IZA Institute of Labor Economics
Initiated by Deutsche Post Foundation

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

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

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

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

* 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

Climate change poses arguably one of the most existential threats to society. In 2015, policymakers of 195 countries acknowledged this challenge in the Paris Agreement and committed to limiting global warming below 2 degrees relative to pre-industrial levels. While many policies that could bring the drastic change necessary to meet this target are not implemented, firms and households have started to engage in *voluntary* reductions of carbon emissions. Overall, the amount of voluntary carbon compensation by consumers and firms has increased by over 100% from 2017 to 2020.[1] Much of this reduction comes from investments into "carbon offset" projects that engage in reforestation, which removes carbon dioxide (CO_2) from the air. Firms increasingly offer consumers the possibility to directly compensate the carbon emissions of their consumption, e.g., for flights or product shipping.

What does this market for voluntary carbon mitigation tell us about people's preferences for climate protection? Learning about people's valuation of environmental protection is crucial for understanding the welfare implications of environmental policies. In a standard economic framework, the larger a household's willingness to pay (WTP) for carbon mitigation, the larger are 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.[3] The growing market for voluntary climate protection provides a unique opportunity to obtain first data points of households' revealed preferences.

Since voluntary contributions to a public good should be below marginal benefits (Samuelson 1954), WTP estimates from the carbon-offsetting market should be interpreted as *lower bounds* of people's marginal benefits of climate protection. These lower bounds provide important information for the current policy debates about the social benefits of carbon mitigation and the size of the optimal carbon price.

[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 5, 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, such as of 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). One alternative approach to WTP elicitations are Integrated Assessment Models (IAMs), which combine economic and physical climate change models (see, e.g., Nordhaus 2010, Tol 2002a, Tol 2002b). These models directly allow for an estimation of the social cost of carbon, but rely on many assumptions, such as future population growth, technological change, and climate sensitivity. Simple WTP elicitations rely on fewer assumptions and provide important information to complement these models.

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 other differences in offset characteristics that explain offset demand. For instance, low-priced offset projects have faced criticism for not planting enough trees to compensate for the promised amount of carbon emissions. Additionally, some carbon offset programs finance activities that may have taken place regardless of the donation, rendering the impact of the donation on emissions negligible (Calel et al. 2021). The second challenge is that even exogenous price variation of the same offset is not sufficient for identification of WTP. The reason is that consumers might value other attributes of the carbon-offsetting project, such as where in the world it plants trees and creates new jobs. 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. 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 vary both the price of the offset and the quantity of carbon that is compensated by the offset. Specifically, the baseline offset compensates the average emissions of a delivery: 2.4kg of CO_2 for a price of 24 cents (i.e., mitigating 1kg of carbon costs 10 cents). In order to vary price and quantity, either the price of the offset is reduced 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.

I cross-randomize whether the subsidy or match are made salient through an additional information treatment. In the standard treatments, henceforth STANDARD, consumers who receive a subsidy simply see a lower offset price, but are not informed that the true price is larger and that the firm subsidizes the offset. Analogously, consumers who receive a quantity match simply see a higher compensation amount. If consumers are able to make a fully-informed trade-off between price and quantity, this variation suffices to elicit WTP in the standard economic framework. However, if consumers have a poor understanding of carbon offsetting, they may not realize when the relative price of carbon compensation changes. In the information treatments, henceforth INFORMATION, the firm provides salient information to the consumer that the price has been subsidized or the quantity has been matched. INFORMATION creates two key differences relative to STANDARD. First, beliefs about and attention to the offset itself may be different. Consumers might only realize that the price per unit of carbon has changed when the firm makes subsidies and matches salient. Since "kilograms of carbon" is an abstract measure of environmental protection, quantity variations may be hard to comprehend for consumers. In addition, consumers might trust the quality of the underlying carbon-offsetting project more when they learn the delivery company is willing to invest its own resources into the offset. Second, in INFORMATION consumers learn that the firm contributes to the offset, which may be perceived as fairer and result in increased willingness to pay.

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 would mean that consumers buy the carbon offset but not because of its impact on environmental protection. This conclusion would be consistent with "warm glow" (Andreoni 1990) and "scope-insensitivity" (Kahneman and Knetsch 1992), in which consumers care about the act of giving to a public good but not about the size of the contribution. However, the result is also consistent with the notion that consumers do not understand kilograms of carbon as a measure of environmental protection and, therefore, *appear* indifferent between different compensated quantities.

Preferences change completely under INFORMATION. When consumers are actively informed that the compensated quantity of the offset is larger, 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. Valuations in STANDARD could, therefore, give rise to environmental policies that undervalue the true benefits of carbon mitigation to consumers.

What explains the effect of INFORMATION on price and quantity elasticities? I explore this question in two different surveys. In a first post-purchase opinion survey, customers are asked how effective they consider the offset to be in reducing externalities. When the compensation amount is increased in STANDARD, consumers fail to realize that the offset is more effective. By contrast, in INFORMATION, consumers' perceived effectiveness of the offset increases as the compensated amount is elevated. This can explain why offsetting demand is fully quantity-inelastic in STANDARD but elastic in INFORMATION. In fact, changes in beliefs can also capture why demand becomes more price-elastic in INFORMATION. When the price falls in STANDARD, consumers believe the offset to be less effective. This is not the case in INFORMATION: consumers maintain a firm belief of the offset's effectiveness when they know the true price of the offset is higher than what they pay. In a second survey, I explore the role of fairness preferences and altruism. I find that most consumers consider it fair when the firm covers 50% of the compensation costs. Thus, the effect of INFORMATION seems to be driven by both changes in beliefs about the offset effectiveness, as well as by fairness preferences.

In the survey, I also study how much *hypothetical* WTP deviates from revealed preferences. This is important because most existing estimates of WTP for carbon mitigation come from "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. The mean stated WTP in the survey is 200 EUR/ton of CO2. This is 1,150% larger than the revealed preference estimate of 16 EUR/tCO2. Therefore, in my setting, the contingent valuation method does not provide a good approximation of revealed preferences.

The revealed preference estimates could also provide lower bounds for the social cost of carbon (SCC).[4] The estimate of 16 EUR/tCO2 ($\approx$ 18 USD/tCO2 in 2020-USD) is larger than the SCC estimate used by the former Trump administration, which went as low as 1 USD/tCO2. Reassuringly, my estimates are substantially smaller than the SCC used under the current Biden administration of 51 USD (Interagency Working Group 2021).[5] My estimates are also lower than the prices of pollution permits traded under the EU Emissions Trading System (EU ETS)—the cap-and-trade system of the EU. Prices during the experiment in February 2020 were 28 EUR/tCO2 and are 76 EUR/tCO2 as of November 2022, i.e., considerably larger than the lower bounds provided in this paper. In sum, current (but not former) assumptions that guide policy-making are largely supported by the results of this experiment.

What does offering carbon offsets imply for the firm? Despite the substantial demand for carbon offsets, I do not find a causal effect on deliveries: Offering carbon offsets in the shop

[4] In the first best, a consumer's WTP would equal the marginal benefit she receives from one unit less of carbon. WTP would then equal the marginal rate of substitution between income and utility from lower carbon emissions.

[5] This estimate is likely to be corrected further upward soon, partially due to a decrease in the headline interest rate from 3% to 2%. See also Carleton and Greenstone (2021) for a discussion.

(relative to not offering them) has a precisely estimated null effect on the probability of ordering a delivery. Importantly, this is true independent of the firm's contribution to the offset. On the one hand, this result suggests that incentives for firms to engage in these programs may be small. On the other hand, it also suggests that offering carbon offsets does not have a negative effect on delivery demand by steering consumers' attention to the polluting attribute of a delivery.

I then analyze which intervention is the most cost-effective in reducing carbon emissions. The first important insight is that quantity matches are always more cost-effective than subsidies. The reason is that, with subsidies, the only incremental increase in mitigation comes from marginal consumers, whereas, with matches, the incremental increase also comes from inframarginal consumers. Price elasticities would have to be substantially larger for subsidies to break even with matches in terms of cost-effectiveness. 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).

The paper makes three main contributions to the existing literature: i) It provides the first revealed preference estimates of WTP for carbon mitigation from a natural field experiment, ii) it illustrates the crucial role of making subsidies and matches salient and explores the underlying mechanisms of the information effect (beliefs, attention, and fairness), iii) it estimates the effects of offering carbon offsets on firm performance.

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.[6][7] The revealed-preference estimates in this paper are an order of magnitude smaller, but I obtain similarly large values for stated preferences in the comple-

[6]Some 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.

[7]Some studies estimate WTP per year (instead of per ton of CO_2). 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.

mentary survey. This result highlights the importance of eliciting revealed preferences and is consistent with evidence on hypothetical bias in stated preferences (Cummings, Harrison, and Rutström 1995, List and Gallet 2001).[8]

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). Different from my setting, these studies elicit WTP for the *permit*, which is generally not the same as WTP for the carbon it mitigates. In particular, the studies vary the permit price but do not have exogenous variation in the compensated quantity. This difference turns out to be pivotal: WTP for the offset is vastly different from WTP for the carbon it compensates in my study. I show that using price variation alone would overstate WTP for carbon mitigation by a factor of 19 or more. Thus, identification of WTP for carbon mitigation requires exogenous variation in compensated quantities.

My paper also improves upon these prior studies by collecting data from a "natural field experiment" (Harrison and List 2004): consumers in their natural market environment make decisions, not knowing they are being observed by a researcher. This approach may offer more accurate measures of consumer preferences in real world markets (Levitt and List 2007).

Another contribution of this study is to show that the salience of subsidies and matches is a crucial driver of WTP. An insight from my findings is that consumers only realize that the impact of the offset is larger when matches are made salient. These findings highlight the importance of making matches salient for the elicitation of WTP for environmental quality (Hanemann 1984). They are also relevant for models of attention formation, which study the role of salience for consumer preferences (Chetty, Looney, and Kroft 2009, Bordalo, Gennaioli, and Shleifer 2013, Kőszegi and Szeidl 2013, Bordalo, Gennaioli, and Shleifer 2022).

Another insight from the information treatments is that consumer beliefs about the offset effectiveness depend on the offset's price. These results particularly relate 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. My findings provide empirical support for this hypothesis: Information that the firm contributes to the offset increases both donations and perceived quality of the offset.[9]

This paper also relates to the literature studying the effectiveness of subsidies and matching mechanisms in increasing charitable giving (e.g., Eckel and Grossman 2003, Karlan and List

[8]While 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.

[9]The importance of beliefs for consumer choices also relate to recent studies suggesting that beliefs about the carbon emissions of consumer goods matter for consumption choices (Pace and van der Weele 2020, Imai et al. 2022).

