Willingness to Pay for Carbon Mitigation: Field Evidence from the Market for Carbon Offsets

Matthias Rodemeier

September 2024

Abstract

This paper estimates willingness to pay (WTP) for carbon mitigation from the demand for carbon offsets in a field experiment with an online supermarket (N=255,000). Initially, consumers are price-elastic but fully inelastic to the impact of offsets—consistent with warm glow utility. However, repeated exposure to varying impact levels makes subjects slightly impact-elastic, indicating that the initial insensitivity was partially driven by inexperience. Further, WTP increases when the firm shares the offsetting costs with consumers. This implies that firms can play an important role in encouraging consumers to lower their carbon footprint. Revealed WTP estimates range from 13–16 EUR/tCO_2.

JEL Codes: D61, D82, H21, Q50, Q58
Keywords: climate change, corporate sustainability, willingness to pay, carbon offsets, fairness preferences

---

Affiliations: Bocconi University, IGIER, BAFFI, and IZA. matthias.rodemeier@unibocconi.it.
Acknowledgments: I thank Nicola Gennaioli, Nicola Limodio, Florian Nagler, Davide Pace, Gregory Sun, Luigi Zingales, and participants at the NBER SI for insightful discussions about this paper. I also thank the editor, Tarun Ramadorai, and knowledgable referees for many helpful comments. Silvia Amalia Meneghesso, Andrea Calgaro, and Adamya Singh provided outstanding research assistance.
First version: December 2022. The paper was selected for the 2024 IZA Best Paper Award for Innovative Research on the Economics of Climate Change (IRECC).

1 Introduction

The market for voluntary carbon mitigation has doubled in size from 2017 to 2020 and is projected to reach USD 50 billion by 2030.[1] Much of this reduction comes from investments into "carbon offset" projects that engage in reforestation, which removes carbon dioxide (CO_2) from the air. Firms increasingly offer consumers the possibility to directly compensate the carbon emissions of their consumption, such as for flights or product shipping. This remarkable trend towards voluntary carbon offsetting raises an important question: What does demand for these offsets reveal about household's valuations of environmental protection?

Understanding this question is crucial for i) cost-benefit analyses of environmental policies and ii) the value of corporate social responsibility for firm fundamentals. First, in a standard economic framework, the larger a household's willingness to pay (WTP) for carbon mitigation, the larger her benefits from policies that reduce emissions.[2] For decades, environmental economists had to resort to *hypothetical* WTP measures from contingent valuation methods in surveys. More recently, economists have started to elicit stated preferences to understand demand for sustainable assets (e.g., Stroebel and Wurgler 2021, Giglio et al. 2023b). The growing market for voluntary climate protection may provide a unique opportunity to obtain first data points of households' *revealed* preferences (see also Pace et al. 2023).

Second, observational data has been used to understand how sustainable firm behavior and ESG rankings affect firm valuations (e.g, Hartzmark and Sussman 2019, Berg et al. 2023). Theoretical work has studied how firm stakeholders—such as consumers—

[1] See, https://www.mckinsey.com/capabilities/sustainability/our-insights/a-blueprint-for-scaling-voluntary-carbon-markets-to-meet-the-climate-challenge.

[2] A special case under which this statement is not true is if those people who voluntarily contribute to the climate dislike environmental regulation. In Section 4, I provide evidence against this hypothesis: people with a larger stated WTP are more likely to vote for a carbon tax.

can reduce externalities by either engaging with polluting firms or by boycotting them (Broccardo, Hart, and Zingales 2022).[3] However, there is little evidence as to how much sustainable corporate activities actually affect consumer behavior. This paper bridges the current empirical disconnect between sustainable firm investments and consumer markets.

To answer the main question of this paper, I partner with one of the largest online supermarkets in Germany and implement an experiment in their online shop, observing over 250,000 consumers and 400,000 visits to the website. In the experiment, consumers can offset the carbon emissions of grocery deliveries by buying carbon offsets. To estimate WTP, I exogenously vary both the price of the offset and the quantity of carbon compensated by the offset. Specifically, the baseline offset compensates the average emissions of a delivery: 2.4kg of CO_2 for a price of 24 cents (i.e., mitigating 1kg of carbon costs 10 cents). In order to vary price and quantity, either the price of the offset is *subsidized* by $x \in \{50\%, 75\%\}$ or the amount of carbon that the offset compensates is *matched* by $z \in \{100\%, 300\%\}$. For example, a consumer that receives a 75% price reduction only needs to pay 6 cents, and the firm covers the remaining 18 cents of the costs. A consumer that receives a 300% quantity match can offset 9.6kg for 24 cents (instead of just 2.4kg), and the firm covers the remaining 72 cents of the costs.

While subsidies and matches offer exogenous variation in price and impact, they also imply that the firm pays the difference between the cost of the offset and the price charged to the consumer. The increase in offset demand may then not just be driven by "intrinsic" preferences for carbon mitigation, but also by a preference to split the compensation costs with the firm. For instance, consumers might consider it fair if the firm contributes to the cost of the offset since it also benefits from the polluting transaction. Therefore, I cross-

[3] See also Green and Roth (2021) and Oehmke and Opp (2024) for theories on socially responsible investing.

randomize whether the firm informs the consumer that it shared the cost of the offset with the consumer. In the standard treatments, henceforth STANDARD, consumers who receive a subsidy or a match simply see a lower offset price or a higher offset quantity. In the information treatments, henceforth INFORMATION, the firm provides salient information to the consumer that the firm has financed the subsidy or match. This allows me to isolate the role of preferences to share the costs with the firm from the "intrinsic" valuation of carbon mitigation.

Household Preferences for Carbon Offsetting. The experiment produces a number of insights. First, in STANDARD, consumers increase demand for the offset when the price falls but are *completely* inelastic to increases in the compensated quantity. Even when the offset compensates 300% more carbon than the baseline offset, demand does not increase. These results suggest that consumers buy the carbon offset but not because of its impact on environmental protection. The conclusion is consistent with canonical theories of warm glow (Andreoni 1990) and "scope-insensitivity" (Kahneman and Knetsch 1992). However, exploiting *within-subject* variation in offsets, I find that consumers become slightly sensitive to impact when they are repeatedly exposed to offsets with varying degrees of impact. This indicates that subjects positively value the attribute "kilograms of carbon" when they can compare different quantities right after each other. This finding is supported by theories of "contrast effects" and memory-based reference points (Bordalo, Gennaioli, and Shleifer 2020), and may provide a new micro foundation for scope-insensitivity.

Second, in INFORMATION, where the firms' contribution is advertised, demand becomes both more price- and quantity-elastic. The effect of a price reduction on offsetting demand increases by up to 250% due to information provision. Doubling the impact of the offset increases demand by 11%, and quadrupling it increases demand by 22%.

This finding highlights that firms can play an important role in encouraging consumers to lower their carbon footprint by leveraging consumers' "equity preferences" (Fehr and Schmidt 1999) for a fair division of compensation costs between consumer and firm.[4]

The difference between STANDARD and INFORMATION delivers largely different conclusions about consumers' valuation of carbon mitigation. Using between-subject differences, I estimate that average WTP is zero in STANDARD but 16 EUR per ton of CO_2 ($p < 0.01$) in INFORMATION. I then estimate a structural model, using both between- and within-subject variation, in order to isolate the "intrinsic preferences" for carbon mitigation from other behavioral factors such as warm glow, inattention to scope, and equity preferences. This model yields my final estimate of intrinsic WTP for carbon mitigation of 13 EUR/tCO_2. This is far below the current estimate by climate scientists of the Social Cost of Carbon (SCC) of 185 USD/tCO_2 (Rennert et al. 2022), as well as the SCC of the Biden administration of 51 USD (Interagency Working Group 2021) that is used for cost-benefit analysis. This finding emphasizes the need for environmental regulation as the market for voluntary climate protection internalizes only a very small fraction of climate externalities.

Cost-Effectiveness of Sustainable Firm Practices: Subsidies vs. Matches. Comparing subsidies and matches, I analyze which intervention is the most cost-effective in reducing carbon emissions and find an arguably surprising result: *Quantity matches are always more cost-effective than subsidies even when matches do not affect demand*. This is because subsidies reduce the price for all consumers, but the only incremental increase in mitigation comes from *marginal* consumers—i.e., from those that buy the offset due

[4]Related studies have investigated the relative effectiveness of subsidies and matching mechanisms in increasing public good provision (e.g., Eckel and Grossman 2003, Kesternich, Löschel, and Römer 2016, Karlan and List 2007, Feldman 2010). Since my study varies the perceived contribution by the firm through an information treatment, it is the first to show that positive match elasticities are largely driven by a preference to share the costs with the firm rather than by intrinsic preferences for the public good.

to the lower price. By contrast, matches also cause *inframarginal* consumers to mitigate more carbon—i.e., those consumers that would have bought the offset even without the match. For subsidies to break even with matches, price elasticities would have to be substantially larger than they turn out to be empirically. The second result is that matches combined with INFORMATION have a "multiplier effect": Since they increase consumers' willingness to offset, every EUR spent by the firm on a match produces a larger reduction in total carbon emissions than if the firm used the same EUR to buy an offset directly.

Stated Versus Revealed Preferences. Much of the existing evidence on consumers and investor preferences comes from surveys in which respondents give hypothetical answers. Through a complementary survey with customers from the same online shop, I study how much hypothetical WTP deviates from revealed preferences in the field. The mean stated WTP in the survey is 238 EUR/tCO_2. This is 1,388% larger than even my largest estimate of 16 EUR/tCO_2 in the information treatment. This result cautions against taking widely-used survey responses at face value.

Contributions to the Literature Contemporaneous surveys and lab experiments in economics and finance have investigated why households demand sustainable goods and invest in "green" assets. Perhaps most closely related to my study, Pace et al. (2023) finds in a survey that WTP is very concave, but not fully flat, in emissions. Another survey experiment by Heeb et al. (2023) shows that investors prefer sustainable assets but are often (although not always) insensitive to the impact of the investment on the environment.[5] My study improves upon these important papers by offering the first evidence from a

[5]Earlier lab experiments measure people's preferences for retiring pollution permits but do not vary the impact of the permits (Löschel, Sturm, and Vogt 2013, Diederich and Goeschl 2011, Diederich 2013). This difference turns out to be pivotal in my setting: I show that ignoring the impact variation overstates true WTP for carbon mitigation *by a factor of 19* or more because it does not account for warm glow utility.

large-scale natural field experiment: real market participants make choices in their natural environment, not knowing they are being observed by a researcher. This may yield more accurate measures of agents' preferences in real-world markets (List 2007).

A related strand of the literature studies the role of nonstandard preferences and cognitive constraints for sustainable behavior (e.g., Andre et al. 2022, Imai et al. 2022, List et al. 2022, Rodemeier and Löschel 2024, Löschel, Rodemeier, and Werthschulte 2023, Tilling 2023). My findings illustrate that voluntary environmental contributions are partially driven by warm glow and a fairness preference to share the contribution costs with the firm. This emphasizes an important role of corporate social responsibility in promoting households' support for carbon mitigation.[6]

Finally, the paper ties into a large literature that elicits stated preferences for carbon mitigation and social discount rates in surveys (see Nemet and Johnson 2010 for an overview of the literature). While some studies report modest WTP values of 40 USD/tCO_2 (measured in 2020-USD), many studies imply large values between 100 and 350 USD/tCO_2. My revealed preference estimates are at least an order of magnitude smaller, but I obtain similarly inflated values for stated preferences in the complementary survey. This illustrates the importance of developing tools to mitigate hypothetical bias in surveys.

The rest of this paper is structured as follows. Section 2 presents the experimental design. Results are discussed in Section 3. In Section 4, I present insights from a complementary survey. Section 5 concludes.

[6]On a broader level, the paper connects to the literature in climate finance (Giglio, Kelly, and Stroebel 2021) that studies whether investors value sustainability (e.g.,Hartzmark and Sussman 2019, Bauer, Ruof, and Smeets 2021, Giglio et al. 2023a, Gormsen, Huber, and Oh 2023) and the role of beliefs about climate change and regulation (Stroebel and Wurgler 2021, Giglio et al. 2023b, Ramadorai and Zeni 2024). In Appendix C, I provide evidence that offering carbon offsets at checkout had no noticeable effects on measures of firm performance. However, since the offsets were only offered at checkout, the present results do not speak to alternative marketing techniques where offsets are saliently advertised at the outset of the website visit.

2 Experimental Design

The experiment takes place in the webshop of one of the largest delivery services for groceries and beverages in Germany.[7] When a subject visits the website, she gets randomized into one of 10 experimental groups with equal probability. A subject is identified based on her HTTP-cookie. The experimental design involves both between- and within-subject variation in treatment. On follow-up visits, subjects are randomized again into one of the 10 groups.

Figure 1 visualizes the experimental design. In the treatment groups, subjects can compensate carbon emissions by buying a carbon offset. The *baseline offset* compensates 2.4kg of CO_2 for a price of 24 Cents. In the other treatments, either the price of the offset is subsidized by $x \in \{50\%, 75\%\}$ or the amount of carbon that the offset compensates is matched by $z \in \{100\%, 300\%\}$.[8]

The experimental design intentionally features a symmetry between matches and subsidies. Both a 50% subsidy and a 100% match imply that the firm splits the total offset costs with the consumer 50:50. Analogously, the 75% subsidy and the 300% match imply a 25:75 split in costs between consumer and firm. This symmetry is useful because it holds the cost of mitigation per kg of carbon constant between a subsidy and its respective match. For instance, the carbon price for the consumer is 50 EUR/tCO_2 both when the firm subsidizes the offset by 50% and when it matches the quantity by 100%.

I also vary whether the firm advertises its own contribution to the carbon offset through an information treatment. I elaborate on this treatment further below.

Finally, after a subject makes a purchase, she is forwarded to a page that confirms the

[7] The experiment was implemented during February 2020.

[8] Since these amounts are small, income effects and liquidity constraints are unlikely to affect demand for offsets. However, I should note that once we scale WTP for a *kilogram* of carbon to WTP for a *ton* of carbon, these factors may become important and are not considered in this paper. See Berkouwer and Dean (2022) for an example where limited access to credit reduced the demand for energy-efficient cook stoves.

order and, in addition, asks her two questions about carbon offsetting.

2.1 Treatments

Figure 2 provides a screenshot of the baseline offset, henceforth "BASELINE." The offset is always displayed in the shopping basket of the shop, next to the list of products the subject has selected. Subjects get to that page either because they want to verify which goods they put into the shopping basket, or to finalize the purchase.

The offset can be added to the shopping basket by ticking the respective box next to the text *"Yes, I would like to support environmental protection and offset 2.4kg of* CO_2 *for 24 Cents."* The text below informs subjects to which carbon-offsetting project the amount is donated.[9] In addition, subjects are informed that 2.4kg of CO_2 correspond to the average emissions of one delivery.[10] This gives a reference point to consumers and helps them relate deliveries to carbon emissions. While the provided information may still be relatively abstract to consumers, we closely followed other shops when designing this treatment to replicate the typical carbon offset product in the market.

The donation goes to a reforestation project that plants trees to compensate for carbon emissions. At the time of the experiment, it cost 0.10 EUR to compensate one kg of CO_2 (i.e., 100 EUR/tCO_2). Thus, one average delivery that emits 2.4kg can be compensated by 0.24 EUR.

Examples of the price and quantity variations are shown in Figure 3. Panel a) shows the simple price reduction of the offset by 50%. Subjects in this group pay 12 Cents for 2.4kg of carbon instead of 24 Cents. The rest of the text is identical to the baseline offset.

Panel b) shows the INFORMATION treatment where the firm explicitly informs the consumer that the firm has subsidized the price by 12 Cents. The additional information

[9] The project name is not mentioned in this paper to protect the company's anonymity.
[10] Average emissions were calculated from historical trip data.

provides two potentially important differences relative to STANDARD. The first difference is that the consumer learns that the firm is contributing its own resources to the offsetting project and shares the burden of compensation with the consumer. This might be considered fairer by consumers and, thereby, increase demand elasticities.

Second, the information may change attention to the offset and beliefs about the offset's effectiveness. The lower price in STANDARD relative to BASELINE may signal to consumers that the offset project is of low quality and not effective at compensating carbon. A low offset price might also signal that the environmental damage of a delivery is negligible since it costs little to compensate for it. By contrast, INFORMATION should avoid this negative signal of low prices because subjects should be aware that the actual price of the offset is higher than the costs they have to cover. In addition, consumers might trust the offset project more if they learn that the company donates its own resources to the project.

