

DISCUSSION PAPER SERIES

IZA DP No. 15939

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

Matthias Rodemeier

FEBRUARY 2023 (THIS VERSION: MARCH 2024)

DISCUSSION PAPER SERIES

IZA DP No. 15939

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

Matthias Rodemeier
Bocconi University and IZA

FEBRUARY 2023 (THIS VERSION: MARCH 2024)

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

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

What do markets for voluntary climate protection imply about people's valuations of environmental protection? I study this question in a large-scale field experiment (N=255,000) with a delivery service, where customers are offered carbon offsets that compensate for emissions. To estimate demand for carbon mitigation, I randomize whether the delivery service subsidizes the price of the offset or matches the offset's impact on carbon mitigation. I find that consumers are price-elastic but *fully* impact-inelastic. This would imply that consumers buy offsets but their willingness to pay (WTP) for the carbon it mitigates is zero. However, I show that consumers can be made sensitive to impact through a simple information treatment that increases the salience of subsidies and matches. Salient information increases average WTP for carbon mitigation from zero to 16 EUR/tCO₂. Two complementary surveys reveal that consumers have a limited comprehension of the carbon-mitigating attribute of offsets and, as a result, *appear* indifferent to impact variations in the absence of information. Finally, I show that the widely-used contingent valuation approach poorly captures revealed preferences: Average *hypothetical* WTP in a survey is 200 EUR/tCO₂, i.e., 1,150% above the revealed preference estimate.

JEL Classification: D61, D82, H21, Q51, Q58

Keywords: climate change, carbon mitigation, willingness to pay, carbon offsets, contingent valuation, nudging

Corresponding author:

Matthias Rodemeier
Bocconi University
Via Roberto Sarfatti, 25
20100 Milano MI
Italy

E-mail: rodemeier.matthias@gmail.com

* The paper was selected for the 2024 IZA Award for Innovative Research on the Economics of Climate Change (IRECC). I thank Nicola Gennaioli, Katrin Gödker, Timm Gries, Nicola Limodio, John List, Andreas Löschel, Christian Rodemeier, Nicolas Serrano-Velarde, and Greg Sun for insightful discussions about this paper. Silvia Amalia Meneghesso and Cristobal Ruiz-Tagle Coloma provided outstanding research assistance.

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 (CO₂) 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 carbon offsetting raises an important question: What does demand for these offsets reveal about people’s valuations of environmental protection?

This question is crucial for cost-benefit analyses of environmental policies. 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 Environmental economists have long understood the importance of obtaining WTP measures for cost-benefit analyses but had to resort to *hypothetical* WTP measures from contingent valuation methods in surveys. More recently, researchers also elicit people’s narratives that motivate stated sustainability preferences (Wekhof and Houde 2023).³ The growing market for voluntary climate protection may provide a unique opportunity to obtain first data points of households’ *revealed* preferences.

However, obtaining WTP estimates from carbon offset demand is complicated by several challenges. The first challenge is that we require exogenous variation in the price of the offset, which is hard to obtain from observational market data. Price variation across different offset projects is often driven by offset quality. Low-cost programs often finance activities that may have taken place regardless of the donation, rendering the impact of the donation on emissions negligible. The second challenge is that even exogenous price variation alone is not sufficient for identification of WTP. This is because consumers might buy the offset simply because of “warm glow” utility: they get utility from the act of giving to a public good but do not care about the impact of that donation.

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.

³ Governments rely on, often hypothetical, WTP estimates to quantify economic damages associated with the environmental impact of human activities, such as the Exxon Valdez oil spill in Alaska in 1989 (see, e.g., Carson et al. 1992) or the BP Deepwater Horizon oil spill in the Gulf of Mexico in 2010 (see, e.g., Bishop et al. 2017).

In order to isolate WTP for carbon mitigation, we also require variation in the *quantity* of carbon that is compensated by the same offset, holding fixed all other attributes. The ideal dataset, therefore, involves exogenous variation in both the offset price and the compensated quantity of carbon.

In this paper, I run a field experiment to generate this data and provide the first revealed preference estimate of WTP from market data. I partner with one of the largest grocery delivery services in Germany and implement an experiment in their online shop, observing over 250,000 consumers. In the experiment, consumers can offset the carbon emissions of grocery deliveries by buying carbon offsets. 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 CO₂ 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-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 fairness preferences from the “intrinsic” valuation of carbon mitigation.

Preferences for Carbon Offsetting. The experiment produces a number of important results. The first one is that 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 theories of warm glow (Andreoni 1990, Karlan and List 2007, Ottoni-Wilhelm, Vesterlund, and Xie 2017) and “scope-insensitivity” (Kahneman and Knetsch 1992).

When consumers are actively informed that the firm has subsidized or matched the offset, demand becomes quantity-elastic. Doubling the compensation amount of the offset increases its demand by 11%, and quadrupling the amount increases demand by 22%. In other words, a minimally invasive information treatment makes consumers sensitive to scope. INFORMATION also increases price elasticities. The effect of a price reduction on offsetting demand increases by up to 250% due to information provision. Thus, information more than triples the effectiveness of carbon offset subsidies.

The difference between STANDARD and INFORMATION delivers largely different conclusions about consumers’ valuation of carbon mitigation. Using the usual random utility model, I find that WTP is zero in STANDARD, but it is 16 EUR per ton of CO_2 ($p < 0.01$) in INFORMATION.

Mechanisms: Fairness Preferences vs. Perceived Effectiveness. The stark difference in WTP between STANDARD and INFORMATION raises the question of why price and quantity elasticities increase when the consumer learns about the firm’s participation in the offset. There are two possible mechanisms the experiment is able to separate from each other. First, consumers may value that the firm is contributing to the offset. Second, and more nuanced, consumers may perceive the quality of the offset differently when they learn that the firm invests its own resources into the project.

I first explore the role of quality beliefs in a post-purchase opinion survey. I find that when the firm matches the quantity, consumers fail to realize that the offset is more effective. However, this is true not just for subjects in STANDARD but also in INFORMATION. Thus, beliefs cannot explain why offsetting demand is fully quantity-inelastic in STANDARD but elastic in INFORMATION.

To isolate the role of fairness preferences, I leverage the experimental variation to accommodate two different fairness models. In the first specification, I assume that consumers care about the *relative* share that the firm contributes to the offset. Note that a 50% subsidy and a 100% match both imply that the firm is splitting the total offset costs

50:50 with the consumer. Similarly, both a 75% subsidy and a 300% match imply that the firm pays two-thirds of the total costs, i.e., the split with the consumer is 25:75. We can identify by how much this relative split between the consumer and the firm shift demand because we have more than one empirical moment for each split. Regressing demand on price, quantity, and the relative contribution of the firm, I find a large effect of fairness preferences on demand. Importantly, once we control for the relative split, consumers are fully quantity inelastic, again. This suggests that fairness preferences explain the entire treatment effect of information and consumers are still insensitive to scope. In the second specification, I estimate an alternative model in which the consumer receives (potentially nonlinear) utility from the absolute (instead of the relative) contribution by the firm. Once we control for this mechanism, I again find a null effect of compensated quantities on demand. Thus, consumers remain inelastic to impact under both model specifications, and implied WTP for carbon mitigation is zero.

Stated Preferences. Decades of research in environmental economics have elicited WTP estimates through “contingent valuation surveys” (Mitchell and Carson 1989) in which subjects are asked how much, hypothetically, they would be willing to pay to avoid the emission of one ton of carbon. Would we have gotten the same results from the field experiment had we just asked people for their preferences? 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/ton of CO₂. This is 1,388% larger than even my largest estimate of 16 EUR/tCO₂ in the information treatment. In addition, I exogenously vary the hypothetical compensation amount across subjects and find scope-insensitivity even for hypothetical choices. Only if compensation amounts are varied *within-subject* rather than *between-subject* do consumers increase their stated WTP as impact increases. This within-subject effect may imply that consumers are “relative thinkers” (Bushong, Rabin, and Schwartzstein 2021) and can only grasp differences in compensated quantities if options are presented directly next to each other (see also Bordalo, Gennaioli, and Shleifer 2013, Kőszegi and Szeidl 2013, Bordalo, Gennaioli, and Shleifer 2021). This finding calls for the development of new survey methods in environmental valuation studies to take cognitive limitations explicitly into account.⁴

⁴Surveys have proven an invaluable tool in eliciting economic beliefs (Stantcheva 2023) and could play a promising role in informing environmental policies.