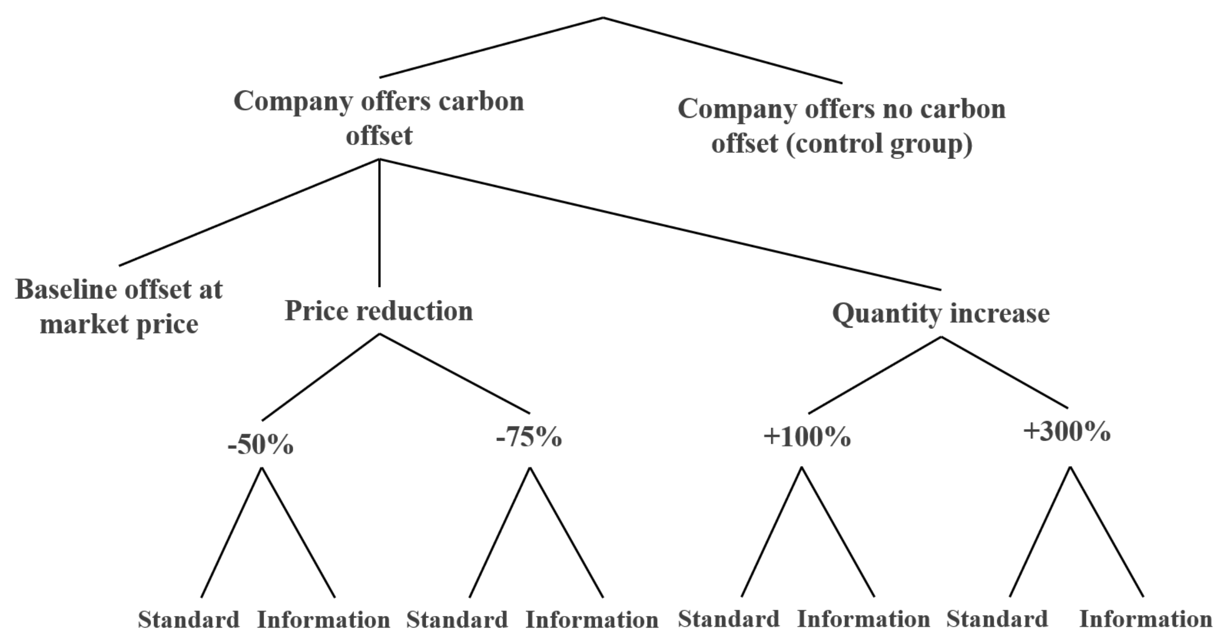
2007), including donations to climate protection (Kesternich, Löschel, and Römer 2016). The contribution of the paper to this line of work is to study how information provision changes donation elasticities and the cost-effectiveness of both subsidies and matches.[10]

Finally, this paper relates to a small set of prior studies that have used observational data to identify revealed-preference estimates of other public goods, such as clean air (Chay and Greenstone 2005, Ito and Zhang 2020) and water quality (Kremer et al. 2011).

The rest of this paper is structured as follows. Section 2 presents the experimental design. Reduced-form results are discussed in Section 3. In Section 4, I estimate WTP for carbon mitigation from the data. Results from a complementary survey are shown in Section 5. Section 6 concludes.

2 Experimental Design

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.

[10]The finding also relates to papers studying how "nudges," such as informational interventions, complement or substitute traditional price interventions (Rodemeier and Löschel 2022, Löschel, Rodemeier, and Werthschulte 2022, List et al. 2022)

The experiment takes place in the webshop of one of the largest delivery services for groceries and beverages in Germany.[11] 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 reduced by $x \in \{50\%, 75\%\}$ or the amount of carbon that the offset compensates is increased by $z \in \{100\%, 300\%\}$. Following the terminology in the literature, I refer to the first set of treatments as *subsidies* and to the second set of treatments as *matches*.

As I discuss in more detail below, I also vary whether subsidies or quantity matches are simply applied to the offset or whether they are made salient through an additional information treatment.

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 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.[12] In addition, subjects are informed that 2.4kg of CO_2 correspond to the average emissions of one delivery.[13] 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. I will later return to the question of whether consumers understand these measures by leveraging two post-purchase surveys.

The donation goes to a reforestation project that plants trees to compensate for carbon emis-

[11] The time span of the experiment was 2 weeks in February 2020.

[12] The project name is not mentioned in this paper to protect the company's anonymity.

[13] Average emissions were calculated from historical trip data.

sions. At the time of the experiment, it cost 0.10 EUR to compensate one kg of CO_2 (i.e., 100 EUR/tCO_2). Thus, one average delivery that emits 2.4kg can be compensated by 0.24 EUR.

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

Panel b) shows the INFORMATION treatment where the firm explicitly informs the consumer that the true offset price of 2.4kg is larger and that the firm has subsidized the price by 12 Cents. The additional information provides a number of 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 make consumers more aware that they can compensate for emissions.

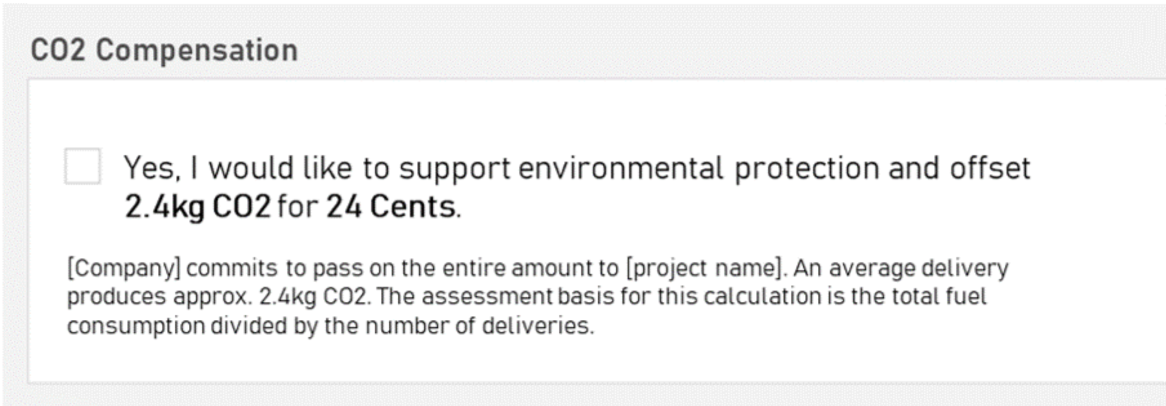
Third, information may change beliefs about the offset's effectiveness. The lower price in STANDARD relative to BASELINE may signal to consumers that the offset project is not effective at compensating carbon, for example, because they plant fewer trees than would actually be necessary. A low offset price might also signal that the environmental damage of a delivery is negligible since it costs little to compensate it. By contrast, in INFORMATION, subjects should be aware that the actual price of the offset is higher than the costs they have to cover. Subjects should, therefore, be less likely to use the lower price as a signal that the quality of the offset or the environmental damage of a delivery are low. In addition, consumers might trust the offset project more if they learn that the company donates its own resources to the project.

Panel c) shows an example of a quantity match. The price is equal to the one of the baseline offset, i.e., 24 Cents. However, the quantity is doubled from 2.4kg to 4.8kg of CO_2. Therefore, this treatment provides exogenous variation in the *impact* of the offset. Note that any exogenous change in quantities implies, by definition, that the compensation amount deviates from emissions of the average delivery. Subjects in this group, therefore, compensate 2 instead of 1 delivery in expectation.[14]

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

[14]Note that the carbon offset is a donation to environmental protection, and the quantity increase raises the effectiveness of the donation per EUR spent. One might be concerned that subjects have an aversion to donating more than they are emitting. This aversion should be mitigated in the INFORMATION treatments where subjects learn that it is actually the company who is compensating the second delivery. In appendix D, I further show that there are no heterogeneous treatment effects based on how much a delivery emits in carbon, as measured by the distance of the customer's home to the warehouse of the company.

Figure 2: Carbon Offset

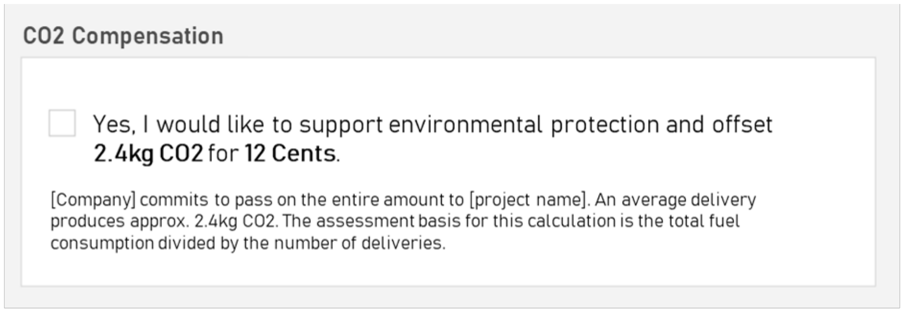


Note: This figure shows the baseline offset.

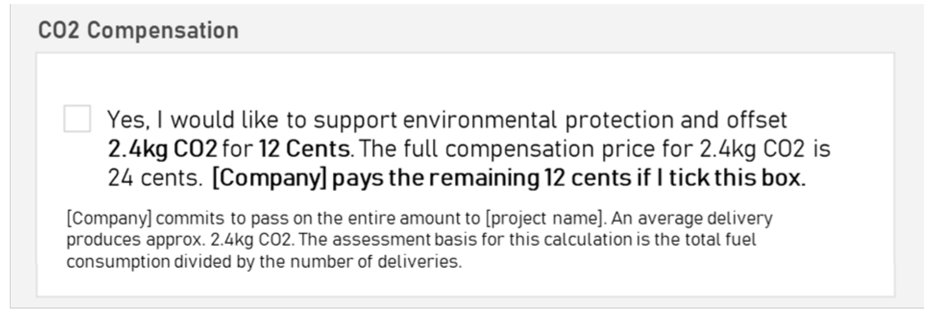
Cents. The reason they are paying half of the amount is that the company pays the remaining 24 Cents.

INFORMATION may be particularly important for the quantity match because kilograms of compensated carbon is a measure that may be abstract to consumers and difficult to understand. If consumers have an imperfect understanding of carbon offsetting, they might not realize in panel c) that the impact of the offset is higher. The information treatment in Panel d) increases this awareness as consumers are actively informed that the company has raised the impact.

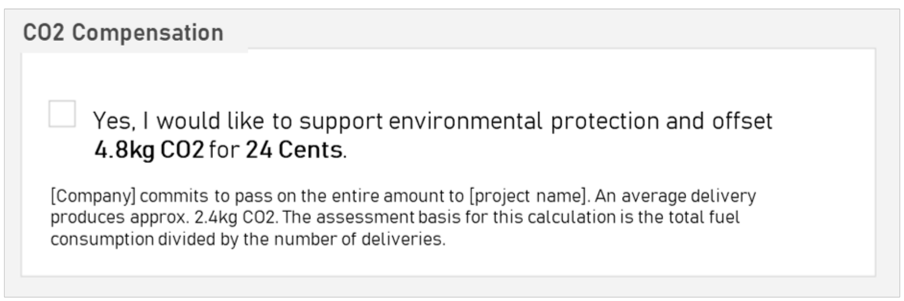
Figure 3: Examples of Treatment Variation



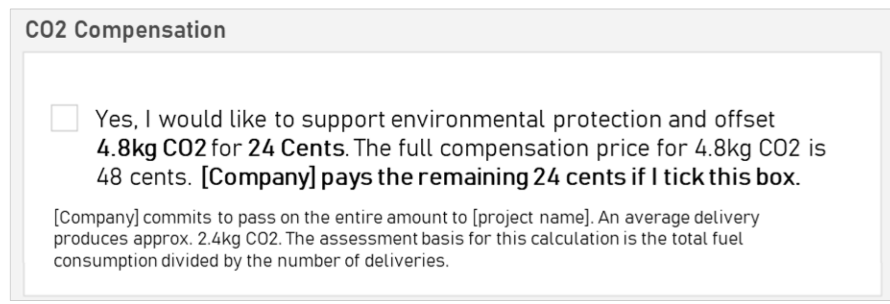
a) Price reduction by 50%



b) Price reduction by 50%, with salient information



c) Quantity increase by 100%



d) Quantity increase by 100%, with salient information

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

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 9 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 see whether subjects interpret a low price as a signal of low effectiveness of the offset. The question also allows us to observe 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 group that has been offered the baseline offset, 12.9% of all subjects who placed an order also chose to buy the offset. In the treatment groups that receive a subsidy or a match, the offsetting probability varies. I analyze these treatment effects using linear regressions in the next section.

