
Note: This figure shows treatment effects of the carbon labels on meal choices over time. The top panel shows effects on the total carbon footprint of meals that customers ordered each week, while the second panel shows effects on the number of Fair meals that customers ordered each week. The x-axis shows fiscal weeks. These effects are estimated via Equation 2.4, restricting the data to a single week. All controls and outcomes are as described in the note to Table 2. Note that the labels were added to HelloFresh menus midway through week 20 for menus for weeks 21 through 28. Customers can view menus and select meals up to 5 weeks in advance, so customers could have chosen meals without seeing carbon labels up to week 25. This attenuation of effects is likely strongest for weeks 21 and 22. The capped and uncapped lines denote 90% and 95% confidence intervals, respectively. Standard errors are clustered by customer.

Figure A3: Robustness of meal-choice effects



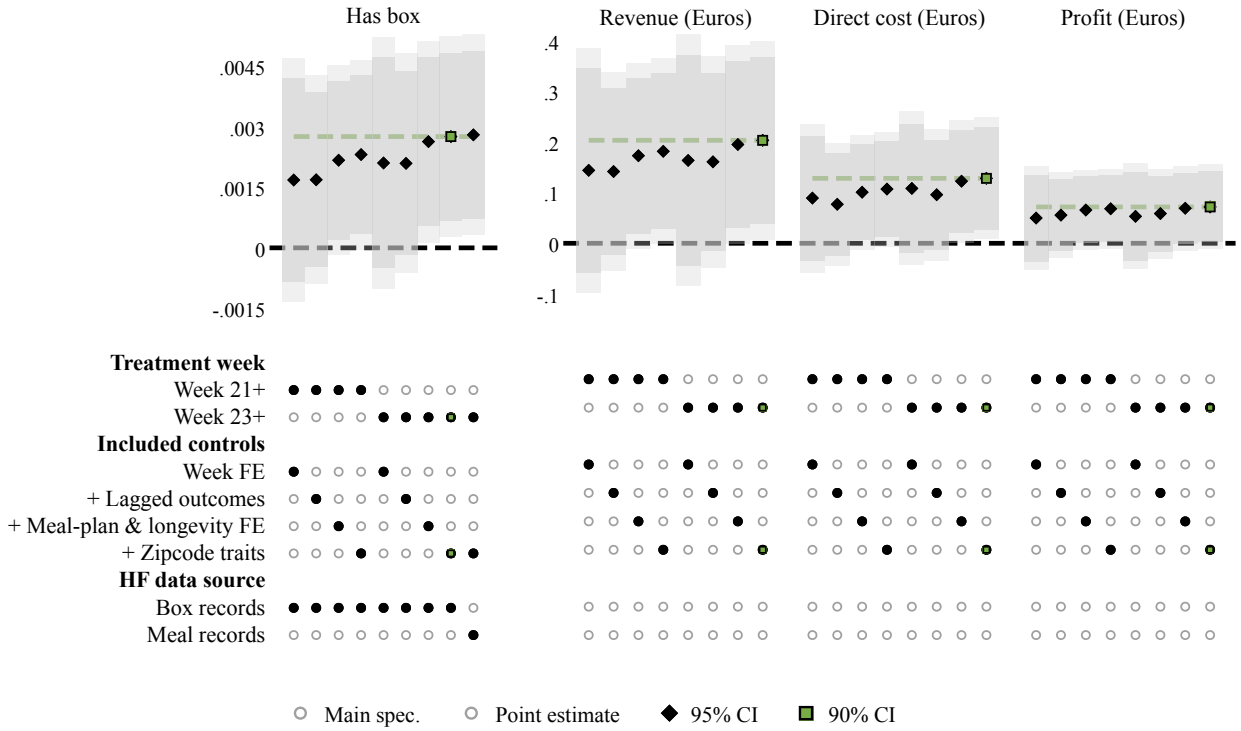
Note: This figure tests the robustness of the labels' impacts on four outcomes: number of *Climate Superstar*, *Good*, and *Fair* meals ordered each week, and total estimated carbon footprint of meals ordered in kg of CO_2e. We present point estimates for versions of Equation 2.4 where we vary whether we define the treatment weeks as starting in weeks 21 or 23, the controls included, and how we treat duplicated observations when defining customers' meal choices (Appendix C.3). The control variants first include only week fixed effects, then add in lagged outcomes for pre-experimental weeks, fixed effects for customers' meal plans and longevity at HelloFresh, and zipcode-level demographics. Appendix C.4 describes these controls in detail.

Figure A4: Treatment effects on retention and profit over time



Note: This figure parallels Appendix Figure A2, except that we estimate treatment effects over time on whether customers order a meal box in a given week and on profit per customer. Appendix C.3 details these outcomes.

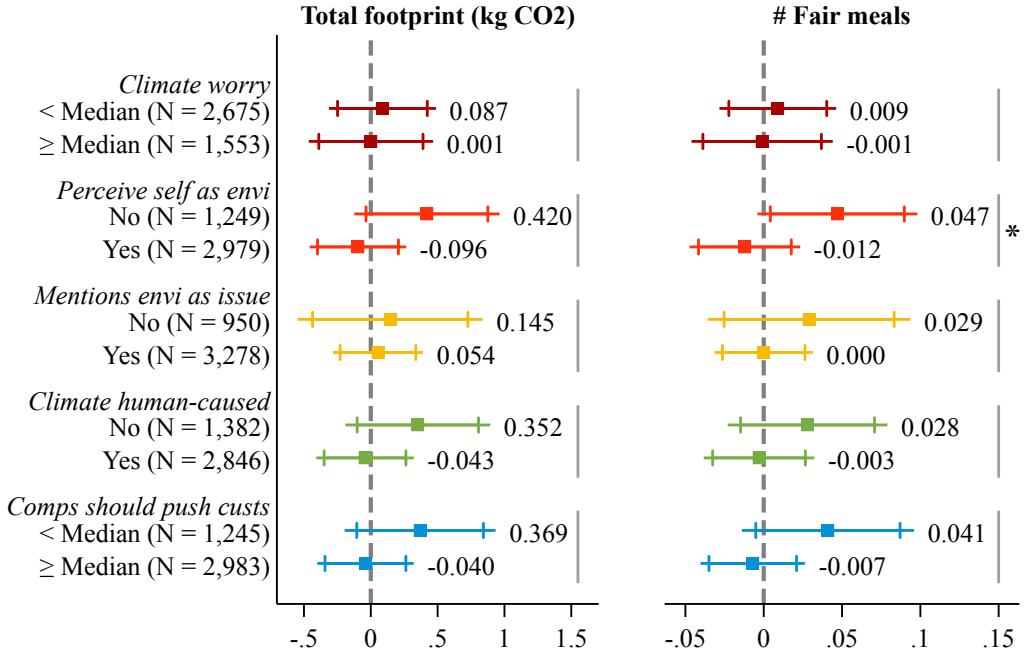
Figure A5: Robustness of retention and profit effects



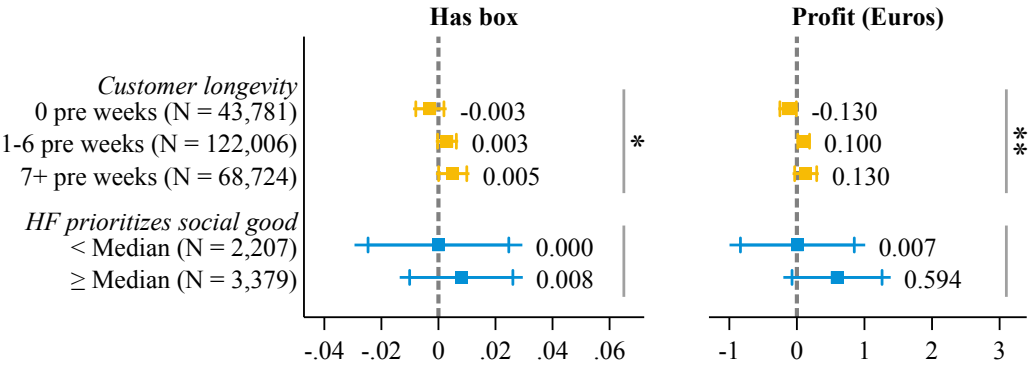
Note: This figure parallels Appendix Figure A3, except that we test robustness of treatment effects on whether customers order a meal box in a given week, revenue per customer, direct meal cost per customer, and profit per customer. Appendix C.3 details these outcomes. We do not test robustness of these outcomes to how we handle duplicate meal-order observations, as in Appendix Figure A3, since we observe all of these outcomes in administrative data at the customer-week level rather than constructing them from customer-meal-level data. We test robustness of the estimated retention effects to whether we observe whether customers order a meal box from box-value records or meal-choice records obtained from HelloFresh, as these records differ in about 3% of customer-weeks in the experimental period (Appendix C.3).