Cost-Effectiveness: 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 if matches have no effect on demand.* This is because subsidies reduce the price for all consumers, but the only incremental increase in mitigation comes from *marginal* consumers. By contrast, matches also cause *inframarginal* consumers to mitigate more carbon. For subsidies to break even with matches, price elasticities would have to be substantially larger than they turn out to be empirically. The second important result is that matches have a “multiplier effect” when they are made salient through information: Every EUR spent by the firm on a quantity match produces a larger reduction in carbon than if that same EUR were directly invested into a carbon offset. These results may provide a motivation for governments to provide financial incentives to firms that offer carbon offsets. Recent global survey evidence shows that such targeted investment programs may receive more political support than traditional policies such as a general carbon tax (Dechezleprêtre et al. 2022).

Contributions to the Literature The paper makes two main contributions to the existing literature: i) It provides the first revealed preference estimates of WTP for carbon mitigation from a natural field experiment, and ii) it identifies warm glow utility and fairness preferences as underlying mechanisms of demand.

The existing literature in environmental economics has mostly used contingent valuation methods to elicit stated preferences for carbon mitigation (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). While some studies report modest values of 40 USD/tCO₂ (measured in 2020-USD), many studies imply large values between 100 and 350 USD/tCO₂.⁵⁶ My revealed preference estimates are at least an order of magnitude smaller, but I obtain similarly large values for stated preferences in the complementary survey. This result highlights the

⁵Some of these studies simply ask for subjects’ WTP to avoid carbon emissions, while others ask for WTP for a particular policy that mitigates carbon emissions. The revealed preference estimates in my study are more comparable to the former because I observe voluntary donations to climate protection (conditional on others free riding) rather than WTP for a policy.

⁶Some studies estimate WTP per year (instead of per ton of CO₂). Most estimates from these studies fall between 50 USD and 300 USD, with a mean of 167 USD (see Nemet and Johnson (2010) for a review.). To put this into perspective, the current carbon footprint per capita in the United States is estimated to be around 15 tCO₂ per year. This number used to be even larger in prior years when some of the studies were implemented.

importance of developing tools to mitigate hypothetical bias (Cummings, Harrison, and Rutström 1995, List and Gallet 2001).⁷ I show that one potential technique to align revealed and stated preferences is to randomize impact in the survey, as well, and estimate WTP based on this variation instead of using subjects' stated values.

A related literature has used lab and survey experiments to measure people's preferences for retiring pollution permits that trade under the EU ETS (Löschel, Sturm, and Vogt 2013, Diederich and Goeschl 2011, Diederich 2013). While these studies are important first steps towards obtaining revealed preferences, they only vary the permit price without modifying the compensated quantity. This difference turns out to be pivotal in my setting since consumers are fully impact-inelastic. I show that equating WTP for the offset with WTP for the carbon it mitigates overstates true WTP for carbon mitigation *by a factor of 19* or more because it ignores warm glow utility. My paper also improves upon these prior studies by collecting data from a natural field experiment, which may offer more accurate measures of consumer preferences in real-world markets.⁸

Another strand of literature studies the role of nonstandard preferences and cognitive constraints for sustainable consumption (e.g., Gosnell, List, and Metcalfe 2020, Pace and van der Weele 2020, Andre et al. 2022, Löschel, Rodemeier, and Werthschulte 2022, Rodemeier and Löschel 2022, List et al. 2022, Bilén 2022, Imai et al. 2022, Tilling 2023, Semken 2023). I add to this literature by providing the first evidence that environmental contributions are driven by warm glow. In addition, I show that many consumers are “conditional cooperators” (Gächter 2006) that can be persuaded to donate if the firm contributes, as well. This relates to a separate literature studying the role of subsidies and matching mechanisms in increasing charitable giving (e.g., Eckel and Grossman 2003, Kesternich, Löschel, and Römer 2016, Karlan and List 2007, Feldman 2010). Since my study varies fairness perceptions, it is the first to show that positive match elasticities may be entirely driven by fairness preferences rather than intrinsic preferences for the charitable good. The implication is that—different from price elasticities in markets with private goods—donation elasticities are unlikely to offer reliable measures of consumer surplus from the public good.

The rest of this paper is structured as follows. Section 2 presents the experimen-

⁷While List and Gallet (2001) find that hypothetical bias inflates WTP by a factor of 3, in my setting estimates are inflated by a factor of up to 12.

⁸A small set of studies have used observational market data to study preferences of other important public goods, such as clean air (Chay and Greenstone 2005, Ito and Zhang 2020) and water quality (Kremer et al. 2011).

tal design. Results are discussed in Section 3. In Section 4, I present insights from a complementary survey. Section 5 concludes.

2 Experimental Design

The experiment takes place in the webshop of one of the largest delivery services for groceries and beverages in Germany.⁹ 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 $z \in \{100\%, 300\%\}$.

The experimental design intentionally features an important 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. As will become more clear throughout the paper, this symmetry can be used to isolate subjects' fairness preferences from their intrinsic utility of mitigating carbon.¹⁰

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 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

⁹The time span of the experiment was 2 weeks in February 2020.

¹⁰In addition, the 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/tCO₂ both when the firm subsidizes the offset by 50% and when it matches the quantity by 100%.

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 CO₂ for 24 Cents.*” The text below informs subjects to which carbon-offsetting project the amount is donated.¹¹ In addition, subjects are informed that 2.4kg of CO₂ correspond to the average emissions of one delivery.¹² 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 CO₂ (i.e., 100 EUR/tCO₂). 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 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.¹³ A low offset price might also signal that the environmental damage of a delivery is negligible since it costs little to compensate 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

¹¹The project name is not mentioned in this paper to protect the company’s anonymity.

¹²Average emissions were calculated from historical trip data.

¹³See [Calel et al. \(2021\)](#) for empirical evidence about adverse selection in the carbon offset market.

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. Such behavior could be a problem for the identification of WTP if we made the mistake to interpret 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 context 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.

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 neither have an effect on the probability of buying at the store, nor on the type of goods that are being purchased. Therefore, differences in offsetting demand across treatments should 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 Effects on Offsetting Probability

I present differences in offsetting probabilities across treatments in Figure 4a. 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. 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 large 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 significant responsiveness to the impact of the offset.

An open question is whether the impact elasticity 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. The colored arrows in the graph may already foreshadow an answer to this question. As measured by the blue arrows, the effect of information is

identical between the 50% subsidy and the 100% match. This is precisely where the firm splits the total offset costs 50:50 with the consumer. If the firm raises its contribution to 25:75, demand increases incrementally by the size of the orange arrows. While this incremental distance is slightly larger for the match than for the subsidy, the magnitudes are again remarkably similar. Further, the difference between the orange arrows is not statistically significant. This may suggest that fairness preferences for the relative split in costs between the firm and the consumer are one of the main drivers underlying the effect of information. I investigate this mechanism further below by testing different models of fairness preferences and by studying effects on beliefs.

3.2 The Role of 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 slightly increases.¹⁴ However, the effect is very small: The 100% match increases perceived effectiveness in INFORMATION by only 6% (0.2 points on the Likert scale). Further, we do not observe any effect of information for the much larger 300% match. Even though consumers should clearly understand that the offset is more effective when the firm saliently matches the quantity, the effect on beliefs is non-monotonic in impact. These results make it unlikely that the previously observed insensitivity to scope in STANDARD and the increase in elasticities due to INFORMATION are explained by consumers' inattention towards the compensation

¹⁴These 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.

amount.

Overall, both Panel a) and b) point towards a limited role of changes in beliefs as the underlying forces of the information effect.

3.3 The Role of Fairness Preferences

To isolate fairness preferences from changes in beliefs, I leverage specific features of the experimental design. As I discuss below, the experiment allows us to test two general types of fairness models. In the first one, consumers care about the relative share that the firm pays. In the second one, the *absolute* size of the contribution (in EUR) rather than the relative share matters.

