Got Beef with Beef? Large-scale evidence that carbon labels yield limited climate benefits and cause backlash

Lucy Page[‡*] University of Pittsburgh

Lisa Ho[†*] Columbia University

Food systems contribute approximately one-third of total greenhouse gas emissions, with wide variation in emissions across food products. Beef, for example, generates disproportionately high emissions: 10 times that of pork or chicken per kilogram and 30 times that of tofu. Without systematic carbon pricing, carbon labeling has emerged as a prominent tool to shift consumers towards lower-emissions foods. However, whether such labels effectively shift behavior remains unclear, particularly in settings where climate change is politically contentious. Here we report results from the largest evaluation of carbon labeling to date and the first in a commercial US setting—a randomized field experiment with over 230,000 customers of a major food company. Carbon-footprint labels yielded modest average emission reductions of 0.6% but caused backlash: customers opposed to environmental nudges increased their emissions and beef consumption. Despite limited environmental impact, the labels had financial returns for our partner company, increasing customer retention by 1.0% and average per-customer-week profits by 0.9%. Our results show that while climate labels will likely remain a popular sustainability nudge, they have only small environmental benefits and may backfire in polarized settings.

[‡]230 S Bouquet St, Pittsburgh, PA 15260. Email: lep147@pitt.edu.
[*]Corresponding authors.
[†]420 W 118th St, New York, NY 10027. Email: lisa.ho@columbia.edu.

Anthropogenic climate change has already raised global average temperatures by over 1.1C from pre-industrial levels, and limiting warming to global targets requires stark cuts to emissions in the next decade. Given the political constraints to systemic climate policy in the US, consumer-facing carbon labels have emerged as a prominent alternative to drive consumers towards lower-emissions choices. These labels display products' emissions either as numeric values or categorical ratings, premised on the idea that making environmental externalities visible will shift consumers' purchasing behavior (Taufique et al., 2022). Carbon-footprint labeling is and will likely remain voluntary for firms in the US, while the EU—and France in particular—has moved towards mandatory environmental labeling. Even so, American firms across diverse sectors have chosen to implement labels, from food (e.g. Panera Bread) to footwear (e.g. Allbirds) to travel (e.g. Google Flights). Whether voluntarily-implemented carbon labels deliver genuine environmental benefits at scale remains unknown, however. These effects are particularly uncertain given the intense politicization of climate change in the US, which could polarize responses to the labels; in 2024, partisan belief gaps were larger on climate change than on any other issue (Pew Research, 2024).

Carbon labeling is particularly common in the food sector, which accounts for approximately one third of total greenhouse gas (GHG) emissions (Lee and Romero, 2023, Crippa et al., 2021, Lamb et al., 2021). Emissions vary widely across foods, with animal products—and beef in particular—accounting for a disproportionate share of food-sector emissions (Poore and Nemecek, 2018, Xu et al., 2021). Then, relatively simple shifts in dietary patterns could be a critical lever for emissions mitigation (Bajželj et al., 2014, Hedenus et al., 2014, Ripple et al., 2014, Springmann et al., 2016, Clark et al., 2020, Li et al., 2024). This lever is particularly relevant in the United States, where per-capita emissions and meat consumption both rank among the world's highest (Crippa et al., 2023, Johns Hopkins Center for a Livable Future, 2023, Jones et al., 2023, International Energy Agency, 2024).

We present results from the largest field experiment to date on the effectiveness of carbon-footprint labeling. In a randomized controlled trial with over 230,000 geographically diverse American consumers, we worked with HelloFresh—a major food company—to examine both the

environmental impacts and economic returns to carbon labeling. Customers were randomized either to choose meals from status-quo menus or from menus with one of two labeling treatments: (i) *All-Tiers*, which labeled all meals as *Climate Superstars* ($<$ 2 kg CO_2e/meal), *Good* (2-7 kg CO_2e/meal), or *Fair* ($>$7 kg CO_2e/meal), or (ii) *Superstars-Only*, which only labeled the *Climate Superstar* meals. In addition to observing administrative data for all customers, a detailed baseline survey of 5,592 customers lets us test how pre-existing climate attitudes shaped responses to the labels.

We find that on average the labels reduced customers' carbon footprints by 0.6% ($p < 0.01$), driven by a 1.5% reduction in consumption of beef meals. However, they generated substantial backlash among customers who do not support climate action. These customers—identified in our baseline sample as reporting below-median beliefs that individuals and companies have a moral duty to address climate change—increased their carbon footprints by 4.1% ($p = 0.04$) in response to labels, driven by a 6.7% increase in the consumption of beef meals ($p = 0.08$). Those with above-median agreement in the baseline survey (non-significantly) decreased their footprints by 2.4% ($p = 0.20$) and beef consumption by 4.0% ($p = 0.32$). Despite the labels' muted and inconsistent environmental benefits, they proved financially beneficial for HelloFresh, increasing customer retention by 1.0% ($p = 0.03$) and profits by 0.9% ($p = 0.09$).

Our findings have several broad implications for climate policy and behavioral science. First, the labels' small overall effects mean that carbon-footprint labeling can achieve only a small share of the dietary emissions cuts required to address climate change; labels must remain a small component of a much broader policy response. Second, customers' starkly heterogeneous response to the labels—with below-median climate-concerned consumers actually increasing their carbon footprints—show that climate nudges can backfire, particularly in the American context where climate action is politically contentious. Together with growing concern that the prevalence of nudges aimed at greening individuals' behavior may divert attention from the systemic policy change required to address climate change (Hagmann et al., 2019, Chater and Loewenstein, 2023), our results caution that carbon labeling may be counterproductive to climate progress in some

settings. Nevertheless, the profit returns of carbon labels suggest that they will likely remain a common sustainability tool.

Finally, our results underline that interventions effective in controlled settings may fail—or even backfire—when deployed at scale in diverse populations. We find much smaller label effects than previous smaller-scale or laboratory studies (Vlaeminck et al., 2014, Elofsson et al., 2016, Camilleri et al., 2019, Muller et al., 2019, Lohmann et al., 2022, Bilén, 2024, Tilling, 2025), which may systematically overestimate effectiveness by sampling from motivated subgroups and creating artificial label salience. In contrast, we test the impacts of carbon labels as implemented by a profit-maximizing company in a fully natural setting. The much smaller effect sizes that we estimate reinforce the importance of large-scale field experiments with realistic implementation in guiding policy design (DellaVigna and Linos, 2022).

Research design

We worked with software engineers and environmental scientists at HelloFresh to design and implement the carbon-labeling experiment. HelloFresh is the largest company globally and in the United States that sells meal kits, or sets of pre-apportioned ingredients along with recipes for consumers to cook at home. Meal-kit services are a large and growing industry; HelloFresh alone sold over 511 million meals in North America in 2024 (hel, 2025), and market researchers estimate that 23% of Americans had tried a meal kit by 2022 (Commisso, 2022). Meal-kit companies have several advantages as a setting in which to test the efficacy of comprehensive food-based carbon labeling. Unlike supermarkets, meal-kit providers have information on the ingredients of—and thus can estimate footprints for—all foods they sell. Unlike individual food brands, they can label the full range of 25 to 35 meals from which customers choose each week.

The experimental intervention added environmental labels to HelloFresh menus over eight weeks in May through July 2022. Environmental scientists at HelloFresh estimated the carbon footprints of meals' ingredients and then categorized them into three grades: *Climate Superstar* ($<$2 kg CO_2e), *Good* (2-7 kg CO_2e), and *Fair* ($>$7 kg CO_2e). The resulting classifications largely

reflected meals' protein content, as seen in Figure 1a, which plots the distribution of estimated carbon footprints for all recipes in the experimental period. Vegetarian meals typically had lower carbon footprints than meals containing poultry, pork, and fish, while beef stands out as a disproportionate source of GHG emissions. All meals with a primary protein other than beef were classified as *Climate Superstar* or *Good*, while all beef meals were classified as *Fair*. On average, 37% of meals were classified as *Climate Superstars* each week, 43% as *Good*, and 20% as *Fair*.

We randomized 234,511 customers across four labeling arms, shown in Figure 1b. First, a control arm saw standard menus without carbon information. Second, customers in the *All-Tiers* arms saw menus where all meals were labeled with carbon-footprint categories, with two variations on label icons. Finally, we test whether labeling schemes that only provide "good" news about low-carbon foods differentially affect customer choices by randomizing other customers to a *Superstars-Only* arm where menus only labeled *Climate Superstars*. We find that treatment effects are quite similar across the label variations, so we pool them for our main analysis. However, we show disaggregated treatment effects in Appendix Figures A.11 and A.12.

The 234,511 customers enrolled in the experiment are drawn from a geographically diverse sample of the American population (Appendix Figure A.1). While we do not observe customer identity or demographics, we observe customer zip codes and merge these with data on vote shares in the 2020 presidential election and demographics from the 2020 American Community Survey. HelloFresh customers live in areas that are somewhat whiter, more educated, and wealthier than the national population (Appendix Table B.1), so we may expect them to express above-average environmental concern (Angrist et al., 2024a, Franzen and Vogl, 2013). Given that our heterogeneity analysis shows that the labels are more effective among those with more concern about climate change, our estimated treatment effects—which are already small and show evidence of backlash—may be an upper bound for the environmental benefits of carbon labels in a representative US setting.

From the full experimental sample, we recruited a subsample of 5,592 customers to complete a detailed baseline survey. This subsample lets us study how customers' baseline attitudes on

climate change mediate label effects. Appendix Table B.2 shows that baseline-sample participants are longer-standing customers of HelloFresh but are otherwise qualitatively similar to the full study sample; note, though, that the small differences between these groups are typically statistically significant due to our very large sample. Both the full customer sample and baseline samples are largely balanced across treatment arms (Appendix Tables B.1 and B.3), and we control for any remaining imbalances.

Main sample summary and balance

HelloFresh allocated any US customer who visited their website during fiscal weeks 18, 19, or 20 of 2022 to the experimental sample, yielding a total sample of 234,511 customers. Appendix Table B.1 presents baseline summary statistics and tests for balance in the full experimental sample between the label and control groups. Several facts bear noting. First, by merging customer zip codes with data from the 2020 American Community Survey (ACS) 5-Year estimates, we see that the HelloFresh customer population live 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, 40% of residents 25 and over have a bachelor's degree or more, and 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. 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., 2024b, Franzen and Vogl, 2013).

Second, our sample is largely balanced across treatment groups on observable baseline characteristics, such as longevity as customers at HelloFresh, baseline meal choices, and zip-code 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.

Recipe-level carbon footprint estimation

HelloFresh estimated the carbon footprints of individual recipes using life cycle assessment (LCA) with values from the Agribalyse database. Agribalyse provides comprehensive environmental impact assessments for food products, accounting for all production stages from agricultural inputs through final consumption. The database is widely cited and is used in official reports by organizations including the European Commission and the United Nations. As Agribalyse data reflect French agricultural systems and supply chains, there may be discrepancies between the carbon footprint estimates and the true emissions associated with each HelloFresh meal. However, the broad categorical rankings used in HelloFresh's carbon labels should be largely robust to these discrepancies, as protein source typically dominates total meal carbon footprint (Poore and Nemecek, 2018). For each HelloFresh recipe, carbon footprints were calculated by first disaggregating recipes into constituent ingredients with specified quantities, matching ingredients to corresponding Agribalyse products, and summing the impacts of individual ingredients to generate recipe-level estimates of CO_2e emissions.

Carbon footprints were calculated for all meals offered during weeks 17 through 28 of 2022. The distribution of estimated carbon footprints for all recipes, separately by protein content, is plotted in Figure 1b. 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*.

Implementation of labels on menus

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 describing the meal (e.g. low-calorie, vegetarian, spicy). The new labels introduced in the experiment are shown in Figure 1b, and Appendix Figure A.10 shows how

the labels appeared on the menu itself. On average, each meal was labeled with 1.2 non-climate labels. From the main menu page, participants can click on each meal to pull up a more detailed meal card including nutritional content and larger renditions of any labels.

HelloFresh randomized the experimental sample equally across four arms, shown in Figure 1b: a control arm with no carbon-footprint information and three treatment arms with different carbon-labeling schemes. Two *All-Tiers* groups saw menus on which meals in the *Climate Superstar*, *Good*, or *Fair* footprint categories were each labeled with their corresponding category. These customers were randomized to two variants on these label appearances. While half saw meals labeled with abstract globe symbols, the other half saw menus where *Climate Superstar*, *Good*, and *Fair* meals were labeled with A, B, or C, respectively. Finally, the *Climate Superstar* group saw menus on which only top-tier meals were labeled with an abstract globe symbol and category name. In each of the labeling arms, menu cards also included short descriptions of what the climate labels meant (Appendix Figure A.10). 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." Other pre-existing labels, none of which were environmentally related, do not have an explanation.

Labels were randomly added to the menu for fiscal weeks 21 through 28 of 2022, first appearing on future menus midway through week 20. Customers select meals at least one week in advance of delivery, so the addition of these labels could not have affected meal choices from the week 20 menu. Customers can preview menus 5 weeks in advance of delivery. Throughout our main analysis, we will define the pre-experimental period as weeks 8 through 20 and the experimental period as weeks 23 through 28, because many customers would have selected meals for weeks 21 and 22 before the labels appeared. However, we also show that our main results are robust to alternative definitions of treatment timing (Appendix Figures A.2 and A.8).