Figure A6: Heterogeneous main effects by other climate beliefs and customer traits

Panel A. Other relevant heterogeneity in meal-choice effects

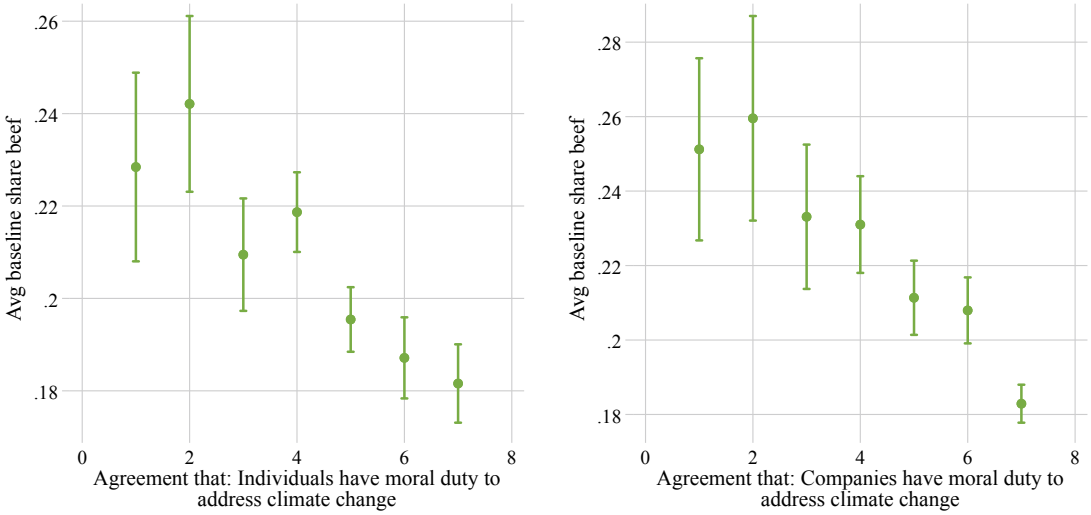


Panel B. Other relevant heterogeneity in retention effects



Note: This figure plots heterogeneous impacts of climate labels on meal choices (Panel A) and retention/profit (Panel B) by dimensions of heterogeneity not shown in Figures 2 and 3. In Panel A, we test heterogeneity by: (i) whether customers score above or below the median of climate worry, (ii) whether they say that they perceive themselves as environmentalists "Moderately" or "Very Much", (iii) whether they mention an environmental issue as a social or political issue that is personally important to them, (iv) whether they report believing that climate change is mostly human-caused, and (v) whether they report above or below median belief that companies should push customers to make more climate-friendly choices. In Panel B, we test heterogeneity by customer longevity, measured as the number of pre-experimental weeks in which they ordered a meal box, and whether they express above- or below-median baseline belief that HelloFresh prioritizes social good in its operations. Appendix C.4.3 details these dimensions of heterogeneity. Spiked and capped error bars show 90% and 95% confidence intervals for these estimates, respectively, with standard errors clustered by customer. We test that coefficients are equal across subgroups, indicating that we can reject equality at the 10%, 5%, or 1% levels with *, **, and ***, respectively. When there are three subgroups, the null hypothesis is that all three coefficients are equal.

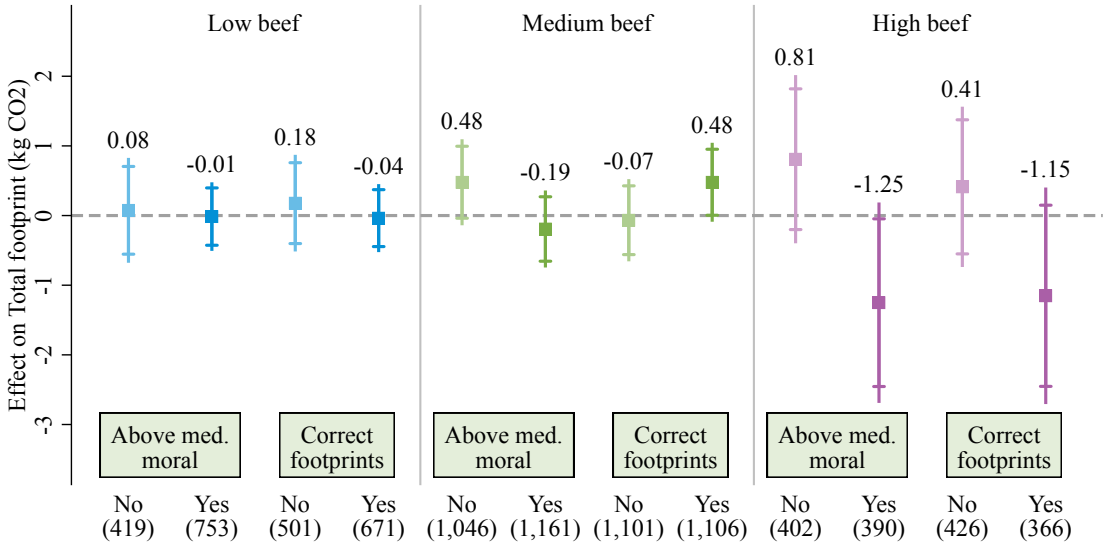
Figure A7: Correlation of baseline climate beliefs and beef consumption



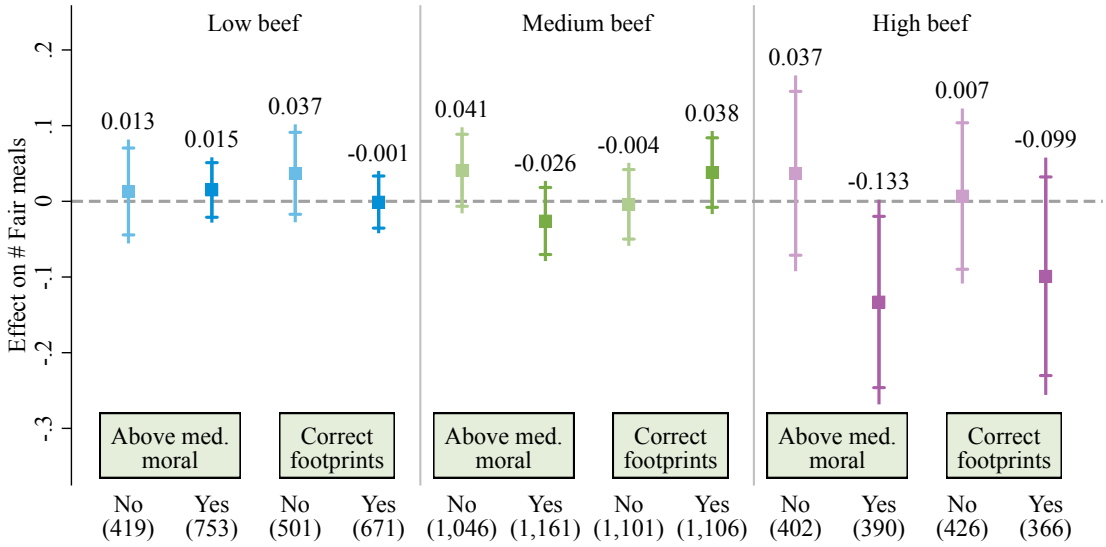
Note: This figure shows the correlation of climate beliefs and baseline beef consumption in the baseline sample, plotting the average share of baseline meals that contained beef among baseline-survey respondents with a given stated agreement with each of the climate-belief statements shown. The capped lines on each point value show the 95% confidence intervals for each value, with width calculated by multiplying the standard error of the mean in each bin by 1.96. See Appendix C.4 for details on the climate-belief elicitations. 174 baseline-survey participants ordered no pre-period meals, so we cannot measure their baseline beef consumption and exclude them from this figure. This restriction leaves a sample of 5,418 customers.