Panel c) shows an example of a quantity match. The price is equal to the one of the baseline offset, i.e., 24 Cents. However, the quantity is doubled from 2.4kg to 4.8kg of CO_2. Therefore, this treatment provides exogenous variation in the *impact* of the offset. Consumers are still informed that an average delivery produces 2.4kg, such that they have the same reference point as in BASELINE. This should help them realize that they compensate 2 instead of 1 delivery in expectation. In general, note that any exogenous change in quantities implies, by definition, that the compensation amount deviates from the emissions of the average subject. However, this is precisely the required variation in order to identify WTP for the compensated amount of carbon.

Panel d) shows the corresponding quantity match in INFORMATION. Subjects receiving the salient quantity match are informed that the full compensation price for 4.8kg of CO_2 is 48 Cents. The reason they are paying half of the amount is that the company pays the remaining 24 Cents.

The role of the outside option for identification. Even if consumers do not choose to offset carbon emissions in the experiment, they may still reduce their carbon footprint through alternative measures outside of the webshop. This could include buying offsets on other platforms or avoiding other emission-intensive activities.[11] Such behavior could be a problem for the identification of WTP if we made the mistake of interpreting a consumer's probability to offset as the reduced-form analog to her willingness to pay. For instance, we could falsely assume that consumers with a low offset probability have a lower willingness to pay than those with a high offset probability, even though the former group might choose to offset much more carbon outside of the web shop.

My experimental design is robust to these misinterpretations and identifies WTP for carbon mitigation despite the fact that consumers have individual-specific outside options. As I explain further below, I identify WTP by the (absolute) ratio of the aggregate quantity and price elasticity. These elasticities are unambiguously identified in my setting because the treatment assignment is, by randomization, orthogonal to both subjects' preferences and their individual outside options. The fact that consumers may choose to reduce their carbon footprint in other contexts is consequently no threat to identification in our experiment. A more formal version of this argument is presented in Section 3.4, where I estimate WTP and explicitly allow for any arbitrary outside option.[12]

[11] It should also be emphasized that the experimental design is specific to carbon offsetting as an *add-on service* offered by firms. While this practice is widespread in real-world consumer markets, the experimental design and findings may not directly be portable to settings where carbon emissions are directly embedded in products or financial assets, such as in Heeb et al. (2023).

[12] Relative to the previous literature, the method employed in this paper differs in the way it identifies WTP. Heeb et al. (2023) and Pace et al. (2023) use the Becker–DeGroot–Marschak mechanism where subjects make multiple choices between options with varying degrees of carbon emissions, and one choice is randomly implemented at the end. This method, while widely accepted, is not applicable in natural field experiments or observational data. To overcome this challenge, I randomize prices and carbon emissions and then use this variation to estimate a random utility model in the spirit of McFadden et al. (1973) and Hanemann (1984). Understanding how different elicitation methods across studies affect WTP would be an important avenue for future research.

2.2 Post-Purchase Survey

If a subject has placed a delivery, she gets forwarded to the order-confirmation page, where she is asked two questions (see Figure A1 in the Appendix). The first question elicits subjects' belief about the environmental damage of a delivery if the emissions are not compensated:

"How large do you think are the negative consequences of your delivery for the environment if the carbon emissions of the delivery are not compensated?"

Possible answers are presented on a 7-point Likert scale from 1 ("very low") to 7 ("very high"). The idea behind the question is that consumers might interpret a low offset price as a signal that the environmental damage of a delivery is low because it costs little to compensate a delivery.

The second question elicits beliefs about the effectiveness of the offset:

"How effective do you think our carbon offset program is in reducing these negative consequences?"

Possible answers are presented on a scale from 1 ("not helpful at all") to 7 ("very helpful").

This question is intended to test i) whether subjects interpret a low price as a signal of low effectiveness of the offset, and ii) whether effectiveness beliefs increase as the compensated quantity increases.

Due to technical reasons, subjects using a mobile device are not forwarded to these questions after placing an order. In addition, subjects in the control group who are not offered carbon offsets cannot answer the two survey questions because they have not been offered the offset previously.

2.3 Sample

I observe 406,984 website visits by 255,376 subjects. These subjects place a total of 108,478 orders during the experimental period. Table 1 reports summary statistics for the 10 experimental groups. Here, a subject's treatment group is defined as the one she has been assigned to during her first visit during the experimental period. Each of the 10 experimental groups consists of approximately 25,000 subjects. The balance in the number of subjects across treatments provides support for successful randomization.

The expected travel time of a delivery van is around 14 minutes across groups. The expected service time refers to the time the driver is expected to need in order to unload the delivery van. This number is larger for orders with a larger number of goods or more bulky products. Expected service time is approximately 7 minutes and balanced across experimental groups.

In the control group, the purchase probability is around 27%, and the average subject visits the website 1.6 times during the experimental period. Both of these numbers are roughly the same for the treatment groups. Note that these differences do not need to balance because they are potentially endogenous to the treatment variation.

The reported offsetting probabilities are conditional on placing an order. For the control group, the offsetting probability is zero by construction. In the other groups, the offsetting probability is positive and varies substantially across treatments.

In Appendix C, I show that the treatments have no effect on the probability of buying at the store. Therefore, differences in offsetting demand across treatments have a causal interpretation since the treatments do not induce selection from the sample of website visitors into the subsample of buyers.

3 Results

3.1 Between-Subject Effects on Offsetting Behavior

Figure 4 presents between-subject effects on the offsetting probability among the subsample of buyers. I condition on subjects' first visit to the website during the experimental period since this is their first exposure to an offset treatment. The grey bars indicate the offsetting probabilities for standard price and quantity variations, as well as for the baseline offset. The transparent bars show offsetting probabilities for the salient price and quantity variations. Figure 4b shows the cost effectiveness of each intervention, which I discuss later in Section 3.5.

At the baseline price of 24 cents for 2.4kg, 13.5% of customers choose to buy the offset. If the offset price falls by 12 and 18 cents, the offsetting probability increases by 0.8 and 2.2 percentage points, respectively. This implies a convex demand curve with price elasticities of -0.12 and -0.31. In other words, the incremental demand response to a given price reduction is larger at lower price levels. The larger price reduction is statistically significant at conventional levels.

We observe an even more pronounced pattern for quantity variations. Increasing the amount of carbon compensated by the offset does not increase demand in STANDARD. The offsetting probabilities are even slightly lower than baseline when quantities increase but these differences are not statistically different from zero. Using these results to identify elasticities would imply that consumers are completely inelastic to compensated quantities. Even when the compensated quantity is increased by 7.2kg, which is a large relative increase of 300% relative to BASELINE, the offsetting probability does not change. Taking these point estimates at face value yields the conclusion that consumers buy carbon offsets but not because of how much carbon they offset. WTP for voluntary carbon mitigation is zero. This conclusion is in line with models of "warm glow" (Andreoni 1990) in which people receive binary utility from the act of giving but do not care about the impact of their donation. Similarly, the results are in line with Kahneman and Knetsch (1992)'s finding that people's *hypothetical* willingness to pay for a public good is "insensitive to scope" in hypothetical choices (e.g., rescuing a bird vs. rescuing an entire species). However, results may also suggest that consumers have an imperfect understanding of the product they are buying and do not understand kilograms of carbon as a measure of impact. In addition, consumers may simply be inattentive to the impact of the offset.

Offsetting behavior changes substantially when consumers are explicitly informed about the firm contribution. Demand becomes more price-elastic and consumers suddenly become sensitive to scope. The price reductions now increase demand by 2.8 and 5 percentage points, respectively. Both effects are highly statistically significant with $p < 0.01$. Put differently, making the price variations salient increases its effects by 250% and 127% for the 12 and 18 Cent reductions. The price elasticities are now -0.53 and -0.84. In INFORMATION, increasing the compensated carbon by 2.4kg and 7.2kg raises the offsetting probabilities by 1.5 and 3 percentage points (both at $p < 0.01$). These responses are relative treatment effects of 11% and 22% compared to baseline. The point estimates imply quantity elasticities of 0.22 and 0.19. Consequently, in the presence of salient matches, consumers exhibit a responsiveness to the impact of the offset. Note, however, that these responses are still relatively small, considering that the firm doubled and quadrupled the amount of compensated carbon.

An open question is whether the increase in the impact elasticity in INFORMATION is due to consumers' recognition of the greater impact of the offset, or whether it stems from their appreciation of the firm's contribution to the offset. I explore these aspects next.

3.2 Effect of Information on Beliefs

Figure 5 illustrates differences across treatments in consumer perceptions elicited in the post-purchase survey. Looking at Panel a), we do not find statistically significant differences between STANDARD and INFORMATION in the perceived environmental damage of an uncompensated delivery. However, a general tendency seems to be that the perceived damage decreases as the costs for the consumer fall. Surprisingly, this is also true when consumers know that the true offset price is higher than what they pay. Overall effects are relatively noisy and small in absolute size.

Panel b) plots the perceived effectiveness of an offset across treatments. While it is hard to draw stark conclusions from the figure, a couple of tendencies emerge. Perhaps most importantly, in STANDARD, the perceived effectiveness of the offset barely increases as the quantity of mitigated carbon increases. This implies that consumers may not understand quantity increases. Information seems to reduce this misperception for the 100% match as perceived effectiveness increases by 6% (0.2 points on the Likert scale).[13] The effect of information for the much larger 300% match is qualitatively similar but not statistically significant. These results are not fully conclusive but point to the possibility that the previously observed insensitivity to scope in STANDARD may partially be explained by consumers' inattention towards the compensation amount.

[13]These results are relevant to an early model in philanthropy by Vesterlund (2003) arguing that information about a fundraiser's own contribution to the charity increases donors' perceived quality of the charity.

3.3 Within-Subject Effects and Learning about Impact

To further shed light on underlying mechanisms, I study how treatment effects vary within- as opposed to between-subject. One potential mechanism by which subjects are impact-inelastic is that they may have a difficult time understanding kilograms of carbon as a measure of impact. However, subjects that return to the shop have a possibility to learn about this measure as they see offsets with varying compensation amounts. A particular challenge in studying within-subject variation is that we do not know whether subjects who did not buy anything at a given point in time saw the offset and might have familiarized themselves with it.[14] To adjust for this possibility, I now consider the entire sample of website visitors in the analysis instead of conditioning on the subsample of buyers. I assign a zero to the outcome variable (offsetting) whenever subjects i) bought at the shop but did not choose the offset, or ii) did not buy at the shop. The resulting analysis represents within-subject Intent-to-Treat (ITT) effects.[15]

To increase precision, I do not compare all 9 treatments but run a regression that pools across the two subsidies and across the two matches. I further generate an indicator $PriorExposure_{it} \in \{0, 1\}$ that equals one if the subject has been treated with an offset during the prior visit. Interacting this indicator with the match treatment allows me to study whether being exposed to an offset previously changes the impact elasticity. Specifically, I estimate the following panel regression:

$$\begin{aligned}\text{Offset}_{it} &= \eta_1 Price_{it} + \eta_2 \times Price_{it} \times I_{it} + \beta_1 Impact_{it} \\ &+ \beta_2 \times Impact_{it} \times I_{it} + \beta_3 \times Impact_{it} \times PriorExposure_{it} + \lambda_i + \mu_t + \epsilon_{it} \quad (1)\end{aligned}$$

where $\text{Offset}_{it} \in \{0, 1\}$ is one if subject i chose the offset at time t, and zero oth-

[14] In the previous between-subject analysis, this was not an issue because we considered only subjects' first possible exposure with the offset, i.e., we conditioned on their first visit during the experimental period.

[15] These provide a lower bound of the Treatment Effect on the Treated (TOT). See Appendix C.1 for a formal argument.

erwise. Price and impact variation are captured by $Price_{it} \in \{6, 12, 24\}$ Cent and $Impact_{it} \in \{2.4, 4.8, 9.6\}$kg of mitigated carbon. The interactions with information $I_{it} \in \{0, 1\}$ measure how price and quantity elasticities change with information. The coefficient β_3 reflects how the elasticity of impact changes when a subject was exposed to another offset during their previous visit. I include subject-fixed effects, λ_i, as well as time-specific effects μ_t.[16] Therefore, the corresponding treatment effects constitute behavioral changes *within-subject* rather than between-subject. Residuals are ϵ_{it}.

Table 2 reports resulting coefficients in column 2. To allow for comparability, column 1 shows between-subject coefficients for the first visit of each subject in the data. At the bottom of the table, I also include the offsetting probability for the BASELINE offset. Comparing columns 1 and 2 illustrates that this probability is lower for the panel data (2.47% as opposed to 4.5%). Further, within-subject variation in subsidies increases demand by around 0.4 percentage points (pp), which is slightly lower than the between-subject effect of 0.5 in column 1. INFORMATION increases the price elasticity by, on average, 0.3pp within-subject, but 0.8pp between-subject. The INFORMATION effect for matches is also lower for within- than between-subject variation (0.6 versus 0.8pp).

If subjects had no prior exposure to offsets, the impact elasticity remains negative and statistically indifferent from zero in the panel data. However, prior exposure to an offset increases the impact elasticity by 0.6pp ($p < 0.01$). A 1kg increase in impact now increases the offsetting probability in STANDARD by 0.3pp, or 12% relative to control. These effects are still not economically large considering that matches doubled or quadrupled the mitigated impact. But they indicate a fundamental change in how subjects understand the impact of the offsets. Repeated interactions with varying levels of carbon emissions may have helped subjects more effectively compare the impact across

[16] Time fixed effects control for the day of the visit as well as for the number of visits. The latter effect is absorbed by constructing a variable that, for each observation, indicates how many website visits have been done by the subject.

different offsets.

By contrast, the effect of INFORMATION about a match is unaffected by whether subjects have been exposed to an offset before. If anything, the effect may be negative, indicating that, over time, INFORMATION may lose its importance in making people more impact-elastic. In column 3, I include an additional interaction term in the regression that equals one whenever subjects were in the INFORMATION treatment during their prior visit: "Match $\times$ previous exposure to info." This allows me to test whether the increase in the impact elasticity is due to i) prior exposure to any offset, or due to ii) the INFORMATION treatment itself. I estimate a small and insignificant interaction term of -0.03 for "Match $\times$ previous exposure to info", while the term "Match $\times$ previous exposure to offset" remains unchanged and highly significant. This implies that it is not INFORMATION that helps consumers learn about impact but rather repeated exposure to varying compensation amounts.

Column 4 tests whether the same patterns can be observed for subsidies. Since subjects are familiar with the attribute "price" we do not expect an experience effect on the price elasticity. In line with this interpretation, prior exposure to offsets neither changes the price nor the INFORMATION elasticity for subsidies. Both coefficients are relatively small and insignificant.

Since the previous effects only represent ITT effects, column 4 conditions on the subsample of 9,112 buyers for which repeated purchases can be observed. Since these subjects made several purchases, they must have gone through checkout multiple times. I do not include interaction terms for prior treatment exposure since prior exposure can happen during visits in which the subject does not make a purchase. I find that the within-subject effect of matches among buyers amounts to a large increase in offsetting of 1.8pp. This effect is even larger than the INFORMATION coefficients in the same column, and not far away from the Subsidy coefficient. Since the sample is much smaller, coefficients are generally less precisely estimated. Yet, the Match coefficient has a p-value of $p = 0.054$, placing it just at the edge of statistical significance at conventional levels.

In conclusion, the within-subject evidence points at two forces that are at play: i) the impact of INFORMATION diminishes over time, ii) subjects learn about differences in impact and start appreciating it even in STANDARD. The second result is consistent with models about contrast effects regarding complex and unfamiliar attributes. For instance, Bordalo, Gennaioli, and Shleifer (2020) hypothesize that experiences with an attribute in the past generate an anchor for valuations in the future. Interestingly, contrast effects are also consistent with one of the robustness tests in the study by Heeb et al. (2023) on investor preferences for impact-investing. They found evidence that investors become *slightly* elastic to the impact of their investment when the varying impacts are presented side-by-side, which creates a contrast and allows for comparability. The robustness of this finding across studies and different settings (consumption versus investing) points to a deep psychological process that deserves further investigation. Specifically, the observation that cognitive mechanisms mute the response to quantities may provide a new micro foundation for scope-insensitivity. The interpretation obviously matters for policy and welfare: If consumers are insensitive to scope due to cognitive limitations, then their choices may not reflect true preferences. This challenges more traditional models in which consumers do not respond to scope because they do not receive any utility from impact and instead only receive warm glow utility from donating (Andreoni 1990). We will find additional evidence in favor of cognitive mechanisms in the complementary email survey in Section 4.

3.4 Structural Model

In this section, I organize the previous findings through a unified framework. I use this framework to quantify the relative importance for carbon offsetting of i) intrinsic WTP for carbon mitigation, ii) warm glow from donating, iii) inattention to impact, and iv) preferences for a fair division of costs between consumer and firm.