Table 1: Summary Table

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

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

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

3 Reduced-Form Results

3.1 Effects on Delivery Demand

Before studying differences in offsetting behavior across treatments, I analyze whether the treatments affected demand for deliveries. This is important for two reasons. The first reason is an economic one: companies may want to offer carbon offsets to convince more consumers to buy at their stores. Consumers with stronger environmental preferences may only be willing to order online if the delivery is carbon-neutral. However, firms may hesitate to offer offsets because it could steer attention to the emissions of a delivery and, thereby, reduce sales.[15] Initially inattentive customers may become aware of the polluting attribute of the delivery and may decide not to order online but rather purchase goods in a more environmentally sustainable way (e.g., by walking or biking to the supermarket).[16]

The second reason relates to causal inference: If treatments affect whether subjects place an order, differences in offsetting behavior conditional on placing an order do not necessarily have a causal interpretation. This happens when the treatments change the type of subjects that select into the subsample of customers, and the type correlates with the offsetting probability. Thus, to identify treatment effects on offsetting demand (without any additional assumptions), we require no effect of the treatments on delivery demand.

To address these questions, I estimate a linear probability model, regressing whether subject i placed an order, $y_i \in \{0, 1\}$, on the treatment vectors:

$$y_i^{order} = \alpha + \rho_1 P_i + \rho_2 P_i \times information + \mu_1 Q_i + \mu_2 Q_i \times information + \xi_i. \qquad (1)$$

Here, $P_i = (\mathbb{1}_{\Delta p=-0.12}, \mathbb{1}_{\Delta p=-0.18})$ is a vector with indicators, where each indicator equals one if subject i received the respective price reduction, and zero otherwise. Analogously, $Q_i = (\mathbb{1}_{\Delta q=+2.4}, \mathbb{1}_{\Delta q=+7.2})$ indicates quantity increases. The indicator $information$ equals one if the respective subsidy or quantity match have been made salient.

Table 2 reports the regression results. In column 1, I include both between- and within-subject variation. This means that all 406,984 website visits, including follow-up visits by the

[15] In discussions with other companies I contacted, many were hesitant to run this field experiment for that particular concern.

[16] Selection on the extensive margin can be quantitatively important. In a different context, Rodemeier (2020) shows that consumers select out of an online shop when the shop tries to exploit cognitive biases, such as inattention.

same subject, are included in the regression. 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.[17]

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. In Appendix B I also show that offsets have no effect on product demand and revenues. Importantly, the previously mentioned concern that subjects may be deterred from ordering because they become more attentive to the environmental consequences of a delivery is not confirmed for this sample. The flip side of the results is that offsets did not convince more website visitors to place an order. These are important result since they suggest that firms may not have a sufficient incentive to provide carbon offsets to consumers. Importantly, not even INFORMATION increased demand, despite the firm highlighting its contribution to the offset. It may, of course, still be possible that a more salient advertisement of the offset could have increased delivery demand.

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.

[17] In 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%.

Table 2: Probability to Place an Order

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

3.2 Effects on Offsetting Probability

I present differences in offsetting probabilities across treatments visually in Figure 4 and in terms of regression coefficients in Table 3. Results in the table are produced by the linear probability model:

$$y_i^{offset} = \zeta + \tau_1 P_i + \tau_2 P_i \times information + \omega Q_i + \omega_2 Q_i \times information + \epsilon_i \qquad (2)$$

where $y_i^{offset} \in \{0, 1\}$ is one if subject i bought the offset, and zero otherwise.

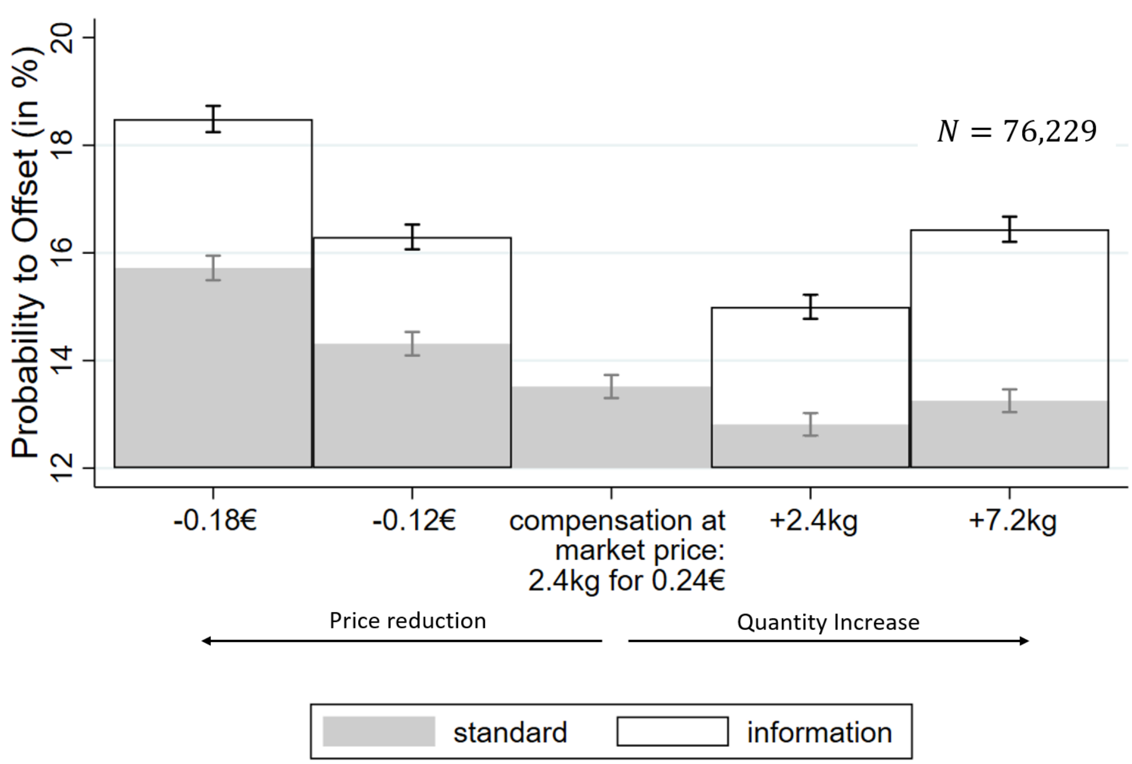
Figure 4 illustrates the offsetting probabilities across treatments. The grey bars indicate the offsetting probabilities for standard price and quantity variations, as well as for the baseline offset. The transparent bars show offsetting probabilities for the salient price and quantity variations.

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. Table 3 shows that only the larger price reduction is statistically significant at conventional levels.

When information is added to the same price variations, demand becomes substantially more price-elastic. 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.

We observe an even more pronounced pattern for quantity variations. As Figure 4 shows, 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. Table 3 indicates that these decreases 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 would yield an odd conclusion: Consumers buy carbon offsets but not because of how much carbon they offset. WTP for carbon mitigation would be zero. This conclusion would be 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

Figure 4: Offsetting Probability among Customers



Note: The figure represents the offsetting probabilities across treatments. The gray bars refer to the standard treatments, while the transparent one refer to the information treatments. The error bars represent standard errors.

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" (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.

In fact, results change entirely when consumers are explicitly informed that the firm has matched the compensation amount. 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. Thus, in the information treatments, consumers become quantity-elastic.

An open question that I explore later with survey answers is whether consumers become quantity-elastic only because they realize that the impact of the offset is larger or also because they value that the firm is contributing to the offset.

Table 3: Offsetting Probability

	(1) Offsetting Probability $\times 100$
-0.12€	0.796
	(0.532)
$\times$ information	2.779***
	(0.547)
-0.18€	2.203***
	(0.546)
$\times$ information	4.970***
	(0.564)
+2.4kg	-0.704
	(0.520)
$\times$ information	1.481***
	(0.538)
+7.2kg	-0.264
	(0.522)
$\times$ information	2.922***
	(0.551)
Constant: Baseline offset, 2.4kg for 24 Cents	13.517***
	(0.373)
N	76,229

Note: This table reports average treatment effects on the offsetting probability. Robust standard errors are in parentheses. *,**,***: significant at $p < 0.1$, $p < 0.05$, $p < 0.01$, respectively.

Demand Functions

To illustrate the stark difference in price- and quantity-elasticities, I estimate demand functions based on STANDARD and INFORMATION. For each subsample that received either STANDARD or INFORMATION treatment, I run the regression

$$y_i^{offset} = \alpha + \eta_1 p_i + \eta_2 p_i^2 + \beta_1 q_i + \beta_2 q_i^2 + \epsilon \tag{3}$$

where p_i is the price subject i has been offered in cents, and q_i the compensated quantity of the offset in kilograms.[18] I add squared terms to capture nonlinearities in demand. I then predict values of offsetting demand for different prices and quantities. Figure 5 plots the results. Demand functions for STANDARD are shown in blue, demand functions for INFORMATION in red.

The left panel shows demand as a function of price for an offset that compensates 2.4kg of carbon. Demand is more price-elastic with INFORMATION, and the difference becomes particularly pronounced for larger price reductions. Both functions are fairly linear, with the demand curve in STANDARD appearing slightly convex.

The right panel shows demand as a function of compensated quantities for an offset that costs 24 Cents. In STANDARD, demand is slightly decreasing in quantities but statistically flat. When quantity matches are made salient through INFORMATION, the demand elasticity becomes positive and economically large. While the demand function in INFORMATION looks slightly nonlinear in quantities, I cannot reject that it is linear statistically.