Figure A8: Crossed heterogeneous meal-choice effects by baseline beef vs. climate beliefs and knowledge

Panel A. Total carbon footprint (kg CO_2)



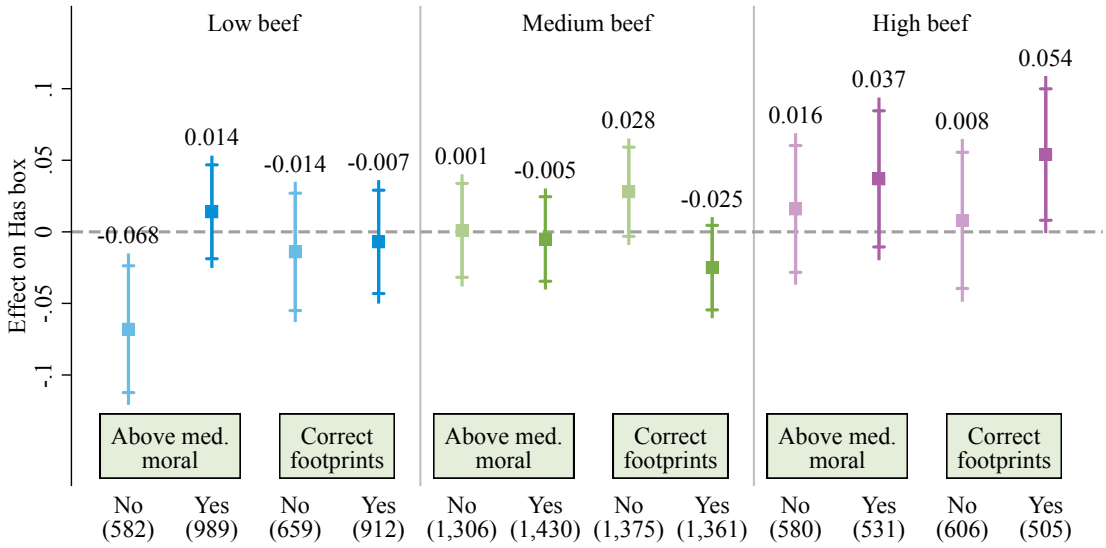
Panel B. # Fair meals



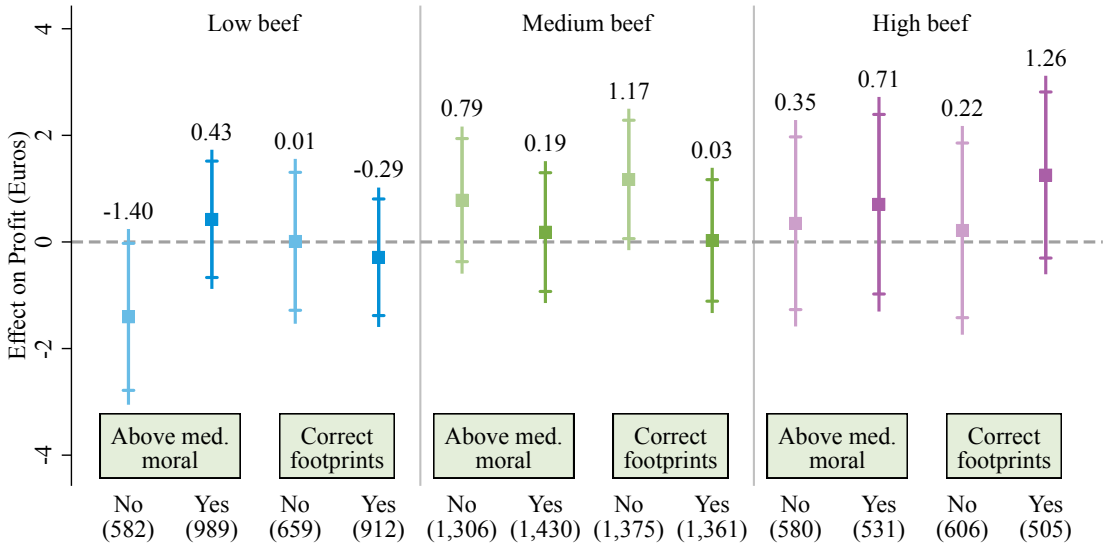
Note: This figure estimates treatment effects on customers' total carbon footprints (Panel A) and number of Fair meals ordered (Panel B) by crossed heterogeneity categories. In particular, we separately estimate Equation 2.4 among customers in a given category of baseline beef consumption and a given category on one of two dimensions of climate beliefs: above- or below-median values for the climate-morality belief index or whether customers correctly or incorrectly rank beef and tomatoes' carbon footprints. Appendix C.4.3 describes these heterogeneity dimensions in more detail. We show the sample size of baseline customers in each crossed heterogeneity cell in parentheses below each estimate. The capped and uncapped lines present 90% and 95% confidence intervals, respectively, with standard errors clustered by customer.

Figure A9: Crossed heterogeneous profit and retention effects by baseline beef vs. climate beliefs and knowledge

Panel A. Has a box

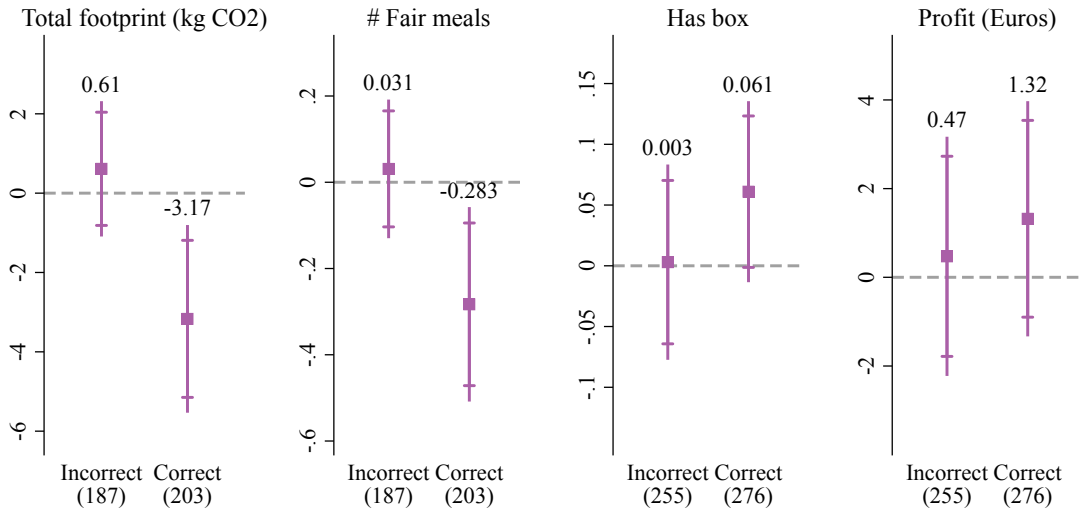


Panel B. Profit (Euros)



Note: This figure parallels Appendix Figure A8, except that we estimate crossed-heterogeneity treatment effects on whether customers order a meal box in a given week (Panel A) and on profit per customer, measured in Euros (Panel B). Appendix C.3 details these outcomes.