Baseline survey sample

We recruited the baseline sample by emailing all customers allocated to the experimental sample with an invitation to complete a short survey in exchange for being entered into lotteries to win a $100 or $200 gift card. Importantly, the email and recruitment occurred before the start of treatment, so the labels could not have affected enrollment into the baseline survey sample. 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 https://www.dropbox.com/s/bx0ort8ntdnx3qq/Baseline_Survey_Formatted.pdf?dl=0.

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 B.2 compares the baseline sample to the full experimental sample, and Appendix Table B.3 tests for balance across treatment arms in the baseline sample. The baseline sample is broadly balanced across treatment arms, except that those assigned to a carbon-labeling arm have slightly higher baseline carbon footprints and state that they consider the environment in their food somewhat less than do those in the control group. We control for observable baseline traits throughout our analysis (see specifications below).

Regression specifications

Our primary analysis uses the following simple regression model:

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

where Y_{it} is an outcome variable for customer i in week t and $AnyLabel_i$ indicates being assigned to see carbon-footprint labels. We control for lagged outcomes in weeks 18, 19, and 20, represented as $\{Y_{iw}\}_{w=18}^{20}$, and indicators that these values are missing, $\{\mathbb{1}(MiY_{iw})\}_{w=18}^{20}$. For some outcomes, these lagged controls are missing because that customer did not order a meal box in

that week. For other administrative variables, such as customer revenue, these values are missing because we observe the variables for a given customer only after they were allocated to the experimental sample; customers were allocated throughout weeks 18 through 20.

The variables δ_t are week fixed effects, and X_i is a vector of other customer control variables. In our specifications in the full experimental sample, these additional controls include indicators for customers' meal plan with HelloFresh (i.e. veggie, chef's choice, premium, etc.), zip-code-level demographics, and customers' longevity at HelloFresh. We also show robustness to specifications in which we iteratively drop these controls in Appendix Figures A.2 and A.8. We control in all baseline-sample regressions for customers' educational attainment, gender, political affiliation, self reports of environmentalism, and how much they consider environmental impacts in their food choices. We describe these controls in more detail below.

We also estimate regressions in which we separately estimate the impacts of the three label variants, as follows:

$$Y_{it} = \sum_{j=1}^{4} \beta_j LabelTreat_{ij} + \sum_{w=18}^{20} \gamma_{1,w} Y_{iw} + \sum_{w=18}^{20} \gamma_{2,w} \mathbb{1}(MiY_{iw}) + \delta_t + \Phi X_i + \varepsilon_{it} \tag{2}$$

where $\{LabelTreat_j\}_{j=1}^{4}$ are indicators for being in the control group or each of the three label groups, and $\{\beta_j\}_{j=1}^{4}$ are our primary coefficients of interest.

Finally, we estimate regressions in which we test for heterogeneous treatment effects, both in the main experimental sample and the baseline sample. We estimate heterogeneous treatment effects by fully stratifying the sample by these baseline traits, described below.

Outcome variables

Meal choices. We use administrative data from HelloFresh on customers' meal boxes to define a series of outcome variables for customers' meal choices. We observe the meals that each customer ordered each week from weeks 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.) 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.

For weeks 17 through 28, we observe the carbon footprints of all meals ordered. These data allow us to define outcomes for customers' total mealbox carbon footprints each week, as well as the number of meals they order in each footprint category.

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.

Financial outcomes. We calculate 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 test effects on an alternate profit variable set to missing in weeks in which customers did not order a meal box.

Control variables

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. 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' zip codes to merge them to county-level data on presidential vote shares in the 2020 election and to demographic data from the 2020 American Community Survey (ACS) 5-year estimates by Zip Code Tabulation Area (ZCTA). We merge customers' zip codes to county FIPS codes using a crosswalk downloaded from the U.S. Department of Housing and Urban Development. About 2% of customers have zip codes that do not merge to either the HUD crosswalk or ACS ZCTAs. Our main regressions control for the following:

  - County-level vote share for President Biden in the 2020 presidential election. Where zip codes span multiple counties, we construct the weighted average of vote shares, weighted by the share of zip code 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.

Baseline survey sample. In regressions with the baseline survey sample, we remove the controls for zip code 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 degree 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.

Dimensions of heterogeneity

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

- *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 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.

  For each of the dimensions above, we test robustness in Appendix Figures A.5 and A.9 to alternative splits on these scales. We also explore the robustness of these results by testing heterogeneous treatment effects by these correlated measures of climate concern:

  - **Climate worry:** We ask participants to rate how worried they are about climate change, from the following: "Not at all worried;" "A bit worried;" "Moderately worried;" and "Very worried." We test heterogeneity by whether customers place themselves 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: "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.

– **Self-perceived environmentalism:** We elicit the extent to which baseline participants identify as environmentally conscious (see above), testing for heterogeneous treatment effects by whether place themselves above or equal to versus below the median response.

– **Role of environment in food choices:** We elicit the extent to which baseline participants consider environmental impacts in their food choices (see above), testing for heterogeneous treatment effects by whether customers place themselves 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.

– **County-level predicted climate beliefs in the full sample:** We get as close as possible to testing for comparable heterogeneity in the full experimental sample by testing for heterogeneous treatment effects by county-level climate beliefs, estimated in the Yale Climate Opinion Maps (Howe et al., 2015). We estimate heterogeneity by terciles of predicted county-level agreement that climate change is caused mostly by human activities, that Congress and corporations should do more to address climate change, that climate change should be a high priority for the next president and Congress, and that they are worried about climate change. We merge customers to these data by zip

codes, which are missing for 2% of customers.

- *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. Appendix Figure A.7 tests robustness to alternative definitions of carbon-footprint knowledge.

Effects on carbon footprints

Adding carbon-footprint labels to the HelloFresh menu induced small average shifts towards lower-carbon diets (Figures 2a-c). On average, labels reduced weekly carbon footprints by 0.075 kg CO2e ($p < 0.01$), a 0.6% decrease from the control mean of 11.8 kg CO2e. This aggregate effect reflects a 1.5% reduction in consumption of *Fair* meals ($p < 0.01$) and a 0.8% increase in consumption of *Climate Superstar* meals ($p = 0.06$). Mapping to protein choices, the labels caused a 1.5% decrease in beef meals ($p < 0.01$), while moderately increasing purchases of vegetarian (1.0%, $p = 0.14$), poultry (0.3%, $p = 0.37$), and pork meals (0.3%, $p = 0.44$). These effects are robust to changing our definition of the treatment period, altering the set of control variables we include, and restricting to customers who historically made active meal choices rather than accepting default meals (Appendix Figure A.2).

These average effects are substantially smaller than those estimated in previous evaluations of carbon-footprint labels. Past studies were implemented in the lab or in smaller-scale commercial or university settings, and they tend to find that labels reduce carbon footprints by about 4% (Potter et al., 2021, Taufique et al., 2022, Brunner et al., 2018, Lohmann et al., 2022, Tilling, 2025), roughly six times our estimates. This discrepancy may arise for several reasons. First, prior pa-

pers test the impacts of large and salient labels (e.g. that mark high-carbon meals in red), while HelloFresh implemented labels that were unobtrusive and relatively positive, categorizing even the highest-carbon meals as *Fair* and in yellow. HelloFresh's more subtle labels passed rounds of internal review and are likely more representative of labeling schemes that would be voluntarily enacted by profit-maximizing companies for long-term use. Second, our study took place in the US with a population that was not selected to be concerned about climate change, while previous studies took place in Europe and often with university students, two factors associated with higher concern about climate change (Gubbala, 2022, Dechezleprêtre et al., 2025). This difference in effect size underscores the critical importance of testing the impacts of labels in realistic settings, at scale, and as designed by profit-maximizing companies, their primary real-world implementers.

Backlash to the carbon labels

The small average label effects mask important heterogeneity in customers' response to environmental nudging (Figure 2d). Among the 5,592 customers surveyed at baseline, we measure a climate-morality index capturing the degree to which respondents believe that individuals or companies have a moral duty to address climate change. Those with below-median scores on the climate-morality index *increased* their carbon footprints by 4.1% ($p = 0.04$) in response to labels, driven by a 6.7% increase in beef consumption ($p = 0.08$), while those with above-median climate morality (non-significantly) decreased their carbon footprints and beef consumption by 2.4% ($p = 0.20$) and 4.0% ($p = 0.32$), respectively. This divergence in the effect of the labels by consumers' climate beliefs increased emissions polarization, widening rather than narrowing the emissions gap between consumers with below- versus above-median scores on the climate morality index (Appendix Figure A.3).

Similar patterns emerge across other correlated climate attitudes, including self-reported environmentalism and belief in anthropogenic climate change (Appendix Figure A.4). These results are also robust to alternate splits of our primary climate-attitudes variables (Appendix Figure A.5) and to simultaneously controlling for heterogeneous effects by other customer traits (Appendix

Table B.4).

These individual-level patterns in the baseline-survey sample are consistent with similar heterogeneity in our full experimental sample. While we lack data on individual climate beliefs for the full sample, we match customers' zip codes to estimates of county-level average climate beliefs in 2024 across the US from the Yale Climate Opinion Maps, derived from nationally-representative surveys of over 32,000 Americans (Howe et al., 2015). Across a range of beliefs, the labels' impacts decrease with the estimated share of zip-code residents with low levels of concern about climate change, consistent with stark heterogeneity in label effects at the individual level (Appendix Figure A.6).

Mechanisms

To understand why labels produced such modest average effects—and triggered backlash among some consumers—we examine whether the labels work by providing new information or by making customers' existing knowledge salient at the moment of choice. Consistent with prior work (Tilling, 2025, Imai et al., 2022), several pieces of evidence suggest that the labels worked largely through salience.

First, the *All-Tiers* and *Superstars-Only* labeling schemes had highly similar impacts on meal choices (Figure 3a). The *Superstars-Only* labeling arm only provided information to distinguish the *Climate-Superstar* meals from all others, and yet it substantially reduced consumption of *Fair* meals relative to *Good* meals. This pattern suggests that the labels worked 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 had larger effects among those without baseline carbon-footprint knowledge, as we would expect if the labels primarily worked through information provision. We elicited knowledge by asking baseline-survey respondents to rank six foods from highest to lowest emissions per pound on average: beef, cheese, farmed shrimp, pork, chicken, and tomatoes (listed here in the correct order by average emissions). We define baseline knowledge by whether participants correctly answered that beef and tomatoes have the highest and

lowest carbon footprints, respectively.

Across all customers in the baseline sample, we find no gap in estimated effects on carbon footprint or number of *Fair* meals by baseline footprint knowledge (Figure 3b). This null heterogeneity is robust to alternative definitions of footprint knowledge (Appendix Figure A.7). When we instead split the sample by both climate-morality beliefs and baseline carbon-footprint knowledge, we find that carbon-footprint reductions and increases among those with above- versus below-median morality beliefs, respectively, are entirely driven by those with correct baseline footprint knowledge (Figure 3c). Thus, reactions to the labels—whether in backlash to or in line with the labels' goals—arose by making customers' existing knowledge salient, not by correcting customers' misperceptions about the climate impacts of different foods.

Effects on retention and profit

If carbon-footprint labels have such modest—and in some cases negative—environmental impacts, why are they increasingly common sustainability tools across the private sector? The answer may lie in their impacts on company profits, which no study has previously examined. Our direct collaboration with HelloFresh allows us to conduct novel tests of the impacts of the climate labels on back-end financial outcomes: customer retention, revenue, total direct costs, and profits.

We find that the carbon labels were financially beneficial for HelloFresh overall, driven by increases in customer retention (Figures 4a-b). Carbon labels increased the probability that customers ordered a meal box in a given week by 0.3 percentage points (1.0%, $p = 0.03$). This retention increase then cascades through HelloFresh's financial metrics. Setting outcomes to zero for customers who did not order a box in a given week, the labels increased direct costs per customer-week by €0.13 (1.0%, $p = 0.04$) and increased revenue per customer-week by €0.20 (1.0% , $p = 0.04$). On net, profit increased by €0.07 per customer-week (0.9%, $p = 0.09$). These effects are robust across a large set of specifications, including changing our definition of the treatment period, altering the set of control variables we include, and restricting to customers who historically made active meal choices rather than accepting defaults (Appendix Figure A.8). Note also that

these treatment effects on retention—which shape the pool of customers among whom we can estimate meal-choice effects each week—do not explain the small impacts of labels on meal-choices that we observe (Appendix Section C). If anything, differential customer retention may cause us to over-estimate the labels' effects on meal choices.

Notably, there is no evidence that backlash to the labels among ideologically opposed customers extended to their meal choices. We detect no significant decrease in customer retention among those who disagree that companies or individuals have a moral duty to address climate change (Figure 4c). These results help to explain the proliferation of voluntary carbon labeling despite limited evidence of environmental impact. From a financial perspective, companies face compelling reasons to implement climate labels: they signal environmental commitment to sustainability-minded consumers without alienating skeptical consumers enough that they abandon the product entirely.