To see how we can test the first model, recall that there exists a symmetry in experimental design between matches and subsidies. Specifically, the experiment has been designed such that each subsidy level has a match for which the split between the consumer and the firm is identical (50:50 or 25:75). Therefore, there is a common factor between a subsidy and its respective match that we can control for. Put differently, the price the consumer pays is *not* perfectly collinear with the split between the firm and the consumer.¹⁵ This allows us to isolate the effect of fairness utility on demand from other factors, such as changes in beliefs about the offset quality. The identifying assumption is that for a given split (e.g., 50:50) a subsidy does not change beliefs differently than the respective match. This assumption seems reasonable since consumers learn essentially the same information in both cases. Specifically, they learn the split between the firm and the consumer and that the price of one kilogram of carbon is 10 Cents.

Given the identifying assumption, we can isolate the underlying mechanisms of the information effect. The empirical specification for this model is¹⁶

$$y_i = \alpha + \eta p_i + \beta q_i + \underbrace{F_1 \times \mathbb{1}_i(50 : 50) \times I_i}_{\text{Fairness from 50:50 split}} + \underbrace{F_2 \times \mathbb{1}_i(25 : 75) \times I_i}_{\text{Fairness from 25:75 split}} + \epsilon_i. \quad (1)$$

¹⁵While the price *per ton of CO₂* is perfectly collinear with the split between firm and consumer, the price the consumer pays for the offset is not. The latter price is what matters for identification.

¹⁶Alternatively, one could also specify a utility function that incorporates fairness preferences and estimate the underlying parameters, e.g. by estimating a conditional logit with utility function $u_i = \alpha + \eta p_i + \beta q_i + F_1 \times \mathbb{1}_i(50 : 50) \times I_i + F_2 \times \mathbb{1}_i(25 : 75) \times I_i + \epsilon_i$. This section focuses on the reduced-form linear model because it is primarily interested in approximating effect sizes of fairness preferences rather than quantifying structural parameters. Section 3.4 estimates structural parameters and also shows how the linear model above can be used for a structural interpretation.

Here, F_1 and F_2 measure the change in demand due to the 50:50 and 25:75 split, respectively. Note that these coefficients represent the *isolated* effect on demand that is solely driven by fairness preferences, not by changes in effectiveness beliefs. If the information treatment raised consumers' awareness of the increased effectiveness of a matched offset, then this change in belief should result in an increased quantity coefficient β , even after controlling for fairness preferences. If, on the other hand, information only increased quantity-elasticities due to fairness utility, then β should become zero when controlling for the split between consumers and firm.

The second fairness model follows a similar strategy. If consumers care about the absolute monetary amount spent by the firm, the experiment is rich enough to control for this, even if fairness utility is nonlinear in absolute contributions. With some abuse of notation, the following specification represents the second model:

$$y_i = \alpha + \eta p_i + \beta q_i + \underbrace{(F_1 \times C_i + F_2 \times C_i^2)}_{\text{Fairness from absolute contribution}} \times I_i + \epsilon_i.$$

Here, C is the absolute contribution by the firm measured in EUR. F_1 and F_2 now represent the linear and quadratic terms of the fairness utility function, respectively.

Regression results of both models are presented in Table 3. In the first model, the 50:50 and 25:75 split increase demand by, on average, 1.7 and 3.2 percentage points, respectively. Both effects are highly significant ($p < 0.01$). Consulting the price coefficient of -0.12 , the fairness effects correspond to respective subsidies of 0.14 and 0.26 EUR. To put this into perspective, the latter represents 27% of the total offset costs of the 25:75 match (0.96 EUR).

Importantly, once we control for the split between firm and customer, the quantity coefficient becomes statistically indistinguishable from zero. This suggests that even though consumers become quantity-elastic under INFORMATION, this is entirely driven by fairness utility and not by a perceived higher impact.

We obtain similar results for the second fairness model. The first-order effect is an increase in demand of 94 basis points for every 10 Cents the firm spends. The slope falls by 31 basis points for every further 10 Cents increase, implying demand is convex in price. Both fairness coefficients are highly significant ($p < 0.01$). Controlling for fairness utility again yields a quantity coefficient indistinguishable from zero (and even negative in sign).

Together with the limited effect of INFORMATION on beliefs (discussed in Section

3.2), these results provide strong evidence in favor of scope-insensitivity in carbon off-setting. Consumers only increase demand in INFORMATION because they appreciate the firm’s participation in the offsetting, not because they respond to the larger impact.

3.4 Willingness to Pay

In this section, I quantify WTP for carbon mitigation in STANDARD and INFORMA-TION. Given the prior results, the difference between WTP in the two groups is likely to measure consumer welfare from the firm’s contribution instead of intrinsic preferences for mitigation.

Estimation To estimate WTP, I rely on two estimation approaches. First, I use the standard random utility model with a logit error term (McFadden et al. 1973), which has also been extensively used in contingent valuation studies for public goods (Hanemann 1984). Second, I use a local approximation of WTP that does not rely on any distri-butional assumptions. The latter approach is commonly used in the sufficient statistics literature (Chetty 2009).

Consumer $i \in \{1, 2, \dots, I\}$ 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 at a total price of p_i . I make the usual assumption that p_i and γ_i enter linearly into utility.¹⁷ Utility is given by:

$$u_i = \alpha + \beta\gamma_i + \eta p_i + \epsilon_i. \quad (2)$$

The parameter β is the marginal utility of mitigating one ton of carbon, and η is the marginal disutility of price. α is an intercept. Idiosyncratic preferences are given by ϵ_i . If γ is measured in tons of carbon and p in euros, willingness to pay to mitigate one ton of carbon is given by $WTP = -\frac{\partial u}{\partial \gamma} / \frac{\partial u}{\partial p} = -\frac{\beta}{\eta}$ EUR. The consumer decides to buy the offset iff $u_i \geq 0$, meaning aggregate demand for the offset is given by $D(p, \gamma) = 1 - G(-\beta\gamma - \eta p)$.

Logistic Distribution Under the usual assumption that ϵ_i follows a logistic distribu-tion, the probability that consumer i chooses to buy the offset, denoted π_i , can be written

¹⁷In an unreported regression, I allow for nonlinearities in the utility function and cannot reject that they are statistically zero.

in closed form as

$$\pi_i = \frac{1}{1 + \exp(-\beta\gamma_i - \eta p_i)}. \quad (3)$$

The model parameters β and η can be estimated by maximum likelihood.

Linear Approximation An alternative approach that does not require a distributional assumption about ϵ_i is to linearly *approximate* WTP by reduced-form elasticities. Specifically, note that the derivatives of aggregate demand with respect to price and carbon quantity are $\frac{\partial D}{\partial p} = \eta g(-\beta\gamma + \eta p)$ and $\frac{\partial D}{\partial \gamma} = \beta g(-\beta\gamma + \eta p)$, such that WTP is given by $WTP = -\frac{\beta}{\eta} = -\frac{\partial D}{\partial \gamma} \frac{\partial D}{\partial p}$.

Denote the demand responses to price and carbon quantity variations by $\Delta_p D$ and $\Delta_\gamma D$, respectively. The demand derivatives can be approximated by $\Delta_p D / \Delta p \approx \eta g(-\beta\gamma + \eta p)$ and $\Delta_\gamma D / \Delta \gamma \approx \beta g(-\beta\gamma + \eta p)$, where the approximation requires that Δp and $\Delta \gamma$ are small, or alternatively, that demand is locally linear, in which case $g(\epsilon)$ is locally constant. WTP can therefore be approximated by

$$WTP \approx -(\Delta_\gamma D / \Delta \gamma) / (\Delta_p D / \Delta p). \quad (4)$$

WTP Estimates Estimation results for both the logistic regression and OLS are shown in Table 4. I estimate the model for STANDARD and INFORMATION separately. Subjects with the baseline offset are included in both estimations. Regression coefficients in columns 1 and 2 come from a logistic regression, while coefficients in columns 3 and 4 are produced by OLS. Implied WTP is the ratio of the quantity and price coefficient, multiplied by (minus) 1.

As shown in column 1, using the variation in STANDARD to estimate utility parameters in the logistic regression, we find that only the disutility of price, η , is significant. Utility from the compensation amount, β , is indistinguishable from zero, suggesting consumers do not value the carbon-mitigating attribute of the offset. As a result, WTP for mitigating a ton of carbon is statistically zero. Column 2 shows that consumers receive larger price disutility and larger, statistically significant utility from the carbon-offsetting attribute of the offset. The utility parameters translate into a WTP estimate of 16.44 EUR/tCO₂. This estimate is highly statistically significant with $p < 0.01$.