Figure A10: Heterogeneous treatment effects by baseline carbon-footprint knowledge among high-baseline-beef consumers with pro-climate morality beliefs



Note: This figure estimates the main treatment effects among high baseline beef consumers with above-median climate-morality beliefs, separately by whether they correctly or incorrectly ranked beef and tomatoes' carbon footprints in the baseline survey. We separately estimate Equation 2.4 among customers in each of these categories, showing the sample size of baseline customers in each crossed heterogeneity cell in parentheses below each estimate. Appendix C.4.3 describes these heterogeneity dimensions in more detail. We plot estimated treatment effects on customers' total carbon footprints, number of *Fair* meals ordered, whether customers ordered a meal box, and profit per customer. Appendix C.3 details these outcomes. The capped and uncapped lines present 90% and 95% confidence intervals, respectively, with standard errors clustered by customer.

B Additional Tables

Table A1: Joint baseline-sample heterogeneity

	(1)	(2)	(3)	(4)
	Meal-choice outcomes		Retention outcomes	
	Footprint (kg CO2)	# Fair meals	Has box	Profit (Euros)
Any label	0.490	0.054	-0.014	-0.037
	(0.357)	(0.033)	(0.022)	(0.722)
Baseline beef heterogeneity:				
Medium base beef	0.740**	0.150***	-0.005	0.328
	(0.288)	(0.027)	(0.019)	(0.631)
Any label * Medium base beef	-0.051	-0.017	0.017	0.824
	(0.314)	(0.028)	(0.021)	(0.709)
High base beef	0.220	0.214***	-0.040*	-0.548
	(0.483)	(0.049)	(0.023)	(0.845)
Any label * High base beef	-0.253	-0.056	0.046*	0.783
	(0.536)	(0.052)	(0.026)	(0.936)
Moral belief heterogeneity:				
$\geq$ Median moral index	0.496*	0.027	-0.021	-0.184
	(0.281)	(0.026)	(0.016)	(0.585)
Any label * $\geq$ Median moral index	-0.709**	-0.054*	0.024	0.303
	(0.315)	(0.030)	(0.018)	(0.646)
Knowledge heterogeneity:				
Correct footprint rank	-0.226	-0.018	0.011	0.583
	(0.276)	(0.026)	(0.016)	(0.562)
Any label * Correct footprint rank	0.048	-0.004	-0.024	-0.621
	(0.315)	(0.030)	(0.018)	(0.641)
Control means	11.366	0.532	0.423	11.788
# Customers	14,108	14,108	33,552	33,552
# Cust-weeks	4,228	4,228	5,592	5,592

Note: This table estimates treatment effects on key outcomes in the baseline-survey sample. Outcomes are defined as in Table 2. These regressions include the control variables detailed for baseline-sample regressions in Appendix C.4: lagged outcome variables, week fixed effects, meal-plan fixed effects, customer longevity, individual demographics, and baseline beliefs about climate change. This table jointly estimates heterogeneous treatment effects by the margins of heterogeneity tested separately in Figures 2 and 3. Dimensions of heterogeneity are as described in those figure notes and detailed in Appendix C.4.3. The sample in columns 1 and 2 is restricted to weeks in which baseline customers ordered meal boxes during the experimental period, while the sample in columns 3 and 4 is all customer-weeks during the experimental period. We cluster standard errors by customer. We indicate statistical significance at the 10%, 5%, and 1% levels by *, **, and ***, respectively.

Table A2: Correlations in baseline climate beliefs

	(1) Companies have moral duty (std)	(2) Individuals have moral duty (std)
Footprint cuts are effective (std)	0.366	0.606
Worry about climate change (std)	0.602	0.523
Self-percept as envi conscious (std)	0.360	0.420
Mentions envi as personally important	0.315	0.291
Believes climate human-caused	0.517	0.407
Thinks companies should push customers (std)	0.638	0.613

Note: This table shows correlations between baseline climate beliefs in the baseline-survey sample. Columns 1 and 2 show the correlations of each measure of climate beliefs with the extent to which customers think that companies have a moral duty to address climate change and that individuals have a moral duty to make lifestyle changes to reduce their carbon footprints, respectively. We elicit these two beliefs, customers' beliefs that individual carbon-footprint cuts could effectively address climate change, and customers' beliefs that companies should push customers to make more climate-friendly choices on integer scales from 1 to 7. We elicit climate worry and self-perception as environmentally conscious on 4-point scales. We convert all of these to numeric scales and standardize them in the baseline sample. We define binary variables for whether customers mention the environment in free-text responses about issues that matter to them and whether customers say that they believe climate change is human-caused. We detail all of these measures in Appendix C.4.

Table A3: Main treatment effects: Unpooled label treatments

Panel A. Meal-choice outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Footprint	# Meals by footprint tier			# Meals by protein				
	(kg CO2)	Superstar	Good	Fair	Veggie	Poultry	Pork	Fish	Beef
All Tiers-Letters	-0.076**	0.005	0.003	-0.009***	0.006	0.001	0.005	-0.003	-0.009***
	(0.035)	(0.004)	(0.005)	(0.003)	(0.004)	(0.004)	(0.004)	(0.002)	(0.003)
All Tiers	-0.075**	0.012***	-0.001	-0.008***	0.005	0.007*	0.001	-0.001	-0.008***
	(0.035)	(0.004)	(0.005)	(0.003)	(0.004)	(0.004)	(0.004)	(0.002)	(0.003)
Superstars Only	-0.075**	0.004	0.003	-0.009***	0.003	0.001	0.002	0.000	-0.009***
	(0.034)	(0.004)	(0.005)	(0.003)	(0.004)	(0.004)	(0.004)	(0.002)	(0.003)
Cont. mean	11.775	0.856	1.635	0.589	0.439	1.070	0.851	0.167	0.589
# Customers	131,017	131,017	131,017	131,017	131,017	131,017	131,017	131,017	131,017
# Cust-weeks	397,514	397,514	397,514	397,514	397,514	397,514	397,514	397,514	397,514

Panel B. Retention and profit outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
		# Meals	Revenue	Cost	Profit (Euros)	
	Has box	/ box	(Euros)	(Euros)	All	If box
All Tiers-Letters	0.003	-0.001	0.182	0.119	0.067	-0.078
	(0.002)	(0.005)	(0.123)	(0.076)	(0.052)	(0.093)
All Tiers	0.003	0.003	0.220*	0.133*	0.083	-0.023
	(0.002)	(0.005)	(0.123)	(0.075)	(0.052)	(0.093)
Superstars Only	0.003	0.002	0.211*	0.135*	0.066	-0.113
	(0.002)	(0.005)	(0.123)	(0.075)	(0.052)	(0.093)
Cont. mean	0.281	3.078	20.628	12.720	7.975	28.384
# Customers	234,511	131,049	234,511	234,511	234,511	131,049
# Cust-weeks	1,407,066	397,565	1,407,066	1,407,066	1,407,066	397,565

Note: This table parallels Table 2, except that we separately test for treatment effects of each of the label variations shown in Table 1. All outcomes, samples, and control variables are as defined in Table 2. Standard errors are clustered by customer. We indicate statistical significance at the 10%, 5%, and 1% levels by *, **, and ***, respectively.

Table A4: Revenue, costs, and profits by box composition

	(1) Net revenue (Euros)	(2) Direct costs (Euros)	(3) Profit (Euros)
Number of meals:			
Vegetarian	9.467***	6.709***	3.707***
	(0.394)	(0.208)	(0.275)
Poultry	10.454***	7.526***	4.049***
	(0.322)	(0.164)	(0.198)
Pork	9.682***	6.504***	4.099***
	(0.457)	(0.188)	(0.339)
Fish	14.306***	9.220***	6.931***
	(0.834)	(0.591)	(0.898)
Beef	11.698***	8.119***	4.987***
	(0.551)	(0.285)	(0.346)
Box means	70.964	43.793	27.380
N	779,744	779,744	779,744

Note: This table estimates how meal-box composition affects HelloFresh's net revenue, direct costs, and profit per box ordered, all in Euros. The sample here is restricted to weeks during which customers ordered a meal box across all pre-experimental and experimental weeks, i.e. across fiscal weeks 8 through 28. We regress these box-level financial outcomes on the number of meals with a given primary protein in that box, controlling also for week fixed effects. Standard errors are clustered by week. We indicate statistical significance at the 10%, 5%, and 1% levels by *, **, and ***, respectively.