Indeed, HelloFresh faced a profit tradeoff that many other firms are unlikely to encounter and yet still found the climate labels profitable overall. Given HelloFresh's pricing structure, the carbon labels if anything reduced profit *per meal sold*. While vegetarian and other low-carbon meals have lower production costs, HelloFresh more than recoups higher costs for beef meals with a premium-protein surcharge. By shifting consumption away from higher-profit beef meals towards lower-margin vegetarian options, the labels if anything reduced per-meal profitability by 0.25% ($p = 0.35$). While the higher-emissions meals at HelloFresh happened to also be the higher-margin choices, many other companies profit *more* from lower-emissions meals and so may have even steeper financial incentives to implement climate labels (Ellis, 2023).

Discussion

We evaluate the impacts of carbon-footprint labeling in a field experiment with over 230,000 customers of a major food company in the U.S. The labels have small environmental benefits, reducing customers' carbon footprints by 0.6% per week. Moreover, ideologically opposed customers *increase* their footprints in response to the labels, resulting in increased dietary polarization by cli-

mate beliefs. Despite small and inconsistent environmental benefits, the labels increased customer retention and profit for HelloFresh by 1.0% and 0.9%, respectively.

Addressing climate change will require stark reductions in human dietary emissions. The labels' minimal 0.6% average effect on emissions, coupled with backlash effects among certain consumer segments, highlight that policymakers cannot rely on voluntary carbon labeling to drive these shifts. If widespread carbon labeling reduced Americans' per-capita total food emissions by 0.6%, this reduction would comprise 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).

While information provision has successfully changed behavior in a range of domains, carbon-footprint labels may face unique challenges because they have become entangled with political identity. For example, Republican lawmakers came out in 2021 against an (imagined) Democratic plan to reduce beef consumption, issuing broad warnings to "stay out of my kitchen" (Washington Post Staff, 2021). Among individuals opposed to climate action, our results show that climate nudges can backfire, widening rather than narrowing the emissions gap between different consumer segments. Coupled with concerns that behavioral-change interventions might divert attention from the broad-scale policy required to address climate change (Chater and Loewenstein, 2023), these backlash effects suggest that carbon labeling may be counterproductive in some settings.

Despite these limited environmental benefits, the positive effects of labels on company profits suggest that carbon labeling and other interventions to reduce customers' carbon footprints may continue to proliferate. In those cases, it is crucial to consider how to increase label effectiveness while minimizing backlash. First, making carbon information opt-in rather than default could prevent triggering resistance among hostile audiences while still serving interested consumers. Second, the heterogeneous responses we document underscore the importance of segmented approaches. Among the 62% of Americans who say that individuals should do more to address climate change, (Leiserowitz et al., 2025) more aggressive interventions—such as defaulting to low-carbon options or implementing higher prices for higher-emissions choices—might prove ac-

ceptable and effective. Among resistant populations, however, indirect approaches that emphasize co-benefits (e.g. health or cost) rather than environmental impacts might avoid triggering identity-based opposition.

Several study limitations merit consideration. First, HelloFresh customers likely differ from the general population in ways that could alter label effects. If anything, we expect label effects to be larger among HelloFresh customers—who live in disproportionately wealthy and well-educated areas—than in nationally representative samples. Second, our eight-week intervention captures short-term responses but cannot assess whether effects persist, decay, or potentially reverse over longer periods. Third, while we observe purchasing behavior through the meal-kit platform, we cannot measure purchasing behavior through other mediums. However, we expect these spillover effects to be small: the labels seem to work via salience, which likely decays quickly over time.

In conclusion, our results provide sobering evidence about the limits of climate nudges in polarized contexts. While carbon labeling generates modest benefits—small emissions reductions among supportive populations and positive returns for companies—it falls far short of driving the transformative dietary changes required to address climate change. Our results suggest that achieving meaningful reductions in food-system emissions will require more comprehensive approaches. As the climate crisis intensifies, our results underscore both the urgency of dietary transitions and the inadequacy of current tools to achieve them.

References

(2025). Annual report 2024. Technical report, HelloFresh SE, Berlin, Germany. Accessed: 14 Aug 2025. (document)

Angrist, N., Winseck, K., Patrinos, H. A., and Zivin, J. G. (2024a). Human Capital and Climate Change. *The Review of Economics and Statistics*, pages 1–28. (document)

Angrist, N., Winseck, K., Patrinos, H. A., and Zivin, J. G. (2024b). Human capital and climate change. *Review of Economics and Statistics*, pages 1–28. (document)

Bajželj, B., Richards, K. S., Allwood, J. M., Smith, P., Dennis, J. S., Curmi, E., and Gilligan, C. A. (2014). Importance of food-demand management for climate mitigation. *Nature Climate Change*, 4(10):924–929. (document)

Bassi, C., Maysels, R., and Anex, R. (2022). Declining greenhouse gas emissions in the us diet (2003–2018): Drivers and demographic trends. *Journal of Cleaner Production*, 351:131465. (document)

Bilén, D. (2024). Do carbon labels cause consumers to reduce their emissions? evidence from a large scale natural experiment. *Working Paper*. (document)

Brunner, F., Kurz, V., Bryngelsson, D., and Hedenus, F. (2018). Carbon Label at a University Restaurant – Label Implementation and Evaluation. *Ecological Economics*, 146:658–667. (document)

Camilleri, A. R., Larrick, R. P., Hossain, S., and Patino-Echeverri, D. (2019). Consumers underestimate the emissions associated with food but are aided by labels. *Nature Climate Change*, 9:53–58. (document)

Chater, N. and Loewenstein, G. (2023). The i-frame and the s-frame: How focusing on individual-

level solutions has led behavioral public policy astray. *Behavioral and Brain Sciences*, 46. (document)

Clark, M. A., Domingo, N. G., Colgan, K., Thakrar, S. K., Tilman, D., Lynch, J., Azevedo, I. L., and Hill, J. D. (2020). Global food system emissions could preclude achieving the 1.5 and 2 c climate change targets. *Science*, 370(6517):705–708. (document)

Commisso, D. (2022). Meal kit delivery continues to thrive in the stay-at-home economy. Technical report, Civic Science. (document)

Crippa, M. et al. (2023). GHG emissions of all world countries - JRC/IEA 2023 report. Technical report, Publications Office of the European Union. (document)

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

Dechezleprêtre, A., Fabre, A., Kruse, T., Planterose, B., Sanchez Chico, A., and Stantcheva, S. (2025). Fighting Climate Change: International Attitudes toward Climate Policies. *American Economic Review*, 115(4):1258–1300. (document)

DellaVigna, S. and Linos, E. (2022). Rcts to scale: Comprehensive evidence from two nudge units. *Econometrica*, 90(1):81–116. (document)

Ellis, M. (2023). How plant-based menu items meet demand and drive profits. Full Service Restaurant (FSR) Magazine. Accessed August 2025. (document)

Elofsson, K., Bengtsson, N., Matsdotter, E., and Arntyr, J. (2016). The impact of climate information on milk demand: Evidence from a field experiment. *Food policy*, 58:14–23. (document)

Franzen, A. and Vogl, D. (2013). Two decades of measuring environmental attitudes: A comparative analysis of 33 countries. *Global Environmental Change*, 23(5):1001–1008. (document)

Gubbala, Moira Fagan and Sneha, J. P. (2022). Climate Change Remains Top Global Threat Across 19-Country Survey. (document)

Hagmann, D., Ho, E. H., and Loewenstein, G. (2019). Nudging out support for a carbon tax. *Nature Climate Change*, 9:484–489. (document)

Hedenus, F., Wirsenius, S., and Johansson, D. J. (2014). The importance of reduced meat and dairy consumption for meeting stringent climate change targets. *Climatic change*, 124(1):79–91. (document)

Howe, P. D., Mildenberger, M., Marlon, J. R., and Leiserowitz, A. (2015). Geographic variation in opinions on climate change at state and local scales in the USA. *Nature Climate Change*, 5(6):596–603. Publisher: Nature Publishing Group. (document), A.6

Imai, T., Pace, D. D., Schwardmann, P., and van der Weele, J. J. (2022). Correcting consumer misperceptions about co 2 emissions. *CESifo working paper 10138*. (document)

International Energy Agency (2024). CO2 emissions in 2023. Technical report, IEA. (document)

Johns Hopkins Center for a Livable Future (2023). Meat consumption: trends and health implications. Technical report, Johns Hopkins Bloomberg School of Public Health. (document)

Jones, M. W., Peters, G. P., Gasser, T., Andrew, R. M., Schwingshackl, C., Gütschow, J., Houghton, R. A., Friedlingstein, P., Pongratz, J., and Le Quéré, C. (2023). National contributions to climate change due to historical emissions of carbon dioxide, methane, and nitrous oxide since 1850. *Scientific Data*, 10(1):155. (document)

Lamb, W. F., Wiedmann, T., Pongratz, J., Andrew, R., Crippa, M., Olivier, J. G., Wiedenhofer, D., Mattioli, G., Al Khourdajie, A., House, J., et al. (2021). A review of trends and drivers of greenhouse gas emissions by sector from 1990 to 2018. *Environmental Research Letters*, 16(7):073005. (document)

Lee, H. and Romero, J. (2023). Climate change 2023: Synthesis report. *IPCC 2023: A Report of the Intergovernmental Panel on Climate Change*. (document)

Leiserowitz, A., Maibach, E., Rosenthal, S., Kotcher, J., Goddard, E., Carman, J., Myers, T., Verner, M., Marlon, J., Goldberg, M., Ettinger, J., Fine, J., and Thier, K. (2025). Climate Change in the American Mind: Politics & Policy, Spring 2025. Technical report, Yale University and George Mason University. (document)

Li, Y., He, P., Shan, Y., Li, Y., Hang, Y., Shao, S., Ruzzenenti, F., and Hubacek, K. (2024). Reducing climate change impacts from the global food system through diet shifts. *Nature Climate Change*, 14:943–953. (document)

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

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

Pew Research (2024). Americans' Top Policy Priority for 2024: Strengthening the Economy. Technical report, Pew Research Center. (document)

Poore, J. and Nemecek, T. (2018). Reducing food's environmental impacts through producers and consumers. *Science*, 360(6392):987–992. (document)

Potter, C., Bastounis, A., Hartmann-Boyce, J., Stewart, C., Frie, K., Tudor, K., Bianchi, F., Cartwright, E., Cook, B., Rayner, M., et al. (2021). The effects of environmental sustainability labels on selection, purchase, and consumption of food and drink products: a systematic review. *Environment and behavior*, 53(8):891–925. (document)

Ripple, W. J., Smith, P., Haberl, H., Montzka, S. A., McAlpine, C., and Boucher, D. H. (2014). Ruminants, climate change and climate policy. *Nature Climate Change*, 4(1):2–5. (document)

Springmann, M., Godfray, H. C. J., Rayner, M., and Scarborough, P. (2016). Analysis and valuation of the health and climate change cobenefits of dietary change. *Proceedings of the National Academy of Sciences*, 113(15):4146–4151. (document)

Taufique, K. M., Nielsen, K. S., Dietz, T., Shwom, R., Stern, P. C., and Vandenbergh, M. P. (2022). Revisiting the promise of carbon labelling. *Nature Climate Change*, 12(2):132–140. (document)

Tilling, A. S. (2025). Changing consumption behavior with carbon labels: Causal evidence on behavioral channels and effectiveness. *Working Paper*. (document)

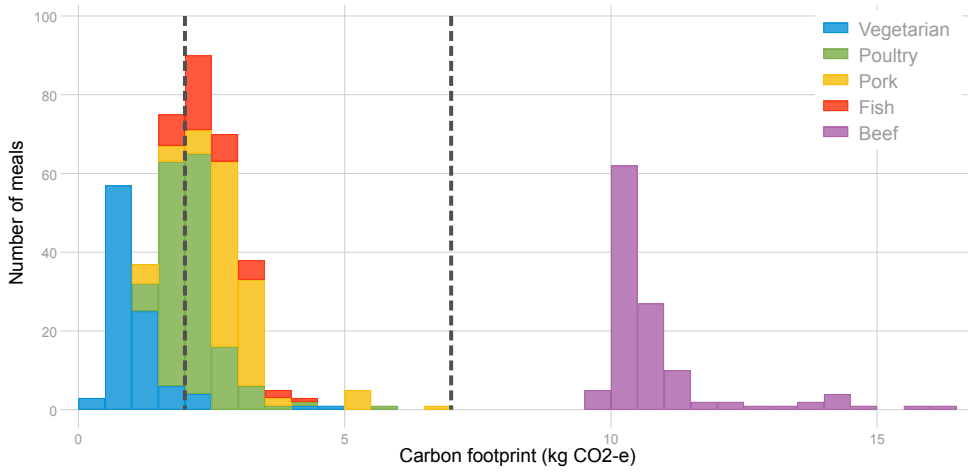
Vlaeminck, P., Jiang, T., and Vranken, L. (2014). Food labeling and eco-friendly consumption: Experimental evidence from a belgian supermarket. *Ecological Economics*, 108:180–190. (document)