As a baseline model, I use the usual random utility model (McFadden et al. 1973), which has also been extensively used in contingent valuation studies for public goods (Hanemann 1984). I then add behavioral extensions to this model such as inattention to scope and contrast effects as in Bordalo, Gennaioli, and Shleifer (2020), as well as fairness preferences about the division of compensation costs, following influential work by Fehr and Schmidt (1999).[17]

The baseline model Consumer $i \in \{1, 2, ...I\}$ can choose between buying a carbon offset and an outside option, where utility from the outside option is normalized to zero. The carbon offset compensates γ_i units of carbon. The consumer pays price $p_i = \rho - s_i$ where ρ is the market price of the offset and s_i the price reduction sponsored by the firm. I make the usual assumption that p_i and γ_i enter linearly into utility.[18] I let ω measure the warm glow utility from donating which is independent of the donation's effectiveness (Andreoni 1990). Idiosyncratic preferences are given by ϵ_i and follow a smooth distribution function G. Utility in the benchmark model is then given by:

$$u_i = \omega + \beta\gamma_i + \eta p_i + \epsilon_i. \tag{2}$$

The consumer buys the offset iff $u_i \geq 0$. Aggregate demand for the offset is $D = 1 - G(\xi)$ with $\xi = -\beta\gamma - \eta p - \omega$. I do not impose a distributional assumption on ϵ_i and rather

[17] Bushong, Rabin, and Schwartzstein (2021) provide another important theory of contrast effects. Their model is not directly applicable to my setting as they model static rather than sequential choices.

[18] In an unreported regression, I allow for nonlinearities in the utility function and cannot reject that they are statistically zero. See also Appendix B.

linearly *approximate* WTP by reduced-form elasticities (see Chetty 2009).[19] Specifically, the derivatives of aggregate demand with respect to price and carbon quantity are $\frac{\partial D}{\partial p} = \eta g(\xi)$ and $\frac{\partial D}{\partial \gamma} = \beta g(\xi)$, such that WTP is given by $WTP = -\frac{\beta}{\eta} = -\frac{\partial D}{\partial \gamma} / \frac{\partial D}{\partial p}$. Note that this approach makes the assumption that the consumer believes that the offset in fact compensates the promised amount. In Appendix B, I relax this assumption by introducing subjective uncertainty over the offset's effectiveness (e.g., due to distrust in offsets) and by allowing for risk aversion. I find that the main estimates derived in this section are robust to a reasonable range of subjective probabilities that are in line with survey responses.[20]

Finally, the money-metric utility gain from warm glow is given by $-\frac{\omega}{\eta}$. We can estimate the normalized utility parameters through a linear model of the form:

$$D_i = \alpha + \tau\gamma_i + \zeta p_i + \xi_i \tag{3}$$

$$\approx -\frac{\omega}{\eta}\zeta - \frac{\beta}{\eta}\zeta\gamma_i + \zeta p_i + \xi_i. \tag{4}$$

This baseline model is estimated by ordinary least squares.

Incorporating inattention and fairness preferences Motivated by our previous findings, I extend the model by considering the possibility that subjects pay attention to impact only with probability $\theta \in [0, 1]$. Any $\theta < 1$ mutes the unbiased offsetting elasticity with respect to γ_i.

Further, I allow consumers to have equity concerns $F(p_i, s_i, I_i) = \phi \max\{s_i I_i - p_i, 0\} + \rho \max\{p_i - s_i I_i, 0\}$ about how costs should be divided between the firm and the consumer. The parameterization of $F(p_i, s_i, I_i)$ follows Fehr and Schmidt (1999) but,

[19] The traditional approach is to assume a logit distribution for ϵ_i and estimate the model by maximum likelihood. If demand is low, as is the case in the experiment, a logit distribution imposes too much probability mass on negative values, implying that a large share of consumers would get negative utility from buying the offset.

[20] See Calel et al. (2021) for empirical evidence on adverse selection problems in the offsetting market.

in addition, accommodates our empirical setting where the firm's contribution is either salient $I_i = 1$ or shrouded $I_i = 0$.[21] If the firm's contribution is salient and exceeds the consumer price, $s > p_i$, the consumer gets marginal utility ϕ for every EUR by which the firm's subsidy exceeds the consumer price. Conversely, for every EUR by the firm for which a salient subsidy is below the consumer price, $s_i < p_i$, the consumer gets marginal utility ρ. Both ϕ and ρ may be positive or negative, either representing a gain or a loss from an unequal division of costs. For equal splits, $p_i = s$, the consumer gets no fairness utility, $F(p_i, p_i, 1) = 0$. If the firm's contribution is shrouded, we have $F(p_i, s_i, 0) = \rho p_i$.

The final specification that incorporates these behavioral extensions is

$$u_i = \omega + \theta\beta\gamma_i + \eta p_i + F(p_i, s_i, I_i) + \epsilon_i. \tag{5}$$

The utility parameters can again be approximated by a linear model:

$$D_i = \alpha + \tau\gamma_i + \zeta p_i + \psi \max\{s_i I_i - p_i, 0\} + \varphi \max\{p_i - s_i I_i, 0\} + \xi_i \tag{6}$$

$$= -\frac{\omega}{\eta}\zeta - \theta\frac{\beta}{\eta}\zeta\gamma_i + \zeta p_i - \frac{\phi}{\eta}\zeta \max\{s_i I_i - p_i, 0\} - \frac{\rho}{\eta}\zeta \max\{p_i - s_i I_i, 0\} + \xi_i \tag{7}$$

The term $-\frac{\rho}{\eta}$ is of particular policy interest: It measures the consumer's willingness to donate (beyond their baseline donation) for every EUR donated by the firm. Put differently, $-\frac{\rho}{\eta}$ is a multiplier effect of sustainable firm engagement: Every EUR invested by the firm causes an additional donation of $-\frac{\rho}{\eta}$ EUR by consumers (conditional on the firm paying weakly more than the consumer).

Identification and Estimation For identification, I leverage both between- and within-subject variation. I assume that subjects are inattentive to impact when it is varied between-subject, $\theta = 0$, but attentive in within-subject variation, $\theta = 1$, because the impact of the previously seen offset serves as an anchor that helps subjects to value the

[21] Another difference is that I hypothesize that consumers have fairness preferences over the division of donation costs, while Fehr and Schmidt (1999) study a subject's preference over the division of money for own consumption.

attribute (Bordalo, Gennaioli, and Shleifer 2020).[22] Put differently, β is identified only through within-subject variation.

Warm glow utility, ω, is the constant in the utility function measuring utility from offsetting that is unrelated to the impact γ. The fairness term $\max\{s_i I_i - p_i, 0\}$ is identified from treatments in which the firm's contribution (either through subsidies or matches) is salient, i.e. from INFORMATION groups. The second term, $\max\{p_i - s_i I_i, 0\}$ is identified from price variation in STANDARD where the consumer perceives to pay more than the firm.[23]

I estimate utility parameters jointly from demand moments between-subject, D_i^b, and within-subject variation, D_{it}^w, using the following two equations:

$$D_i^b = \zeta p_i - \underbrace{\frac{\omega}{\eta}\zeta}_{\text{warm glow}} \underbrace{-\frac{\phi}{\eta}\zeta \max\{s_i I_i - p_i, 0\} - \frac{\rho}{\eta}\zeta \max\{p_i - s_i I_i, 0\}}_{\text{equity concerns}} + \xi_i^b \tag{8}$$

$$D_{it}^w = \left(\zeta p_i - \underbrace{\frac{\beta}{\eta}\zeta}_{\text{WTP}} \gamma_i \underbrace{-\frac{\phi}{\eta}\zeta \max\{s_{it} I_i - p_{it}, 0\} - \frac{\rho}{\eta}\zeta \max\{p_{it} - s_{it} I_{it}, 0\}}_{\text{equity concerns}} \right) \underbrace{\Omega}_{\text{attenuation effect}} + \underbrace{\lambda_i + \mu_t}_{\text{fixed effects}} + \xi_{it}^w \tag{9}$$

The important differences between the two equations are that i) only D_i^w is used to estimate WTP, $\frac{\beta}{\eta}$, ii) D_i^w includes subject- and time-fixed effects, λ_i and μ_i, which means that $\frac{\beta}{\eta}$ represents a within-subject treatment effect, iii) D_i^w does not include the term $\frac{\omega}{\eta}$

[22] If this assumption is violated because consumers are still not fully attentive to within-subject variation, my estimates represent an upper bound of θ.

[23] Note that the term is always zero in INFORMATION because, in the experiment, whenever the firm provides information, the subsidy is at least as large as the price paid by the consumer ($s_i \geq p_i$ if $I_i = 1$). Since the experimental design only varies information in which the firm saliently pays more than the consumer, one might argue that consumers in STANDARD never think about the fact that they pay more than the firm. In this case, a better model would be one in which $F = \max\{s_i I_i - p_i, 0\}$, so the term $\max\{p_i - s_i I_i, 0\}$ is dropped from the estimation. In an unreported estimation, I find that this alternative specification yields almost identical results.

because constant terms are absorbed by the fixed effects, and iv) D_i^w controls for a general attenuation in treatment effects over time by multiplying all coefficients by Ω.

Point iv) deserves additional explanation. As we have seen in Table 2, coefficients *in general* tend to become smaller over time. This decrease over time may not be due to a change in subjects' underlying deep primitives. Instead, attenuation could be caused by other idiosyncratic factors that are unrelated to economic fundamentals. For example, the INFORMATION treatment may be more salient to consumers on the first visit than on follow-up visits to the shop (see, e.g., Allcott and Rogers 2014). The parameter Ω absorbs the average attenuation effect across treatments.[24]

I jointly estimate equations 8 and 9 by the Generalized Method of Moments (GMM) and find the optimal weight matrix using the two-step GMM estimator.

Results Estimation results are shown in Table 3. In the baseline specification, I estimate the model for STANDARD and INFORMATION separately to illustrate the difference in WTP estimates. Subjects with the baseline offset are included in both estimations.

As shown in column 1, using the variation in STANDARD to estimate utility parameters, we find that WTP for carbon mitigation is indistinguishable from zero, suggesting consumers do not value the carbon-mitigating attribute of the offset. Instead, warm glow utility is highly statistically significant and amounts to, on average, 1.27 EUR ($p < 0.01$). Column 2 uses variation in INFORMATION and yields a WTP estimate of 16 EUR/tCO_2 ($p < 0.01$). Warm glow utility amounts to 0.74 EUR ($p < 0.01$) per offset. The empirical moments in STANDARD and INFORMATION, therefore, deliver vastly different implications for why consumers buy carbon offsets: In one case consumers seem to be entirely driven by warm glow, while in the other case they exhibit strong intrinsic preferences for

[24]This adjustment is not crucial to the estimation results. An alternative estimation without allowing for an dynamic attenuation effect (i.e., forcing $\Omega = 1$) yields a WTP of 12.92EUR/tCO_2, which is virtually identical to the estimate in Table 3.

the effectiveness of the offset.[25]

Column 3 suggests that inattention to the impact of the offset might have muted consumers' true WTP in STANDARD. Identifying WTP from our within-subject variation, we obtain a WTP of 12.84 EUR/tCO_2 ($p < 0.05$). Warm glow now amounts to 0.86 EUR per offset ($p < 0.01$), which is somewhere between the estimate in STANDARD and INFORMATION. In terms of fairness preferences, the coefficient of $-\frac{\phi}{\eta}$ is close to zero and statistically insignificant, suggesting that subjects do not get direct disutility from paying more than the firm. Instead, there is strong evidence that consumers appreciate when the firm pays weakly more. The coefficient $-\frac{\rho}{\eta}$ implies that every EUR donated by the firm above the consumer's contribution increases willingness to pay for the offset by 0.21 EUR ($p < 0.01$). This is a sizeable multiplier effect of sustainable firm investment.

Finally, we observe a significant dynamic attenuation of treatment effects in general. The parameter $\Omega = 0.84$ implies that within-subject elasticities are 84% of between-subject elasticities. This result further emphasizes the uniqueness of the quantity elasticity which is the only one that *increases* for the within-subject variation.

In conclusion, the results indicate that i) scope-insensitivity in STANDARD is at least partially driven by inattention to impact, and ii) the increased responsiveness to scope in INFORMATION is partially driven by consumers' equity concerns. Point i) is broadly in line with the results discussed in Section 3.2 where perceived effectiveness did not change between-subject as we varied the impact. Point ii) suggests an important role of corporate sustainable engagement in promoting voluntary contributions to climate protection among households.

[25] The reason warm glow utility is lower in Column 2 is that INFORMATION increases the price elasticity (as reported in the second-to-last column), which in turn implies that the utility value of the regression constant α becomes smaller.

3.5 Cost-Effectiveness of Sustainable Firm Practices: Subsidies vs. Matches.

What is the cost-effectiveness profile of subsidies and quantity matches, and how does information change this profile? This question is not just important for policy makers but also for stakeholders who seek to maximize the impact of sustainable firm investments.

To quantify cost-effectiveness, I calculate the difference in compensated carbon between an intervention (subsidy or match) and the baseline offset. I then divide this number by the total monetary contributions made by the firm on that intervention. We can interpret this number as the incremental increase in compensated carbon of the intervention per EUR spent by the firm.

Panel B in Figure 4 visualizes the results, again conditioning on the first interaction that each buyer had with a treatment during the experiment. The dotted gray line marks the market price if the firm directly buys the offset instead of offering it to consumers (i.e., the baseline price of 10kg/EUR). Perhaps surprisingly, quantity matches are always more cost-effective than subsidies, *even when matches have no impact on demand*. The reason for this stark result is that with subsidies, the only incremental increase in compensated carbon comes from marginal consumers. By contrast, with matches, the increase in compensated carbon also comes from inframarginal consumers since every offset now compensates a larger amount. Price elasticities would have to be much larger for subsidies to be more cost-effective than quantity matches.

In terms of magnitudes, we see that subsidies in STANDARD increase the compensated quantity by approximately 1kg/EUR and 2kg/EUR per invested EUR for the 12 and 18 cent subsidies, respectively. Only the latter is statistically significant from zero. INFORMATION, instead, increases the benefit-cost ratio of both subsidies substantially. The cost-effectiveness ratio becomes 3.40kg/EUR and 3.60kg/EUR, respectively. The ratio is always below the market price of 10kg/EUR. This means the firm could offset more carbon if they used the money spent on subsidies and purchased carbon offsets directly instead.

By contrast, quantity matches in STANDARD just break-even with the market price of 10kg/EUR and are thereby more than 2.5 times more cost-effective than subsidies. The quantity matches in INFORMATION are able to offset more carbon per EUR spent, implying that matches can have a multiplier effect. In particular, every EUR spent by the firm compensates around 11kg of carbon, i.e., 10% more than if the same EUR were invested directly into the baseline carbon offset.

This result suggests that firms may leverage consumer preferences for corporate social responsibility to more efficiently invest into offset projects.

4 Revealed Versus Stated Household Preferences

To further understand consumers' preferences, I implement a second survey several months after the field experiment. Customers receive an email from the company inviting them to take an opinion survey. The survey investigates how stated preferences for carbon mitigation respond to changes in the impact of carbon offsets, to an education treatment about carbon offsetting, as well as to the firm's contribution to the offset. It also sheds light on people's preferences for a carbon tax as an alternative protective policy.[26] A translated version of the survey can be found in Appendix H.

Survey Design In order to elicit subjects' stated preferences, they receive two questions that elicit their hypothetical WTP. First, they are asked how much they are willing to pay to compensate $x \in \{2.4, 4.8\} kg$ of CO_2, where the amount they see is randomly

[26] For privacy reasons, I cannot match survey participants to the observations in the field experiment.

assigned. Directly after that, they are asked how much they would be willing to pay to compensate a higher amount $y \in \{4.8, 9.6\}kg$ of CO_2. Subjects who saw 2.4kg in the first question, see 4.8kg in the second. Analogously, subjects who first saw 4.8kg, next see 9.6kg. This creates both within- and between-subject variation in the compensation amount and allows me i) to estimate the distribution of stated WTP, and ii) to test if subjects are inattentive to scope between- and within-subject.

In addition, I randomize a treatment in which subjects receive additional information in the second question on WTP that the firm *matches the compensation amount on its own cost* to $Y \in \{4.8, 9.6\}$kg of CO_2. This treatment allows us to investigate the effect on stated WTP of a quantity match by the company.

Finally, I investigate whether education about carbon offsetting affects WTP. I randomize a treatment in which subjects see three stylized facts about carbon emissions before answering the WTP questions. Treatment subjects are informed i) that an average delivery emits 2.4kg of CO_2 (as in the field experiment), ii) that one would have to drive 11km in an average car to emit the same amount of carbon as the delivery, iii) that one would have to plant 5 beech trees, on average, to compensate 2,000 deliveries. Subjects are then randomly asked about one of these facts in a follow-up question to test their understanding.