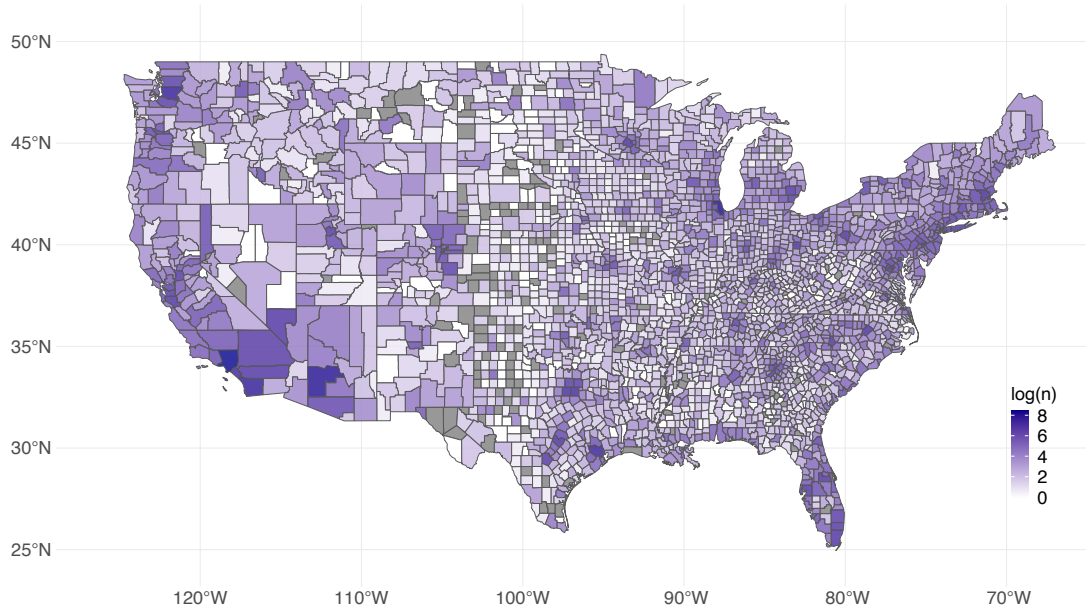
C Data appendix

C.1 Full experimental sample: Data summary and balance

Table A5 presents baseline summary statistics and balance regressions for the main experimental survey, split across the four treatment groups. Several facts bear noting. First, the HelloFresh customer population lives in areas that are somewhat whiter, more educated, and wealthier than the national population. In particular, customers live in zip codes where 67% of residents are non-Hispanic whites on average, where 40% of residents 25 and over have a bachelor's degree or more, and where on average 11% live below the US poverty level. In contrast, 60% of Americans are non-Hispanic whites, 33% of Americans over age 25 have a bachelor's degree or more, and 13% of Americans nationally live below the poverty line.[16] Studies have found robust evidence that environmental concern rises with education, so HelloFresh's pool of customers may be somewhat more responsive to carbon labels than the average US population (Angrist et al., 2024; Dechezleprêtre et al., 2024).

Second, our sample is largely balanced on observable baseline characteristics, such as longevity as customers at HelloFresh, baseline meal choices, and zipcode-level demographics. To adjust for the small differences that remain, we control for baseline meal choices, demographics, and customer longevity in our main specifications. We also show that our results are robust to excluding the demographic and longevity controls.

Figure A11: Geographic distribution of HelloFresh customers



Note: This map shows the geographic distribution of HelloFresh customers who were enrolled in our study at the county level. The intensity of color for a given county corresponds to the natural log of the number of customers in that county.

[16] We calculate all national statistics using data from the 2020 American Community Survey (ACS) 5-Year estimates. We merge HelloFresh customers to the same data by Zip Code Tabulation Area (ZCTA).

Table A5: Summary statistics and balance: Main experimental sample

	N	Control Mean	Letters Score	Abstract Score	Climate Superstar
Longevity at HelloFresh					
Had any pre-period meal	234,511	0.813	-0.001	0.002	0.000
			(0.002)	(0.002)	(0.002)
Week of first pre meal	190,730	11.930	0.018	0.051*	0.066**
			(0.028)	(0.028)	(0.028)
# Pre weeks with meals	234,511	4.449	-0.015	-0.017	-0.028
			(0.024)	(0.024)	(0.024)
Total # pre-period meals	234,511	13.917	-0.048	-0.027	-0.062
			(0.080)	(0.080)	(0.080)
Baseline meal choices					
Avg box carbon footprint	160,958	12.966	0.020	0.062	-0.010
			(0.044)	(0.044)	(0.044)
Share of baseline meals:					
Vegetarian	190,730	0.149	-0.002**	-0.001	0.000
			(0.001)	(0.001)	(0.001)
Poultry	190,730	0.365	0.000	-0.000	-0.000
			(0.001)	(0.001)	(0.001)
Pork	190,730	0.232	0.000	0.001	0.000
			(0.001)	(0.001)	(0.001)
Fish	190,730	0.052	0.001	-0.000	0.000
			(0.001)	(0.001)	(0.001)
Beef	190,730	0.216	0.001	0.001	0.000
			(0.001)	(0.001)	(0.001)
Demographics by zipcode					
Share below US poverty line	229,084	0.108	0.001	0.000	0.000
			(0.000)	(0.000)	(0.000)
Share adults 25+ with bachelor's	229,103	0.395	-0.001	0.001	0.001
			(0.001)	(0.001)	(0.001)
Shares by race or ethnicity:					
White, non-hispanic	229,106	0.670	-0.002**	-0.000	-0.001
			(0.001)	(0.001)	(0.001)
Black	229,106	0.101	0.000	-0.000	-0.001
			(0.001)	(0.001)	(0.001)
Hispanic	229,106	0.138	0.001	-0.000	0.001
			(0.001)	(0.001)	(0.001)
2020 Democratic vote share	229,805	0.530	0.000	-0.000	0.000
			(0.001)	(0.001)	(0.001)

Note: This table compares average characteristics of the study participants by treatment arm. Column 1 presents the number of participants in the study for whom we have information on the given characteristic. Column (2) is the average for the characteristic in the control group. Columns 3-5 present the difference in means between participants in the three labelled treatment arms (*Letters*, *Abstract*, and *Climate Superstar*) and the control group. Standard errors for the estimated difference from the control group are in parentheses. The variables shown in the table are: (i) whether or not the participant ordered any HelloFresh meals during the pre-period we observe in the administrative data before the start of the experiment (Week 8 to Week 20), (ii) the week in which the participant ordered their first HelloFresh meal (for participants who ordered at least one meal during the pre-period), (iii) the number of weeks during the pre-period in which the participant ordered at least one HelloFresh meal, (iv) the total number of meals that the participant ordered during the pre-period from HelloFresh, (v) the average carbon footprint of a HelloFresh box ordered by the participant in the pre-period for which we observe carbon footprints (Week 17 to Week 20), (vi) the share of meals ordered during the pre-period which had a main protein which was vegetarian, poultry, pork, fish, or beef, (vii) the share of people in the participant's county below the US poverty line (according to 2020 5-year summary data from the American Community Survey (ACS), for HelloFresh customers with zipcodes that matched to an ACS zipcode), (viii) the share of adults aged 25 or over in the participant's county who have earned at least a bachelor's degree (according to the ACS), (ix) the share of people in the participant's county who are classified as non-hispanic white, black, or hispanic (according to the ACS), and (x) the share of people in the participant's county who voted for the Democratic party candidate in the 2020 US presidential election.