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

Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., Garnett, T., Tilman, D., DeClerck, F., Wood, A., et al. (2019). Food in the anthropocene: the eat–lancet commission on healthy diets from sustainable food systems. *The lancet*, 393(10170):447–492. (document)

Xu, X. et al. (2021). Global greenhouse gas emissions from animal-based foods are twice those of plant-based foods. *Nature Food*, 2:724–732. (document)

(a)



(b)

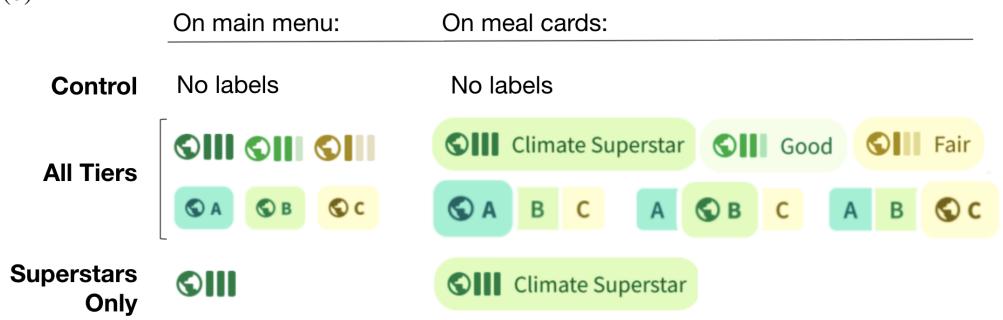


Fig.1 | **Adding climate-footprint labels to meals. a,** The distribution of estimated carbon footprints for meals during the experimental period. The dotted lines indicate the emissions cutoffs between *Climate Superstar* and *Good* meals, and *Good* and *Fair* meals, respectively. This figure groups meals according to their main protein source. **b,** The carbon-labeling treatments, as displayed on customers' main menus and on more detailed menu cards about each meal. The Control group saw menus with no carbon-footprint labels. In the All-Tiers arms, environmental grades were shown for all meals, but with two variants on the label appearance. In the Superstars-Only treatment arm, only the environmental grade for *Climate Superstar* meals were shown. Appendix Figure A.10 shows examples of how these labels were placed on menus.

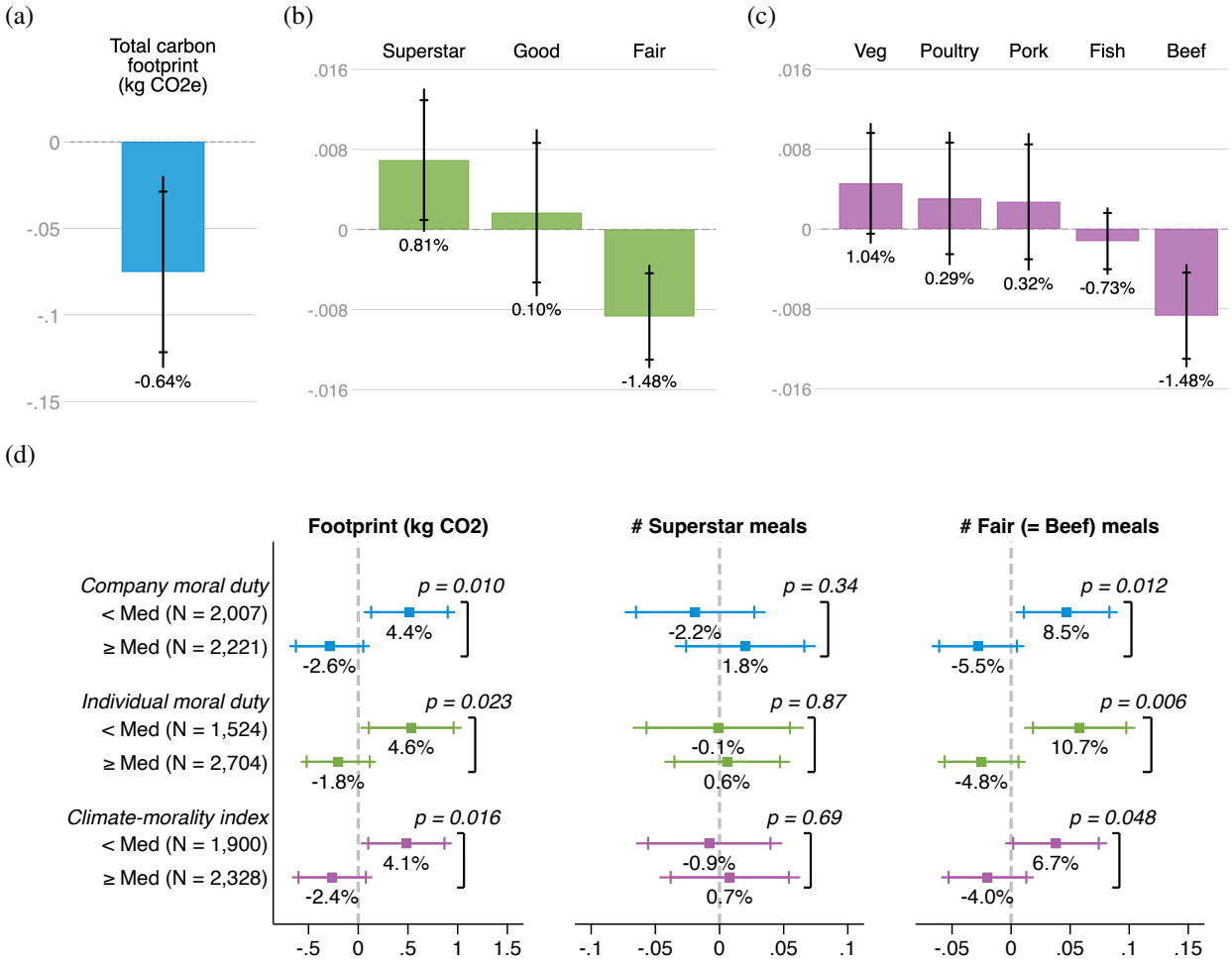


Fig.2| **Effects of labels on meal choices. a,** The average treatment effect of the climate labels on customers' carbon footprint per week, conditional on ordering meals. This figure pools together all treatment arms (effects shown separately in Appendix Figure A.11). Regressions include controls for lagged outcomes, fixed effects for customers' meal plan with HelloFresh (e.g. vegetarian or chef's choice), zip-code demographics, customer longevity at HelloFresh, and week fixed effects. Standard errors are clustered by customer. The sample is 397,514 customer-week observations across 131,017 customers who ordered at least one meal kit during the study period. Capped and uncapped error bars indicate 90% and 95% confidence intervals, respectively. **b,** The average treatment effect of the climate labels on number of meals ordered per customer per week in each footprint category, conditional on ordering meals and using the same regression set-up as subfigure a. **c,** The average treatment effect of the climate labels on customer's meal choices by protein, conditional on ordering meals, again using the same regression set-up as subfigure a. **d,** The heterogeneous treatment effect of the climate labels on customers' meal choices by whether customers have below- or above-median beliefs that individuals and companies have a moral duty to take action on climate change, measured in the baseline survey. These regressions replace zip-code demographics with controls for individual education, gender, political affiliation, and stated environmentalism, measured in the baseline survey. We show *p*-values for testing the null of equal treatment effects across the groups. We show the number of customers in each heterogeneity category in parentheses beside each label. The total sample comprises 14,108 customer-week observations across 4,228 baseline-survey customers.

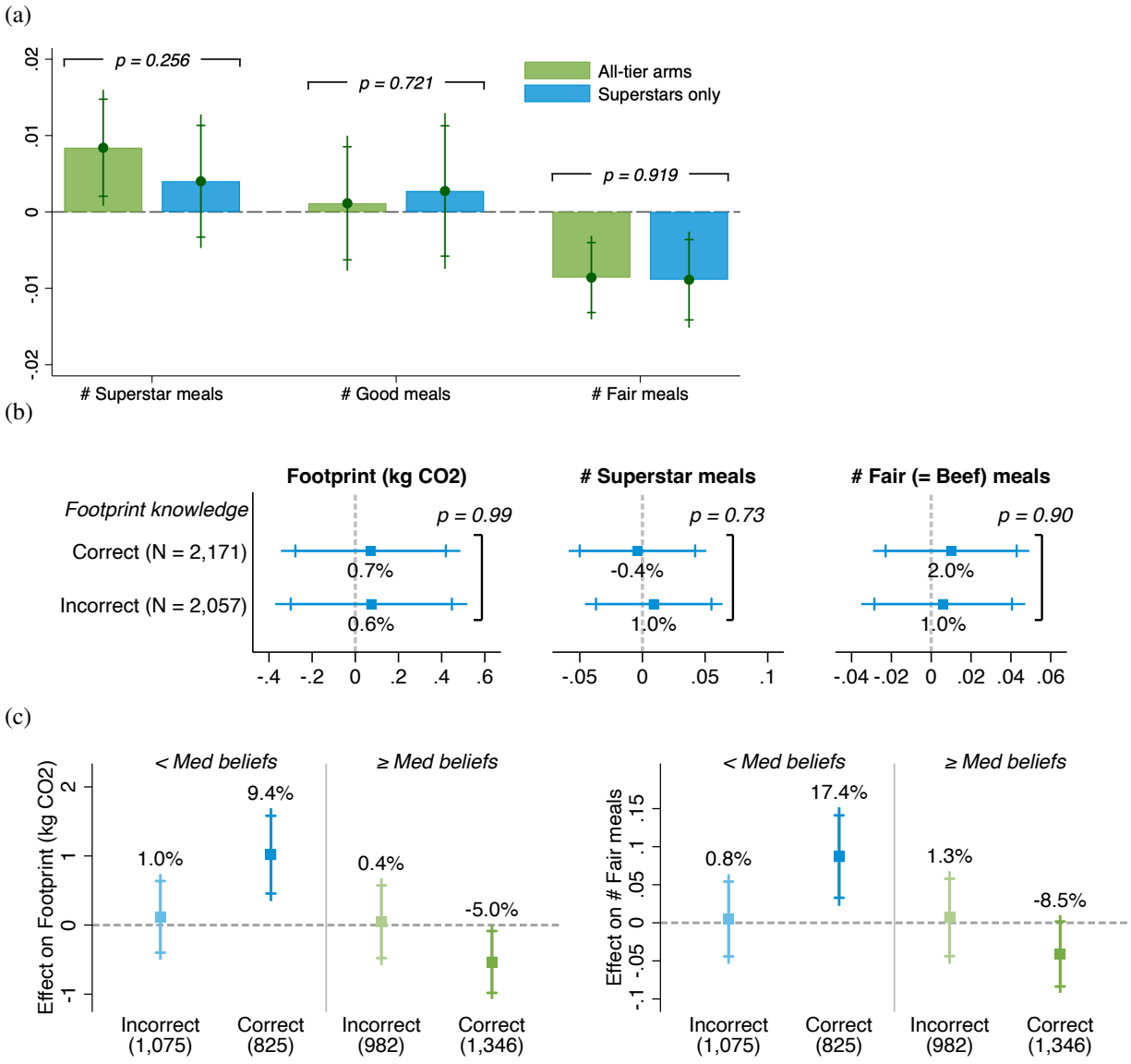


Fig.3| Testing the role of information versus salience. a, Comparing the treatment effect of climate labels by whether only *Climate Superstar* meals were labeled (the Superstars-Only arm) or whether all meals were labeled (the All-Tiers arms, pooling the two label variants shown in Figure 1b). These regressions use the same controls and error clustering as in Figures 2a-2c. We randomized a total of 58,645, 116,826, and 59,040 customers to the Control, All-tiers, and Superstars-only arms, yielding samples of 32,685 Control customers, 65,301 All-tiers customers, and 33,031 Superstars-only customers who ever ordered a meal box and are included in these regressions. **b,** Comparing the treatment effect of climate labels by baseline knowledge about carbon footprints of different types of food, measured in the baseline sample. We identify those with correct or incorrect baseline knowledge as those who were able (versus unable) to correctly identify that beef and tomatoes have the highest and lowest carbon footprints, respectively, in a list of 6 foods. These regressions use the same controls and error clustering as in Figure 2d. **c,** Comparing the treatment effect of climate labels jointly by baseline footprint knowledge and scoring above or below median on the climate-morality index. These regressions use the same controls and error clustering as in Figure 2d. We show the number of customers in each heterogeneity category in parentheses beside each label.