In Appendix G, I describe the sample in more detail and discuss observable characteristics. Subjects are more likely to be male, slightly younger than the average German citizen, and less likely to be unemployed. While I cannot exclude that subjects select on unobservables into the survey, observable statistics are fairly representative of the firm's customer population. They also match the typical profile of a US customer that shops groceries online, according to market research data (Capital One 2024).

Results Table 4 reports results from an OLS regression of WTP on the treatments. As is common in the literature that measures WTP with open-ended questions, I adjust for outliers by only considering the 90th percentile of WTP answers.[27] Column 1 is stated WTP in Cents. The constant implies that subjects in the first question state a WTP of 57 Cents. This translates into 238 EUR/tCO_2 as reported at the bottom of the table. The estimate falls into the range of prior estimates from contingent valuation studies (e.g., Hersch and Viscusi 2006, Viscusi and Zeckhauser 2006, Nemet and Johnson 2010, Brouwer, Brander, and Van Beukering 2008, Nemet and Johnson 2010 Carlsson et al. 2012, Achtnicht 2012): numbers range from 40 to 350 USD/tCO_2 (in 2020-USD). Overall, the stated preference approach used in the survey does not capture the revealed preference estimate from the experiment. If we were to take 16 EUR/tCO_2 as our preferred estimate, the survey results would overstate WTP by 1,388%.[28]

Stated preferences do not significantly change when consumers receive the information that the firm contributes to the offset. One coefficient is even marginally significantly negative, although this is not a robust finding as other coefficients are positive. In a follow-up question, subjects were asked what share of the carbon compensation costs of the delivery should be paid by the firm. Possible answers were between 0% and 100%. Figure 6a illustrates that the modal consumer thinks the company should pay half the compensation costs, indicating that consumers do value the firm's contribution positively.

The education treatment generally has positive coefficients, although none of them is

[27] More specifically, I use the 90th percentile of WTP *per t*CO_2. It is important to normalize in this context as subjects have been offered different compensation amounts. If we do not exclude outliers, stated WTP estimates become more inflated due to some unreasonably large extreme values.

[28] A limitation is that I do not observe which customers answered the survey because participation was fully anonymous. However, even if there is systematic selection into the survey, the results provide an important insight: A survey with stated preferences yields estimates 11 times larger than estimates from a field experiment with the entire customer base that makes actual consumption choices. Whether this is driven by hypothetical bias or selection, we can conclude that the survey yields inflated estimates for the sample of interest.

statistically significant. This suggests a limited role of pedagogic information provision for WTP in line with prior studies (Imai et al. 2022, Pace et al. 2023).[29]

There is no statistically significant effect of raising the compensation amount by 2.4kg of CO_2 *between-subject*. This again implies that consumers are fully quantity-inelastic even for hypothetical choices, in line with the seminal result by Kahneman and Knetsch (1992). However, WTP increases by 65% (+32 Cents) when the compensation amount is raised *within-subject*. This is true for both the increase to 4.8kg and to 9.6kg (both $p < 0.01$). Thus, consumers again become quantity-elastic when they realize that the compensation amount is larger. Another interesting observation is that even in the within-subject design, consumers are scope-insensitive *in differences*: the effect of the quantity increase seems to be the same for 4.8kg as for 9.6kg. While this could point to extreme concavity in the WTP function, it is likely another symptom of the same cognitive manifestation related to contrast effects. The stated values support the findings from the field experiment that consumers do not seem to be able to compare magnitudes unless they are presented right after each other. This behavioral phenomenon seems to be the one where the field and survey data agree the most with each other. The robustness of this result calls for further research to develop methods that allow us to estimate environmental preferences while accounting for inattention to scope.

Finally, I investigate how preferences for voluntary climate protection relate to political support for a carbon tax. At the end of the survey, subjects were asked whether they would support a carbon tax. 33% of subjects oppose a carbon tax, while 67% endorse it. Subjects' political preference for carbon taxation is a strong predictor of hypothetical

[29]To complement this result, Appendix D shows subjects' answers to the belief questions and suggests that, without the education treatment, subjects overestimate the carbon emissions of the average delivery, the equivalent kilometers that one needs to drive with a conventional car, and the number of trees necessary to compensate for 2,000 deliveries. The education treatment reduces the average overestimation for the last two questions. Consequently, subjects realize that it takes less to compensate for a delivery than they thought, which may explain the positive coefficients on WTP.

WTP. Figure 6b plots the empirical distribution of WTP in EUR/ton of CO_2 for supporters and opponents of the tax. I exclude values above the 90th percentile to adjust for outliers and increase the readability of the graph. Around 55% of subjects who oppose a carbon tax have a WTP below 20 EUR/tCO_2, while 32% have a WTP of zero. By contrast, only 20% of carbon tax supporters have a WTP below 20 EUR/tCO_2 and 6% a WTP of zero. The modal opponent of a carbon tax has a stated WTP of zero, while the modal supporter has a stated WTP of around 208 EUR/tCO_2. Overall, the probability distribution is shifted to the right for supporters relative to opponents of the tax. This suggests that hypothetical WTP—while overstating true WTP—still has strong predictive power regarding stated political preferences for environmental policies.

5 Conclusion

What does the market for voluntary climate protection imply about people's environmental preferences? How can this market be leveraged by firms? This paper investigates these questions with a large-scale natural field experiment to estimate how demand for carbon offsets responds to exogenous variations in subsidies and matches by the firm.

I find that consumers are elastic to price but fully inelastic to between-subject variations in impact. At first sight, this result indicates that consumers buy the offset but do not value the carbon it mitigates. Using additional within-subject variation, I find that this scope-insensitivity may not be driven by indifference but rather by inattention to impact. A simple but powerful intervention that advertises the firm's participation in the offset makes subjects sensitive to impact and implies a WTP of 16 EUR/tCO_2. This underscores that firms can play an important role in encouraging consumers to lower their carbon footprint.

The paper further shows how firms can mitigate carbon cost-effectively. I show that demand is too price-inelastic for subsidies to be cost-effective. Inelastic demand implies that subsidies induce transfers to all inframarginal consumers but do not induce a sufficiently large increase in mitigation. By contrast, matches induce sufficient increases in mitigation because even inframarginal consumers mitigate more. The takeaway for firms and shareholders is that cost-effective carbon offsetting likely involves matches rather than subsidies.

Finally, stated preferences from a complementary survey heavily diverge from revealed preferences in the experiment. Additional tests of scope-insensitivity point to models of contrast effects and memory-based anchoring as new and unexplored mechanisms. The development of techniques aimed at obtaining agents' environmental valuations in the presence of behavioral models is an important avenue for future research.

References

Achtnicht, Martin. 2012. "German car buyers' willingness to pay to reduce CO2 emissions." *Climatic change* 113 (3): 679–697.

Alan, Sule, Mehmet Cemalcilar, Dean Karlan, and Jonathan Zinman. 2018. "Unshrouding: Evidence from bank overdrafts in Turkey." *The Journal of Finance* 73 (2): 481–522.

Allcott, Hunt, and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *The American Economic Review* 104 (10): 3003–3037.

Andre, Peter, Teodora Boneva, Felix Chopra, and Armin Falk. 2022. "Misperceived Social Norms and Willingness to Act Against Climate Change." *Econtribute Discuss. Pap*.

Andreoni, James. 1990. "Impure Altruism and Donations to Public-Goods: A Theory of Warm-Glow Giving." *Economic Journal* 100 (401): 464–477.

Bauer, Rob, Tobias Ruof, and Paul Smeets. 2021. "Get real! Individuals prefer more sustainable investments." *The Review of Financial Studies* 34 (8): 3976–4043.

Berg, Florian, Julian F Koelbel, Anna Pavlova, and Roberto Rigobon. 2023. "ESG confusion and stock returns: Tackling the problem of noise." Technical Report, National Bureau of Economic Research.

Berkouwer, Susanna B, and Joshua T Dean. 2022. "Credit, attention, and externalities in the adoption of energy efficient technologies by low-income households." *American Economic Review* 112 (10): 3291–3330.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2020. "Memory, Attention, and Choice." *The Quarterly Journal of Economics* 135 (3): 1399–1442.

Boston Consulting Group. 2023. The Next Frontier in Carbon Credits: Consumers.

Broccardo, Eleonora, Oliver Hart, and Luigi Zingales. 2022. "Exit versus voice." *Journal of Political Economy* 130 (12): 3101–3145.

Brouwer, Roy, Luke Brander, and Pieter Van Beukering. 2008. "A convenient truth: air travel passengers' willingness to pay to offset their CO2 emissions." *Climatic change* 90 (3): 299–313.

Bushong, Benjamin, Matthew Rabin, and Joshua Schwartzstein. 2021. "A model of relative thinking." *The Review of Economic Studies* 88 (1): 162–191.

Calel, Raphael, Jonathan Colmer, Antoine Dechezleprêtre, and Matthieu Glachant. 2021. "Do carbon offsets offset carbon?" Working Paper.

Capital One. 2024. Online Grocery Shopping Statistics. URL: https://capitaloneshopping.com/research/online-grocery-shopping-statistics/.

Carlsson, Fredrik, Mitesh Kataria, Alan Krupnick, Elina Lampi, Åsa Löfgren, Ping Qin, Susie Chung, and Thomas Sterner. 2012. “Paying for mitigation: A multiple country study.” *Land Economics* 88 (2): 326–340.

Chetty, Raj. 2009. “Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods.” *Annual Review of Economics* 1 (1): 451–488.

Clever Carbon. 2024. URL: https://clevercarbon.io/carbon-footprint-of-common-items/ (Retrieved: 08/27/2024).

CO2 Everything. 2024. URL: https://www.co2everything.com (Retrieved: 08/27/2024).

Diederich, Johannes, and Timo Goeschl. 2011. “Willingness to pay for individual greenhouse gas emissions reductions: evidence from a large field experiment.” Discussion paper series 517, University of Heidelberg, Department of Economics, Heidelberg.

Diederich, Johannes Heiko. 2013. “Determinants of the Voluntary Provision of Public Goods: Experimental Investigations.” Ph.D. diss., Ruprecht Karl University of Heidelberg.

Eckel, Catherine C, and Philip J Grossman. 2003. “Rebate versus matching: does how we subsidize charitable contributions matter?” *Journal of Public Economics* 87 (3-4): 681–701.

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2022. “The preference survey module: A validated instrument for measuring risk, time, and social preferences.” *Management Science*.

Fehr, E, and Klaus M Schmidt. 1999. “A Theory Of Fairness, Competition, and Cooperation.” *Quarterly Journal of Economics* 114 (3): 817–868.

Feldman, Naomi E. 2010. “Time is money: Choosing between charitable activities.” *American Economic Journal: Economic Policy* 2 (1): 103–130.

Giglio, Stefano, Bryan Kelly, and Johannes Stroebel. 2021. “Climate finance.” *Annual Review of Financial Economics* 13:15–36.

Giglio, Stefano, Theresa Kuchler, Johannes Stroebel, and Xuran Zeng. 2023a. “Biodiversity risk.” Technical Report, National Bureau of Economic Research.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, Zhenhao Tan, Stephen Utkus, and Xiao Xu. 2023b. “Four facts about ESG beliefs and investor portfolios.” Technical Report, National Bureau of Economic Research.

Gormsen, Niels Joachim, Kilian Huber, and Sangmin Oh. 2023. “Climate capitalists.” *Available at SSRN*.

Green, Daniel, and Benjamin Roth. 2021. “The allocation of socially responsible capital.” *Available at SSRN 3737772*.

Hanemann, W Michael. 1984. “Welfare evaluations in contingent valuation experiments with discrete responses.” *American journal of agricultural economics* 66 (3): 332–341.

Hartzmark, Samuel M, and Abigail B Sussman. 2019. “Do investors value sustainability? A natural experiment examining ranking and fund flows.” *The Journal of Finance* 74 (6): 2789–2837.

Heeb, Florian, Julian F Kölbel, Falko Paetzold, and Stefan Zeisberger. 2023. “Do investors care about impact?” *The Review of Financial Studies* 36 (5): 1737–1787.

Hersch, Joni, and W Kip Viscusi. 2006. “The generational divide in support for environmental policies: European evidence.” *Climatic Change* 77 (1): 121–136.

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

Interagency Working Group. 2021. "Technical Support Document: Social Cost of Carbon, Methane, and Nitrous Oxide Interim Estimates under Executive Order 13990." Technical Report, Interagency Working Group on Social Cost of Greenhouse Gases, United States Government.

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

Karlan, Dean, and John A List. 2007. "Does price matter in charitable giving? Evidence from a large-scale natural field experiment." *American Economic Review* 97 (5): 1774–1793.

Karlan, Dean, and Jonathan Zinman. 2009. "Observing unobservables: Identifying information asymmetries with a consumer credit field experiment." *Econometrica* 77 (6): 1993–2008.

Kesternich, Martin, Andreas Löschel, and Daniel Römer. 2016. "The long-term impact of matching and rebate subsidies when public goods are impure: Field experimental evidence from the carbon offsetting market." *Journal of Public Economics* 137:70–78.

List, John, Matthias Rodemeier, Sutanuka Roy, Gregory Sun, et al. 2022. "Judging Nudging: Understanding of the Welfare Effects of Nudges Versus Taxes." NBER Working Paper.

List, John A. 2007. "Field experiments: a bridge between lab and naturally occurring data." *The BE Journal of Economic Analysis & Policy* 6 (2): 0000102202153806371747.

Löschel, Andreas, Matthias Rodemeier, and Madeline Werthschulte. 2023. "Can self-set goals encourage resource conservation? Field experimental evidence from a smartphone app." *European Economic Review* 160:104612.

Löschel, Andreas, Bodo Sturm, and Carsten Vogt. 2013. "The demand for climate protection—Empirical evidence from Germany." *Economics Letters* 118 (3): 415–418.

McFadden, Daniel, et al. 1973. "Conditional logit analysis of qualitative choice behavior."

Nemet, Gregory F, and Evan Johnson. 2010. "Willingness to pay for climate policy: a review of estimates." no. 2010-011.

Oehmke, Martin, and Marcus M Opp. 2024. "A theory of socially responsible investment." *Review of Economic Studies*, p. rdae048.

Pace, Davide, Taisuke Imai, Peter Schwardmann, and Joel van der Weele. 2023. "Uncertainty about Carbon Impact and the Willingness to Avoid CO2 Emissions." Working paper.

Ramadorai, Tarun, and Federica Zeni. 2024. "Climate regulation and emissions abatement: Theory and evidence from firms' disclosures." *Management Science*.

Rennert, Kevin, Frank Errickson, Brian C Prest, Lisa Rennels, Richard G Newell, William Pizer, Cora Kingdon, Jordan Wingenroth, Roger Cooke, Bryan Parthum, et al. 2022. "Comprehensive evidence implies a higher social cost of CO2." *Nature* 610 (7933): 687–692.

Rodemeier, Matthias. 2024. "Buy Baits and Consumer Sophistication: Field Evidence from Instant Rebates." *American Economic Journal: Economic Policy*.

Rodemeier, Matthias, and Andreas Löschel. 2024. "Information Nudges, Subsidies, and Crowding Out of Attention: Field Evidence from Energy Efficiency Investments." *Journal of the European Economic Association*.

Stroebel, Johannes, and Jeffrey Wurgler. 2021. What do you think about climate finance?

Tappwater. 2024. URL: https://tappwater.co/blogs/blog/carbon-footprint-bottled-water (Retrieved: 08/27/2024).

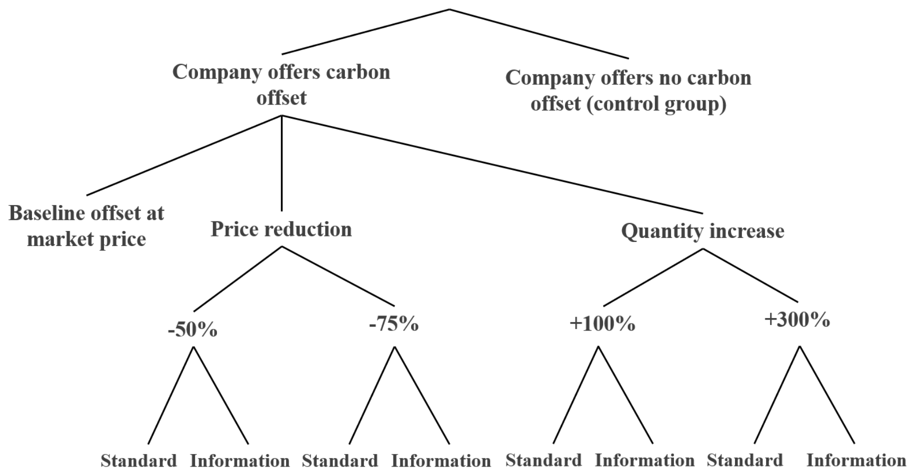
Tilling, Anna Schulze. 2023. "Changing consumption behavior with carbon labels: Causal evidence on behavioral channels and effectiveness."

