ECONOMIC STUDIES

DEPARTMENT OF ECONOMICS
SCHOOL OF BUSINESS, ECONOMICS AND LAW
UNIVERSITY OF GOTHENBURG
256

Sustainable Consumption and Prosocial Actions

David Bilén

UNIVERSITY OF
GOTHENBURG

ISBN 978-91-88199-71-3 (printed)
ISBN 978-91-88199-72-0 (pdf)
ISSN 1651-4289 (printed)
ISSN 1651-4297 (online)

Printed in Sweden,
Stema 2023

SVANENMÄRKET
Trycksak
3041 0234

ACKNOWLEDGMENTS

First, I would like to express my gratitude to my supervisors, Olof Stenman-Johansson and Eva Ranehill. Both have been remarkably generous with their time and guidance in refining my ideas, drafts, presentations, and papers. They have provided clear feedback on areas that require enhancement while acknowledging my progress and commenting on what I do well. Olof's door has always been open whenever I had a question or concern I wanted to discuss, and he guided me to see the broader picture of my research.[1] Eva has been particularly important in guiding me through the academic profession and helping me to communicate my research clearly and concisely.

A special thanks to my final seminar discussant, Linda Thunström, for her feedback on the first two papers of this thesis, which significantly improved the quality, especially of the second chapter. I would also like to thank my co-authors, Anna Dreber and Magnus Johannesson, both brilliant researchers with whom I was fortunate to write my first paper.

I am truly grateful to the supermarkets I collaborated with, which made the two chapters of this thesis possible, and to all the researchers who generously shared their data, which allowed the third chapter to be possible.

I extend warm thanks to all seminar participants at conferences I attended during my fourth and fifth years in the Ph.D. program. Having worked on my papers in isolation during the covid pandemic, these conferences and the engaged curiosity and interest lifted my perspective and appreciation of the scientific contribution. I would like to thank all the seminar participants at the University of Gothenburg for their interest, constructive comments, and questions that helped me refine my understanding and improve my work. A special thanks to Fredrik Carlsson, Mitesh Kataria, and Katarina Nordblom, whose feedback substantially improved the first chapter.

A special thanks to my cohort of 2018 – Dominik, Magnus, Ville, and Vu – for making the coursework much more enjoyable than merely passing exams and helping me learn more about the material, broadening my perspective, and sharing the journey of finding what path we should explore.

[1] Sometimes Olof's door was open, but he was not there though.

I would also like to thank everyone I have interacted with at the Department of Economics for their support throughout this journey and the administration staff for their help with everything from teaching to reimbursements and travel reporting. I thank Hans Bjurek, Katarina Nordblom, Pelle Ahlerup, and Claes Ek for their teaching quality and for helping me to develop as a teacher. A special thanks to Randi Hjalmarsson and Andreas Dzemski for all their assistance during the job market process, providing invaluable support and guidance.

Lastly, I would like to thank everyone I have met throughout these years who I felt were genuinely curious about whatever topic they were interested in, those who were curious about exploring new ideas, listening to mine or discussing their own, and those whom I felt genuinely cared about the research process and the quality of science. I truly appreciated those interactions.

Contents

Introduction

This thesis contains three chapters that provide large-scale evidence of sustainable consumption and prosocial actions. In the first two chapters, I focus on interventions designed to encourage individuals to make sustainable choices by employing information provision in Chapter One and combining monetary incentives with normative appeals in Chapter Two. The third chapter focuses on unconditional generosity, and, in collaboration with my co-authors, we investigate potential gender differences in generosity and whether such differences relate to the recipient's needs.

The first two chapters utilize natural experiments to study interventions designed to promote sustainable food consumption. The food system is believed to be responsible for about one-third of the annual global greenhouse gas emissions, and changing food consumption is one of the key ways to reduce emissions (Crippa et al. 2021).