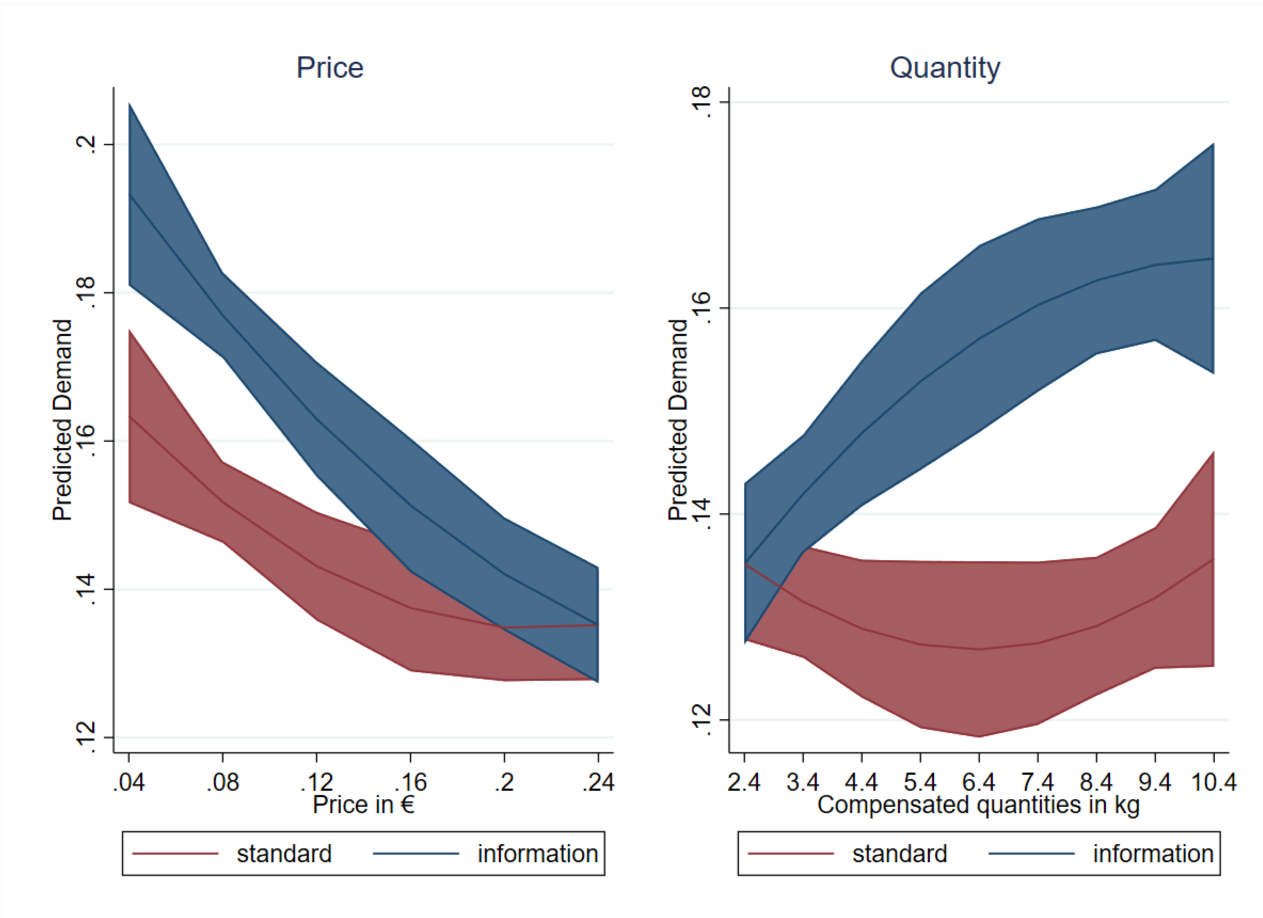
The large differences in elasticities between STANDARD and INFORMATION have important implications. First, if policymakers or firms want to subsidize carbon-offsetting programs, subsidies are substantially more effective if they are made salient through an information treatment. Second, quantity matches are completely ineffective in changing offsetting demand if no additional information is provided. Third, if we want to estimate WTP for mitigating a kilogram of carbon, the demand functions under STANDARD and INFORMATION will result in vastly different values. I explore this latter point in more detail in Section 4.

3.3 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 policy. If carbon offsets offered by firms to consumers are very cost-effective, then it may be optimal

[18]Subjects in the group that receives the baseline offset are included in both regressions.

Figure 5: Demand as a Function of Price and Quantity



Note: The figure plots estimated demand as a function of price and quantity. Demand functions for standard treatments are in red, while demand functions for information treatments are in blue.

for governments to sponsor these programs. This is particularly true if privately-offered carbon offsets are cost-effective but do not increase sales (as we have seen), such that the incentive to offer offsets may be missing in equilibrium.

To quantify cost-effectiveness, I estimate a ratio that allows us to conclude which intervention maximizes compensated carbon per invested EUR by the firm. Specifically, 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. The interpretation of this number is that it is the *incremental* increase in compensated carbon of the intervention per EUR spent by the firm.

Figure 6 visualizes the results. The gray line marks the market price if the firm directly buys the offset instead of offering it to consumers (i.e., the baseline price of 10kg/EUR).

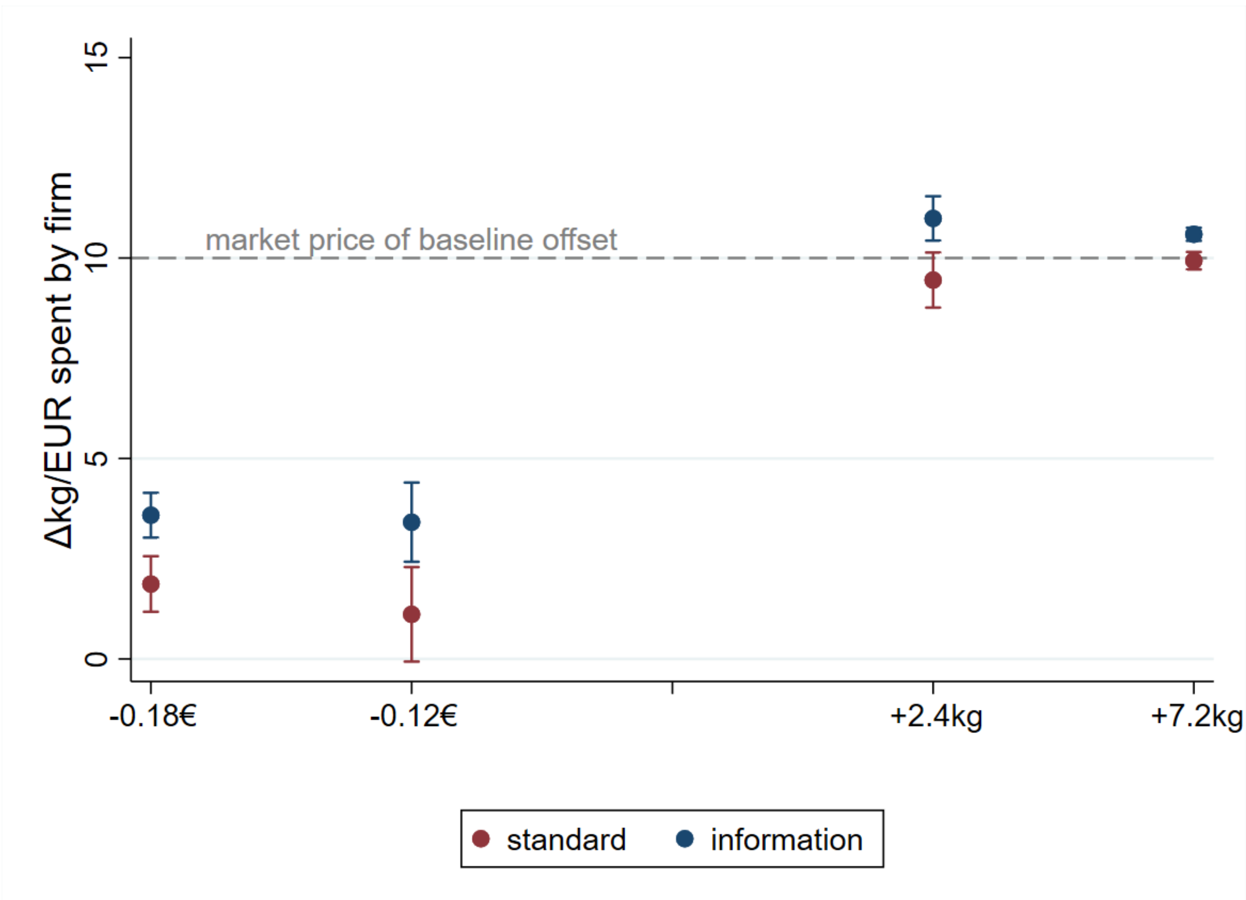
Quantity matches are overall more cost-effective than subsidies. The reason is that with subsidies, the only incremental increase in compensated carbon comes from marginal consumers. By contrast, with matches, the increase in compensated carbon also comes from inframarginal consumers since every offset now compensates a larger amount. Price elasticities would have to be much larger for subsidies to be more cost-effective than quantity matches.

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

By contrast, quantity matches in STANDARD just break-even with the market price of 10kg/EUR and are thereby more than 2.5 times more cost-effective than subsidies. Only the quantity matches in INFORMATION are able to offset more carbon per EUR spent. In particular, 1 EUR spent by the firm compensates around 11kg of carbon, i.e., 10% more than if the euro were invested directly into the baseline carbon offset.

Assume for a moment that these results also applied to offsets offered by the firm that are implicitly subsidized by the government, which may be a strong generalization. Then a takeaway from these results would be that public funds should not be offered to firms to subsidize the offsets they offer to consumers but rather for quantity matches.

Figure 6: Incremental Kilograms Compensated per EUR spent by Firm (Relative to Baseline Offset)



Note: This figure plots the increase in compensated kilograms per EUR spent by the firm, relative to the baseline offset.

3.4 Effects on Beliefs

Next, I study how the two belief measures elicited in the post-purchase survey vary with the treatments. I regress each outcome variable on price, compensated quantity, information, and interaction terms. The regression specification is the same as in equation 3 but with one of the belief measures as the outcome variable. Instead of estimating the regression for STANDARD and INFORMATION separately, I include interaction terms indicating whether subsidies and quantity matches have been made salient.

Table 4 reports the results. The first column shows the effects on the beliefs about the effectiveness of the offset, which is a measure of impact. The second column illustrates the effects on beliefs about the environmental damage of a delivery. In total, 3,600 customers answered the belief questions during their first visit. In the baseline treatment, the quality of the offset and the damage of a delivery are, on average, believed to be 3.79 and 3.87 respectively.

Column 1 shows that decreasing the price in STANDARD causes consumers to believe that the offset is less effective. Reducing the price by 1 Cent reduces the perceived effectiveness of the offset by 2.9% (i.e., 0.11 of 3.79). This negative effect of a lower price on quality perceptions disappears with INFORMATION, as indicated by the interaction term of -0.17. This implies that consumers use the price as a signal of the effectiveness of the offset. These results can explain why INFORMATION increases price elasticities: Demand in STANDARD is not sufficiently price-elastic because consumers incorrectly reduce the perceived effectiveness of the offset.

The coefficient of $Price^2$ is statistically significant and indicates that the change in beliefs is concave in price. This means that the incremental decrease in quality beliefs becomes smaller for larger price reductions. The information treatment fully offsets this concavity.

Increasing the compensated quantity in STANDARD has no effect on quality perceptions. This is consistent with the notion that consumers do not understand compensated kilograms of carbon as a measure of impact. When the quantity increase is made salient in INFORMATION, consumers update their beliefs and understand that the offset has become more effective. The interaction effect "Quantity $\times$ information" equals 0.55 and is highly statistically significant ($p < 0.01$). Under INFORMATION, a 1kg increase in the compensation amount increases the perception of the effectiveness of the offset by 17% relative to baseline (i.e., $\frac{0.08+0.55}{3.79}$).

Under INFORMATION, beliefs are concave in compensated quantities, meaning that the incremental increase in perceived quality gets smaller as we increase the compensated quantities.[19]

These changes in beliefs can explain why demand changes from quantity-inelastic to elastic

[19] Concavities may also be a ceiling effect of the Likert-scale. Once perceived effectiveness is high, it cannot increase much more as the compensated quantity increases.

as we move from STANDARD to INFORMATION. Consumers only understand that the effectiveness of an offset has been increased when the compensated quantity is increased in a salient way. Otherwise, they perceive the offset to be equivalent to the baseline offset. Alternatively, the findings may also imply that consumers trust the offset impact more when they learn that the firm is investing its own funds into the offsets. This is consistent with theoretical predictions in a model by Vesterlund (2003) in which donors' trust in the charity increases as the fundraiser announces its own contributions to the charity.