C.2 Baseline survey sample

We recruited the baseline sample from the full experimental sample by emailing all customers allocated to the experiment with an invitation to complete a short survey, for which all participants would be entered into lotteries to win a $100 or $200 gift card. The survey elicited participants' beliefs about climate change, their self-perceptions on traits like altruism and environmental consciousness, their political affiliation, and their baseline knowledge about the carbon footprints of different foods. The full text of the survey is available at this link.

In total, 7,259 customers completed the baseline survey, and 5,592 provided email addresses that successfully merged with administrative data from HelloFresh. Appendix Table A6 compares the baseline sample to the full experimental sample, and Appendix Table A7 tests for balance across treatment arms in the baseline sample. The baseline sample is broadly balanced across treatment arms, except that those assigned to any carbon-labeling arm state that they consider the environment in their food somewhat less than do those in the control group and have somewhat higher baseline footprints. We will control for these factors throughout our analysis of the baseline sample.

Table A6: Comparing baseline sample and main experimental sample

	In baseline sample?				
	No		**Yes**		
	(1)	(2)	(3)	(4)	(5)
	N	Mean	N	Mean	Δ Means
Longevity at HelloFresh					
Had any pre-period meal	228,919	0.810	5,592	0.969	0.159***
Week of first pre meal	185,312	11.964	5,418	11.970	0.006
# Pre weeks with meals	228,919	4.394	5,592	6.063	1.670***
Total # pre-period meals	228,919	13.745	5,592	19.519	5.774***
Baseline meal choices					
Avg box carbon footprint	155,899	12.997	5,059	12.582	-0.415***
Share of baseline meals:					
Vegetarian	185,312	0.148	5,418	0.174	0.026***
Poultry	185,312	0.366	5,418	0.352	-0.014***
Pork	185,312	0.233	5,418	0.226	-0.007***
Fish	185,312	0.052	5,418	0.064	0.012***
Beef	185,312	0.217	5,418	0.200	-0.017***
Demographics by zipcode					
Share below US poverty line	223,514	0.108	5,570	0.108	-0.000
Share adults 25+ with bachelor's	223,532	0.395	5,571	0.406	0.011***
Shares by race or ethnicity:					
White, non-hispanic	223,535	0.669	5,571	0.679	0.011***
Black	223,535	0.101	5,571	0.096	-0.005**
Hispanic	223,535	0.139	5,571	0.132	-0.007***
2020 Democratic vote share	224,219	0.530	5,586	0.541	0.011***

Note: This table compares customers in the baseline sample with customers in the full experimental sample on all baseline traits included in Appendix Table A5. See the note to Appendix Table A5 for a description of these variables. Columns 1 and 3 present the number of customers from the main experimental sample who are not or are in the baseline sample, respectively, and for whom we observe each baseline trait. Columns 2 and 4 present the mean values in each group, and Column 5 presents the coefficient on an indicator that a given customer is in the baseline sample when we regress each trait on this dummy. We indicate statistical significance at the 10%, 5%, and 1% levels by *, **, and ***, respectively.

Table A7: Baseline sample summary statistics and balance

	Control		Any Label			
	N	Mean	N	Mean	Δ Mean	*p-value*
Longevity at HelloFresh:						
Had any pre-period meal	1,377	0.968	4,215	0.969	0.001	(0.854)
Week of first pre meal	1,333	11.935	4,085	11.981	0.045	(0.738)
# Pre weeks with meals	1,377	6.171	4,215	6.028	-0.143	(0.256)
Total # pre-period meals	1,377	19.610	4,215	19.490	-0.120	(0.785)
Baseline meal choices:						
Avg box carbon footprint	1,249	12.261	3,810	12.688	0.427**	(0.037)
Share of baseline meals:						
Vegetarian	1,333	0.181	4,085	0.171	-0.010	(0.186)
Poultry	1,333	0.352	4,085	0.352	0.000	(1.000)
Pork	1,333	0.223	4,085	0.227	0.004	(0.418)
Fish	1,333	0.064	4,085	0.064	0.000	(1.000)
Beef	1,333	0.195	4,085	0.202	0.007	(0.109)
Demographics						
Woman	1,377	0.827	4,215	0.831	0.004	(0.733)
Age bins:						
18-29	1,377	0.237	4,215	0.212	-0.025*	(0.056)
30-49	1,377	0.513	4,215	0.514	0.001	(0.949)
50-69	1,377	0.211	4,215	0.239	0.029**	(0.024)
70+	1,377	0.038	4,215	0.033	-0.005	(0.390)
Education:						
No 4-year college degree	1,377	0.274	4,215	0.281	0.007	(0.614)
Max 4-year college degree	1,377	0.403	4,215	0.409	0.006	(0.694)
Has graduate degree	1,377	0.323	4,215	0.310	-0.014	(0.334)
Political affiliation:						
Democrat	1,377	0.395	4,215	0.409	0.014	(0.357)
Republican	1,377	0.169	4,215	0.184	0.015	(0.201)
Independent	1,377	0.265	4,215	0.251	-0.014	(0.305)
Prefer not to answer	1,377	0.171	4,215	0.156	-0.015	(0.195)
Climate beliefs and knowledge:						
Think climate change is human-caused	1,377	0.779	4,215	0.767	-0.012	(0.354)
Think companies have climate duty (1-7)	1,377	5.918	4,215	5.906	-0.012	(0.792)
Think companies should push custs (1-7)	1,377	5.225	4,215	5.190	-0.035	(0.496)
Think footprint cuts effective (1-7)	1,377	4.111	4,215	4.073	-0.039	(0.432)
Think have moral duty on footprint (1-7)	1,377	4.934	4,215	4.942	0.008	(0.868)
Identify as envi conscious (1-4)	1,377	2.898	4,215	2.874	-0.023	(0.341)
Envi is factor in food choices (1-4)	1,377	2.253	4,215	2.197	-0.056**	(0.041)
Climate worry (1-4)	1,377	2.954	4,215	2.926	-0.028	(0.351)
Correctly rank beef and tomato	1,377	0.524	4,215	0.509	-0.015	(0.333)

Note: This table compares average characteristics of the baseline-sample study participants by treatment arm. Columns 1 and 3 present number of participants in the control group or any of the label arms, respectively, for whom we observe the given characteristic. Columns 2 and 4 then present mean values for the control and treatment arms, respectively, while column 5 presents the difference in means between these groups for each trait. Column 6 presents heteroskedasticity-robust p-values for a test of whether each difference in means equals zero. We indicate statistical significance at the 10%, 5%, and 1% levels by *, **, and ***, respectively. The variables for longevity at HelloFresh and baseline meal choices are as described in Appendix Table A5, since we observe these for the full experimental sample. We then also test for balance on customer traits observed only in the baseline sample: gender, age, educational attainment, political affiliation, beliefs about anthropogenic climate change, a range of baseline beliefs about the morality of climate action by individuals and companies, self-perception as an environmentalist, the degree to which they consider the environment in meal choices, worry about climate change, and whether they correctly rank baseline carbon footprints. We describe the demographics and environmental self-perceptions in Appendix C.4.2 and describe the climate-morality belief and carbon-footprint knowledge elicitations in Appendix C.4.3. We elicit belief about the causes of climate change by asking participants to state whether they think climate change is mostly caused by human activity, mostly caused by natural changes in the environment, not happening, or other.

C.3 Detail on outcome variables

C.3.1 Meal choices

We use administrative data from HelloFresh on customers' meal boxes to define a series of outcome variables for customers' meal choices.