The linear approximation produces almost identical results. WTP is zero in STANDARD and 15.99 EUR/tCO₂ ($p < 0.01$) in INFORMATION. Results are similar to the

logistic regression because, empirically, aggregate demand turns out to be fairly linear (i.e., $g(\epsilon)$ is locally constant).

Comparison to Policy Assumptions. Given the prior results, it is likely that the estimate of 16 EUR/tCO₂ measures the welfare gains from the firm’s contribution rather than from carbon mitigation. However, one may wonder what this estimate implies for policy in case this interpretation is not correct, and the estimate actually represents intrinsic preferences for mitigation. In this case, we obtain a useful *lower bound* for people’s marginal utility of mitigation that we can compare to policy assumptions.¹⁸ Specifically, the estimate does not support assumptions made by the former Trump administration, which used a social cost of carbon of as low as 1 USD/tCO₂. However, the estimate supports the current assumptions by the Biden administration of 51 USD/tCO₂. It also supports the carbon price implemented through the EU ETS, which, at the time of the experiment, was 28 EUR/tCO₂ and has now increased to 85 EUR/tCO₂.

WTP for Mitigation vs. WTP for Offsets. In the second-to-last column, I report WTP for the offset itself instead of for the carbon it mitigates. This is identified by dividing the regression constant by the absolute price coefficient: $-\alpha/\eta$. This is the strategy that important prior laboratory experiments followed to identify WTP for carbon mitigation: they estimated WTP for carbon mitigation solely based on the offset price (i.e., had no quantity variation).

We can see that this intuitive approach is not applicable in my setting. For the logistic regression, average WTP for the offset is -1.56 EUR in STANDARD and -0.76 EUR in INFORMATION. Negative average WTP is to be expected because a large share of the sample does not buy the offset (even at a price of 6 Cents), and the logistic distribution of the error term, ϵ , allows for probability mass on negative values. If we now falsely equated WTP for the offset with WTP for carbon mitigation, we would incorrectly infer that consumers receive large positive utility from pollution. Since the baseline offset compensates 2.4kg of carbon, STANDARD would imply that average willingness to pay for carbon mitigation is $WTP = \frac{-1.77}{2.4} \times 1000 = -737.50 \text{ EUR/tCO}_2$. Similarly, in INFORMATION it would be $-279.16 \text{ EUR/tCO}_2$. Both values are unrealistic and far from those obtained in Table 4 that exploits quantity variation.

¹⁸The estimate gives a lower bound rather than a point estimate of marginal utility due free riding: the private provision of public goods will be below the social optimum (Samuelson 1954).

The linear approximation yields positive values for WTP but is still off by several orders of magnitude. The implied WTP for carbon mitigation would be $587\text{EUR}/t\text{CO}_2$ and $308\text{EUR}/t\text{CO}_2$. These estimates would dramatically overstate WTP. The estimate under INFORMATION is more than 19 times larger than the $16\text{EUR}/t\text{CO}_2$ obtained from accounting for quantity variations. These results highlight that WTP for carbon mitigation cannot be identified by exogenous variation in the offset price alone.

3.5 Cost-Effectiveness of Subsidies and 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 the firm but also for public policies that support corporate sustainability. If carbon offsets offered by firms to consumers are very cost-effective, then it may be efficient for governments to sponsor these programs.¹⁹

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 4a visualizes the results. 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 $10\text{kg}/\text{EUR}$). 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 $1\text{kg}/\text{EUR}$ and $2\text{kg}/\text{EUR}$ 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 $3.40\text{kg}/\text{EUR}$ and $3.60\text{kg}/\text{EUR}$, respectively. The

¹⁹This is particularly true if privately-offered carbon offsets are cost-effective but do not increase sales, such that the incentive to offer offsets may be missing in equilibrium.

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 policymakers may leverage corporate social responsibility to more efficiently invest into offset projects. A potential takeaway is that public funds should not be offered to firms to subsidize offsets they offer to consumers but rather for quantity matches.

4 Email Survey

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.²⁰ A translated version of the survey can be found in Appendix G.

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 CO₂, where the amount they see is randomly 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 CO₂. 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.

²⁰For privacy reasons, I cannot match survey participants to the observations in the field experiment.

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 CO₂. 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 CO₂ (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 F, I describe the sample in more detail and discuss observable characteristics. While I cannot exclude that subjects select on unobservables into the survey, observable statistics are fairly representative of the firm's customer population.

Results Table 5 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.²¹ 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/tCO₂ 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/tCO₂ (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/tCO₂ as our preferred estimate, the survey results would overstate WTP by 1,388%.²²

²¹More specifically, I use the 90th percentile of WTP *per tCO₂*. 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.

²²A 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

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 statistically significant. This suggests a limited role of information provision for WTP in line with prior studies (Pace and van der Weele 2020, Imai et al. 2022).²³

There is no statistically significant effect of raising the compensation amount by 2.4kg of CO₂ *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 become highly 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 behavioral phenomena. Specifically, the results suggest that consumers might not be able to compare magnitudes unless they are presented right after each other. This would be supported by theories of “relative thinking” (Bushong, Rabin, and Schwartzstein 2021) and salience (Bordalo, Gennaioli, and Shleifer 2013, Kőszegi and Szeidl 2013, Bordalo, Gennaioli, and Shleifer 2013) and may provide a new explanation for scope-insensitivity. The interpretation obviously matters for welfare: If consumers are insensitive to scope due to relative thinking, then their choices do not reflect true preferences. This contrasts with more traditional models of warm glow in which consumers do not respond to scope be-

sample of interest.

²³To complement this result, Appendix E 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.

cause they do not receive any utility from it.

A limitation of this interpretation is that within-subject differences can also be explained by experimenter demand effects in which subjects try to comply with the objective of the researcher. While recent evidence suggests that these effects are likely small (De Quidt, Haushofer, and Roth 2018), they cannot be ruled out within the scope of this survey.

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 WTP. Figure 6b plots the empirical distribution of WTP in EUR/ton of CO₂ 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/tCO₂, while 32% have a WTP of zero. By contrast, only 20% of carbon tax supporters have a WTP below 20 EUR/tCO₂ 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/tCO₂. 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? This paper investigates this question by leveraging a large-scale natural field experiment to estimate how demand for carbon offsets responds to exogenous variations in subsidies and matches.

I find that consumers are elastic to price but fully inelastic to simple variations in impact. This result indicates that consumers buy the offset but do not value the carbon it mitigates. 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/tCO₂. The experimental design allows us to quantify that this effect is mostly driven by fairness preferences rather than by intrinsic preferences for mitigation. Once we correct for fairness considerations, implicit WTP for carbon mitigation is zero.

Given the strong support for environmental policies in the general population, these results cast doubt on whether the market for carbon offsets can yield reliable estimates of people's preferences. Stated preferences from a complementary survey heavily diverge from revealed preferences. Additional tests of scope-insensitivity point to models of relative thinking and salience as new and unexplored mechanisms. The development of techniques aimed at obtaining people's 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.
- 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." *The economic journal* 100 (401): 464–477.
- Bilén, David. 2022. "Do carbon labels cause consumers to reduce their emissions? Evidence from a large-scale natural experiment." Technical Report, Mimeo, Gothenburg University.
- Bishop, Richard C, Kevin J Boyle, Richard T Carson, David Chapman, W Michael Hanemann, Barbara Kanninen, Raymond J Kopp, Jon A Krosnick, John List, Norman Meade, et al. 2017. "Putting a value on injuries to natural assets: The BP oil spill." *Science* 356 (6335): 253–254.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2013. "Salience and Consumer Choice." *Journal of Political Economy* 121 (5): 803–843.
- . 2021. "Salience." Technical Report, National Bureau of Economic Research.
- 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.
- 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.
- Carson, Richard T, Robert C Mitchell, W Michael Hanemann, Raymond J Kopp, Stanley Presser, Paul A Ruud, et al. 1992. “A contingent valuation study of lost passive use values resulting from the Exxon Valdez oil spill.” Technical Report, University Library of Munich, Germany.
- Chay, Kenneth Y, and Michael Greenstone. 2005. “Does air quality matter? Evidence from the housing market.” *Journal of Political Economy* 113 (2): 376–424.
- Chetty, Raj. 2009. “Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods.” *Annual Review of Economics* 1 (1): 451–488.
- Cummings, Ronald G, Glenn W Harrison, and E Elisabet Rutström. 1995. “Home-grown values and hypothetical surveys: is the dichotomous choice approach incentive-compatible?” *The American Economic Review* 85 (1): 260–266.
- Dechezleprêtre, Antoine, Adrien Fabre, Tobias Kruse, Bluebery Planterose, Ana Sanchez Chico, and Stefanie Stantcheva. 2022. “Fighting climate change: International attitudes toward climate policies.” Technical Report, National Bureau of Economic Research.
- De Quidt, Jonathan, Johannes Haushofer, and Christopher Roth. 2018. “Measuring and bounding experimenter demand.” *American Economic Review* 108 (11): 3266–3302.
- 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