Vesterlund, Lise. 2003. "The informational value of sequential fundraising." *Journal of public Economics* 87 (3-4): 627–657.

Viscusi, W Kip, and Richard J Zeckhauser. 2006. "The perception and valuation of the risks of climate change: a rational and behavioral blend." *Climatic change* 77 (1): 151–177.

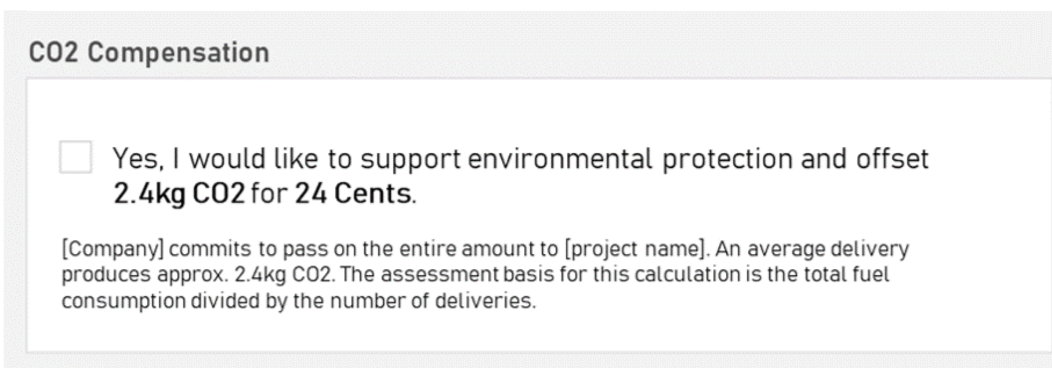
Figures

Figure 1: Experimental Design



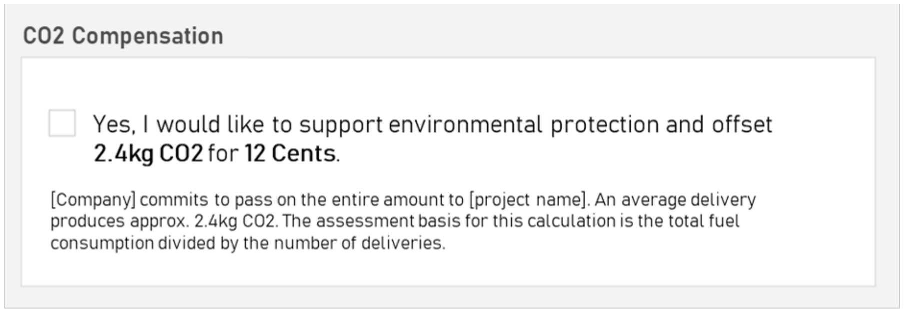
Note: This figure illustrates the experimental design. Subjects are randomized into one of ten groups with equal probability upon visiting the website.

Figure 2: Carbon Offset

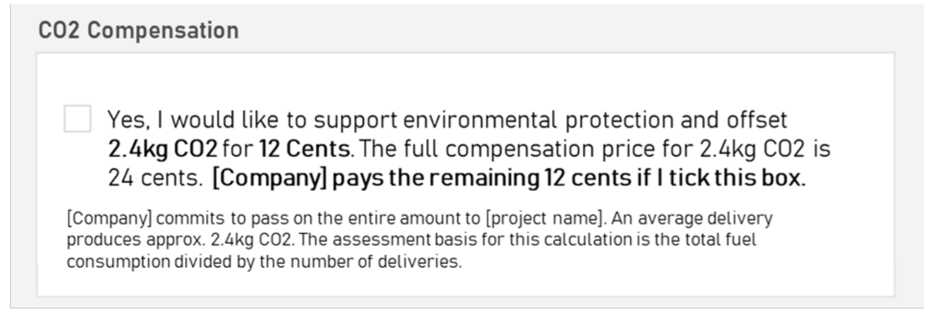


Note: This figure shows the baseline offset.

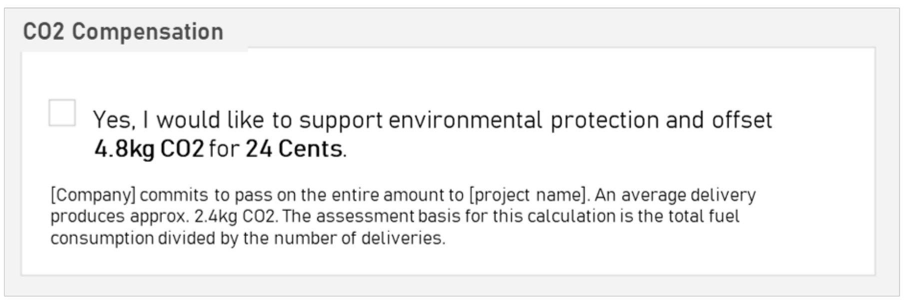
Figure 3: Examples of Treatment Variation



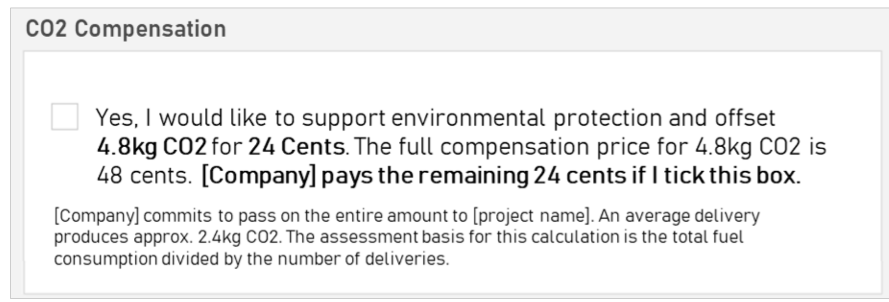
a) Price reduction by 50%



b) Price reduction by 50%, with salient information



c) Quantity increase by 100%

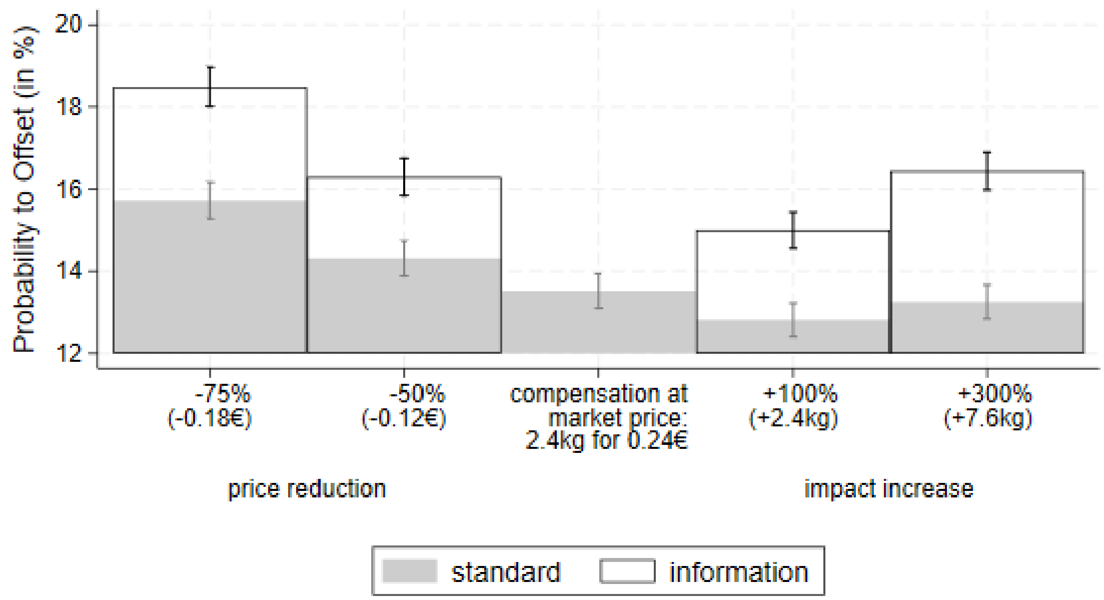


d) Quantity increase by 100%, with salient information

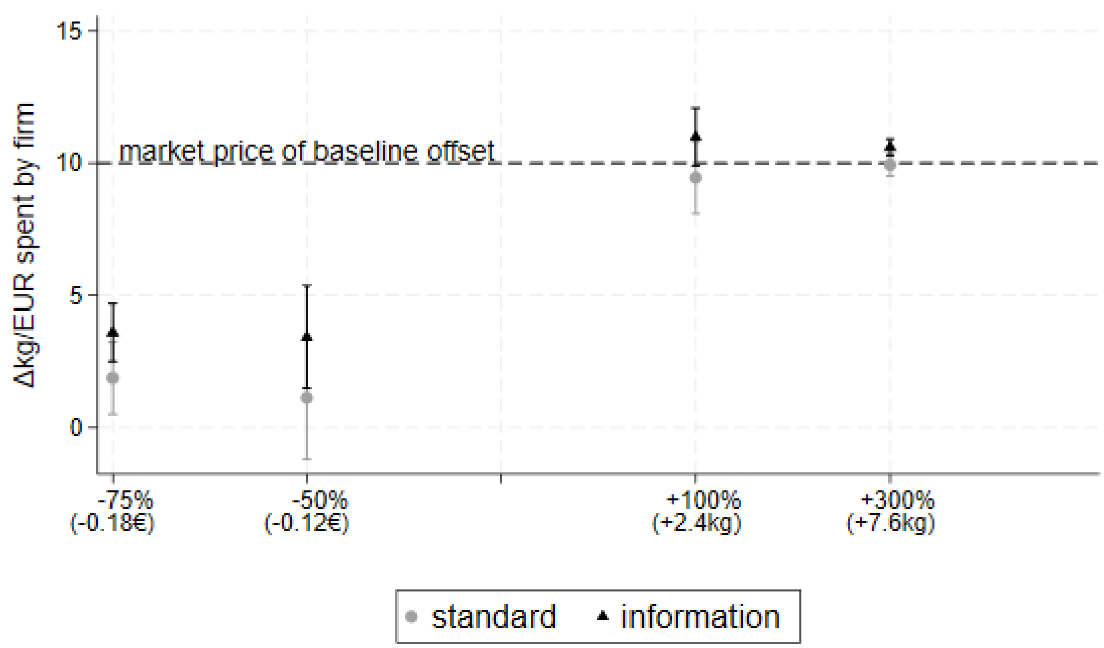
Note: This figure shows examples of price and quantity variations. Panel a) and c) illustrate the variations in the standard treatments, whereas panel b) and d) illustrate the variations in the information treatments.

Figure 4: Main Results

(a) Offsetting Probabilities



(b) Cost Effectiveness

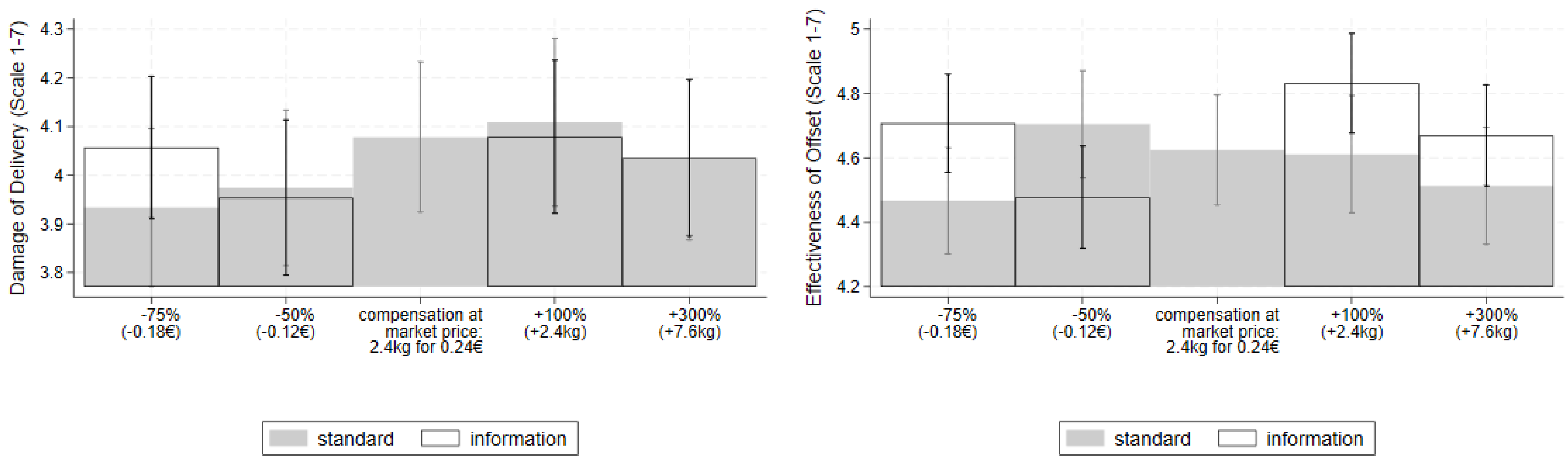


Notes: Panel a) represents the offsetting probabilities across treatments. Gray bars represent standard treatment groups, transparent bars represent information treatment groups. Panel b) plots the increase in compensated kilograms per EUR spent by the firm, relative to the baseline offset. The dotted line indicates the market price of the baseline offset (10kg/EUR).

Figure 5: Post-Experimental Survey

(a) Perceived Environmental Damage of a Delivery w/o Offsetting

(b) Perceived Effectiveness of Offset

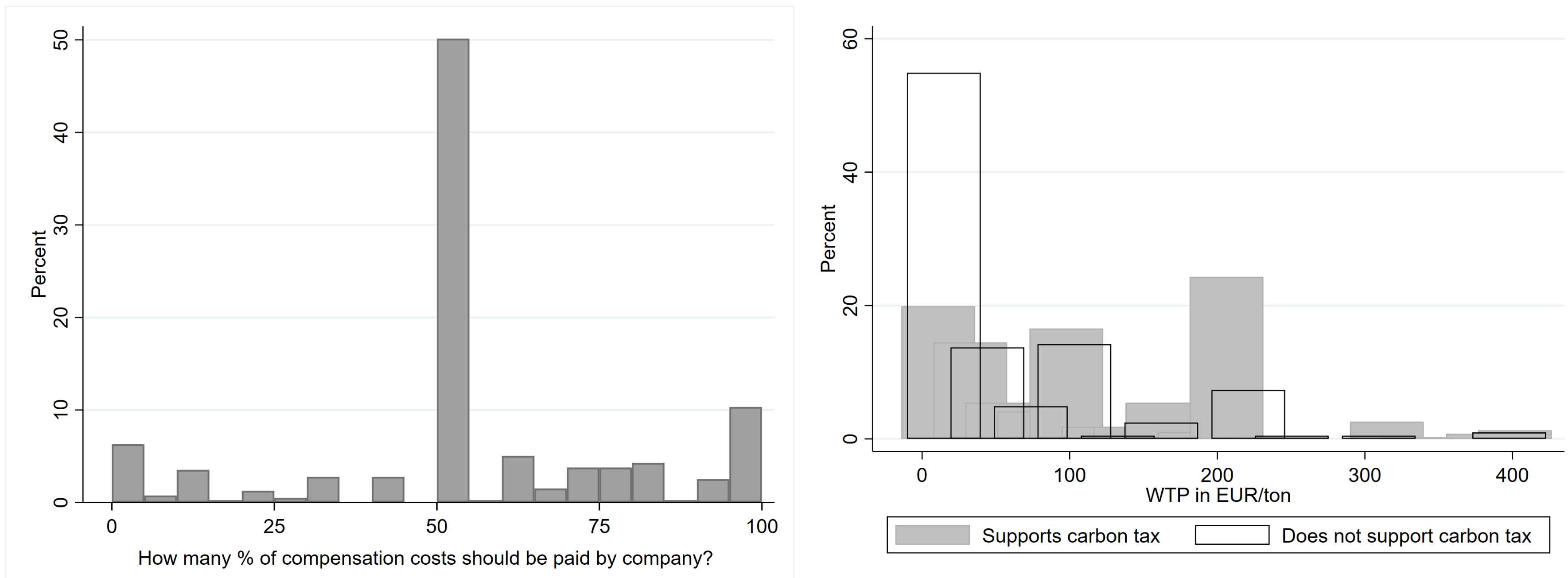


Notes: Panel a) illustrates subjects' beliefs about the size of the environmental damage of one delivery that is not compensated by an offset. Panel b) shows beliefs about the effectiveness of the offset in mitigating environmental damages.