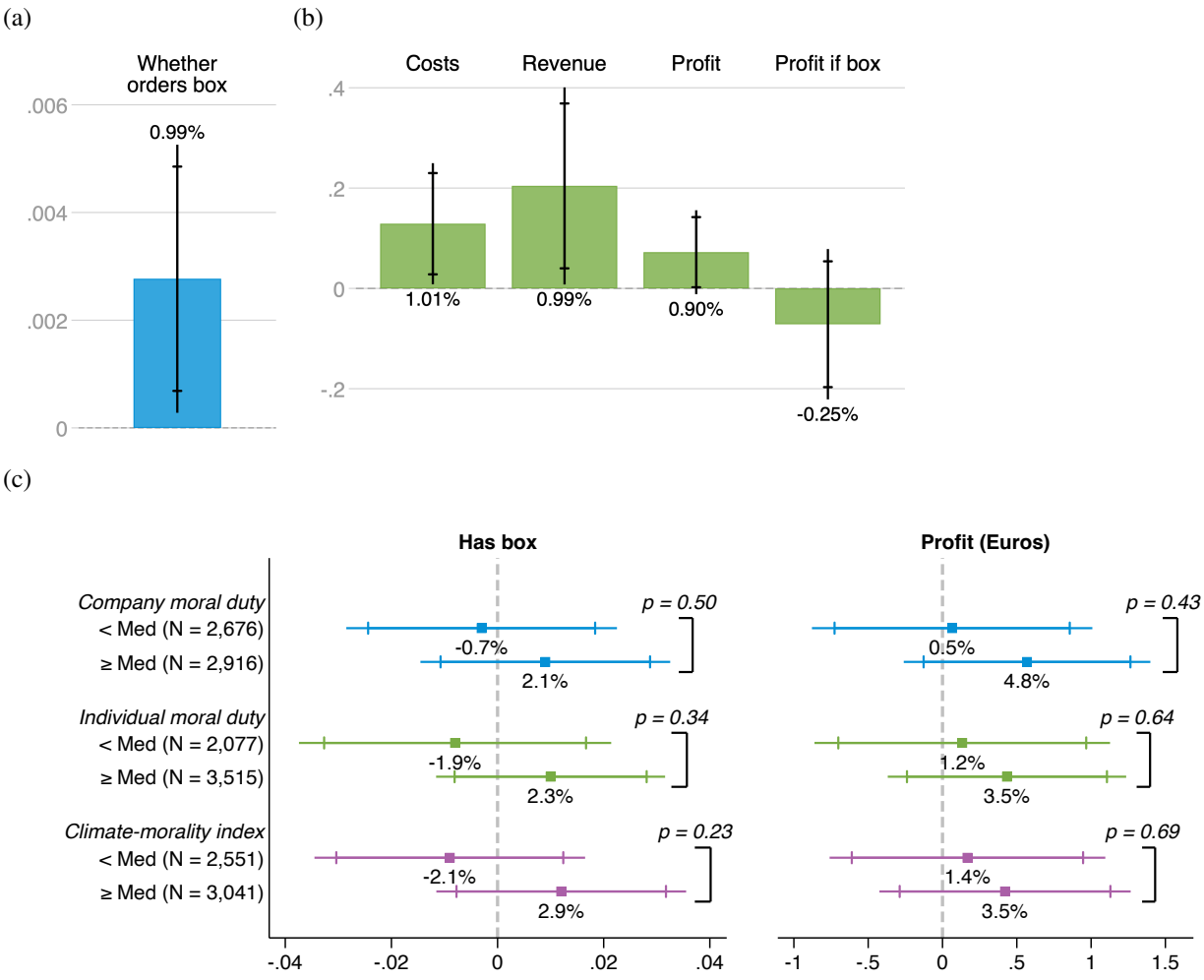
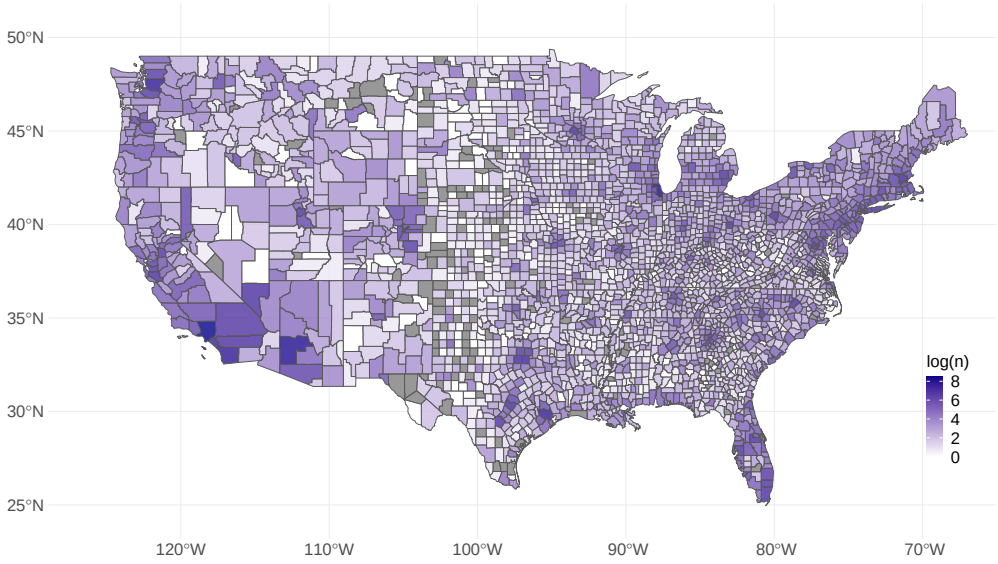


Fig.4| **a,** The average treatment effect of the climate labels on customer retention. For this figure, all treatment arms are pooled together (effects shown separately in Appendix Section A.12). Regressions use the same controls and error clustering as in Figures 2a-2c. The sample is observations for fiscal weeks 23 through 28 for all 234,511 customers, yielding a total sample of 1,407,066 customer-weeks. **b,** The average treatment effect of the climate labels on costs, revenue, and profits per customer-week in Euros, using the same regression set-up as subfigure a. **c,** The heterogeneous treatment effect of the climate labels on customer retention and profits by whether customers have below- or above-median beliefs that individuals and companies have a moral duty to take action on climate change. These regressions use the same controls as in Figure 2d. We show the number of customers in each heterogeneity category in parentheses beside each label. The total sample comprises 33,552 customer-week observations across 5,592 baseline-survey customers.

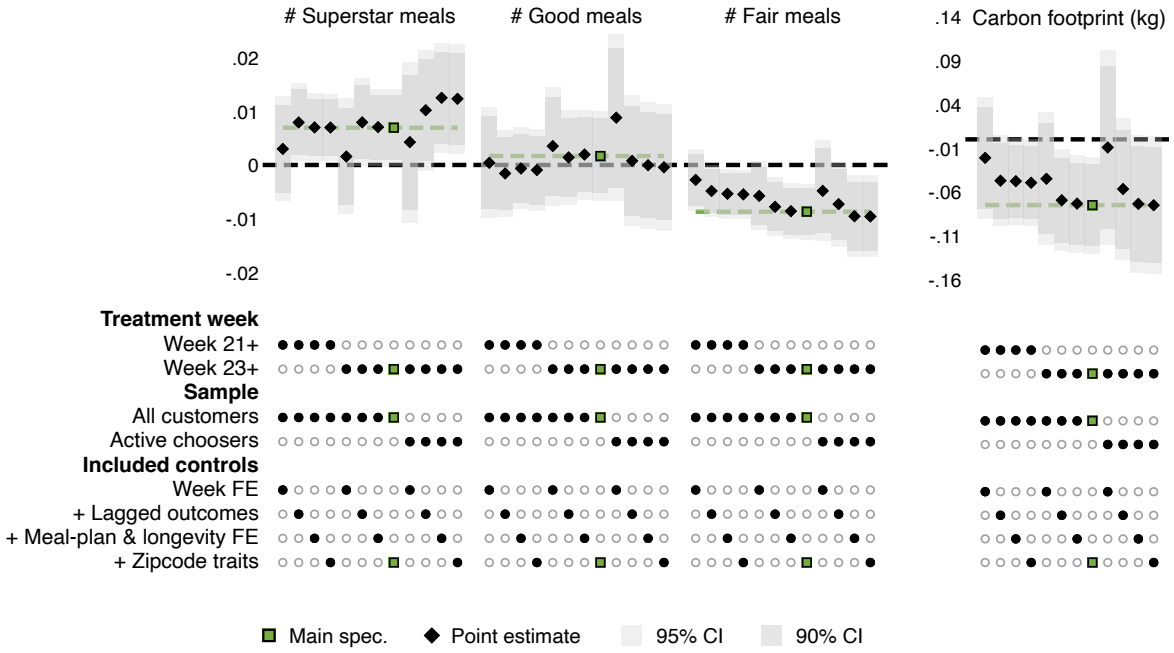
A Appendix Figures

Figure A.1: Geographic distribution of HelloFresh customers



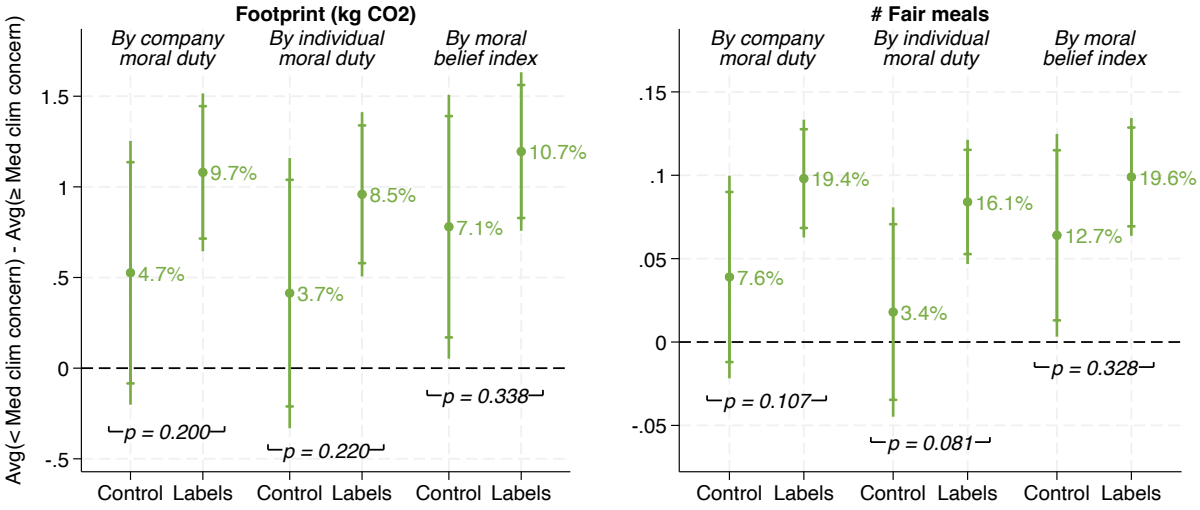
Note: Experiment participants were drawn from HelloFresh customers across the United States. This figure plots the log-distribution of those customers across zip codes, with a darker color indicating a greater number of study participants. Grey indicates that there were no customers from that zip code enrolled in the experiment.

Figure A.2: Specification chart for meal choices



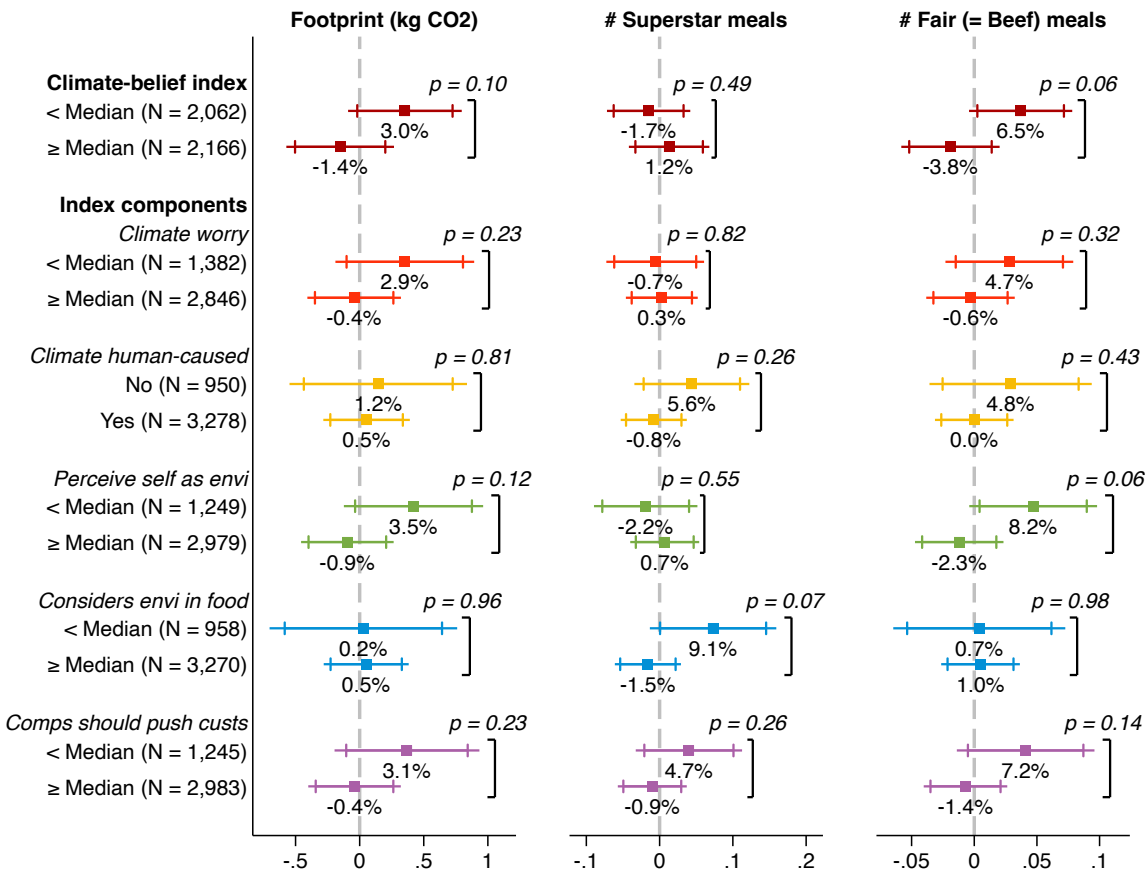
Note: This figure tests the robustness of the labels' impacts on four outcomes: number of *Climate Superstar*, *Good*, and *Fair* meals ordered each week, and total estimated carbon footprint of meals ordered in kg of CO_2-e. We present point estimates for versions of Equation 1 where we vary whether we define the treatment weeks as starting in weeks 21 or 23, whether we define the sample as all customers or only "active choosers," and the controls included. We define active choosers as those who chose a non-defaulted meal in all meal box they ordered in the three weeks before the experimental period. 57% of the sample has at least one meal box in the three pre-experimental weeks, and 61% of these are classified as active choosers. 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. We describe all of these control variables in the Methods section.