Meal choices by protein category: We observe the meals that each experimental customer ordered each week from fiscal week 8 to 28, along with the quantity of each meal that each customer ordered. (Customers can select multiple orders of a single recipe in a given week.) A small share of customer-week-recipe combinations in this data have duplicates (<1%), and we test robustness in Appendix Figure A3 to defining customers' total meals ordered as including or excluding these duplicates. Using HelloFresh's records for the ingredients contained in each recipe, we classify whether each recipe includes chicken, turkey, duck, pork, beef, or fish, or if it is vegetarian. Combining these records with customers' meal choices, we then calculate the number of meals containing each protein that customers ordered each week in fiscal weeks 8 through 28.

Carbon-footprint outcomes: Next, we measure the total carbon footprint of the meals that customers' ordered in a given week by combining records for the meals that customers ordered with HelloFresh's estimates of the carbon footprint of each recipe. HelloFresh estimated these carbon footprints using ingredients' footprints estimated in the Agribalyse dataset, which estimates products' carbon footprint using Life Cycle Assessments (LCA). This method takes into account each stage in food production (i.e. from field to plate). Agribalyse is widely cited, including by the EU and the UN. Because the data are based on products in France, the carbon footprint estimates for HelloFresh US are likely to have some errors, although the meals would likely largely be categorized the same way. For more information, see https://doc.agribalyse.fr/documentation-en/agribalyse-data/documentation.

HelloFresh estimated the carbon footprint of all meals offered in weeks 17 through 28, allowing us to construct customers' total ordered carbon footprint for each week in this range. Figure 1 plots the distribution of estimated carbon footprints for all recipes, separately by protein content. HelloFresh then categorized each meal as falling in one of three carbon-footprint categories: "*Climate Superstars*," with estimated carbon footprints below 2kg CO_2e, "*Good*" meals, with estimated carbon footprints between 2 and 7 kg CO_2e, and "*Fair*" meals, with estimated footprints above 7 kg CO_2e. We use customers' meal choices to define outcomes for the number of meals that customers ordered each week that were *Climate Superstars*, *Good*, or *Fair.*

C.3.2 Profit and retention outcomes

Next, we define outcome variables for customer retention and HelloFresh's costs, revenue, and profits per customer-week.

Customer retention: We define customer retention by whether customers ordered a meal box from HelloFresh in a given week. We observe this outcome from the company's records of the total number of meal boxes they shipped to each customer each week. We also define a measure of intensive-margin customer retention by the number of meals that customers ordered in a given week conditional on ordering a box.

Financial outcomes: Next, we define financial outcomes for the total direct cost and total revenue net of customer discounts for each customer-week. We then define profit for each customer-week as revenue (net of discounts) minus direct costs. Each of these outcomes is measured in Euros. To limit the influence of outliers, we replace values over the 99th percentile with the 99th-percentile value for each of these variables. All values are set to zero if customers did not order a meal box in a given week. To test for impacts on profit conditional on ordering a meal box, we also define a profit variable set to missing in weeks where customers did not order a meal box.

C.4 Detail on control variables and dimensions of heterogeneity

C.4.1 In regressions in the full experimental sample

Our primary regressions in the full experimental sample include the following control variables:

- *Lagged outcome variables:* We control for the values of the outcome variable in weeks 18 through 20, the last three weeks of the pre-experimental period. In some cases, these lagged outcomes are missing. First, we only observe customers' profit and retention outcomes in and following the week during which HelloFresh added them to the experimental sample, which occurred sometime during weeks 18 and 20 for each customer. Moreover, while we observe customers' meal choices for all pre-period weeks in which they ordered a box, these values are missing whenever customers did not order a box. Thus, we also include dummy variables that a given customer is missing this lagged outcome in weeks 18 through 20 (in which case we set the lagged outcome to zero).

- *Week fixed effects:* We include fixed effects for each fiscal week in the experimental period (weeks 21 through 28)

- *Meal-plan fixed effects:* We include fixed effects indicating that customers are enrolled in one of the following meal plans: Chef's choice (60%), Fit (13%), Family (8%), Veggie (5%), Quick (4%), Pescatarian (2%), Pork-free (1.6%), or Other (8%). "Other" includes the 7% of customers with no listed subscription type and a series of other categories that together account for only 1% of customers.

- *Customer-longevity fixed effects:* We include fixed effects for the number of pre-experimental weeks in which customers ordered a meal box from HelloFresh. We define the pre-experimental period as fiscal weeks 8-20, so the total number of pre-period weeks ranges from 0 for 19% of experimental customers to 13 for about 5% of experimental customers. We include separate fixed effects for each integer from 0 to 13 pre-experimental weeks.

- *Zipcode-level demographics:* We use customers' zipcodes to merge them to county-level data on presidential voteshares in the 2020 election and to demographic data from the American Community Survey by Zip Code Tabulation Area (ZCTA). We merge customers' zipcodes to county FIPS codes using a crosswalk downloaded from the U.S. Department of Housing and Urban Development. About 2% of customers have zipcodes that do not merge to either the HUD crosswalk or ACS ZCTAs. Our main regressions control for the following:

- County-level voteshare for President Biden in the 2020 presidential election. Where zipcodes span multiple counties, we construct the weighted average of voteshares, weighted by the share of zipcode residences in each county.
- ZCTA share of the population below 100% of the federal poverty line
- ZCTA share of the population ages 25+ with a bachelor's or graduate degree
- ZCTA share of the population that are non-Hispanic whites.

We control linearly for each of these variables and include dummy variables to indicate missing values (less than 2.5% of customers in each case). We replace missing values with the non-missing sample mean for each variable.

C.4.2 In regressions in the baseline-survey sample

In regressions in the baseline sample, we remove the controls for zipcode-level demographics and instead include individual demographic traits and climate beliefs observed in the baseline survey:

- *Education fixed effects:* We include fixed effects for whether baseline participants report that they have completed the following highest levels of education: Less than high school; High-school degree or GED; Vocational diploma; 2-year college degree; 4-year college degree; Master's degere or higher
- *Gender fixed effects:* We include fixed effects for whether customers identify as a Man, Woman, or Other gender identity.
- *Politics fixed effects:* We include fixed effects for whether customers report that they identify as a Republican, a Democrat, an Independent, or prefer not to answer.
- *Role of environment in food choices:* In the baseline survey, we ask participants how much they consider a range of factors in their food choices: healthiness; religious or cultural customs or restrictions; cost; tastiness; environmental impacts; and animal welfare, presented in a random order. For each factor, participants report whether they consider it "Not at all," "Only a little," "A moderate amount," or "A lot." We include fixed effects for each of these categories for environmental impacts.
- *Self-perception as environmentalists:* We ask baseline participants to report the extent to which they identify as patriotic; environmentally conscious; politically active; health conscious; open minded; and altruistic or public-spirited, presented in a random order. For each dimension, participants choose from "Not at all;" "A bit;" "Moderately;" and "Very much." We include fixed effects for these categories for self-perceived environmental consciousness.

C.4.3 Dimensions of heterogeneity

We test for heterogeneous treatment effects by the following baseline traits:

- *Baseline beef consumption:* We test for heterogeneity by customers' baseline beef consumption, measured by the share of pre-experimental meals that they ordered and that contain beef. We define low, medium, and high baseline beef consumers as those whose share of

pre-period meals containing beef falls below the 25th percentile, between the 25th and 75th percentile, or above the 75th percentile across customers, respectively. We leave baseline-beef category missing for customers who did not order any pre-period meals (about 19% of customers). We test for heterogeneous treatment effects across those with low, medium, and high baseline beef consumption.

- *Customer longevity at HelloFresh:* We next test for heterogeneity by customers' longevity at HelloFresh, measured by the number of weeks during which customers ordered a meal box during the pre-experimental period, defined as weeks 8-20. We categorize customers as ordering meals in 0 pre-experimental weeks, 1-6 pre-experimental weeks, or 7+ pre-experimental weeks and test for heterogeneous treatment effects across these categories. We observe these categories using administrative data for the full experimental sample.