Figure 6: Email Survey

(a) Fairness Preferences

(b) WTP and Support for Carbon Tax



Notes: Panel a) illustrates the distribution of subjects' answers to the question of what share of the carbon compensation costs should be paid by the firm. Panel b) shows the distribution of WTP in EUR/ton of CO_2 among the supporters of the tax (in gray) and the opponents (transparent).

Tables

Table 1: Summary Table

Variable	Control	Baseline: 0.24€ at 2.4kg	-0.12€	-0.18€	-0.12€, information
Number of website visits	1.593	1.596	1.595	1.588	1.602
	(1.365)	(1.828)	(1.390)	(1.350)	(1.415)
Order (1= yes)	0.329	0.330	0.333	0.327	0.333
	(0.470)	(0.470)	(0.471)	(0.469)	(0.471)
Offset (1= yes)	0.000	0.135	0.143	0.157	0.163
	(0.000)	(0.342)	(0.350)	(0.364)	(0.369)
Expected travel time (in min)	14.508	14.366	14.498	14.509	14.561
	(9.397)	(9.433)	(10.110)	(9.582)	(9.825)
Expected service time (in min)	7.201	7.260	7.304	7.282	7.296
	(3.817)	(3.650)	(4.071)	(3.815)	(3.773)
N	25,564	25,427	25,654	25,556	25,643

Variable	-0.18€, information	+2.4kg	+7.2kg	+2.4kg, information	+7.2kg, information
Number of website visits	1.584	1.591	1.598	1.592	1.598
	(1.617)	(1.526)	(1.492)	(1.462)	(1.449)
Order (1= yes)	0.332	0.333	0.334	0.330	0.331
	(0.471)	(0.471)	(0.472)	(0.470)	(0.471)
Offset (1= yes)	0.185	0.128	0.133	0.150	0.164
	(0.388)	(0.334)	(0.339)	(0.357)	(0.371)
Expected travel time (in min)	14.525	14.442	14.428	14.470	14.685
	(9.546)	(9.319)	(9.464)	(9.562)	(9.832)
Expected service time (in min)	7.371	7.334	7.305	7.285	7.302
	(3.855)	(3.781)	(4.048)	(3.921)	(4.159)
N	25,375	25,564	25,762	25,642	25,189

Note: This table presents the mean of observable variables in different treatment conditions. Standard deviations are reported in parentheses.

Table 2: Between- and Within-Subject Treatment Effects

	Offsetting Probability $\times 100$				
	(1)	(2)	(3)	(4)	(5)
Subsidy	0.500***	0.403***	0.403***	0.542**	2.579**
	(0.160)	(0.155)	(0.155)	(0.220)	(1.084)
$\times$ information	0.837***	0.272**	0.267**	0.471*	1.231
	(0.140)	(0.133)	(0.134)	(0.246)	(0.876)
Match	-0.087	-0.319	-0.317	0.041	1.959*
	(0.157)	(0.214)	(0.215)	(0.153)	(1.015)
$\times$ information	0.838***	0.555**	0.546**	0.411***	0.373
	(0.133)	(0.232)	(0.235)	(0.128)	(0.899)
Interactions with Previous Treatment Exposure					
Match $\times$ pevious exposure to offsets		0.561***	0.572**		
		(0.216)	(0.223)		
Match $\times$ Info $\times$ pevious exposure to offsets		-0.226	-0.219		
		(0.278)	(0.280)		
Match $\times$ pevious exposure to info			-0.031		
			(0.157)		
Subsidy $\times$ pevious exposure to offsets				-0.217	
				(0.222)	
Subsidy $\times$ Info $\times$ pevious exposure to offsets				-0.312	
				(0.293)	
Baseline Probability $\times 100$	4.46	2.47	2.47	2.47	11.15
Sample	First Visit	Panel	Panel	Panel	Buyers in Panel
Time and Zip Code FE	Yes	Yes	Yes	Yes	Yes
Subject FE	No	Yes	Yes	Yes	Yes
R^2	0.01202	0.43625	0.43625	0.43624	0.74159
N	229,812	201,258	201,258	201,258	9,102

Note: This table reports treatment effects on the offsetting probability among website visitors. Column 1 reports between-subject effects during the first visit. Columns 2-5 use the subsample for which repeated visits are observed. Column 5 conditions on buyers only. Interaction effects measure how treatment elasticities change when subjects have previously been exposed to offsets and INFORMATION. Standard errors are clustered on the subject level and presented in parenthesis. *,**,***: significant at $p < 0.1$, $p < 0.05$, $p < 0.01$, respectively.

Table 3: Willingness to Pay for Carbon Mitigation, Warm Glow, and Fairness

	(1) Baseline Model	(2)	(3) Behavioral Model
	Standard	Information	Entire Sample
WTP in €/tCO_2 ($-\frac{\gamma}{\eta}$)	-0.18	15.99***	12.84**
	(5.59)	(2.51)	(5.13)
Warm glow utility in € ($-\frac{\omega}{\eta}$)	1.27***	0.74***	0.86***
	(0.24)	(0.07)	(0.14)
Fairness disutility per € above firm contribution ($-\frac{\phi}{\eta}$)			0.08
			(0.16)
Fairness utility per € below firm contribution ($-\frac{\rho}{\eta}$)			0.21***
			(0.06)
Price coefficient (ζ)	0.13***	-0.25***	-0.21***
	(0.03)	(0.03)	(0.04)
Dynamic Attenuation Effect (Ω)			0.84***
			(0.17)
Sample	Buyers' first visit	Buyers' first visit	Buyers in panel
Time & Zip-Code FE	Yes	Yes	Yes
N for between-subject estimation	42440	42186	76229
N for within-subject estimation			9102

Note: This table reports coefficients from the empirical specifications in equation 4 and 7. Coefficients in column 1 and 2 are obtained by estimating the baseline model (equation 4) from the sample in STANDARD and INFORMATION, separately. Column 3 reports coefficients from the extended model that allows for inattention and fairness preferences (equation 7). The first four coefficients are measured in EUR. The final coefficient, ζ, measures the change in demand in percentage points for every one EUR change in price. Standard errors in parentheses. *,**,***: significant at $p < 0.1$, $p < 0.05$, $p < 0.01$, respectively.

Table 4: Hypothetical WTP

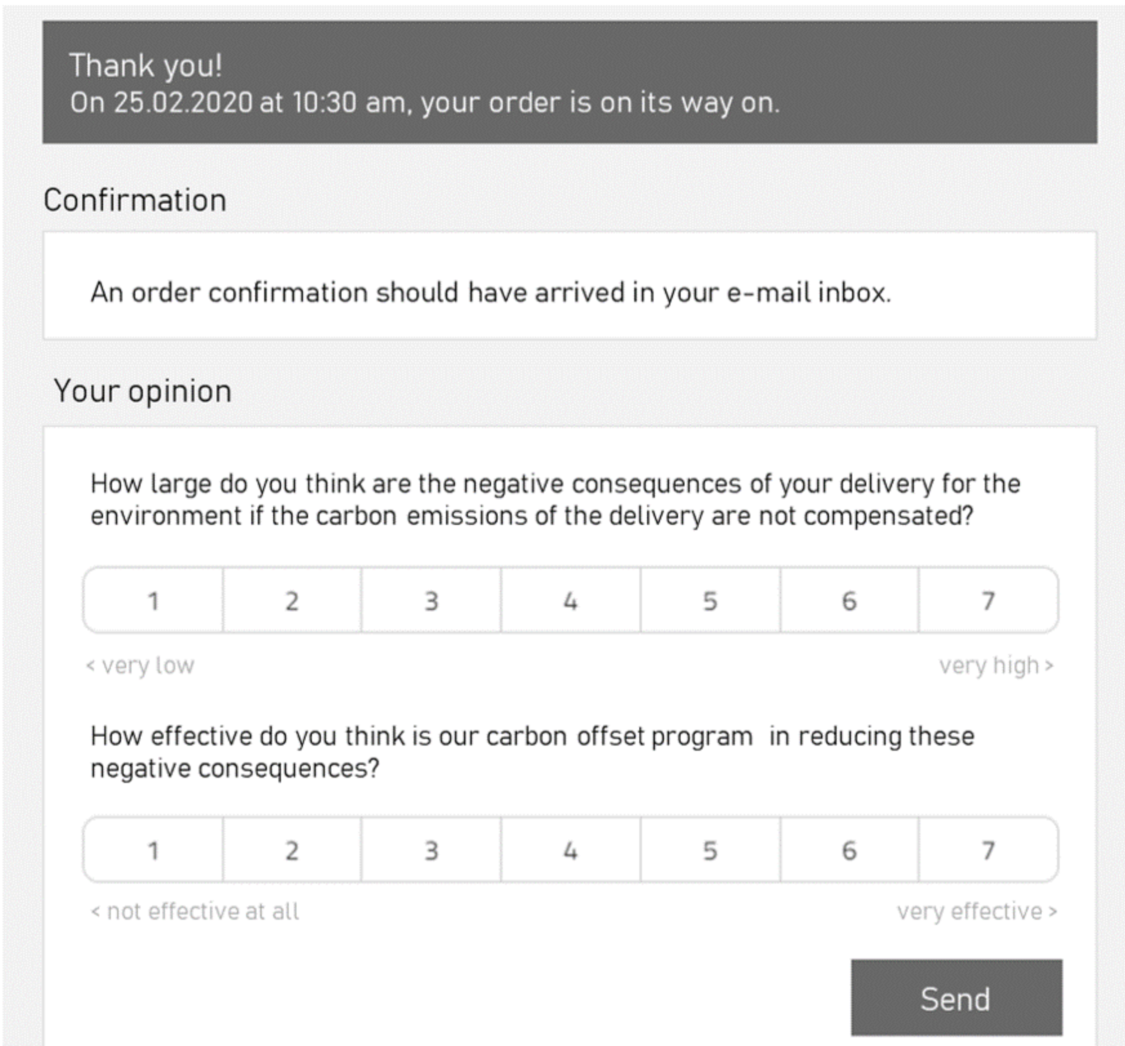
	Total WTP (in Cents)
Quantity increase between-subject:	
+2.4kg	-7.857
	(9.891)
+4.8kg	-0.996
	(9.984)
Quantity increase within-subject:	
+2.4kg	31.243***
	(2.559)
+4.8kg	32.149***
	(3.463)
Between-subject variation in Education and Fairness Treatments:	
+4.8kg, Education	10.457
	(12.936)
+2.4kg, Education & Firm Contribution	-0.184
	(9.977)
+2.4kg, Firm Contribution	-23.774**
	(9.312)
+4.8kg, Firm Contribution	-8.936
	(9.479)
+4.8kg, Education & Firm Contribution	-9.073
	(9.579)
Constant (baseline offset: 2.4kg)	57.001***
	(6.751)
WTP in EUR/tCO_2	237.50***
	(28.13)
R^2	0.0478
N	1,617

Note: This table reports treatment effects on hypothetical WTP as absolute WTP in Cents. Subjects stated their WTP in an open-end question. The second-to-last row shows implied WTP in EUR/ton of CO_2. The treatment "Education" indicates whether subjects received an education treatment about carbon offsetting prior to the WTP elicitation. "Firm contribution" indicates whether subjects were informed that the firm contributes to the match. Robust standard errors are in parentheses. *,**,***: significant at $p < 0.1$, $p < 0.05$, $p < 0.01$, respectively.

Online Appendix

A Additional Figures

Figure A1: Two-Question Survey Directly after Purchase



B Distrust and Risk Aversion

The structural estimates relied on the assumption that consumers trusted the offset compensates for the promised amount. While I did not find evidence that perceived effectiveness varies substantially across treatments in the post-purchase survey, it may well be that there is a general distrust into offsets' effectiveness. In this section, I study how this type of uncertainty about impact affects WTP estimates and how it interacts with consumers' risk preferences.

Let $\mathbb{E}[\gamma]$ denote the consumer's subjective expectation that the offset in fact mitigates γ units of carbon. Under uncertainty about the reliability of the offset, expected utility in the linear case becomes

$$\mathbb{E}[u_i] = \omega + \beta_1 \mathbb{E}[\gamma_i] + \eta p_i + \epsilon_i. \tag{10}$$

If consumers are risk-averse, uncertainty causes a larger reduction in expected utility than for risk neutral consumers. To capture risk aversion over carbon mitigation and over money, I specify a quadratic utility function:

$$\mathbb{E}[u_i] = \omega + \beta_1 \mathbb{E}[\gamma_i] + \frac{\beta_2 \mathbb{E}[\gamma_i^2]}{2} + \eta p_i + \frac{\eta_2 p_i^2}{2} + \epsilon_i. \tag{11}$$

This utility function generates demand derivatives $\frac{\partial D}{\partial E[\gamma]} = (\beta_1 + \beta_2 \mathbb{E}[\gamma])g(\xi)$ and $\frac{\partial D}{\partial p} = (\eta_1 + \eta_2 p)g(\xi)$. Writing demand in terms of these structural parameters, we obtain

$$D_i = \alpha + \tau_1 \gamma_i + \frac{\tau_2 \gamma_i^2}{2} + \zeta p_i + \frac{\zeta_2 p_i^2}{2} + \xi_i \tag{12}$$

$$\approx \alpha + (\beta_1 + \beta_2 \mathbb{E}[\gamma_i]) g(\xi) \gamma_i + (\zeta_1 + \zeta_2 p_i) g(\xi) p_i + \xi_i \tag{13}$$

The experimental variation allows us to estimate the nonlinear terms, β_2 and ζ_2, because we have three price points, $p_i \in \{6, 12, 24\}$ Cents, and three quantity levels, $\gamma_i \in \{2.4, 4, 8, 9.6\}$ kg of CO_2. Willingness to pay for carbon mitigation now varies together with the *levels* in price and carbon quantities: $WTP(p_i, \gamma_i) \approx -\frac{\beta_1+\beta_2 E[\gamma_i]}{\zeta_1+\zeta_2 p_i}$. Relative to WTP under risk neutrality, risk aversion reduces WTP only if $\beta_2 \mathbb{E}[\gamma_i] > \eta_2 p_i$; that is, when risk aversion over carbon mitigation has a larger import than risk aversion over money.

To investigate the impact of distrust in the carbon offset, I vary the consumer's subjective probability. For simplicity, I assume that subjective beliefs are described by a Bernoulli distribution with expectation $\mathbb{E}[\gamma] = \gamma\rho$. In other words, the consumer believes the offset will compensate with probability ρ, and will fail to compensate with probability $1 - \rho$.

I estimate WTP around the baseline offset for different values of ρ. Figure B1 plots WTP in INFORMATION for values of ρ on the interval $[0.2, 1]$.[30] WTP is either computed under risk neutrality (equation 10) or under risk aversion (equation 11).

Note first that the implied willingness to pay (WTP) *increases* as distrust in the offset goes up. While this may seem counter-intuitive at first sight, it is a mechanical result. Specifically, if a consumer believes the offset only compensates with probability 0.5, then any *given* demand response to an increase in γ implies that the consumer's WTP is twice as large as in the alternative case in which she believes the offset compensates with probability 1. To understand this further, consider another numerical example for a risk-neutral consumer. Assume that the consumer is willing to pay 10 EUR for an offset that promises to offset 10kg. If she believes the offset is trustworthy, $\rho = 1$, her implied WTP is 1 EUR/kg. On the other hand, if she believes the offset compensates only with $\rho = 0.5$, the expected reduction in emissions is only 5kg, and her implied WTP is 2

[30] I exclude values close to zero because WTP approaches infinity as perceived effectiveness goes to zero.

EUR/kg.[31]

The next observation is that the point estimates under risk aversion are only slightly lower than under risk neutrality. This is due to the fact that quantity and price elasticities are, at least locally, fairly constant, so that concavities in utility have little import on the point estimates. The nonlinearities are imprecisely estimated, and statistically insignificant, which inflates the WTP confidence intervals under risk aversion.