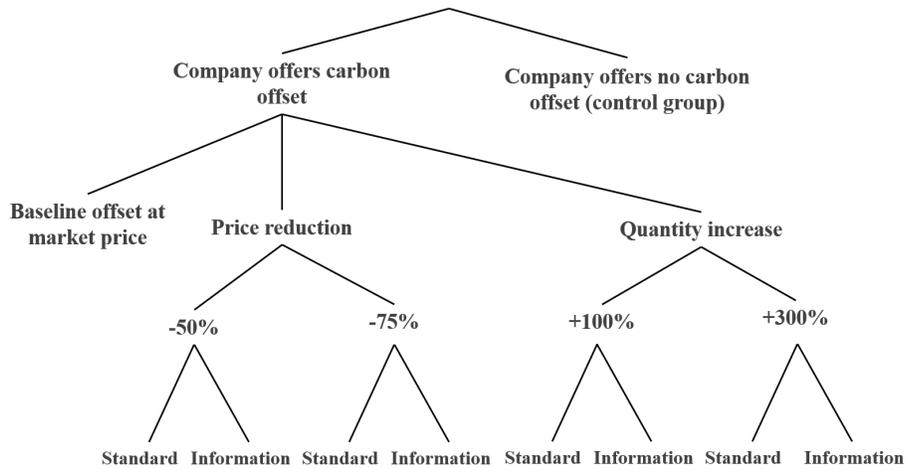
- we subsidize charitable contributions matter?” *Journal of Public Economics* 87 (3-4): 681–701.
- Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2022. “The preference survey module: A validated instrument for measuring risk, time, and social preferences.” *Management Science*.
- Feldman, Naomi E. 2010. “Time is money: Choosing between charitable activities.” *American Economic Journal: Economic Policy* 2 (1): 103–130.
- Gächter, Simon. 2006. “Conditional cooperation: Behavioral regularities from the lab and the field and their policy implications.” Technical Report, CeDEx Discussion Paper Series.
- Gosnell, Greer K, John A List, and Robert D Metcalfe. 2020. “The impact of management practices on employee productivity: A field experiment with airline captains.” *Journal of Political Economy* 128 (4): 1195–1233.
- Hanemann, W Michael. 1984. “Welfare evaluations in contingent valuation experiments with discrete responses.” *American journal of agricultural economics* 66 (3): 332–341.
- 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.”
- Ito, Koichiro, and Shuang Zhang. 2020. “Willingness to pay for clean air: Evidence from air purifier markets in China.” *Journal of Political Economy* 128 (5): 1627–1672.
- 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.
- 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.
- Kőszegi, Botond, and Adam Szeidl. 2013. “A model of focusing in economic choice.” *The Quarterly journal of economics* 128 (1): 53–104.
- Kremer, Michael, Jessica Leino, Edward Miguel, and Alix Peterson Zwane. 2011. “Spring cleaning: Rural water impacts, valuation, and property rights institutions.” *The Quarterly Journal of Economics* 126 (1): 145–205.
- List, John, Matthias Rodemeier, Sutanuka Roy, Gregory Sun, et al. 2022. “Judging Nudging: Toward an Understanding of the Welfare Effects of Nudges Versus Taxes.” Working Paper.
- List, John A, and Craig A Gallet. 2001. “What experimental protocol influence disparities between actual and hypothetical stated values?” *Environmental and resource economics* 20 (3): 241–254.
- Löschel, Andreas, Matthias Rodemeier, and Madeline Werthschulte. 2022. “Can Self-Set Goals Encourage Resource Conservation? Field Experimental Evidence from a Smartphone App.” Working Paper.
- 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.”
- Mitchell, Robert Cameron, and Richard T Carson. 1989. *Using surveys to value public goods: the contingent valuation method*. Rff press.
- Nemet, Gregory F, and Evan Johnson. 2010. “Willingness to pay for climate policy: a review of estimates.” no. 2010-011.
- Ottoni-Wilhelm, Mark, Lise Vesterlund, and Huan Xie. 2017. “Why do people give? Testing pure and impure altruism.” *American Economic Review* 107 (11): 3617–3633.
- Pace, Davide, and Joël J van der Weele. 2020. “Curbing Carbon: An Experiment on Uncertainty and Information About CO₂ Emissions.”

- Rodemeier, Matthias, and Andreas Löschel. 2022. “Information Nudges, Subsidies, and Crowding Out of Attention: Field Evidence from Energy Efficiency Investments.” Working Paper.
- Samuelson, Paul A. 1954. “The pure theory of public expenditure.” *The review of economics and statistics*, pp. 387–389.
- Semken, Christoph. 2023. “The Marginal Impact of Emission Reductions: Estimates, Beliefs and Behavior.” Working paper.
- Stantcheva, Stefanie. 2023. “How to run surveys: A guide to creating your own identifying variation and revealing the invisible.” *Annual Review of Economics* 15:205–234.
- 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.
- Wekhof, Tobias, and Sébastien Houde. 2023. “Using narratives to infer preferences in understanding the energy efficiency gap.” *Nature Energy* 8 (9): 965–977.

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

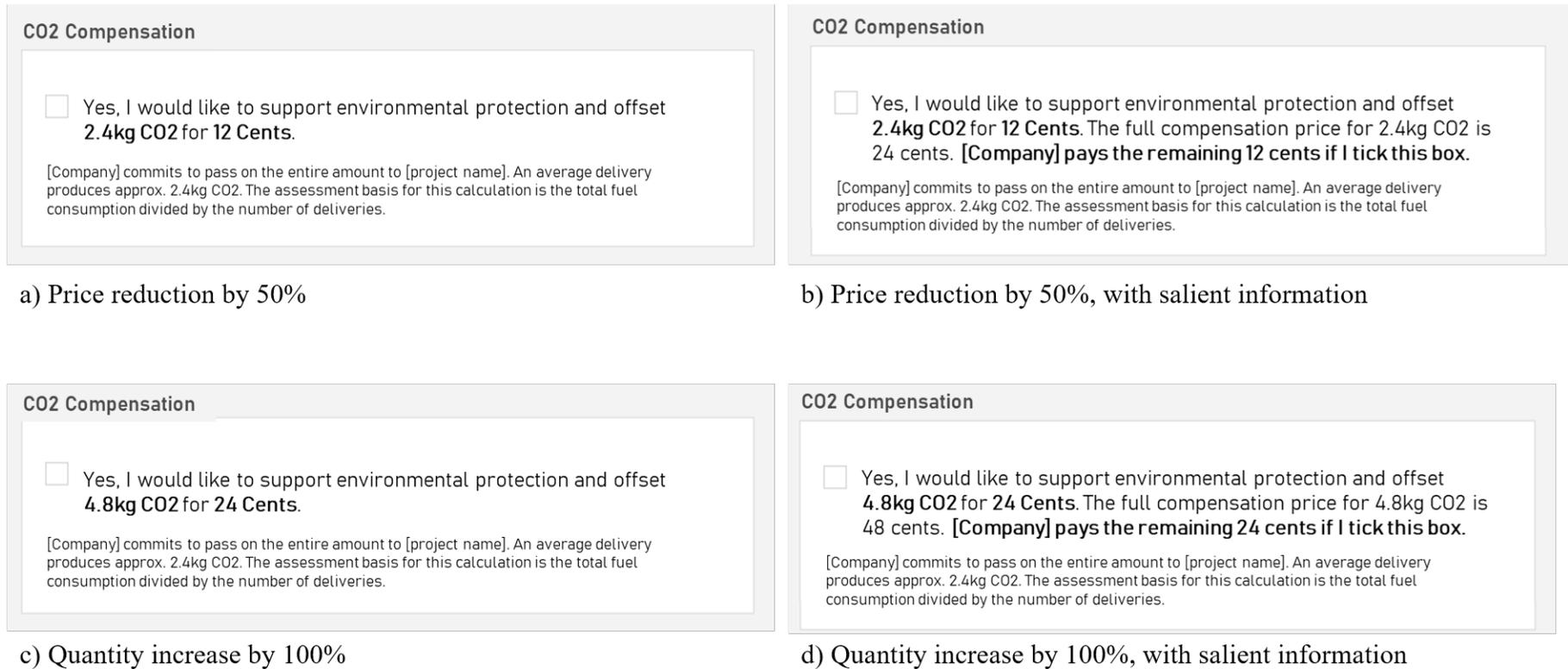
CO2 Compensation

Yes, I would like to support environmental protection and offset **2.4kg CO2 for 24 Cents.**

[Company] commits to pass on the entire amount to [project name]. An average delivery produces approx. 2.4kg CO2. The assessment basis for this calculation is the total fuel consumption divided by the number of deliveries.