- *Baseline climate beliefs:* In the baseline sample, we test for heterogeneous treatment effects by several dimensions of customers' beliefs about climate change:

  - **Companies' moral duty on climate:** We ask baseline participants to rate the extent to which they think "companies have a moral duty to help address climate change," from 1 (No moral duty) to 7 (Morally crucial). We test heterogeneous treatment effects by whether customers report beliefs that are above or equal to versus below the median response.

  - **Individuals' moral duty on climate:** We ask baseline participants to rate the extent to which they "think individuals have a moral duty to make lifestyle changes to reduce their carbon footprints," from 1 (No moral duty) to 7 (Morally crucial). We test heterogeneous treatment effects by whether customers report beliefs that are above or equal to versus below the median response.

  - **Climate-morality belief index:** We calculate a standardized index of climate-morality beliefs by standardizing customers' responses to the questions on companies' moral duty on climate and individuals' moral duty on climate, summing these values, and then standardizing this sum in the baseline sample. We test heterogeneous treatment effects by whether customers score above or at versus below the median on this index.

  - **Effectiveness of lifestyle changes:** We ask baseline participants to rate how effective they "think individual efforts to reduce personal carbon footprints could be in addressing climate change," from 1 (Not effective at all) to 7 (Extremely effective). We test heterogeneous treatment effects by whether customers report beliefs that are above or equal to versus below the median response.

  - **Whether companies should push customers:** We ask baseline participants to answer whether they think "companies should push their customers to make more climate-friendly choices," from 1 (Definitely not) to 4 (No feeling either way) to 7 (Definitely yes). We test heterogeneous treatment effects by whether customers report beliefs that are above or equal to versus below the median response.

  - **Whether climate change is human-caused:** We ask baseline participants to think about "climate change, sometimes called global warming," and then ask them which of the following is closest to their beliefs on climate change: "It is caused mostly by

human activities;" "It is caused mostly by natural changes in the environment;" "Neither, since climate change isn't happening;" or "Other." We test heterogeneity by whether customers select that they think climate change is mostly human-caused or not.

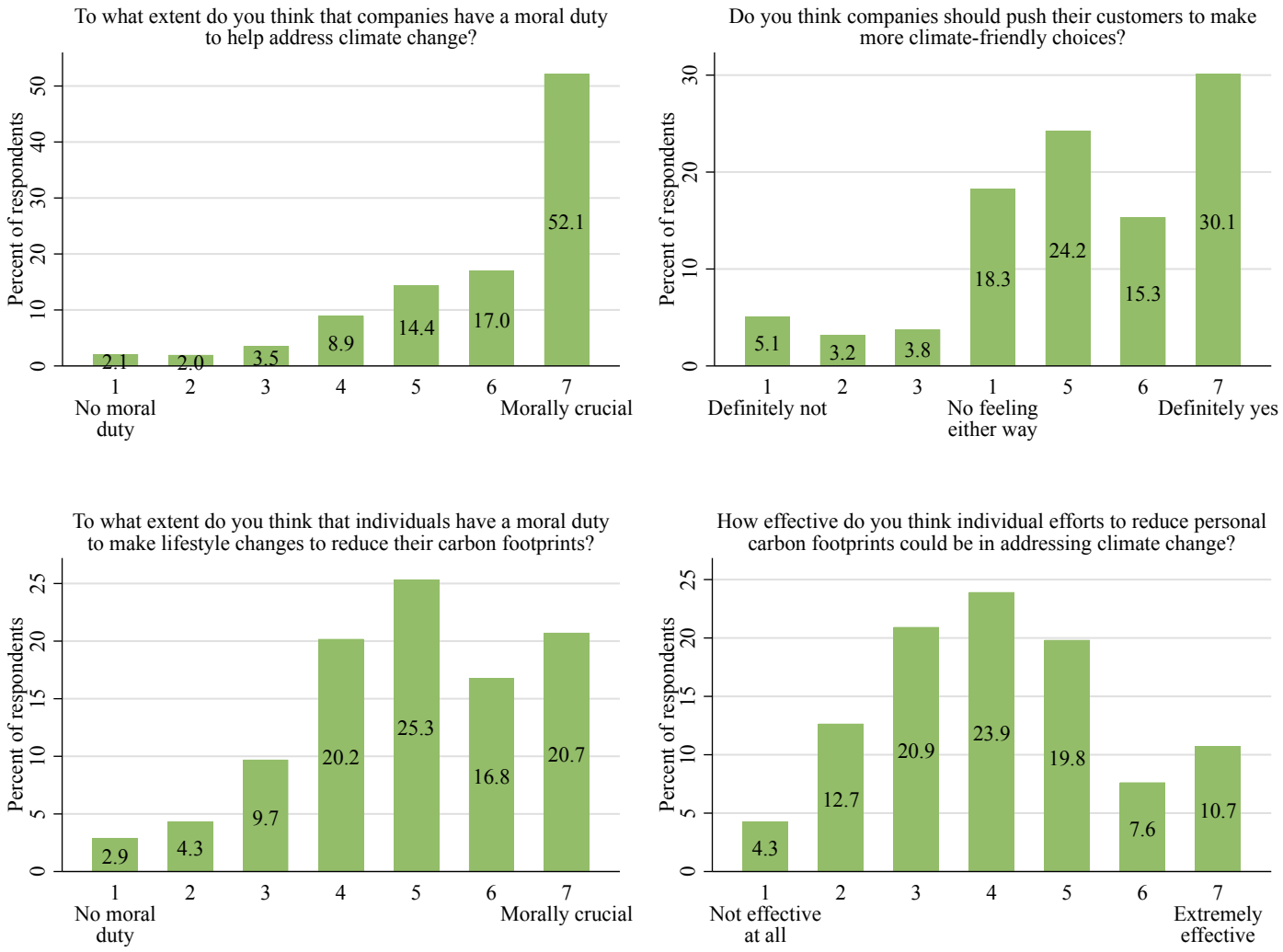
- *Baseline carbon-footprint knowledge:* We measure customers' baseline knowledge on carbon footprints by asking them to rank the following foods from that which creates the highest to lowest emissions per pound on average: beef, cheese, farmed shrimp, pork, chicken, and tomatoes, listed here in the correct order by median emissions (Poore and Nemecek, 2018). We test heterogeneity by whether customers correctly rank tomatoes as having the lowest and beef as having the highest emissions per unit weight. Only 1% of baseline respondents report the fully correct order, while 51% correctly identify tomatoes and beef as the endpoints of the footprint scale.

- *Concern about the environment:* In the baseline sample, we test for heterogeneous treatment effects by several measures of customers' concern about the environment:

  - **Climate worry:** We ask participants to rate how worried they are about climate change, from the following options: "Not at all worried;" "A bit worried;" "Moderately worried;" and "Very worried." We test heterogeneity by whether customers choose "Not at all worried" or "A bit worried" versus choosing "Moderately worried" or "Very worried."
  - **Self-perceived environmentalism:** As described above in Appendix C.4, we elicit the extent to which baseline participants identify as environmentally conscious. We test for heterogeneous treatment effects by whether customers answer "Not at all" or "A bit" versus choosing "Moderately" or "Very much."
  - **Mentioning an environmental issue:** Before we mention the environment at all during the baseline survey, we ask participants to list the 5 or fewer social or political issues or causes that are most personally important to them. We then hand-coded responses to identify participants who mentioned any environmental issue. We test for heterogeneous treatment effects by whether participants did so.

- *Perceptions of HelloFresh:* We ask baseline participants to rate the extent to which they think "HelloFresh prioritizes social good in its operations" on a scale from 1 (Not at all) to 7 (Extremely highly.) We test heterogeneous treatment effects by whether customers report beliefs that are above or equal to versus below the median response.

Figure A12: Baseline-sample climate beliefs



Note: This figure plots the distribution of climate morality beliefs elicited in the baseline survey and described in Appendix C.4.3. These figures plot the frequency of responses from 1 to 7 among the full baseline sample of 5,592 customers. Numbers on each bar show the percentage of the sample that answered the corresponding value.