In sum, I find reduced-form evidence that beliefs about the effectiveness of the offset are a function of price, quantity, and information.

Finally, column 2 provides no evidence that consumer beliefs about the environmental damage of a delivery are a function of the treatments.

Table 4: Treatment Effects on Beliefs

	(1) Effectiveness of Offset	(2) Env. Damage of Delivery
Price	0.110**	-0.012
	(0.050)	(0.050)
Quantity	0.080	0.053
	(0.139)	(0.130)
Price $\times$ information	-0.167***	-0.020
	(0.053)	(0.052)
Quantity $\times$ information	0.546***	0.111
	(0.181)	(0.175)
Price^2	-0.003**	0.001
	(0.002)	(0.002)
Quantity^2	-0.008	-0.006
	(0.011)	(0.011)
Price$^2 \times$ information	0.004***	0.000
	(0.001)	(0.001)
Quantity$^2 \times$ information	-0.038***	-0.008
	(0.013)	(0.013)
Constant	3.789***	3.873***
	(0.419)	(0.407)
N	3,600	3,600

Note: This table presents effects of price, quantity and information on i) beliefs about the quality of the offset and ii) beliefs about the environmental damage from an uncompensated delivery. Coefficients are obtained from the regression specification in equation 3. Robust standard errors are in parentheses. *,**,***: significant at $p < 0.1$, $p < 0.05$, $p < 0.01$, respectively.

4 Willigness to Pay for Carbon Mitigation

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 follow a sufficient statistics approach, which does not rely on any distributional assumptions (Chetty 2009).

Consumer $i \in \{1, 2, ...I\}$ can choose between buying a carbon offset and an outside option. Without loss of generality, 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.[20] Utility is given by:

$$u_i = \alpha + \beta\gamma_i + \eta p_i + \epsilon_i. \tag{4}$$

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 distribution, the probability that consumer i chooses to buy the offset, denoted π_i, can be written in closed form as

$$\pi_i = \frac{1}{1 + \exp(-\beta\gamma_i - \eta p_i)}. \tag{5}$$

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

Sufficient Statistics Approach An alternative approach that does not require a distributional assumption about ϵ_i is to *approximate* WTP by reduced-form elasticities. To see how this works, note that the derivatives of aggregate demand with respect to price and carbon quantity are:

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

$$\frac{\partial D}{\partial p} = \eta g(-\beta\gamma + \eta p) \tag{6}$$
$$\frac{\partial D}{\partial \gamma} = \beta g(-\beta\gamma + \eta p) \tag{7}$$

WTP is given by:

$$WTP = -\frac{\beta}{\eta} = -\frac{\partial D}{\partial \gamma} / \frac{\partial D}{\partial p} \tag{8}$$

.

Denote the demand responses to price and carbon quantity variations by $\Delta_p D$ and $\Delta_\gamma D$, respectively. The demand derivatives in 6 and 7 can be approximated by

$$\Delta_p D/\Delta p \approx \eta g(-\beta\gamma + \eta p) \tag{9}$$
$$\Delta_\gamma D/\Delta \gamma \approx \beta g(-\beta\gamma + \eta p) \tag{10}$$

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

$$WTP \approx -(\Delta_\gamma D/\Delta\gamma)/(\Delta_p D/\Delta p). \tag{11}$$

The parameters can be estimated by OLS as in equation 3. Since the nonlinearity coefficients are statistically insignificant, I do not include them in the main specification.

Estimation results for both the logistic regression and OLS are shown in Table 5. I estimate the model for STANDARD and INFORMATION separately. Subjects with the baseline offset are included in both estimations. Regression coefficients in column 1 and 2 come from a logistic regression, while coefficients in column 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.

Conclusions change once information is provided. Column 2 shows that consumers receive larger disutility of price 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/tCO_2. This estimate is highly statistically significant with $p < 0.01$.

The sufficient statistics approach produces almost identical results. WTP is zero in STANDARD and 15.99 EUR/tCO_2 ($p < 0.01$) in INFORMATION. Results are similar to the logistic regression because, empirically, aggregate demand is linear (i.e., $g(\epsilon)$ is locally constant).

What do these estimates imply for public policy? If we are willing to interpret WTP in INFORMATION as a lower bound for the social cost of carbon, we can compare it to current policy assumptions. The estimate under INFORMATION 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/tCO_2. 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 76 EUR/tCO2.

The preferred estimate by a large group of climate scientist lies at around 185 USD with a large 95%-confidence interval of 44 - 413 USD (Rennert et al. 2022).[21]

Thus, current assumptions by policymakers and climate scientists are largely supported by the estimate from this experiment, while former policy assumptions are not.

Finally, it should be noted that the WTP estimates in my study may also constitute a lower bound because my sample may not be representative. Subjects in my study prefer to order groceries online instead of choosing less carbon-intensive alternatives, such as walking or cycling to the store. One might argue that these types of consumers generally have a lower willingness to protect the climate voluntarily, such that my estimates understate WTP of the entire population.

[21]This large estimate is also supported by expert surveys on the correct size of the social discount rate. Most experts support a relatively low discount rate of 2% (Drupp et al. 2018), which should raise the SCC above 100 USD/tCO2. This has led New York to implement a SSC of 125 USD/tCO_2.

Table 5: Willingness to Pay for Carbon Mitigation

	(1)	(2)	(3)	(4)
	Logistic Distribution		Sufficient Statistics	
	Standard	Information	Standard	Information
Quantity	-0.29	31.42***	-0.02	4.07***
	(6.08)	(5.75)	(0.71)	(0.76)
Price	-1.05***	-1.91***	-0.13***	-0.25***
	(0.22)	(0.21)	(0.03)	(0.03)
Constant	-1.63***	-1.46***	0.16***	0.19***
	(0.04)	(0.03)	(0.04)	(0.03)
WTP in €/tCO_2	-0.27	16.44***	-0.18	15.99***
	(5.83)	(2.47)	(5.59)	(2.51)
N	42440	42186	42440	42186

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 ratio of the quantity and price coefficients, multiplied by minus 1. 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.

4.1 WTP for the Offset Versus WTP for the Carbon it Mitigates

It seems straightforward to assume that WTP for a carbon offset would directly identify WTP for carbon mitigation. If WTP for a carbon offset that compensates 10kg of CO_2 is 0.20 EUR, then WTP for carbon mitigation should be 20 EUR/tCO_2. In this case, we would only require exogenous variation in the price of the offset to identify WTP for carbon mitigation, not in the quantity it compensates. However, this approach implicitly assumes that consumers buy the carbon offset *only* because of its impact on carbon reduction. This approach fails, however, if consumers buy the offset because of warm glow utility or if they do not correctly understand the effectiveness of the offset. Prior laboratory studies have inferred WTP from randomizing the price of retiring pollution permits, implicitly assuming that consumers only buy the permit because of its impact on carbon emissions. The data in this experiment allows us to conclude how much WTP for the carbon offset actually differs from WTP for the carbon it compensates.

In Table 6, I present results from estimating the random utility model in equation 4 but excluding quantity variation. Thus, using the two subsidy levels, I estimate utility for the baseline offset that compensates 2.4kg of carbon. Willingness to pay for the offset itself is then simply the constant term divided by the price coefficient.

For the logistic regression, the average WTP for the offset is negative, both in STANDARD and in INFORMATION. Average WTP for the offset is -1.77 EUR in STANDARD and -0.67 EUR in INFORMATION. Negative average WTP is in part to be expected because a large share of the sample does not buy the offset (even at a price of 6 Cents) and because 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 offset compensates 2.4kg of carbon, STANDARD would imply that average willingness to pay for carbon mitigation is $WTP = \frac{-1.77}{2.4} \times 1000 = -737.50 EUR/tCO_2$. Similarly, in INFORMATION it would be $-279.16 EUR/tCO_2$. Both values are unrealistic and far from the values obtained in Table 5 that exploits quantity variations.

The sufficient statistics approach yields positive values for WTP for the offset of 1.41 EUR and 0.74 EUR for STANDARD and INFORMATION, respectively. The implied WTP for carbon mitigation would be $587 EUR/tCO_2$ and $308 EUR/tCO_2$. Note first that this would yield the incorrect conclusion that WTP for carbon mitigation *decreases* with information provision. The reason is that demand is much more elastic under INFORMATION, such that any positive demand for the offset translates into low money-metric utility. Second, these estimates would dramatically overstate WTP. The estimate under INFORMATION is more than 19 times larger than the $16 EUR/tCO_2$ 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. Instead, identification requires both price and quantity variation.

Table 6: Willingness to Pay for the Offset Itself

	(1)	(2)	(3)	(4)
	Logistic Distribution		Sufficient Statistics	
	Standard	Information	Standard	Information
Price	-0.93***	-2.05***	-0.11***	-0.27***
	(0.24)	(0.24)	(0.03)	(0.03)
Constant	-1.64***	-1.37***	0.16***	0.20***
	(0.04)	(0.04)	(0.04)	(0.04)
WTP for 2.4kg-Offset in EUR	-1.77***	-0.67***	1.41***	0.74***
	(0.50)	(0.09)	(0.33)	(0.07)
N	25,308	25,372	25,308	25,372

Note: This table reports parameter estimates when quantity variations are excluded from the regressions. The second-to-last row reports implied WTP for the offset itself instead of for the carbon it mitigates. 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.

5 Email Survey

To further understand consumers' preferences, I implement a second survey. Several months after the 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 education about carbon offsetting, as well as to the firm's contribution to the offset. It also sheds light on people's preferences for protective policies, in particular, for a a carbon tax. In total, 1,617 subjects participated in the survey.[22] A translated version of the survey can be found in Appendix G.

5.1 Survey Design

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

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

first saw 4.8kg, next see 9.6kg. This creates both within- and between-subject variation in the compensation amount.

Testing Subjects' Awareness about Quantity Increases

The design allows us to test whether increasing the compensation amount has a different effect within- than between-subject. Note that some subjects see 4.8kg in the first question, while others see it in the second. If WTP for 4.8kg is higher within- than between-subject, it may suggest that subjects who saw 4.8kg in the first question did not fully realize the magnitude of that number—i.e., they have a poor understanding of kilograms of carbon as a measure of impact. In the within-subject group, subjects may better understand the quantity increase because they have seen a lower number before and, therefore, are more aware that the environmental impact is larger for y than for x.[23] A caveat of this interpretation is that within-subject effects can also be driven by "experimenter demand effects:" subjects may simply state a larger number on the second question because they think the surveyor expects them to do so. However, current evidence suggests that experimenter demand effects are rather small in magnitude (De Quidt, Haushofer, and Roth 2018).

Preferences for Sharing Compensation with the Company

In addition, I randomize a treatment in which subjects receive additional information in the second question on WTP. Specifically, they are asked how much they would be willing to pay *if the company doubles the compensation amount on its own cost* to $Y \in \{4.8, 9.6\}$kg of CO_2. This treatment allows us to investigate the effect on stated WTP of a quantity match by the company.