Note: This figure shows the baseline offset.

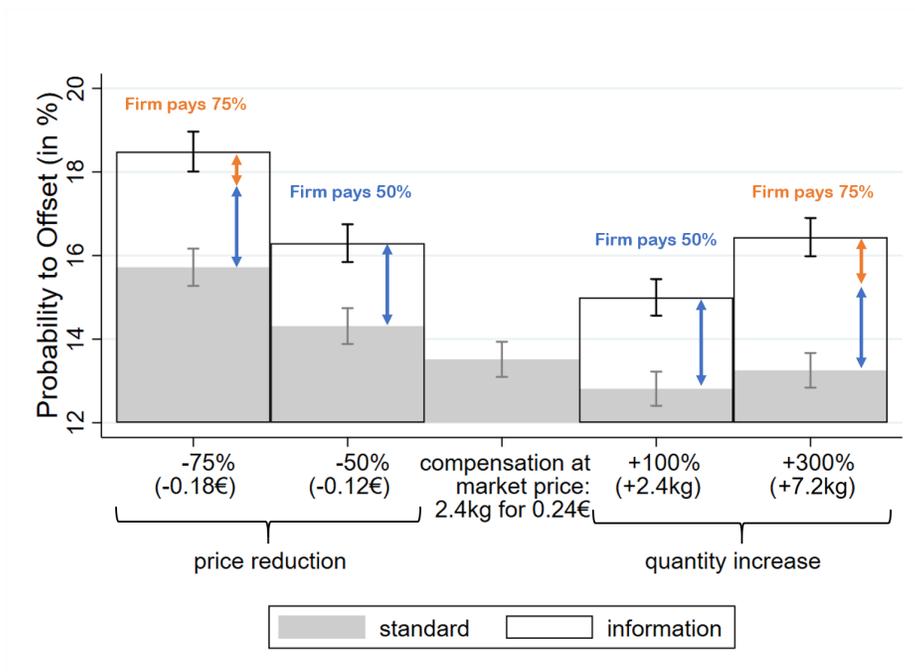
Figure 3: Examples of Treatment Variation



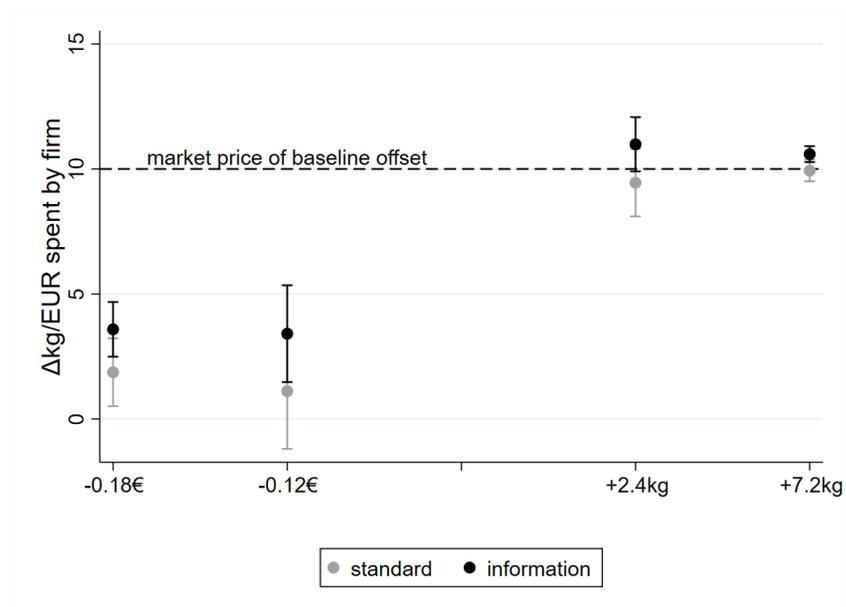
Note: This figure shows examples of price and quantity variations. Panel a) and c) illustrate the variations in the standard treatments, whereas panel b) and d) illustrate the variations in the information treatments.

Figure 4: Main Results

(a) Offsetting Probabilities

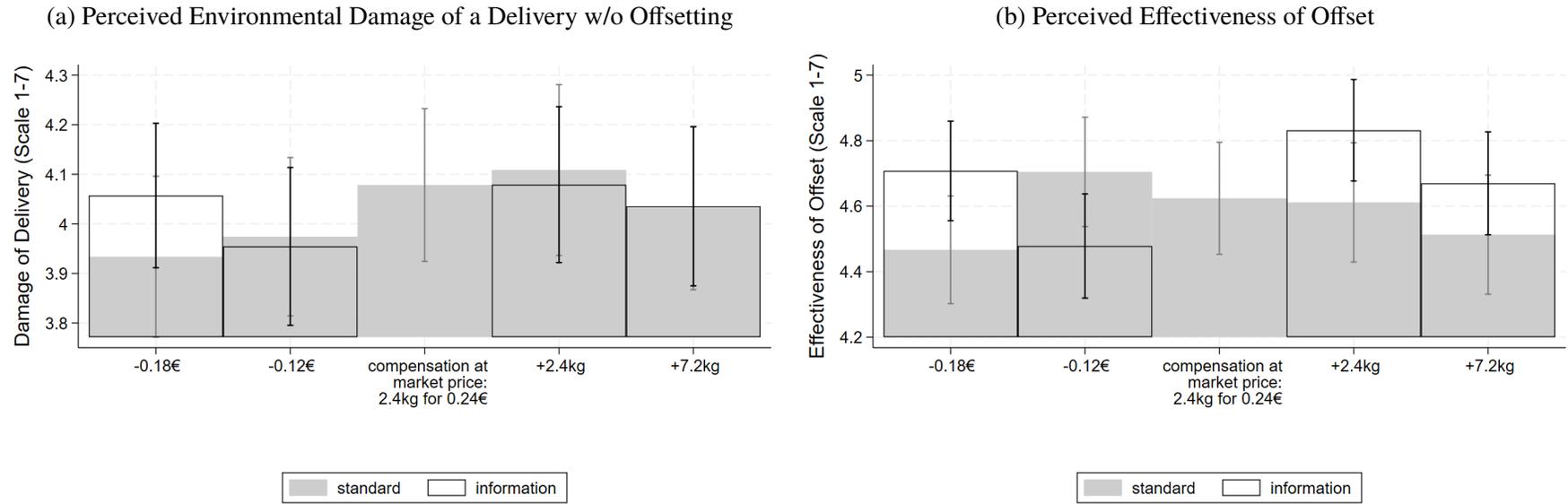


(b) Cost Effectiveness



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).