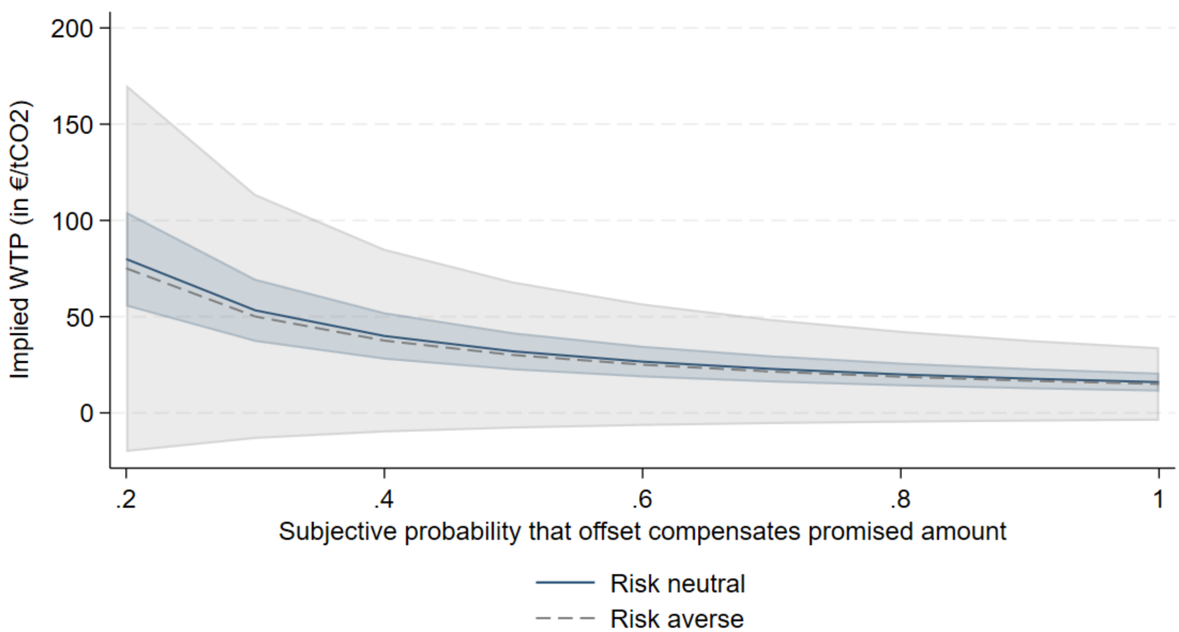
In terms of magnitudes, at $\rho = 1$, we obtain our main estimate of 16 EUR$/tCO_2$, both under risk neutrality and risk aversion. WTP is close and statistically indistinguishable from this estimate for any value above $\rho = 0.7$, i.e., until consumers believe that the offset compensates at least 70% of the promised amount. At $\rho = 0.6$, WTP is 27 EUR$/tCO_2$, and statistically different from 16 EUR for the risk neutral model. Since WTP approaches infinity as ρ goes to zero, very low subjective probabilities yield much higher WTP estimates. For values anywhere below 30%, we obtain estimates ranging from 53 to 160 EUR$/tCO_2$.

While we do not know subjects' actual belief distribution, we may use the post-purchase survey as auxiliary evidence. As discussed in Section 3.2, when asked about the effectiveness of the offset, subjects reported an average value of 5 on a scale from 1 to 7. One approach is to assume that this approximately maps into a subjective probability of $\rho = \frac{5}{7} = 71\%$. This number is also broadly consistent with a representative consumer survey in the US that suggests that around 60% of consumers believe that carbon offsets are effective (Boston Consulting Group 2023). If we take $\rho = 71\%$ as our preferred value, Figure B1 tells us that our main estimate of 16 EUR$/tCO_2$ remains a good approximation—independent of whether consumers are risk neutral or risk averse.

[31] A slightly more formal explanation can be obtained by consulting equation 10. Here, β_1 measures the increase in utility per unit increase in expected γ. If we now fix the demand response by the consumer to an increase in γ, but reduce the probability that γ will truly be compensated, then β must mechanically increase.

In conclusion, WTP estimates increase if we believe subjects in the experiment distrust the offset. Our main estimate is robust to a reasonable range of subjective probabilities. However, the sensitivity at low levels of trust highlights the importance for future research to quantify consumers' belief distributions and to address adverse selection in the offsetting market.

Figure B1: Willingness to Pay as a Function of Perceived Effectiveness



Note: This figure shows how WTP estimates change when consumers might distrust that the offset compensates for the promised amount of carbon. The abscissa varies the subjective probability that the baseline offset compensates 2.4kg of carbon. The blue line is implied WTP if consumers are risk neutral over income and carbon emissions. The dashed gray line shows WTP when consumers are risk averse as estimated from nonlinearities in the price and quantity coefficient. Confidence intervals are given by the shaded areas. Since nonlinearities in price and carbon emissions are statistically insignificant in the estimation, confidence intervals for the risk-averse model are substantially larger than for the risk-neutral model.

C Firm Outcomes

C.1 Demand for Deliveries

To analyze whether the treatments affected demand for deliveries, I first estimate a linear probability model, regressing whether a subject placed an order on the treatment vectors. Column 1 and 2 in Table C1 report the regression results. Column 1 includes both between- and within-subject variation and cluster standard errors on the subject-level. Column 2 conditions on the first visit during the experimental period.

The probability of ordering at the shop is 27% and 33%, respectively. All treatment coefficients are economically small and tightly estimated null effects. This suggests that offering website visitors a carbon offsetting program during checkout does not affect demand for deliveries.

The coefficients constitute "Intent-to-Treat" effects (ITT) because not everyone who visited the website saw the treatment. Instead, only subjects that clicked on the shopping basket and/or checked out were offered the offset. Formally, ITT $= E[Y_1 - Y_0 \mid Z = 1]$ where Y_1 is the outcome if treated, Y_0 is the outcome if not treated, and $Z \in \{0, 1\}$ is being assigned to treatment. This differs from the "treatment effect on the treated" (TOT), which is the effect for those that made it to the checkout: TOT $= E[Y_1 - Y_0 \mid D = 1]$ where $D \in \{0, 1\}$ indicates making it to the checkout in the treatment group. Simple algebraic manipulations show that $TOT = \frac{ITT}{Pr(D=1|Z=1)}$, which is identified if we knew the probability of making it to checkout.[32] Unfortunately, browsing behavior remains unobserved in the data. However, we can obtain an upper bound of the TOT by leveraging the fact that everyone who bought something must have gone through checkout. Denoting the buying decision by $B \in \{0, 1\}$, we have $Pr(B = 1|Z = 1) \leq Pr(D = 1|Z =$

[32] By the law of total probability, we have ITT $= E[Y_1 - Y_0 \mid Z = 1] = \Pr(D = 1 \mid Z = 1) \cdot E[Y_1 - Y_0 \mid D = 1] + \Pr(D = 0 \mid Z = 1) \cdot E[Y_1 - Y_0 \mid D = 0]$. Note that $E[Y_1 - Y_0 \mid D = 0] = 0$ since if no treatment was received, the difference in outcomes is zero. Thus, ITT $= \Pr(D = 1 \mid Z = 1) \cdot E[Y_1 - Y_0 \mid D = 1] = \Pr(D = 1 \mid Z = 1) \cdot TOT$.

$1) \leq 1$. An upper bound for the TOT is thus given by $TOT_{ub} = \frac{ITT}{Pr(B=1|Z=1)}$. In column 1 of Table C1, we can compute this upper bound by scaling each coefficient by 0.33, the probability of buying. In the best case scenario for the firm, this yields a TOT of $0.005/0.33 = 0.015$ percentage points, in the worst case scenario a TOT of $-0.002/0.33 = -0.006$ percentage points. These bounds turn out to be empirically small, suggesting that even the TOT on delivery demand is negligible.

Columns 2 to 4 show whether the treatments had an effect on follow-up purchases during the experimental period. Column 2 considers all individuals and the treatment they were assigned during their first visit. I build a binary variable taking the value of one if they buy again during the experimental period, and zero otherwise. I then regress this outcome variable on the different treatments as assigned during the *first visit*.

Most coefficients are negative but economically very small and statistically insignificant. The exception is the 100% match which constitutes a marginally significant reduction in the probability to buy again by 0.2pp ($p < 0.1$). This likely constitutes a false positive since i) all other treatment effects are zero, including the larger 300% match, ii) the regression includes a large number of treatments, increasing the chance of a false positive, and iii) the effect is only marginally significant. Column 3 pools across all offsets to increase power and includes an interaction term with the INFORMATION treatment. Both effects are small and insignificant. The upper bound for the TOT is a reduction in the probability to purchase again of $-0.001/0.33 = -0.003$pp due to offsets.

Column 4 conditions the sample on subjects who made a purchase during the first visit. Since these subjects must have gone through checkout, they were offered an offset (except for control subjects). Here the coefficients are slightly more negative but still far from statistical significance. The treatment effect on buyers is a 0.003 reduction in the probability to purchase again, which, interestingly, is identical to the upper bound on the

TOT from column 3.[33]

An upside of these results is that offering carbon offsets does not seem to have large negative effects on delivery demand by steering consumers' attention to the polluting attribute of a delivery.[34] This was one of the main concerns of the firm prior to the experiment. However, the presented results do not allow us to conclude whether saliently advertising carbon offsets at the outset could have attracted or deterred consumers.[35]

[33]A sufficient condition for this result is that $P(B = 1|Z = 1) \approx P(D = 1|Z = 1)$, i.e., most people who visit the checkout page end up purchasing.

[34]See Alan et al. (2018) for an empirical example where highlighting add-ons (such as carbon offsets) can steer attention to unfavorable attributes and backfire for firms.

[35]An interesting experimental design could explore such effects by not just randomizing offsets at the checkout but also advertisements of the offset at the outset (e.g., the homepage). This way, one could parse out which consumer types select into the shop due to offsets. See Karlan and Zinman (2009) and Rodemeier (2024) for examples of such experimental designs.

Table C1: Intent-to-Treat Effect on Demand for Deliveries

	Probability to Purchase		Probability to Purchase Again		
	(1)	(2)	(3)	(4)	(5)
Baseline: 24 Cents, 2.4kg	0.001	0.001	-0.001		
	(0.003)	(0.004)	(0.001)		
50% subsidy	0.003	0.004	-0.000		
	(0.003)	(0.004)	(0.001)		
× information	0.000	-0.000	-0.001		
	(0.003)	(0.004)	(0.001)		
75% subsidy	0.000	-0.002	-0.000		
	(0.003)	(0.004)	(0.001)		
× information	0.001	0.005	-0.000		
	(0.003)	(0.004)	(0.001)		
100% match	0.003	0.004	-0.002*		
	(0.003)	(0.004)	(0.001)		
× information	-0.003	-0.003	0.000		
	(0.003)	(0.004)	(0.001)		
300% match	0.003	0.005	-0.001		
	(0.003)	(0.004)	(0.001)		
× information	-0.003	-0.003	-0.000		
	(0.003)	(0.004)	(0.001)		
All offsets pooled				-0.001	-0.003
				(0.001)	(0.003)
× information				-0.000	-0.001
				(0.001)	(0.002)
Constant: No offset offered	0.265***	0.329***	0.021***	0.021***	0.054***
	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)
Sample	All visits	First visit	First visit	First vist	Buyers' first visit
R^2	0.00001	0.00002	0.00002	0.00001	0.00003
N	406,984	255,376	255,376	255,376	84,644

Note: This table reports Intent-to-Treat effects on the probability of placing an order among website visitors in columns 1 and 2. The first column includes all website visits, whereas the second column only includes the first visit of a subject during the experimental period. Columns 3-5 report the probability of purchasing again during the experimental period. Column 4 pools across offsets and information treatments. Column 5 conditions on the subsample of buyers (instead of all visitors). Standard errors are in parentheses; clustered on the subject level in column 1 and heteroskedasticity robust in columns 2-5. *,**,***: significant at $p < 0.1$, $p < 0.05$, $p < 0.01$, respectively.

C.2 Cashflows

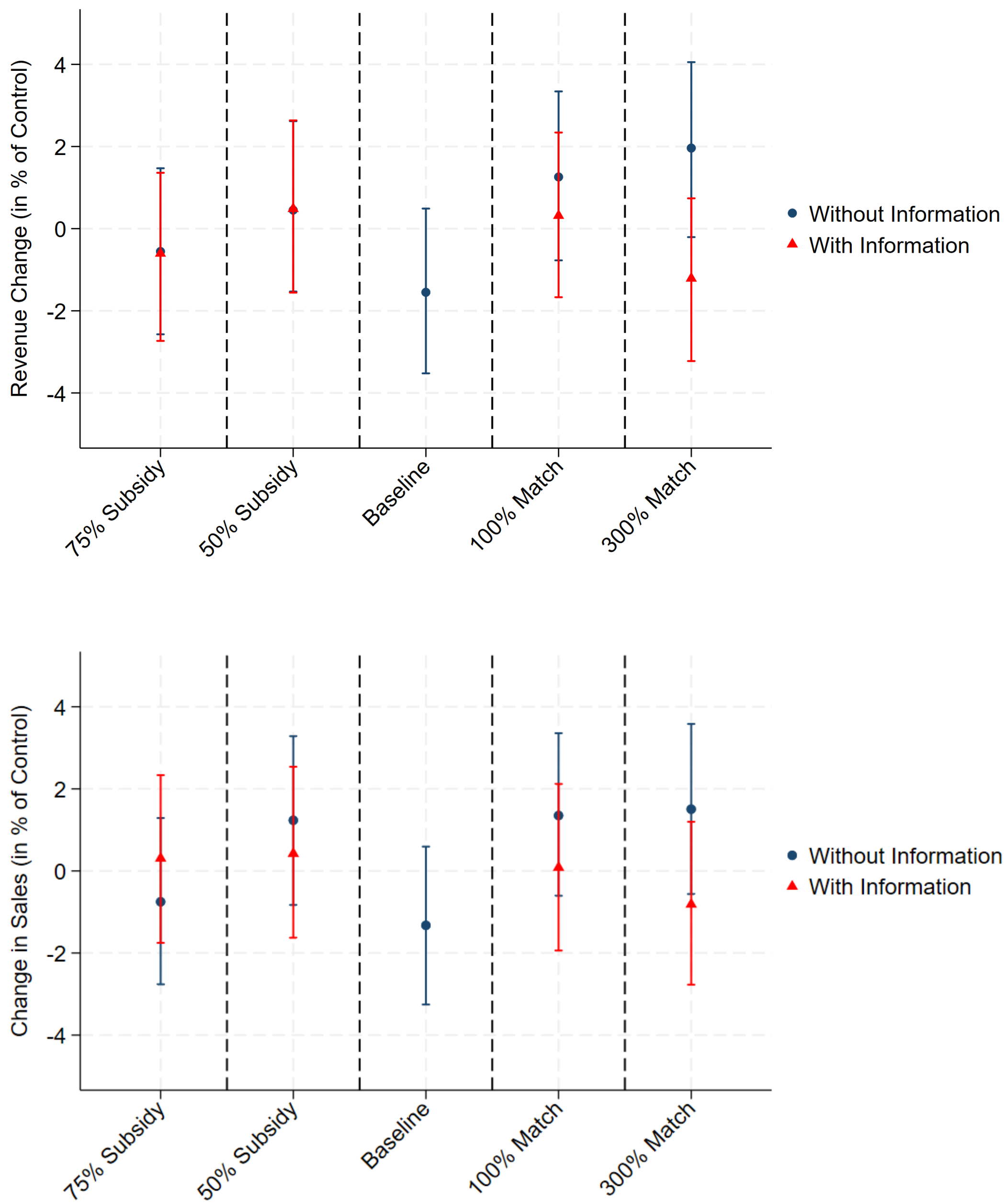
Figure C1 plots differences in sales (quantities sold) and cashflows across treatment groups. In Panel A, I compute the cash flows *net of the offsetting costs to the firm*, relative to control group cash flows. Thus, I explicitly adjust for the costs of sustainable firm engagement. I compute these "net cash flows" nonparametrically through paired bootstrap with one thousand repetitions. The bars in the figure represent 95% confidence intervals from this estimation. Treatments with information are presented in red, treatments without information in blue. The entire sample is included in the estimation.

The baseline offset has a slightly negative but statistically insignificant coefficient of —1.5%. Subsidies and matches tend to have mixed coefficients that vary around zero and none of them is statistically significant.

The effect does also not vary substantially with INFORMATION as most coefficients are indistinguishable from each other. The only exception is the 300% match for which STANDARD has marginally significantly *higher* point estimate—representing a 2% increase in cash flows relative to control. The INFORMATION coefficient itself is, however, *not* significantly different from zero and one should be hesitant to interpret too much into this isolated result. For all other treatments, we find no difference between INFORMATION and STANDARD, which may suggest that the difference for the 300% match constitutes a false positive from sampling variation.

Panel B plots differences in sales (number of items), which are again nonparametrically estimated. We observe a similar picture as for net cashflows. None of the point estimates is statistically significant. The information treatment had no discernible effect on the number of items that subjects bought. The exception is again the 300% match where STANDARD has a larger coefficient of 2%, which is marginally statistically different from INFORMATION, but not significantly different from zero.

Figure C1: Effect of Carbon Offsets on Sales and Cash Flows



Notes: Panel a) represents differences in cash flows net of offsetting costs to the firm. All differences are in percent relative to control. Panel b) plots differences in sales relative to control. Standard errors are computed nonparametrically through paired bootstrap with 1,000 repetitions.