Figure A.3: Labels may induce polarization of meal choices by climate beliefs



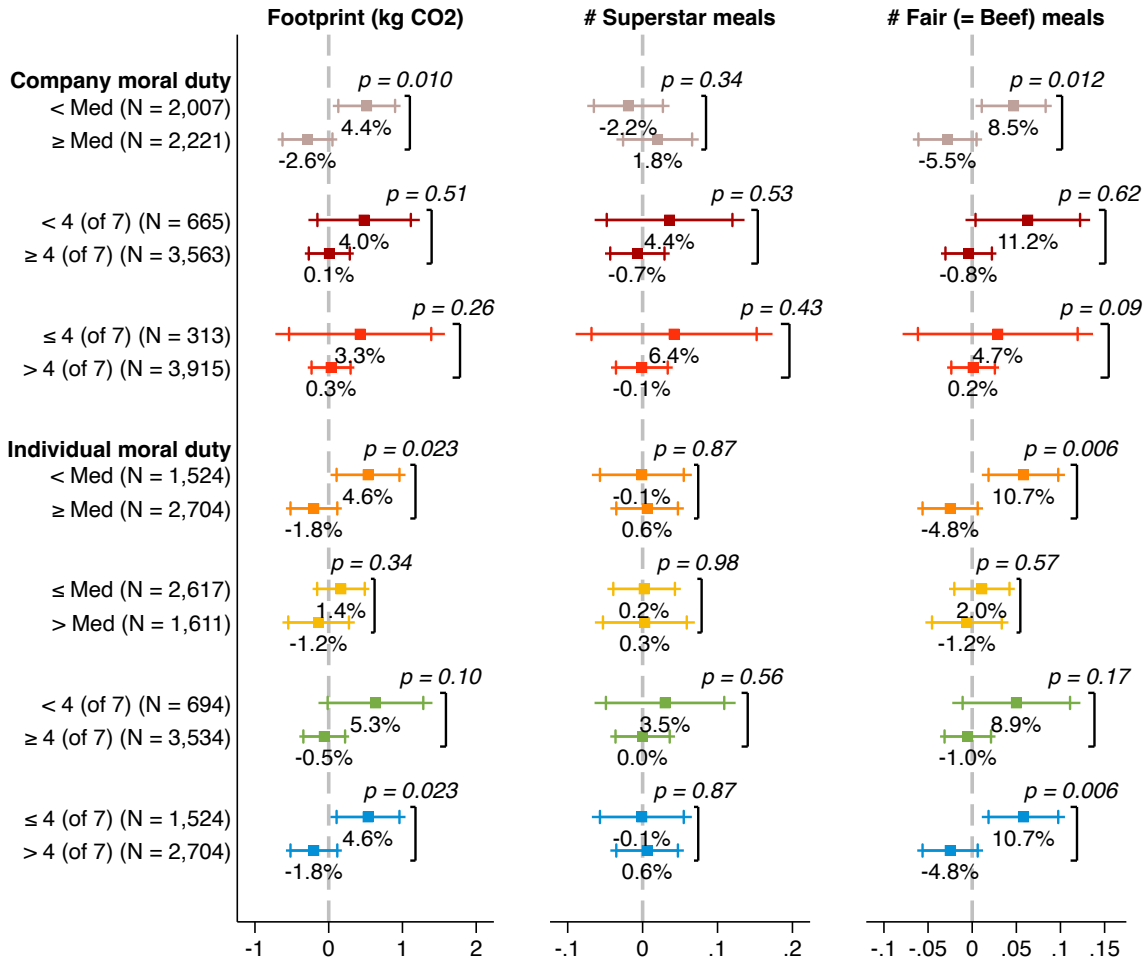
Note: This figure shows that the carbon labels may polarize meal choices by climate beliefs. Within the label treatment and control groups, we iteratively regress total carbon footprint (at left) and number of fair meals ordered per week (at right) on indicators for being below the median on three metrics: belief that companies have a moral duty to address climate change, belief that individuals have a moral duty to make lifestyle changes to reduce their carbon footprints, and the index of these two measures. These are the same measures used in Figure 2d. The point estimates thus show the gap in meal choices between those with above- versus below-median climate-morality beliefs in the samples that did or did not see carbon labels. Capped and uncapped lines denote 90% and 95% confidence intervals, with standard errors clustered by customer. The *p*-values correspond to tests of equality between the gaps in the meal-choice outcomes by climate-morality beliefs in the treatment versus control groups. Gaps in meal choices by beliefs tend to be larger for those who saw carbon labels, though this difference-in-differences tends to be statistically insignificant.

Figure A.4: Heterogeneous meal-choice effects by traits correlated with climate morality



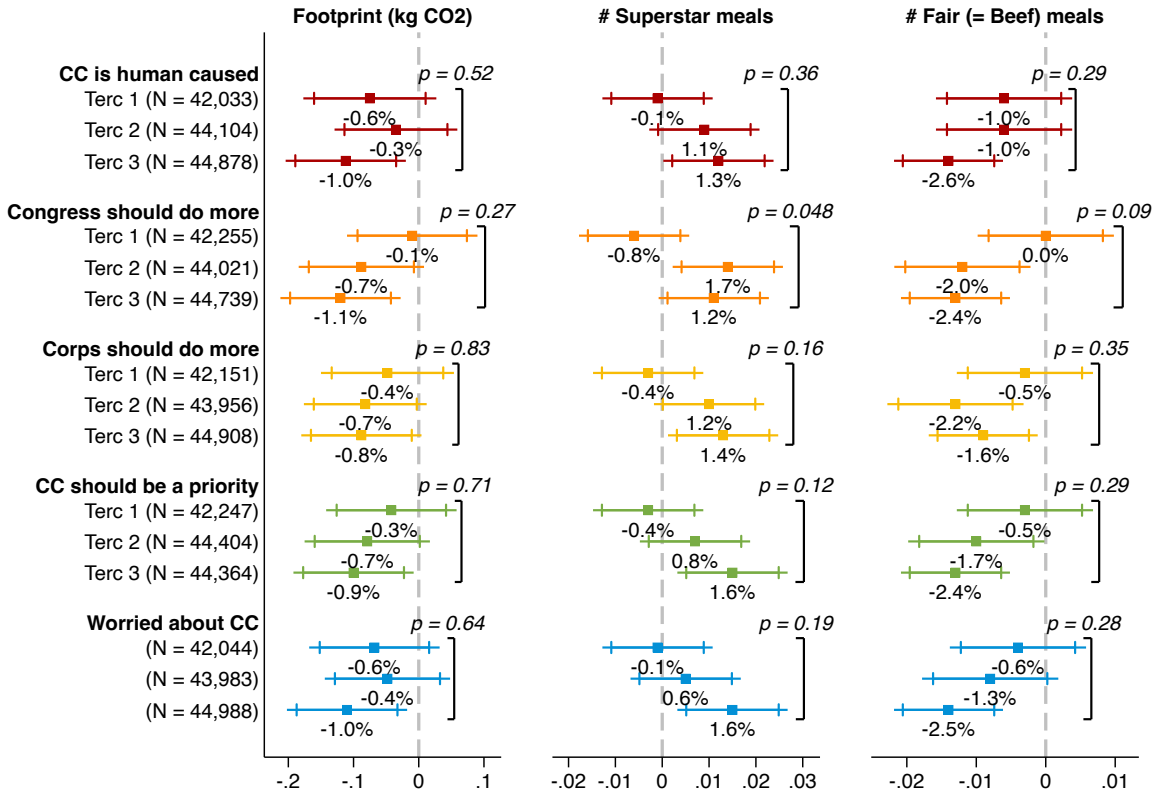
Note: This figure tests for heterogeneous treatment effects of labels on meal choices by other measures of concern about climate change. The top panel in this figure show heterogeneous treatment effects by a climate-belief index constructed by standardizing the climate-morality belief variables shown in Figure 2d and each of the other traits shown here, summing these standardized variables, and then re-standardizing the sum. We test for heterogeneous effects between those who score above versus below the median on this index. The rest of the figure shows heterogeneous treatment effects by scoring above or below the median on each of the index components. The only discrepancy is that the index includes a scale for participants' beliefs about the cause of climate change, standardized from the values 0 (Believe climate change isn't happening), 1 (Believe climate change is naturally caused), and 2 (Believe climate change is human caused), while for power here we test heterogeneity by a binary variable for believing that climate change is human-caused or not. All regressions use the same controls and total sample size as in Figure 2d. Capped and uncapped bars show 90% and 95% confidence intervals, respectively, with standard errors clustered by customer. The p-values correspond to tests of equal treatment effects between groups.

Figure A.5: Robustness of meal-choice regressions to climate-morality definitions



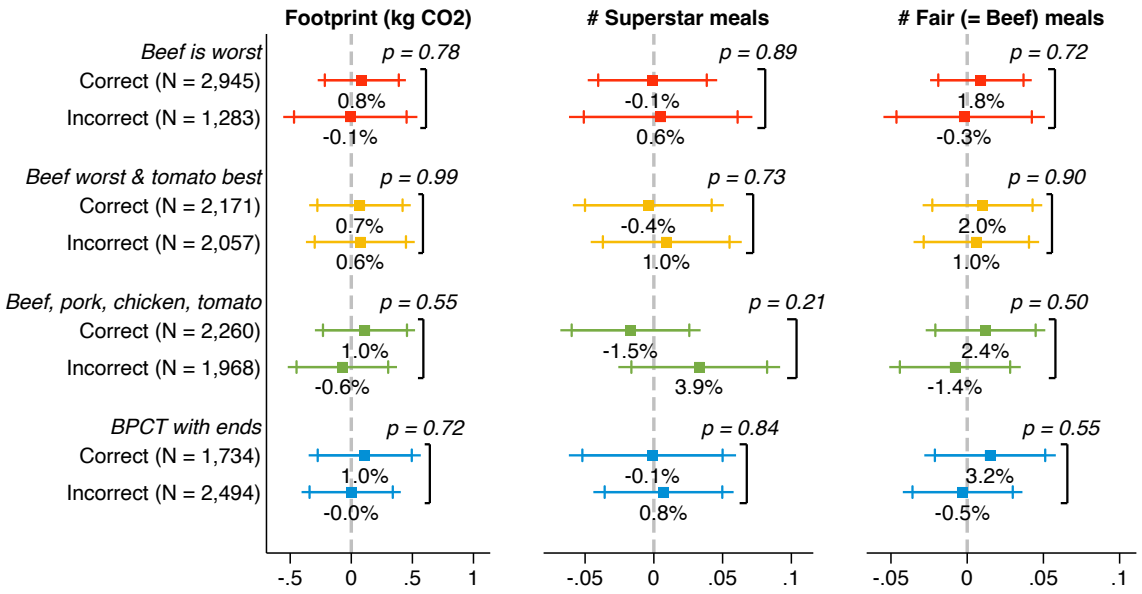
Note: This figure tests for heterogeneous treatment effects of the labels on meal-choice outcomes by alternate definitions of climate-morality beliefs. The top panels for company and individual beliefs reproduce heterogeneous treatment effects when we define these beliefs as in Figure 2d. We then show heterogeneous effects when reclassifying median responses and splitting these scales based on answering above or equal to the scale midpoint. Note that the median response when asked whether companies have a moral duty to address climate change is a 7 of 7, so we cannot test robustness to separating out those with beliefs strictly above the median on this scale. All regressions use the same controls and sample definition as in Figure 2d. Capped and uncapped bars show 90% and 95% confidence intervals, respectively, with standard errors clustered by customer. The p-values correspond to tests of equal treatment effects between groups.