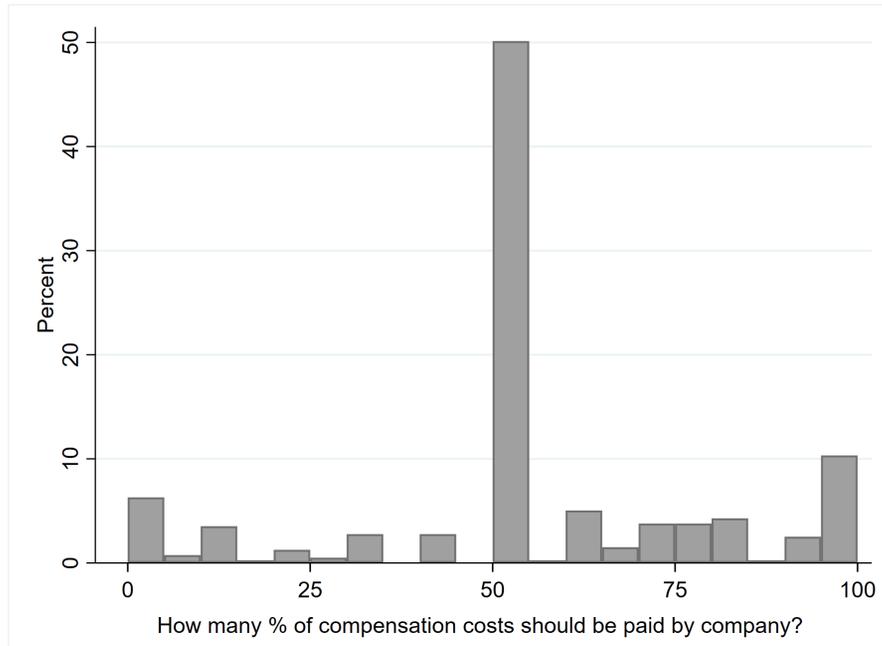
Figure 5: Post-Experimental Survey



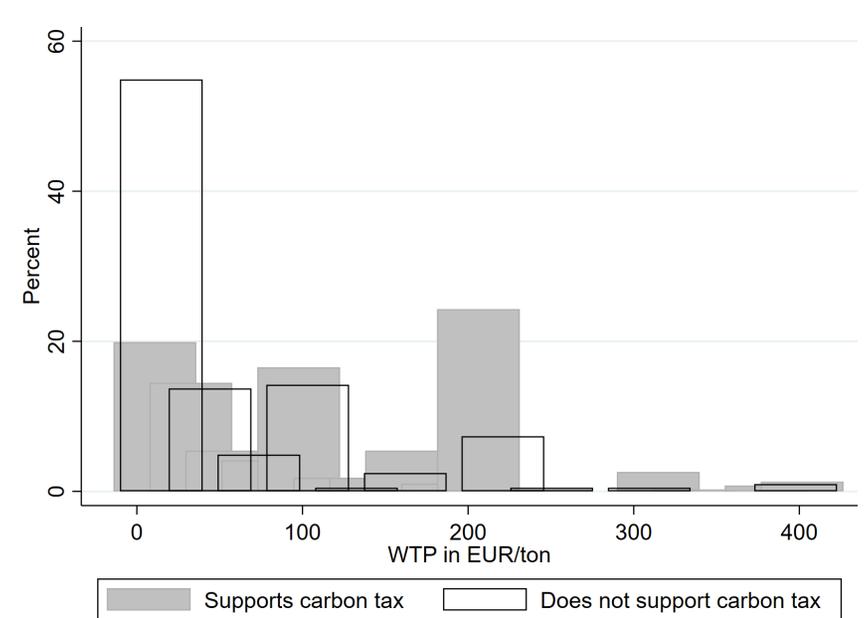
Notes: Panel a) illustrates subjects' beliefs about the size of the environmental damage of one delivery that is not compensated by an offset. Panel b) shows beliefs about the effectiveness of the offset in mitigating environmental damages.

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 CO₂ 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 (1.365)	1.596 (1.828)	1.595 (1.390)	1.588 (1.350)	1.602 (1.415)
Order (1= yes)	0.329 (0.470)	0.330 (0.470)	0.333 (0.471)	0.327 (0.469)	0.333 (0.471)
Offset (1= yes)	0.000 (0.000)	0.135 (0.342)	0.143 (0.350)	0.157 (0.364)	0.163 (0.369)
Expected travel time (in min)	14.508 (9.397)	14.366 (9.433)	14.498 (10.110)	14.509 (9.582)	14.561 (9.825)
Expected service time (in min)	7.201 (3.817)	7.260 (3.650)	7.304 (4.071)	7.282 (3.815)	7.296 (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 (1.617)	1.591 (1.526)	1.598 (1.492)	1.592 (1.462)	1.598 (1.449)
Order (1= yes)	0.332 (0.471)	0.333 (0.471)	0.334 (0.472)	0.330 (0.470)	0.331 (0.471)
Offset (1= yes)	0.185 (0.388)	0.128 (0.334)	0.133 (0.339)	0.150 (0.357)	0.164 (0.371)
Expected travel time (in min)	14.525 (9.546)	14.442 (9.319)	14.428 (9.464)	14.470 (9.562)	14.685 (9.832)
Expected service time (in min)	7.371 (3.855)	7.334 (3.781)	7.305 (4.048)	7.285 (3.921)	7.302 (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: Probability to Place an Order

	Order Probability $\times 100$	
	(1)	(2)
Baseline: 24 Cents, 2.4kg	-0.104 (0.407)	0.107 (0.416)
-0.12€	0.494 (0.434)	0.419 (0.416)
× information	0.063 (0.433)	0.394 (0.416)
-0.18€	0.096 (0.379)	-0.209 (0.415)
× information	-0.062 (0.446)	0.316 (0.417)
+2.4kg	-0.145 (0.303)	0.391 (0.416)
× information	-0.268 (0.435)	0.107 (0.416)
+7.2kg	0.219 (0.466)	0.531 (0.416)
× information	-0.192 (0.368)	0.216 (0.418)
Constant: No offset offered	26.643*** (4.254)	32.917*** (0.294)
N	406,984	255,376

Note: This table reports treatment effects on the probability to place an order among website visitors. The first column includes all website visits, whereas the second column only includes the first visit of a subject during the experimental period. Standard errors are in parentheses. *, **, ***: significant at $p < 0.1$, $p < 0.05$, $p < 0.01$, respectively.

Table 3: Isolating Fairness Preferences

	(1) Relative Contribution	(2) Absolute Contribution
Quantity (β)	0.001 (0.007)	-0.001 (0.007)
Price (η)	-0.122 *** (0.025)	-0.182 *** (0.023)
<i>Fairness Parameters:</i>		
F_1	0.017 *** (0.003)	0.094 *** (0.017)
F_2	0.032 *** (0.004)	-0.031 *** (0.013)
Constant (α)	0.132 *** (0.003)	0.130 *** (0.003)
N	76229	76229

Note: This table reports price and quantity coefficients while holding fixed fairness preferences. Column 1 assumes that consumers receive fairness preferences from the relative share that the firm contributes to the total offset costs. Column 2 assumes that fairness preferences are a (nonlinear) function of the absolute contribution. *, **, ***: significant at $p < 0.1$, $p < 0.05$, $p < 0.01$, respectively.

Table 4: Willingness to Pay for Carbon Mitigation

	(1)	(2)	(3)	(4)
	Logit		OLS	
	Standard	Information	Standard	Information
Quantity	-0.29 (6.08)	31.42*** (5.75)	-0.02 (0.71)	4.07*** (0.76)
Price	-1.05*** (0.22)	-1.91*** (0.21)	-0.13*** (0.03)	-0.25*** (0.03)
Constant	-1.63*** (0.04)	-1.46*** (0.03)	0.16*** (0.04)	0.19*** (0.03)
WTP in €/tCO ₂	-0.27 (5.83)	16.44*** (2.47)	-0.18 (5.59)	15.99*** (2.51)
WTP for offset itself (in €)	-1.56*** (0.35)	-0.76*** (0.10)	1.27*** (0.24)	0.74*** (0.07)
N	42,440	42,186	42,440	42,186

Note: This table reports regression coefficients and implied WTP for carbon mitigation. Coefficients in column 1 and 2 are from a logistic regression, coefficients in column 3 and 4 from OLS. Implied WTP is the absolute ratio of the quantity and price coefficients. The second-to-last row shows WTP for the offset, computed by the absolute ratio between the regression *constant* and the price coefficient. Standard errors for WTP are obtained by the delta method and reported in parentheses. *, **, ***: significant at $p < 0.1$, $p < 0.05$, $p < 0.01$, respectively.

Table 5: Hypothetical WTP

	(1) Total WTP (in Cents)
<i>Quantity increase between-subject:</i>	
+2.4kg	-7.857 (9.891)
+4.8kg	-0.996 (9.984)
<i>Quantity increase within-subject:</i>	
+2.4kg	31.243*** (2.559)
+4.8kg	32.149*** (3.463)
<i>Between-subject variation in Education and Fairness Treatments:</i>	
+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)
N	1,617
WTP in EUR/tCO ₂	237.50*** (28.13)

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 CO₂. 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

The image shows a screenshot of a survey form. At the top, a dark grey box contains the text: "Thank you! On 25.02.2020 at 10:30 am, your order is on its way on." Below this, the form is divided into sections. The "Confirmation" section contains a text box with the message: "An order confirmation should have arrived in your e-mail inbox." The "Your opinion" section contains two questions, each with a 7-point Likert scale. The first question is: "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?" The second question is: "How effective do you think is our carbon offset program in reducing these negative consequences?" Both scales range from 1 to 7, with labels "< very low" and "very high >" for the first question, and "< not effective at all" and "very effective >" for the second. A "Send" button is located at the bottom right of the form.