Education Treatment about Carbon Compensation

Finally, I investigate whether education about carbon offsetting affects WTP. In the field experiment, it would have been difficult to concisely provide sufficient information about carbon offsetting. The survey allows us to convey more (even though still not complete) information and is, therefore, more suitable to study the role of education about carbon mitigation. I randomize a treatment in which subjects see three stylized facts about carbon emissions before they answer 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

[23] However, note that there may also be an experimenter demand effect: Subjects simply may increase WTP in the second question because they feel they are expected to do so.

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 randomly asked about one of these facts in a follow-up question to test their understanding.

5.2 Results

Table 7 reports results from an OLS regression of WTP on the treatments. As is common in the literature that measures WTP with open-end questions, I adjust for outliers by only considering the 90th percentile of WTP answers.[24] Column 1 is the stated WTP in Cents. The constant implies that subjects in the first question state a WTP of 49 Cents. There is no statistically significant effect of raising the compensation amount by 2.4kg of CO_2 *between-subject*. This would again imply consumers are fully quantity-inelastic, as in the field experiment in STANDARD. However, WTP increases by 63% (+31 Cents) when the compensation amount is raised by 2.4kg of CO_2 *within-subject* ($p < 0.01$). Thus, consumers become highly quantity-elastic when they realize that the compensation amount is larger. Consistent with the results from the post-purchase survey in Section 3.4, this suggests that part of the effect of INFORMATION in the field experiment is that consumers did not realize in STANDARD whenever the impact of the offset had been raised.

If we increase the quantity from 4.8kg to 9.6kg, WTP increases by 38 Cents within-subject. Interestingly, this is only slightly larger than when the quantity is raised by 2.4kg within-subject, suggesting that the marginal increase in WTP drops quickly as we increase the compensation amount.

The education treatment about carbon offsetting increases WTP by 20% (+10 Cents), suggesting that imperfect information about carbon mitigation is a driver of lower WTP. This also implies that the revealed preference estimates from the field experiment may constitute a lower bound because subjects in the field did not receive an extensive education treatment. To complement this result, Appendix F 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 amount 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

[24]More specifically, I use the 90th percentile of WTP *per* tCO_2. It is important to normalize in this context as subjects have been offered different compensation amounts. If we do not exclude outliers, stated WTP estimates become more inflated due to some unreasonably large extreme values.

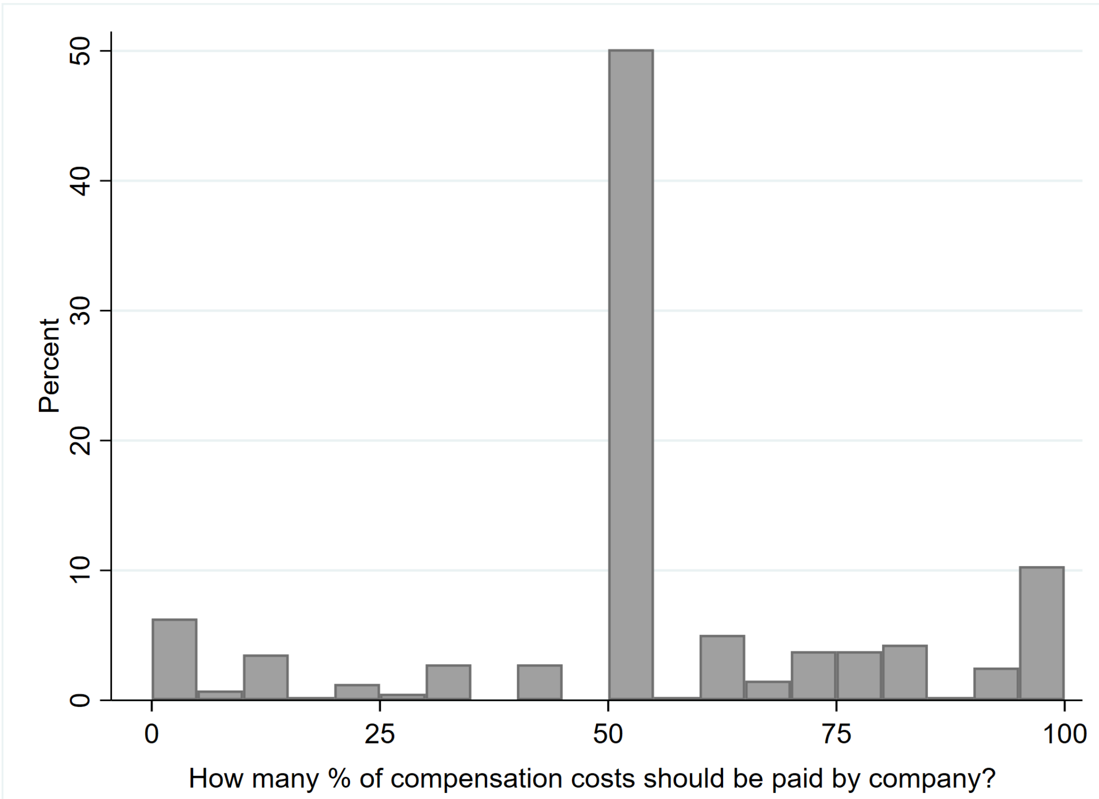
takes less to compensate for a delivery than they thought. This belief change may increase the perceived effectiveness of a donation to carbon offsets and explain why WTP increases with the education treatment.

Another noticeable result in Table 7 is that WTP decreases when the firm offers to match the compensation amount. This may be consistent with models of "pure altruism," in which donors care about the impact of the contribution and reduce their own contribution when the contribution of others increases. If consumers think that a certain amount of funds should suffice to mitigate, say, 2.4kg of carbon, then their willingness to pay might fall if they learn that the company contributes, as well. These results suggest that consumers, in fact, care about the impact of their contribution and are not entirely driven by "warm glow.

Results do *not* imply that consumers prefer to compensate emissions without the company's contribution. In fact, the opposite is the case. 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 7 illustrates that the modal consumer thinks the company should pay half the compensation costs. In conclusion, consumers seem to care about the impact of carbon offsetting, and they decrease WTP when the firm gives more. Still, they positively value that the firm contributes to the costs, which may explain why in the field experiment more subjects offset when they learn about the company's contribution.[25]

[25] In an unreported regression of fairness preferences on the treatments, I do not find significant effects. Most consumers prefer a 50-50 split with the company, independent of the size of the total compensation amount.

Figure 7: Fairness Preferences



Note: The figure illustrates the distribution of subjects' answers to the question of what share of the carbon compensation costs should be paid by the firm.

Column 2 of Table 7 is WTP in EUR/tCO_2, i.e., a normalized version of total WTP. On average, subjects stated WTP implies a WTP of 200 EUR/tCO_2, which is dramatically higher than the revealed preference estimate of 16 EUR/tCO_2 from the field experiment. Note that stated WTP in EUR/tCO_2 decreases as we increase the compensation amount both between- and within-subject. This reduction means that total WTP does not increase proportionally with increases in the compensation amount and would suggest that WTP is concave in compensated quantities.

The stated WTP estimate of 200 EUR/tCO_2 from this survey falls into the range of numbers in other prior 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/tCO_2 as our preferred estimate, the survey results would overstate WTP by 1,150%.

A limitation is that I do not observe which customers answered the survey because participation was fully anonymous. Therefore, I cannot distinguish whether stated preferences are much larger because consumers overstate their preferences or because consumers with extremely high valuations have selected into the survey. The large difference between survey responses and WTP in the experiment makes a pure selection mechanism unlikely. However, even if there is substantial selection, the results provide an important insight: A survey with stated preferences yields estimates that are 11 times larger than estimates from a field experiment with the entire customer base that makes actual consumption choices. Whether this is driven by hypothetical bias or selection, we can conclude that the survey yields inflated estimates for the sample of interest (i.e., for the entire customer base).

Table 7: Hypothetical WTP

	(1) Total WTP (in EUR)	(2) Implied WTP in EUR/tCO
Quantity increase +2.4kg, between-subject	0.056 (0.041)	-90.481*** (12.645)
Quantity increase +2.4kg, within-subject	0.312*** (0.026)	-37.322*** (5.291)
Quantity increase +4.8kg, between-subject	0.377*** (0.060)	-113.392*** (12.107)
Education treatment about carbon offsetting	0.104* (0.054)	26.830** (11.335)
Firm participates in compensation	-0.107** (0.054)	-18.059 (11.238)
Constant (baseline offset: 2.4kg)	0.494*** (0.044)	200.622*** (12.957)
N	1,617	1,617

Note: This table reports treatment effects on hypothetical WTP. Column 1 is absolute WTP in cents, as directly indicated by the subjects in an open-end question. Column 2 is the implied WTP in EUR/ton of CO_2. Robust standard errors are in parentheses. *,**,***: significant at $p < 0.1$, $p < 0.05$, $p < 0.01$, respectively.

5.3 Political Support for Carbon Tax

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 8 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/tCO_2, while 32% have a WTP of zero. By contrast, only 20% of carbon tax supporters have a WTP below 20 EUR/tCO_2 and 6% a WTP of zero. The modal opponent of a carbon tax has a stated WTP of zero, while the modal supporter has a stated WTP of around 208 EUR/tCO_2. Overall, the probability distribution is shifted to the right for supporters relative to opponents of the tax. This suggests that hypothetical WTP—while overstating true WTP—still has strong predictive power regarding stated political preferences for environmental policies.

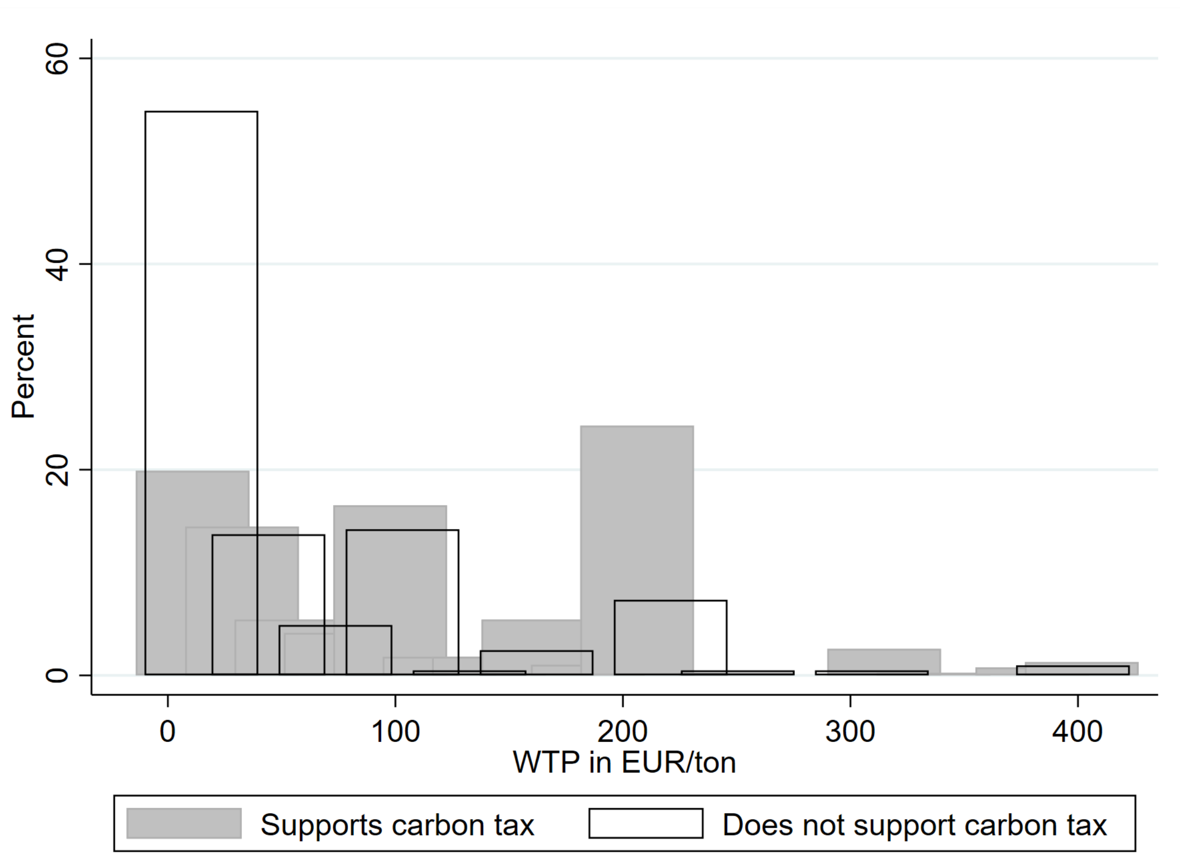
Answers to this question are also important from a welfare perspective. They suggest that people who state to voluntarily support climate protection also have a strong preference for stringent policies that enforce climate protection on others. Put differently, it seems not to be the case that those who voluntarily donate to the environment would receive disutility from policies that force others to donate.