Figure A.6: Heterogeneous treatment effects by zip-code climate beliefs



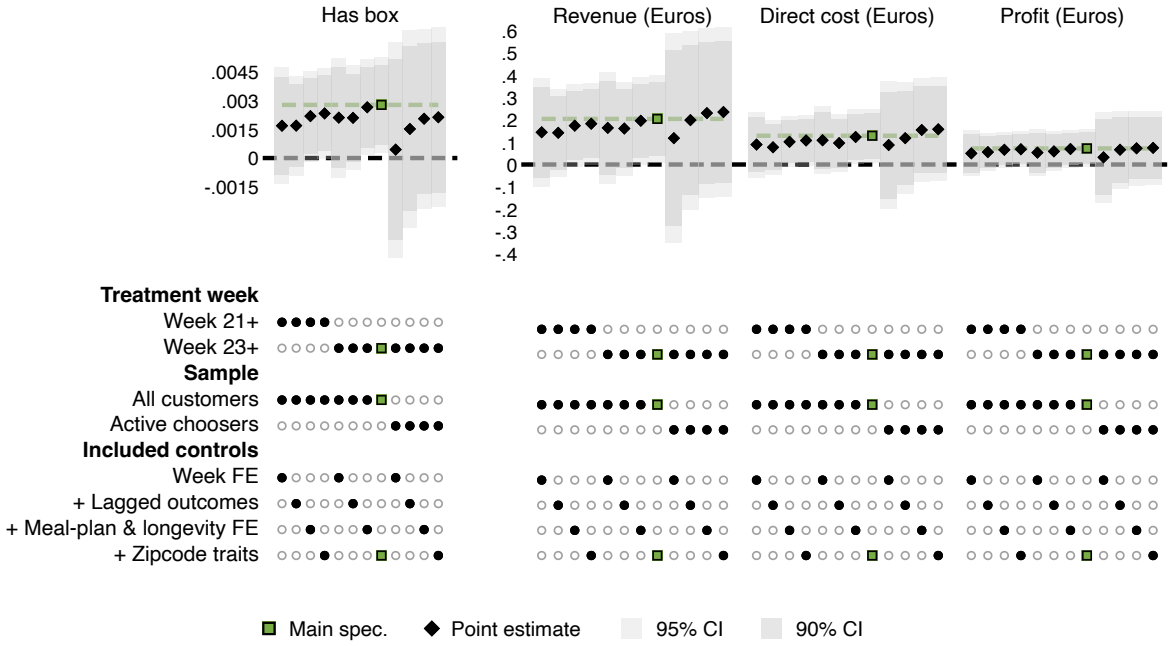
Note: This figure shows heterogeneous label effects by predicted climate beliefs in customers' zip codes, using the full experimental sample. We match customer zip codes to counties using the HUD-USPS zip code crosswalk for September 2020, then match customers to county-level estimated climate beliefs from the 2024 Yale Climate Opinion Maps (Howe et al., 2015). 98% of customers in our sample have zip codes, of whom all but 4 customers have non-missing climate-belief data. We use estimates from these data for the percent of adults who: (1) think that global warming is caused mostly by human activities, (2) think Congress should do more to address global warming, (3) think corporations should do more to address global warming, (4), think global warming should be a high priority for the next president and Congress, and (4) are worried about global warming. For each of these variables, we split the sample with non-missing data into terciles, then estimate parallel regressions to those in Figures 2a-2c separately within each tercile. The capped and uncapped bars show 90% and 95% confidence intervals, respectively, with standard errors clustered by customers. The p-values correspond to tests of equal coefficients across tercile groups.

Figure A.7: Robustness of meal-choice regressions to knowledge definitions



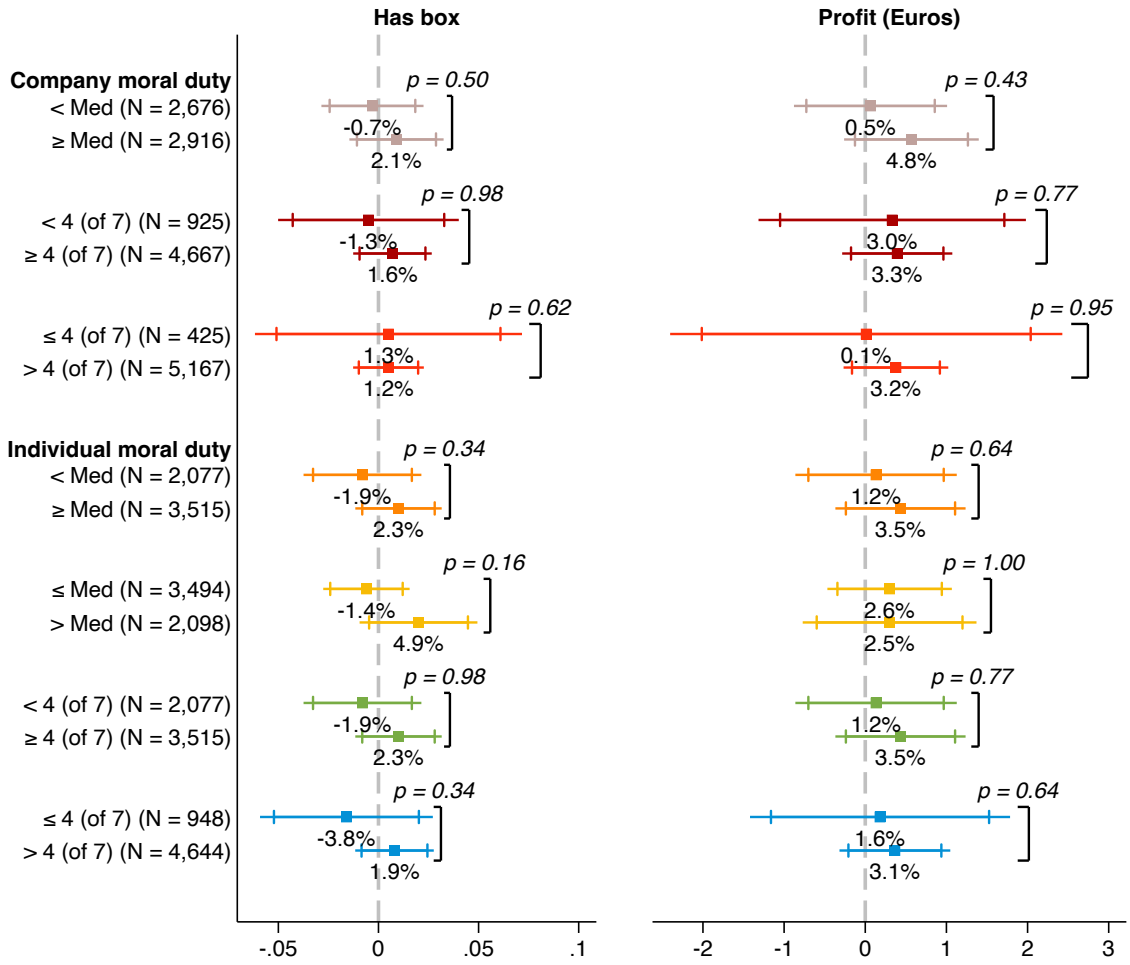
Note: This figure tests for heterogeneous treatment effects of the labels on meal-choice outcomes by alternate definitions of baseline carbon-footprint knowledge. All regressions use the same controls and sample definition as in Figure 3b. All knowledge definitions are based on participants' rankings of the emissions from producing 1kg of the following foods: beef, cheese, farmed shrimp, pork, chicken, and tomatoes (listed here in the correct order by average emissions). We show heterogeneous effects when defining correct baseline knowledge as correctly identifying beef as the highest-emissions food, identifying beef and tomato as the highest- and lowest-emissions foods, respectively (as in Figure 3b), correctly ordering beef, pork, chicken, and tomato, or both correctly ordering these foods and identifying beef and tomato as the absolute highest- and lowest-emitters in the list of 6 foods. Note that only 1% of customers get the entire scale correct. Capped and uncapped bars show 90% and 95% confidence intervals, respectively, with standard errors clustered by customer. The p-values correspond to tests of equal treatment effects between groups.

Figure A.8: Specification chart for retention and profit



Note: This figure parallels Appendix Figure A.2, except that we test the 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, measured in Euros.

Figure A.9: Robustness of retention and profit regressions to climate-morality definitions



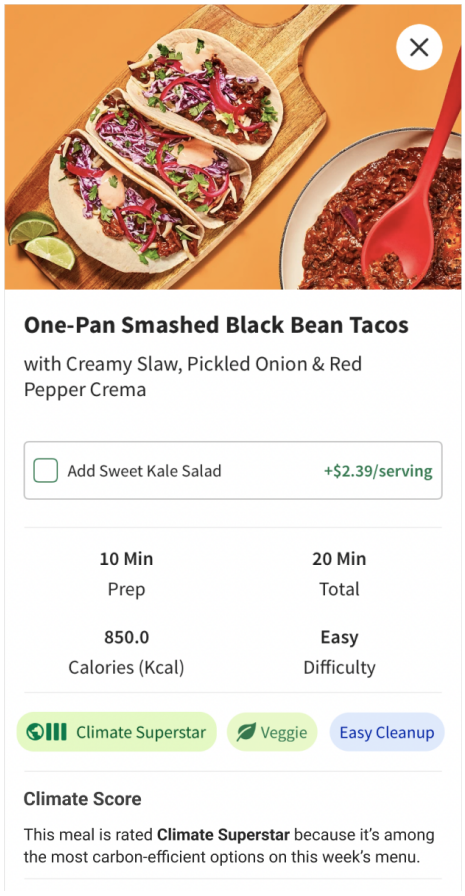
Note: This figure tests for heterogeneous treatment effects of the labels on profit and retention outcomes by alternate definitions of climate-morality beliefs, paralleling Appendix Figure A.5. Outcomes and control variables are defined as in Figure 4c. Capped and uncapped bars show 90% and 95% confidence intervals, respectively, with standard errors clustered by customer. The p-values correspond to tests of equal treatment effects between groups.