Overall, these results indicate that offering carbon offsets during checkout has no negative impact on cashflows and profits. The presented results mitigate the worry that carbon offsetting could decrease demand for deliveries for existing customers. However, existing results do not allow us to conclude that carbon offsets could not attract new customers since we randomized in the shopping basket rather than at the front page of the shop. The current results are still important for current firm practices as most firms that offer carbon offsets do so towards the end of the shopping process (e.g., airlines offer offsets during checkout).

D Additional Results from Email Survey

Table D1 reports results from a regression of the education treatment on the belief questions. Without the education treatment, subjects overestimate the carbon emissions of the average delivery, the equivalent kilometers that one needs to drive with a conventional car, and the amount of trees necessary to compensate for 2,000 deliveries. The last column shows how certain subjects were in their answers, where larger values indicate more certainty. The education treatment results in a substantial and statistically significant rise in subjects' certainty by close to 100% compared to the control group.

Table D1: Answers to Belief Questions in Email Survey

	(1) Delivery	(2) Car	(3) Trees	(4) Certainty
Education treatment about carbon offsetting	-0.409	-11.363***	-85.991***	2.323***
	(0.826)	(3.577)	(22.477)	(0.139)
Constant	5.463***	30.288***	144.844***	2.354***
	(0.543)	(2.471)	(15.893)	(0.095)
R^2	0.0009	0.0368	0.0635	0.2656
N	285	266	218	769

Note: This table reports answers to the belief questions in the email survey. Robust standard errors are in parentheses. *,**,***: significant at $p < 0.1$, $p < 0.05$, $p < 0.01$, respectively.

Demographics and Risk Preferences

I regress hypothetical WTP on basic demographics elicited in the survey. I also include an established measure of risk preferences developed by Falk et al. (2022) that has been shown to predict actual risk preferences in incentivized questions. The question asked to subjects is: *"Please tell us, in general, how willing or unwilling you are to take risks."* Potential answers are on a Likert scale from 1 ("not at all willing to take risks") to 10 ("very willing to take risks").

Table D2 reports results. In terms of demographics, relatively few variables are a strong predictor of WTP. The constant represents WTP for an employed, male subject, between 40-49 years of age. On average, female subjects have a substantially higher WTP by around 31 Cents. In addition, retired subjects have a higher WTP of 20 Cents. This may be surprising as it is often claimed that older people have a lower incentive to protect the climate as they will be less exposed to future damages. Subjects that answered "other" to the employment question have an 18-cents lower WTP.

Interestingly, risk preferences are a strong predictor of hypothetical WTP. For every

1-point increase on the "willingness to take risk"-scale, WTP increases by 4 Cents, a relative increase of 11% relative to the constant. Note that the direction of the relationship between WTP and risk preferences partially depends on how much uncertainty subjects have about climate change versus how uncertain they are about the effectiveness of carbon offsets. On the one hand, it seems reasonable to assume that more risk-averse individuals have a stronger willingness to pay for carbon mitigation since there is large uncertainty about future climate damages. On the other hand, the effectiveness of carbon offsets itself is uncertain, such that more risk-averse individuals may be less willing to donate to these projects. The present results may indicate that the second effect dominates.

To investigate this relationship visually, Figure D1 plots the correlation between risk preferences and average WTP. Specifically, each data point represents average WTP for a given level of risk preferences. The red line provides a linear prediction of the relationship.

While the relationship does not appear linear visually, it seems positive for most intervals. Thus, more risk-seeking consumers state a higher willingness to invest into carbon offsets. While correlations should always be interpreted cautiously, these patterns suggest that uncertainty may constitute a barrier to voluntary climate protection. However, we also saw in Section B, that risk aversion—as revealed by subjects' actual choices—was not very pronounced locally. Studying concavities over larger emission levels would be a promising avenue for future research.

Table D2: WTP and Demographics

	Total WTP (in EUR)
Willingness to take risk	0.043***
	(0.010)
Age:	
18-19	0.016
	(0.287)
20-29	0.097
	(0.077)
30-39	-0.012
	(0.063)
50-59	0.009
	(0.069)
60-79	-0.125
	(0.090)
> 70	-0.064
	(0.156)
Gender:	
diverse	-0.169
	(0.295)
female	0.308***
	(0.049)
Employment Status:	
unemployed	0.001
	(0.186)
apprentice	-0.340
	(0.295)
housewife/husband	-0.277
	(0.181)
retired	0.202**
	(0.096)
other	-0.180**
	(0.079)
student	-0.180
	(0.111)
Constant (40-49 years, male, employed)	0.349***
	(0.080)
R^2	0.0471
N	1,466

Note: This table reports correlations between WTP, risk preferences, and demographics. The constant represents WTP for an employed, male subject, between 40-49 years of age. Robust standard errors are in parentheses. *,**,***: significant at $p < 0.1$, $p < 0.05$, $p < 0.01$, respectively.

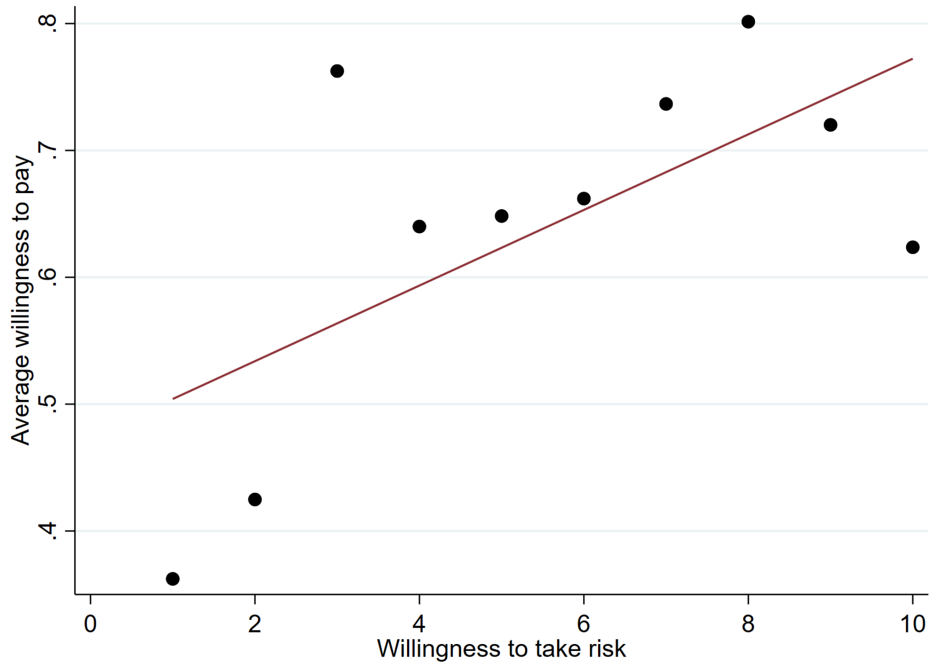


Figure D1: Correlation between Risk Preferences and WTP

E Heterogeneity Based on Emissions

This Section studies heterogeneity in treatment effects across subjects with different emission levels. I use three proxies for emissions: i) the delivery time from the warehouse to the household, ii) the total spending in the shop, and iii) the carbon intensity of products that consumers purchase. For each of these proxies, I generate an indicator, $X_i \in \{0, 1\}$, that equals one if subject i's value for the proxy lies above the sample median. I then regress the offsetting choice on the indicator and its interaction with INFORMATION, using the following specification:

$$\text{offset}_i = \alpha + \eta_1 p_i + \eta_2 p_i \times I_i + \beta_1 \gamma_i + \beta_2 \gamma_i \times I_i + \delta_1 X_i + \delta_2 X_i \times I_i + \epsilon_i \quad (14)$$

The coefficient δ_1 measures the difference in offsetting for above-median households. δ_2 indicates whether these subjects are relatively more or less responsive to INFORMATION. To increase statistical power, I pool across all INFORMATION treatments when estimating the interactions.

To estimate proxy i), I obtain additional data from the logistics department of the firm. This data includes the expected delivery time for each order which is estimated from a firm-internal algorithm. I generate a dummy variable equal to one if the placed order is above the median expected delivery time, and zero otherwise. Column 1 in Table E1 reports regression coefficients, multiplied by 100 for readability. Households with an above median travel distance are 2.6 percentage points *less* likely to choose an offset ($p < 0.01$). Even though these subjects produce higher travel emissions, their willingness to offset is substantially lower. Since the warehouses of the firm are relatively central in each city, this coefficient may imply that rural/suburban customers are generally less likely to compensate for emissions than urban customers. INFORMATION does not seem to reduce this gap as the coefficient is both statistically and economically insignificant.

To study heterogeneity across spending (proxy ii), I generate a dummy that indicates above and below median shopping basket values. Results in column 2 do not suggest that subjects who spent more exhibit a different willingness to offset. However, the effect of INFORMATION is 0.9 percentage points larger for these subjects ($p < 0.1$).

Finally, column 3 shows differences in proxy iii), shopping baskets with relatively high or low carbon-intensity. These results should be interpreted as a coarse approximation since I do not have access to the carbon emissions of each product. Instead, I obtain publicly available information on the carbon emissions of broad categories from educational websites.[36] This allows me to construct a general ranking of categories: For instance, a cup of coffee ($\approx$ 0.4kg CO2) is, on average, more carbon-intensive than a

[36]See Clever Carbon (2024), CO2 Everything (2024), and Tappwater (2024)

330ml bottle of beer ($\approx$ 0.25kg CO2), which is in turn more carbon-intensive than a 150ml glass of wine ($\approx$ 0.13kg of CO2). This approach does not allow me to distinguish carbon emission differences between brands within the same category, but only across categories. To the best of my knowledge, information about brand-specific emissions is indeed mostly unavailable or not reliable. Using the category-specific carbon emissions, I adjust for the "size" of each product: e.g. a *pack* of coffee on average amounts to 140 cups of coffee, which implies a total emission of carbon of $140 \times 0.4kg = 56$ kg of CO2, while a six-pack of beer amounts to $6 \times 0.25 = 4$kg of CO2. Next, I compute aggregate emissions by summing up all product-specific emissions of the shopping basket. Finally, I compute for each order whether the total emissions are above or below the sample median.

Column 3 does not show any differences in offsetting across high- or low-emission shopping baskets. Subjects with high-emission baskets are only 0.15 percentage points more likely to compensate and 0.67 percentage points more responsive to INFORMATION. These estimates are small and insignificant. One concern could be that the shopping basket is endogenous to the treatment variation. In Appendix F, I study the effect of offering an offset in the shop on the basket composition of customers and find no conclusive evidence.

Table E1: Heterogeneity

	Offsetting Probability (in%)		
	(1)	(2)	(3)
Price (in Cents)	-0.173***	-0.172***	-0.172***
	(0.023)	(0.023)	(0.023)
× Information	0.165***	0.119***	0.129***
	(0.045)	(0.044)	(0.043)
Quantity (in kg)	0.017	0.028	0.025
	(0.064)	(0.064)	(0.064)
× Information	0.365***	0.298***	0.313***
	(0.062)	(0.059)	(0.060)
above median travel time	-1.548***		
	(0.328)		
× Information	-0.120		
	(0.486)		
above median spending		-0.047	
		(0.327)	
× Information		0.894*	
		(0.489)	
above median product emissions			0.151
			(0.328)
× Information			0.668
			(0.492)
Constant: Baseline Offset	17.974***	17.054***	16.979***
	(0.441)	(0.436)	(0.435)
Sample	First Visit Only	First Visit Only	First Visit Only
R2	0.0025	0.0021	0.0021
N	76,228	76,228	76,228

Note: This table reports differences in offsetting across proxies for high emission orders. Column 1 includes an indicator for orders that have an above median travel time from the warehouse to the household. Column 2 reports differences for orders with an above median spending. Column 3 shows differences for shopping baskets with above median carbon emissions produced by the products in the basket. Only the first visit of each subject during the experiment is included in the regressions. Standard errors are heteroskedasticity robust and presented in parenthesis. *,**,***: significant at $p < 0.1$, $p < 0.05$, $p < 0.01$, respectively.

F Effect of Offsets on Product Emissions

Following the previous section, I examine the impact of providing carbon offsets in the webshop on the total emissions of each basket. To enhance accuracy, I estimate a pooled regression of the form

$$\xi_i = \alpha + \lambda \cdot \Omega_i + \delta \cdot \Omega_i \cdot I_i + \epsilon_i \tag{15}$$

where ξ_i represents the aggregate emissions of participant i's basket, and Ω_i is an indicator equal to 1 if participant i was presented with a carbon offset option in the shop, and zero otherwise. Thus, λ captures the average treatment effect of carbon offsets on basket emissions compared to the control group without offsets. δ measures the interaction effect with INFORMATION.

Results are shown in Table F1. All coefficients are measured in percent relative to control. Column 1 includes only first visits to the website during the experiment. Column 2 and 3 include the entire sample and add subject fixed effects. The regression in column 3 adds an interaction term for whether the offset had an INFORMATION treatment.

All coefficients for λ are statistically insignificant. They range from a 1.3 to 7 percent reduction in basket emissions, which are meaningful effect sizes. The INFORMATION coefficient is much smaller economically and also imprecise. While these results may provide suggestive evidence of a reduction in emissions, they ultimately do not allow us to draw strong conclusions. This is not just because none of the effects is significant, but also because product emissions are very coarsely measured in the data, such that measurement error may affect results. Since we also did not find any systematic effect on sales and cashflows in Appendix C, it seems unlikely that the offsets induced a large effect on emissions.

The imprecisely estimated results in this Section certainly call for new methods and databases that allow us to more accurately estimate product-specific emissions.

Table F1: Effect of Offering Carbon Offsets on Product Emissions

	Change in Product Emissions (in %)		
	(1)	(2)	(3)
Offering Offset	-1.310	-6.973	-6.922
	(1.149)	(18.527)	(18.191)
× Information			-0.129
			(9.957)
Sample	First Visit Only	All	All
Subject-FE	No	Yes	Yes
R2	0.000	0.964	0.964
N	83,859	107,106	107,106

Note: This table reports the effect of offering a carbon offset in the webshop on the emissions of subjects' shopping baskets. All coefficients are in percent relative to control. Column 1 includes only the first visit of each subjects and shows heteroskedasticity robust standard errors. Column 2 and 3 cluster standard errors on the subject-level. *,**,***: significant at $p < 0.1$, $p < 0.05$, $p < 0.01$, respectively.

G Sample Characteristics

Table G1 reports observable characteristics of the sample of respondents. Around 68% are male, 30% female, 0.5% diverse, and 1.7% do not report a gender. 36% are between 20 and 40 years old, which is similar to the German average (31%).[37] Subjects between 40 and 60 years are slightly over-represented compared to the national average (40% vs. 33%), while subjects between 60-79 are underrepresented (13% vs. 27.5%). Consistent with the age distribution, fewer subjects are retired than in the German population (12% vs. 32%). 5.5% are students compared to the national average of 3.5%. Around 1.3% are unemployed compared to 5% nationally. These statistics are remarkably similar to the socio-demographics of consumers in the United States who buy groceries online (Capital

[37] For national statistics see https://www.destatis.de/EN/.

One 2024).

As we would expect from online shop customers, the sample is overall slightly younger and more likely to have an occupation than the German population. According to the firm, the statistics on gender and age are very representative of their customer base. This is important when we want to compare stated preferences from the survey with revealed preferences from the field experiment. While we cannot exclude that subjects select on unobservables into the survey, it is reassuring that observable statistics are fairly representative of the firm's customer population.

Since the sample is not fully representative of the average household in Germany, we should be careful not to generalize the WTP estimates to the general population.[38]

[38]The field experimental data does not include subject-level sociodemographics, such that I cannot estimate a weighted regression to increase the representativeness of the estimates.

Table G1: Summary Statistics of Survey Sample

	Variable	N	Percent
Gender			
	male	514	67.72
	female	228	30.04
	diverse	4	0.53
	no answer	13	1.71
Age			
	18-19	6	0.79
	20-29	120	15.81
	30-39	195	25.69
	40-49	164	21.61
	50-59	140	18.45
	60-79	102	13.44
	> 70	24	3.16
	no answer	8	1.05
Occupation			
	employed	517	68.12
	unemployed	10	1.32
	apprentice	4	0.53
	homemaker	11	1.45
	retired	91	11.99
	student	42	5.53
	other	65	8.56
	no answer	19	2.50
	N	820	100.00

Note: This table reports frequencies of gender, age, and occupational status among participants in the email survey that are included in the analysis.

H Email Survey

The email survey, with all instructions and questions translated from German into English, can be found under this link: click here.