To investigate whether this relationship also holds for revealed and not just stated preferences, I obtain publicly available data on vote shares for the green party ("Die Grünen") in Germany. The green party aims to enforce many climate protection policies, including a carbon tax. I then merge the data from the 2021 election to the field experimental data based on the zip code. In Appendix E, I find that consumers from districts with an above-median vote share for the green party are 22% (2.6 percentage points) more likely to buy the baseline offset. This result is consistent with the notion that consumers who support climate protection policies are also more likely to engage in voluntary carbon mitigation.

These results imply that people's WTP for *voluntary* climate protection, as elicited in the field experiment, is informative about their welfare gains from enforcing climate policies.[26]

[26] In Appendix G, I further investigate how stated WTP correlates with demographics and risk preferences.

Figure 8: WTP and Support for Carbon Tax



Note: This figure shows the distribution of WTP in EUR/ton of CO_2 among the supporters of the tax (in gray) and the opponents (transparent).

6 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 price reductions and impact increases. I randomize an information treatment that explicitly informs consumers whenever the price has been subsidized or the impact has been raised by the firm.

I find that consumers are price-elastic but fully impact-inelastic in the standard treatments, consistent with the "scope-insensitivity" phenomenon. If consumers were able to make a well-informed trade-off between price and quantity, then these results would mean that consumers do not care about the impact of carbon offsets. I show that a simple information treatment that increases the salience of subsidies and matches changes this interpretation and makes consumers sensitive to impact. I investigate the information effect with two complementary surveys. Results suggest that consumers care about the impact of the offsets but do not know how to interpret the magnitudes of carbon mitigation. These results highlight the importance of attention and incomplete information for the elicitation of environmental preferences. In addition, I find that consumers exhibit a preference for sharing the compensation costs with the firm. Both effects likely explain the increase in elasticities due to information provision.

The preferred estimate of willingness to pay for carbon mitigation under information provision is 16 EUR per ton of CO_2. This estimate provides a first data point of revealed preferences for voluntary carbon mitigation. The estimate is to be interpreted as a *lower bound* of people's preferences for climate protection. As such, it can serve as useful information for policymakers when forming beliefs about public support for climate policies, as well as about the social cost of carbon. The estimated lower bound in my study supports current policy assumptions in the EU and the US, but is not congruent with assumptions by the former Trump administration. The empirical framework developed in this study sets the stage for future studies to build upon, with the goal of obtaining estimates from various samples and in different contexts.

Stated preferences from a complementary survey largely overestimate revealed preferences. This provides a challenge for contingent valuation methods that rely on hypothetical answers. However, such surveys offer many benefits and are easier to implement and scale than the field experiment in this paper. Consequently, an important avenue for future research would be the development of techniques aimed at reducing the disconnect between individuals' statements and their choices.

References

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

Andreoni, James. 1990. "Impure altruism and donations to public goods: A theory of warm-glow giving." *The economic journal* 100 (401): 464–477.

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.

———. 2022. "Salience." *Annual Review of Economics* 14:521–544.

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.

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

Carleton, Tamma, and Michael Greenstone. 2021. "Updating the United States government's social cost of carbon." *University of Chicago, Becker Friedman Institute for Economics Working Paper*, no. 2021-04.

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.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *The American Economic Review* 99 (4): 1145–1177.

Cummings, Ronald G, Glenn W Harrison, and E Elisabet Rutström. 1995. "Homegrown 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.

Drupp, Moritz A., Mark C. Freeman, Ben Groom, and Frikk Nesje. 2018. "Discounting Disentangled." *American Economic Journal: Economic Policy* 10 (4): 109–34 (November).

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*.

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

Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic Literature* 42 (4): 1009–1055.

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

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

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

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.

Levitt, Steven D, and John A List. 2007. "What do laboratory experiments measuring social preferences reveal about the real world?" *Journal of Economic perspectives* 21 (2): 153–174.

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.

Nordhaus, William D. 2010. "Economic aspects of global warming in a post-Copenhagen environment." *Proceedings of the National Academy of Sciences* 107 (26): 11721–11726.

Pace, Davide, and Joël J van der Weele. 2020. "Curbing Carbon: An Experiment on Uncertainty and Information About CO 2 Emissions."

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

Rodemeier, Matthias. 2020. "Buy baits and consumer sophistication: Theory and field evidence from large-scale rebate promotions." CAWM Working Paper 124.

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.

Tol, Richard SJ. 2002a. "Estimates of the damage costs of climate change. Part 1: Benchmark estimates." *Environmental and resource Economics* 21:47–73.

———. 2002b. "Estimates of the damage costs of climate change, Part II. Dynamic estimates." *Environmental and Resource Economics* 21:135–160.

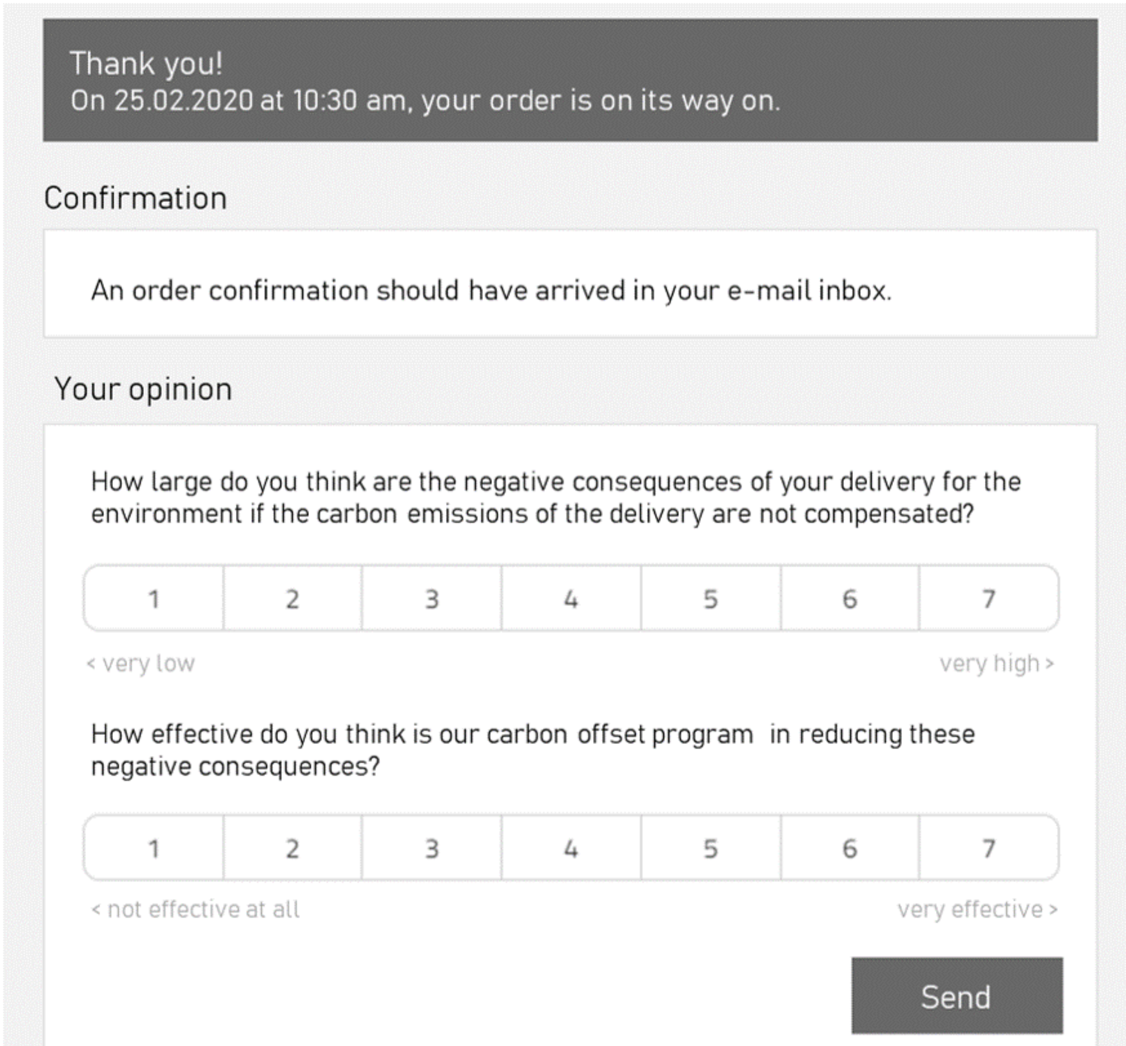
Vesterlund, Lise. 2003. "The informational value of sequential fundraising." *Journal of public Economics* 87 (3-4): 627–657.

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.

Online Appendix

A Additional Figures

Figure 9: Two-Question Survey Directly after Purchase



B Product Demand and Revenue

Subjects might also take into account that the amount of products they order affects the carbon emissions of a delivery. For instance, if subjects order less products, the delivery truck will be lighter and has more space for additional orders. Since subjects in the treatment groups have an additional way to reduce the environmental harm (i.e. by buying offsets), the average number of products ordered might be higher than in the control group. An alternative behavioral effect is that the offered offsets increase attention to the environmental damage of the delivery and that this in turn decreases the number of products ordered by subjects in the treatment groups.

I analyzed treatment effects on aggregate product demand, which is simply defined as the total number of goods purchased. In addition, I estimate effects on revenue.

Table B1 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 results in Section 3, this suggests that offering carbon offsets neither affects customer conversion, nor product demand of existing customers.

Table B1: Treatment Effects on Product Demand and Revenue

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

C 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 that were 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 Table 3, all coefficients are precisely estimated. Any null effect has a 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.