Figure A.10: HelloFresh Menu with All-Tiers labels

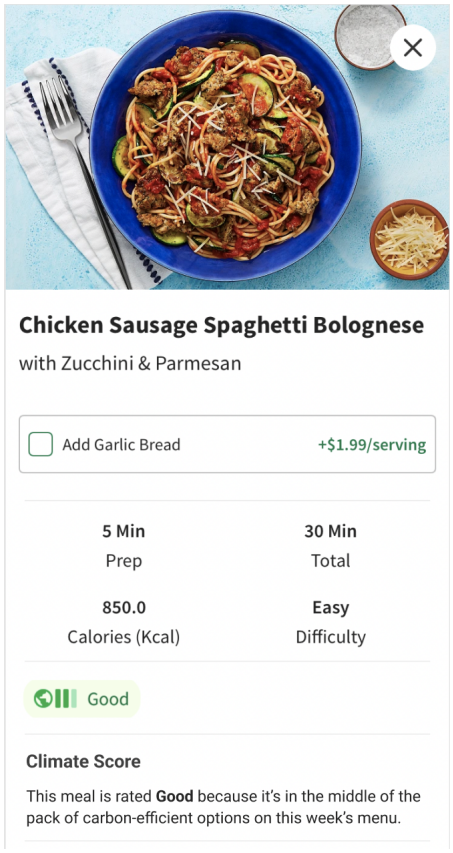
(a) Example row of menu landing page



(b) Menu card example 1

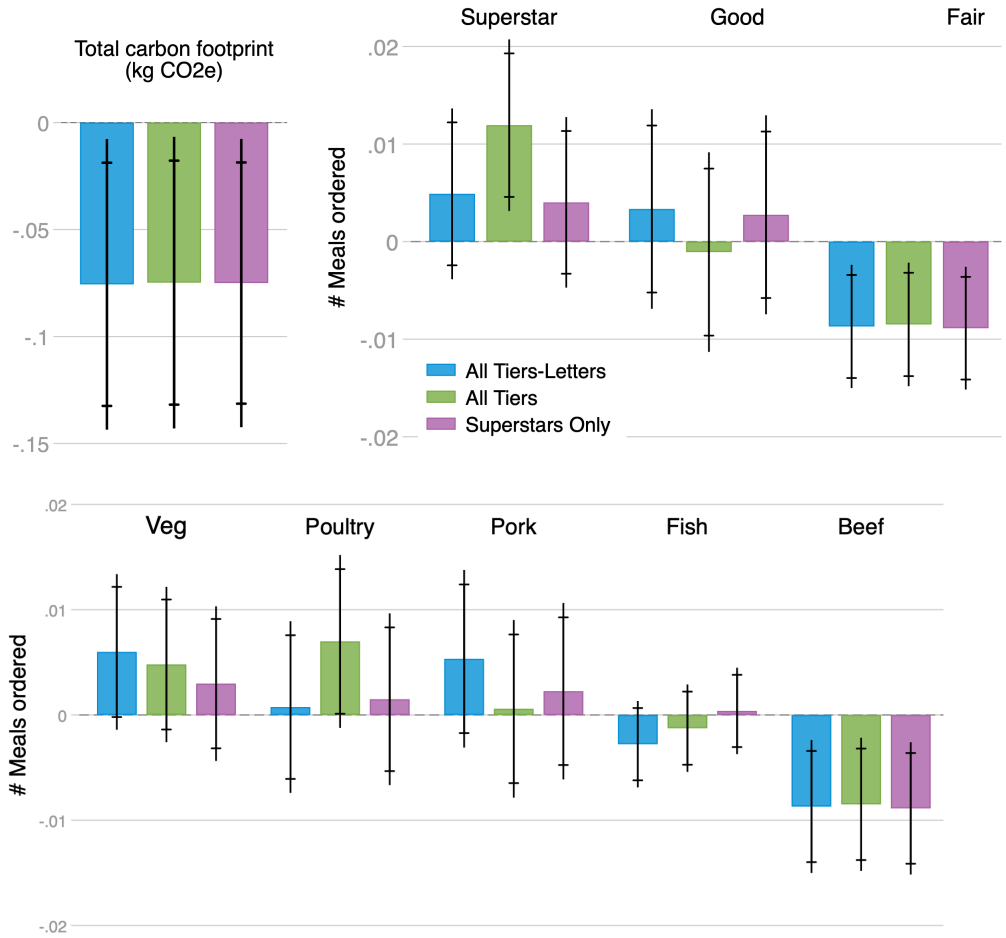


(c) Menu card example 2



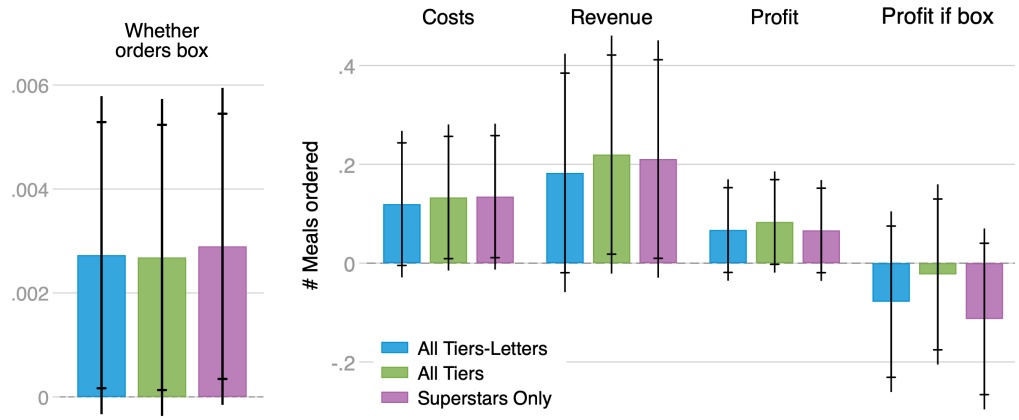
Note: This figure shows excerpts of sample HelloFresh menus, including the carbon labels from the All-Tiers variant (Figure 1b). Subfigure (a) shows an excerpt from a sample main menu, which includes photo entries for each of about 25-35 recipes per week. These small placards show recipe tags including 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 A.11: Meal-choice effects split by label type



Note: This figure plots the impacts of the three label variants (shown in Figure 1b) on meal-choice outcomes and emissions. Figures 2a-2c show the pooled effects of these label types on the same outcomes and details these outcomes in the table notes. The regressions are identical to those in Figures 2a-2c other than including separate variables for each label type. The capped and uncapped lines show 90% and 95% confidence intervals, respectively, with standard errors clustered by customer.

Figure A.12: Retention and profit effects split by label type



Note: This figure plots the impacts of the three label variants (shown in Figure 1b) on the customer retention and profit outcomes. Figures 4a and 4b show the pooled effects of these label types on the same outcomes and details these outcomes in the table notes. The regressions are identical to those in Figures 4a and 4b other than including separate variables for each label type. The capped and uncapped lines show 90% and 95% confidence intervals, respectively, with standard errors clustered by customer.

B Appendix Tables

Table B.1: Summary statistics and balance: Main experimental sample

	US	Control		Any Label			
	Mean	N	Mean	N	Mean	Δ Mean	*p-value*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Longevity at HelloFresh							
Had any pre-period meal		58,645	0.813	175,866	0.813	0.001	(0.591)
Week of first pre meal		47,671	11.930	143,059	11.975	0.045**	(0.046)
# Pre weeks with meals		58,645	4.449	175,866	4.429	-0.020	(0.305)
Total # pre-period meals		58,645	13.917	175,866	13.871	-0.046	(0.480)
Baseline meal choices							
Avg box carbon footprint		40,206	12.966	120,752	12.990	0.024	(0.504)
Share of baseline meals:							
Vegetarian		47,671	0.149	143,059	0.149	-0.001	(0.402)
Poultry		47,671	0.365	143,059	0.365	-0.000	(1.000)
Pork		47,671	0.232	143,059	0.233	0.001	(0.236)
Fish		47,671	0.052	143,059	0.053	0.000	(1.000)
Beef		47,671	0.216	143,059	0.217	0.001	(0.200)
Demographics by zipcode							
Share below US poverty line	0.128	57,280	0.108	171,804	0.108	0.000	(1.000)
Share adults 25+ with bachelor's	0.329	57,284	0.395	171,819	0.395	0.000	(1.000)
Shares by race or ethnicity:							
White, non-hispanic	0.601	57,286	0.670	171,820	0.669	-0.001	(0.353)
Black	0.126	57,286	0.101	171,820	0.101	-0.000	(1.000)
Hispanic	0.182	57,286	0.138	171,820	0.139	0.001	(0.175)
2020 Democratic vote share	0.513	57,474	0.530	172,327	0.530	0.000	(1.000)

Note: This table compares average characteristics of the study participants by treatment arm and to the average US population. Column 1 presents average characteristics for the US population. Columns 2 and 4 are the number of participants in the control group or label group (pooling those assigned to the three labeled treatment arms in Figure 1b) for whom we have information on the given characteristic, respectively. Columns 3 and 5 give the average value of each characteristic for the control and label treatment groups, respectively. Column 6 is the difference in average values for the characteristic between the control group and the labeled treatment arms, and column 7 presents the corresponding *p*-value for the test that this difference in means is significantly different from zero. The variables shown in the table are: (i) whether or not the participant ordered any HelloFresh meals during the pre-period before the start of the experiment (week 8 to 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 from HelloFresh during the pre-period, (v) the average carbon footprint of a HelloFresh box ordered by the participant in the pre-period for which we observe carbon footprints (weeks 17-20), (vi) the share of meals ordered during the pre-period that 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. We describe these traits in more detail in the Methods section.

Table B.2: Comparing baseline sample and main experimental sample

	In baseline sample?				
	No		Yes		
	N	Mean	N	Mean	Δ Means
	(1)	(2)	(3)	(4)	(5)
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	224,215	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 B.1. See the note to Appendix Table B.1 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 B.3: Baseline sample summary statistics and balance

	Control		Any Label			
	N	Mean	N	Mean	Δ Mean	*p-value*
	(1)	(2)	(3)	(4)	(5)	(6)
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 B.1, 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, a range of climate beliefs, and baseline carbon-footprint knowledge. We describe all of these measures in more detail in the Methods section.

Table B.4: Joint heterogeneity regressions in the baseline sample

	(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 heterogeneous treatment effects on key outcomes in the baseline-survey sample jointly across several dimensions of heterogeneity. Outcomes are defined as in Figures 2 and 4. These regressions include the control variables detailed for baseline-sample regressions in the Methods. This table jointly estimates heterogeneous treatment effects by climate-morality beliefs, baseline carbon footprint knowledge, and baseline beef consumption. 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. 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.

C Sample composition effects

In this Appendix section, we explore the role that differential retention by baseline customer traits may play in the treatment effects we observe on meal choices, which we can only estimate among the sample of customers who order a meal box in a given week. The effects of the labels on customer retention (Figure 4) could induce selection effects that shape these estimates.

In Appendix Table C.1, we first test for treatment-induced sample selection on baseline meal choices. In particular, we estimate the effect of the label treatments on the average baseline meal choices of customers who order a box in a given week. While remaining customers have slightly higher baseline consumption of beef meals and lower baseline consumption of poultry meals when we control only for week fixed effects, these gaps disappear when we condition on the other customer controls that we include in our main regressions for treatment effects on meal choices (Figure 2). Thus, selection on baseline average food choices does not drive the treatment effects of labels on meal choices that we observe.

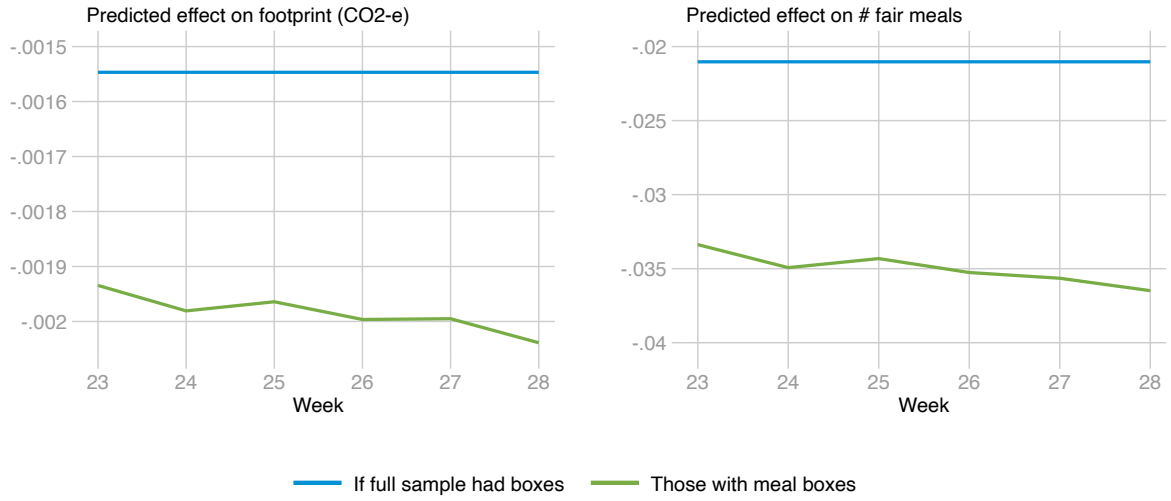
In Appendix Figure C.1, we then test for the role of selection on treatment effects. We estimate heterogeneous treatment effects in the full experimental sample by customers' baseline meal choices, longevity at HelloFresh, and predicted zip-code-level climate beliefs. Using these heterogeneous treatment effects, we then estimate predicted effects of the labels on meal choices for all customers in the experimental sample based on their baseline traits. Appendix Figure C.1 plots average predicted treatment effects among the full sample of customers and among those who ordered a meal box in a given week. These comparisons suggest that, if anything, differential customer retention due to the labels selects for customers with larger (more negative) treatment effects, and that we might over-estimate the environmental benefits of the carbon labels. Of course, we cannot rule out the possibility that the labels cause differential customer selection on unobservable determinants of differential label treatment effects that we do not capture here.

Table C.1: Compositional effects of retention

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Average # baseline meals among those with box									
	Veggie		Poultry		Pork		Fish		Beef	
Any label	-0.005	-0.003	-0.007*	-0.006	0.003	0.002	0.002	0.001	0.005*	0.003
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)
Cont. mean	0.498		1.147		0.742		0.175		0.652	
# Customers	118,889		118,889		118,889		118,889		118,889	
# Cust-weeks	366,657		366,657		366,657		366,657		366,657	
Week FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Other controls		✓		✓		✓		✓		✓

Note: This table estimates the role of differential customer attrition in shaping average meal choices over time during the experimental period. We calculate the average baseline number of meals with a given protein that a customer ordered each week in the pre-experimental period. For each week during the experimental period, we then define outcomes equal to these pre-experimental meal choices in each week in which a customer ordered a meal box; we set this outcome to missing in any week in which customers did not order a meal box. This lets us test whether the labeling treatment changed the average composition of customers by baseline meal choices, separately from treatment effects on customers' meal choices themselves. Columns 1, 3, 5, 7, and 9 regress baseline meal choices among remaining customers on the treatment indicator and fixed effects by week. Columns 2, 4, 6, 8, and 10 then add in all other controls included in our main regressions (e.g. Figure 2) other than controls for baseline meal choices. We cluster standard errors by customer. We indicate statistical significance at the 10%, 5%, and 1% levels by *, **, and ***, respectively.

Figure C.1: Predicted meal-choice effects under full vs. observed retention



Note: This figure plots point estimates for the predicted treatment effects of the labels on meal choices among the full set of customers in our sample versus among the set of customers who order a box in a given week of the experimental period. We generate these estimates by first estimating heterogeneous treatment effects by pre-experimental baseline beef consumption, customer longevity, and predicted zip-code climate beliefs in the full experimental sample. We split baseline beef consumption by those below the 25th percentile, between the 25th and 75th percentiles, and above the 75th percentiles. We split longevity by those with no pre-experimental meals, those who ordered meals in 1 through 6 pre-experimental weeks, and those who ordered meals in at least 7 pre-experimental weeks. We split predicted zip-code climate beliefs into terciles. Then, we use these heterogeneous effects to predict the treatment effect on number of *Fair* meals ordered and total carbon footprint for all customers in the experimental sample versus among the set of customers who ordered a meal box in a given week. Here, we plot the average predicted treatment effect among the full set of customers versus among those who ordered a box in a given week, for experimental weeks 23 through 28.