Thank you!
On 25.02.2020 at 10:30 am, your order is on its way on.

Confirmation

An order confirmation should have arrived in your e-mail inbox.

Your opinion

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?

1 2 3 4 5 6 7

< very low very high >

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

1 2 3 4 5 6 7

< not effective at all very effective >

Send

B Additional Tables

Table B1: 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

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

C Firm Outcomes

C.1 Buying Probability

To analyze whether the treatments affected demand for deliveries, I estimate a linear probability model, regressing whether a subject placed an order on the treatment vectors. Table 2 reports the regression results. In column 1, I include both between- and within-subject variation. I add subject-fixed effects and cluster standard errors on the visit-level (i.e., on the level of randomization). In column 2, I only consider between-subject variation, i.e., a subject's first visit to the website during the experimental period. As pre-registered in the pre-analysis plan, I focus on between-subject variation when analyzing offsetting behavior. This also turns out to be a reasonable approach ex-post since most of the variation in offsetting comes from between-subject and little from within-subject variation.²⁴

The probability of ordering at the shop is 27% for the whole sample and 33% during the first visit. All treatment coefficients in both columns are economically small and tightly estimated null effects. This suggests that offering website visitors a carbon offsetting program does not affect demand for deliveries. Below, I also show that offsets do not affect product demand and revenues.

A reassuring implication from these results is that differences in offsetting behavior conditional on placing an order have a causal interpretation because treatments do not cause systematic selection from the sample of website visitors to the subsample of customers. Therefore, I proceed to analyze the subsample of subjects that placed an order.

C.2 Product Demand and Revenue

In this section, I analyze treatment effects on aggregate product demand and revenue. Table C1 reports treatment coefficients from an OLS regression that includes all observations during the first visit in the experimental period. Treatment effects are to be interpreted in percent relative to control. There is no noticeable treatment effect of any of the offsets on aggregate demand or revenue. Coefficients are economically small and statistically indistinguishable from zero. Together with previous results, this suggests

²⁴In particular, subjects rarely change their offsetting behavior relative to the first visit, meaning that, in a panel regression, subject-fixed effects would absorb most of the variation. Specifically, the between-subject standard deviation for the offsetting probability is 33.9%, whereas the within-subject standard deviation is only 4.8%.

that offering carbon offsets affects neither customer conversion nor product demand of existing customers.

Table C1: Treatment Effects on Product Demand and Revenue

	(1) Aggregate Demand (in %)	(2) Revenue (in %)
P0Q0: 24c, 2.4kg	-1.160 (0.938)	-1.256 (0.986)
P1Q0: 12c, 2.4kg	0.356 (0.978)	-0.519 (0.994)
P2Q0: 6c, 2.4kg	-0.750 (0.957)	-0.727 (1.006)
P1Q0+Info: 12c, 2.4kg	-0.854 (0.931)	-0.994 (0.987)
P2Q0+Info: 6c, 2.4kg	0.224 (0.956)	-1.043 (0.985)
P0Q1: 24c, 4.8kg	0.073 (0.934)	0.021 (1.000)
P0Q2: 24c, 9.6kg	0.803 (0.973)	0.921 (1.045)
P0Q1+Info: 24c, 4.8kg	0.274 (0.958)	0.653 (1.013)
P0Q2+Info: 24c, 9.6kg	-0.546 (0.962)	-0.605 (0.997)
N	83,859	83,859

Note: This table presents treatment effects on product demand in percent relative to control. Aggregate demand is defined as the total number of quantities (of all goods) purchased. Robust standard errors are in parentheses.

D Deviations from Pre-Analysis Plan

In this section, I document any deviations from the Pre-Analysis Plan (PAP). The experiment, including the PAP, were registered at the AEA RCT Registry under trial number AEARCTR-0005375.

The experiment lasted for a period of 2 weeks but was initially scheduled to last for 4 weeks. The company had to shorten the experimental time frame due to unexpected business commitments unrelated to the experiment. The company had forecasted around 300,000 transactions for a time period of 4 weeks. Based on an ex-ante assumed offsetting probability of 0.3, the experiment was powered to detect effects larger than 1.05 percentage points.

The shortening of the experimental time period reduced the number of transactions to around 100,000 subjects. Fortunately, this large number still provided a well-powered study. As can be inferred from the confidence intervals in Figure 4a, all coefficients are precisely estimated. Any null effect has an upper bound of the 95%-CI that is below 1.8 percentage points, which is close to the ex-ante targeted minimum detectable effect size.

The PAP also included extensions of the structural estimation of preference parameters, which I decided to exclude from the paper for brevity.

E Additional Results from Email Survey

Beliefs about Carbon Emissions and Offsetting

Table E1 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 E1: Answers to Belief Questions in Email Survey

	(1) Delivery	(2) Car	(3) Trees	(4) Certainty
Education treatment about carbon offsetting	-0.409 (0.826)	-11.363*** (3.577)	-85.991*** (22.477)	2.323*** (0.139)
Constant	5.463*** (0.543)	30.288*** (2.471)	144.844*** (15.893)	2.354*** (0.095)
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 E2 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 E1 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 an important barrier to voluntary climate protection.

Table E2: WTP and Demographics

	(1) Total WTP (in EUR)
Willingness to take risk	0.043*** (0.010)
<i>Age:</i>	
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)
<i>Gender:</i>	
diverse	-0.169 (0.295)
female	0.308*** (0.049)
<i>Employment Status:</i>	
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)
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.

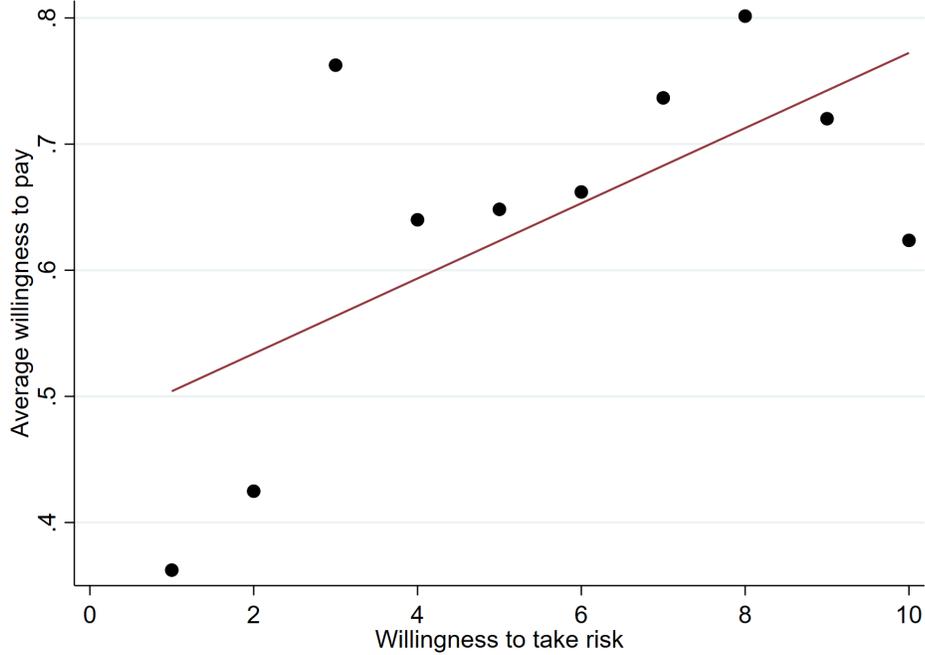


Figure E1: Correlation between Risk Preferences and WTP

F Sample Characteristics

Table B1 in the Appendix reports observable characteristics of the sample of respondents.²⁵ 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%).²⁶ 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.

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

²⁵I present statistics for the sample included in the analysis, excluding outliers, as described further below.

²⁶For national statistics see <https://www.destatis.de/EN/>.

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.

G Email Survey

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