D Travel Distance of Delivery

Since some customers live farther away from the company's warehouse than others and order at different times of the day, the true carbon emission of each delivery is heterogeneous. This also means that the amount that is compensated by the carbon offset corresponds more or less to the individually true emissions of the delivery. Consumers might take this into account when deciding to buy the offset. For instance, consumers that live close to the warehouse might perceive the amount of carbon that is compensated to be too high for their delivery. The opposite might be true for customers living far away from the warehouse. To allow for heterogeneous responses to the treatments, I re-run regression 2 but include an indicator that equals one if subject i's distance to the warehouse in kilometers is above the median distance, and zero otherwise. I interact the indicator with each treatment. I further include zip code and day fixed effects to increase precision. Table D1 reports results.

Subjects with an above-median distance to the warehouse are no more likely to choose the baseline offset than subjects living closer to the warehouse. The estimate is economically small, with around –0.64 percentage points, and statistically insignificant.

Subjects living farther away from the warehouse also exhibit no different treatment effects, as implied by the interaction terms with the dummy variable.

In conclusion, travel distance to the warehouse does not seem to be important for the offsetting probability and any of the treatment responses. This suggests that consumers may either not realize when the emissions of their delivery differ from the emissions compensated by the offset; or they do not care that the compensated amount of the offset differs from their individual emissions.

Table D1: Heterogeneous Treatment Effects by Distance to Warehouse

	Offseting Probability x 100
-0.12€	0.760
	(0.694)
× information	2.732***
	(0.748)
-0.18€	1.473*
	(0.723)
× information	5.728***
	(0.808)
+2.4kg	-0.527
	(0.737)
× information	1.474*
	(0.712)
+7.2kg	-0.268
	(0.721)
× information	2.599**
	(0.810)
Above Median Distance	-0.638
	(0.878)
-0.12€× above	0.313
	(0.993)
× information	0.300
	(1.066)
-0.18€× above	1.759
	(1.025)
× information	-1.285
	(1.142)
+2.4kg × above	0.203
	(1.056)
× information	0.171
	(1.016)
+7.2kg × above	0.374
	(1.053)
× information	0.685
	(1.124)
Day FE	Yes
Zip Code FE	Yes
N	76,229

Note: This table presents heterogeneous treatment effects for subjects with an above- and below- median travel distance to the warehouse. Robust standard errors are in parentheses. *,**,***: significant at $p < 0.1$, $p < 0.05$, $p < 0.01$, respectively.

E Political Support for the Green Party

I merge data from the 2021 election to the field experimental data based on the first four digits of a 5-digit zip code.[27] Since some zip codes belong to more than one election district, and vote shares are not available for every zip code, the number of observations is reduced to 21,902 subjects that made at least one purchase. I repeat the basic regression from equation 2 but include a dummy variable that is one if subject i lives in a region with a vote share for the green party that is above the median vote share in Germany. Table E1 reports results. Offsetting demand is 22% (2.6 percentage points) higher when a subject comes from an area with a high support for the green party ($p < 0.01$).

[27] Data on election outcomes are obtained from www.bundeswahlleiter.de.

Table E1: Offsetting Probability and Support for Green Party

	(1) Offsetting Probability $\times 100$
-0.12€	0.764
	(0.977)
$\times$ information	3.631***
	(1.014)
-0.18€	2.275**
	(1.010)
$\times$ information	4.617***
	(1.034)
+2.4kg	-1.217
	(0.948)
$\times$ information	1.170
	(0.984)
+7.2kg	-0.389
	(0.963)
$\times$ information	3.955***
	(1.028)
High support for green party	2.632***
	(0.477)
Constant: Baseline offset, 2.4kg for 24 Cents	11.719***
	(0.727)
N	21,902

Note: This table presents treatment effects on the offsetting probability together with a dummy indicating whether the subject is from a region with a high vote share for the green party ("Die Grünen"). The dummy equals one if the vote share is above the median vote share for the green party in Germany. Robust standard errors are in parentheses.

F Additional Results from Email Survey

Beliefs about Carbon Emissions and Offsetting

Table F1 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 F1: Answers to Belief Questions in Email Survey

	(1) Delivery	(2) Car	(3) Trees	(4) Certainty
Education treatment about carbon offsetting	-0.409	-11.363***	-85.991***	2.323***
	(0.826)	(3.577)	(22.477)	(0.139)
Constant	5.463***	30.288***	144.844***	2.354***
	(0.543)	(2.471)	(15.893)	(0.095)
Observations	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 shown to predict actual risk preferences in incentivzed questions. The question ask subjects: *"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 F2 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 a 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 uncertaint 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 F1 plots the correlation between risk preferences and average WTP. Specificaly, 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 for voluntary climate protection.

Table F2: WTP and Demographics

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

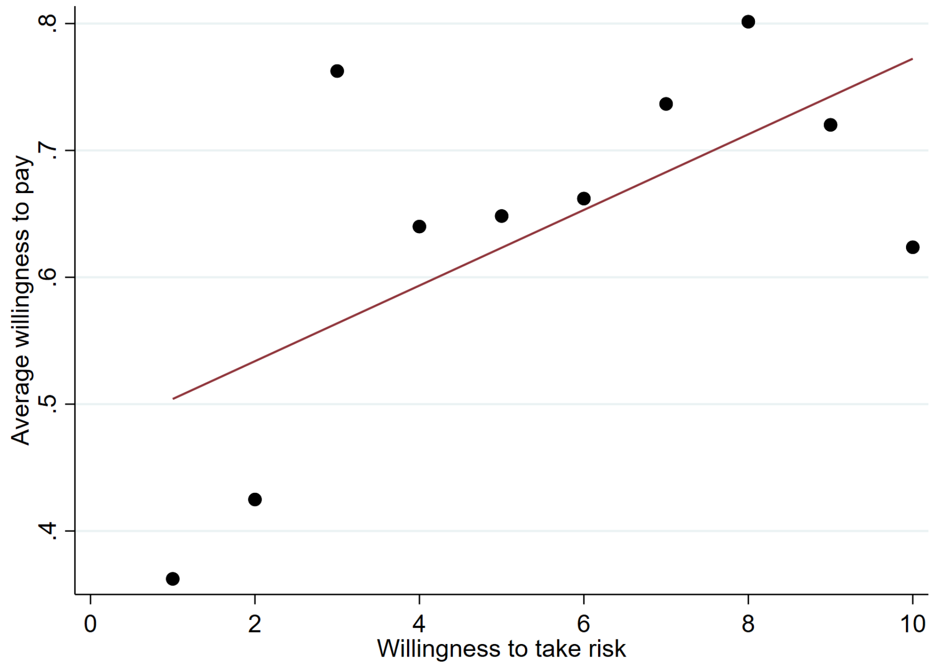


Figure F1: Correlation between Risk Preferences and WTP

G Email Survey

The following pages present the email survey, with all instructions and questions translated from German into English.

Introduction & Consent

Survey on Climate Protection

This survey is a joint project of the University of Münster and [company name]. Through this survey, we want to better understand your knowledge and opinion on climate protection. Please answer all questions truthfully. Your answers will be used anonymously for research purposes at the University of Münster. **Your participation supports research and is enormously valuable to us.** The survey takes about 7 minutes.

From all participants, we will draw five winners. In their name, 100 € will be donated to a non-profit organization of their choice.
Further information can be found here.

Data protection consent and conditions of participation

☐ I agree

Next

Education Treatment

Please read carefully the following three facts about the environmental impact of [company name] deliveries. **We will then ask you a question about them.**

1. An average delivery of [company name] produces approx. 2.4kg CO_2.

2. With a conventional car (fuel consumption: 7.8L/100km) you would have to drive about 11 kilometers to produce the same amount of CO_2 as an average delivery of [company name].

3. Climate protection projects plant trees in Germany to compensate for CO_2 emissions. A beech tree, for example, has a lifespan of about 80 years. In order to compensate for 2,000 [company name] deliveries, 5 beech trees must be planted and maintained.

Next

Belief Question (1 out of 3 was randomly asked)

Let's start with the questions.

A 33.34% How many kilograms of CO2 do you think an average delivery of [company name] produces?
Please enter a number in **kilograms**:

B 33.33% How many kilometers do you think you would have to drive with a conventional car (fuel consumption: 7.8L/100km) to produce the same amount of CO2 as an average delivery of [company name]?
Please enter a value in **kilometers**:

C 33.33% How many beech trees do you think would have to be planted to offset the CO2 emissions of 2,000 [company name] deliveries? Assume that each planted Beech 80 years lives (avg. lifetime).
Please provide your best estimate:

Next

Uncertainty Question

How sure are you about your last answer?

very unsure ○ ○ ○ ○ ○ ○ ○ very confident

Next

First Willingness to Pay Question: 2.4kg

How much would you be willing to pay to offset **2.4 kilograms of CO2 emissions**?

Please enter a value in **cents**:

Next

Second Willingness to Pay Question: 4.8kg

How much would you be willing to pay to compensate for twice as much, i.e. **4.8 kilograms of CO2** emissions?

Please enter a value in **cents**:

Next

Second Willingness to Pay Question: 4.8kg, with Quantity Match by Company

How much would you be willing to **pay if [company name] participated in environmental protection and doubled the compensation amount at its own expense**? The total compensation amount is then **4.8 kilograms of CO2**.

Please enter a value in **cents**:

Next

First Willingness to Pay Question: 4.8kg

How much would you be willing to pay to offset **4.8 kilograms of CO2 emissions**?

Please enter a value in **cents**:

Next

Second Willingness to Pay Question: 9.6kg

How much would you be willing to pay to compensate for twice as much, i.e. **9.6 kilograms of CO2** emissions?

Please enter a value in **cents**:

Next

Second Willingness to Pay Question: 9.6kg, with Quantity Match by Company

How much would you be willing to **pay if [company name] participated in environmental protection and doubled the compensation amount at its own expense**? The total compensation amount is then **9.6 kilograms of CO2**.

Please enter a value in **cents**:

Next

Question Regarding Fairness Preferences for Burden Sharing with Company

Let's assume that [company name] offered you the option to compensate for the CO2 emissions of the delivery with each delivery. As a customer, you would share the costs with [company name]. We are interested in which distribution of costs you consider fair. **What percentage of the costs should [company name] bear?**

Please enter a figure in **percent**:

Next

Question Regarding Support for Carbon Tax

Suppose a referendum is held in which the **introduction of a CO2 tax** is to be decided. A CO2 tax imposes a levy on each product and service intended to correspond to the resulting environmental impact.

How would you decide in the referendum?

- ◯ I would **reject** a CO2 tax.
- ◯ I would **be in favor of** a CO2 tax.
- ◯ Not specified.

Next

Question Regarding Risk Preferences

Please tell us, in general, how willing or unwilling you are to take risks? Please use the following scale from 1 to 10, where 1 means "not at all willing to take risks" and 10 means "very willing to take risks".

not at all willing
to take risk

very willing
to take risk

Next

Demographics

Finally, a few questions about yourself:

How old are you?

- ○ 18-19
- ○ 20-29
- ○ 30-39
- ○ 40-49
- ○ 50-59
- ○ 60-79
- ○ >70
- ○ No answer

What is your gender?

- ○ Female
- ○ Male
- ○ Different
- ○ No answer

Which of the following best describes your employment status?

- ◯ Apprentice
- ◯ Student
- ◯ Employed
- ◯ Unemployed
- ◯ Housewife/househusband
- ◯ Retired
- ◯ Other
- ◯ No answer

Next