The first chapter, titled "Do Carbon Labels Cause Consumers to Reduce Their Emissions? Evidence from a Large-scale Natural Experiment," studies whether introducing carbon labels that provide information on the carbon emissions of products leads consumers to choose more climate-friendly options and reduce their carbon emissions. I exploit a difference-in-differences setting where one supermarket unexpectedly introduced labels on around 3,000 products displaying their carbon footprint, while a similar supermarket did not. I find an average reduction in total emissions of around 2.5 percent resulting from informing consumers about the products' carbon emissions. This effect size is smaller than found in most other studies based on lab or field experiments but still economically meaningful. It corresponds to estimates in the literature of a carbon tax on meat and dairy products of about 20$ per tonne CO_2. The main driver was that customers in the treated store reduced their beef consumption by 8–16 percent and increased their consumption of products with a lower carbon footprint, such as pork, poultry, and vegetarian products.

In the second chapter, "Combining monetary incentives and nudges to promote green consumption: Evidence from a large-scale natural experiment," I study another large-scale natural experiment in which a supermarket, without prior announcement, increased the bonus points on fruits and vegetables and motivated the bonus by awarding "good deeds." The additional bonus points function as an equivalent price reduction of 0.6–2%, and the labeling provides a normative nudge on fruit and vegetables as environmentally friendly products. I use panel data that cover more than 40,000 consumers who place over 800,000 orders several months prior to the implementation of the bonus, throughout the entire year when the bonus is in place, and for several months after the bonus has been removed. The results indicate a larger consumer response than expected solely on the basis of the monetary incentive. The bonus program increased overall fruit and vegetable consumption by, on average, 6–9% per order. I further find no evidence that increasing the monetary incentive from an equivalent price reduction of 0.6% to 2% had any additional impact on consumption.

In the third chapter, titled "Are Women More Generous than Men? A Meta-analysis," my co-authors and I study gender differences in generosity by performing a meta-analysis of the standard (windfall) dictator game (DG). In this game, one player receives a monetary endowment and the choice to donate some of that endowment to a recipient player or charity organization. While giving in this game is not necessarily due to altruistic concerns - for example, the results of List (2007), Krupka & Weber (2013), Bardsley (2008) and Dana et al. (2006) suggest that DG giving is influenced by the strategy space, reference points and expectations of social norms - this is the most commonly studied game in order to understand non-strategic prosocial behavior.

We collect raw individual participation data from 53 studies with 117 conditions, providing 15,016 unique individual observations. We find that women on average give 4 percentage points more than men (Cohen'sCohen's d=0.16) and that this difference decreases to 3.1 percentage points (Cohen'sCohen's d=0.13) if we exclude studies where dictators can only give all or nothing. The gender difference is larger if the recipient in the DG is a charity, compared to the standard DG with an anonymous individual as the recipient (a 10.9 versus a 2.3% points gender difference).

Chapter 1

Do carbon labels cause consumers to reduce their emissions? Evidence from a large-scale natural experiment

Abstract: In this paper, I use consumer transaction data from two supermarkets to study if carbon labels cause consumers to reduce their grocery carbon emissions. I exploit a difference-in-difference setting where one supermarket unexpectedly, from one day to the next, introduced labels on around 3000 products with their carbon footprint, while another similar supermarket did not. My sample includes more than 30,000 consumers making 400,000 orders, and I follow them up to one year post-treatment. I find an average reduction in total emissions of around 2.5 percent from informing consumers about the products' carbon emissions, which corresponds to about 1.4 kg CO_2 less per order. This effect size is smaller than found in most other studies based on lab or field experiments, but still economically meaningful. It corresponds to estimates in the literature of a carbon tax on meat and dairy products of about 20$ per tonne CO_2. The main driver was that customers in the treated store reduced their beef consumption by 8–16 percent and increased their consumption of products with a lower carbon footprint, such as pork, poultry, and vegetarian products.

1.1 Introduction

The food system is believed to be responsible for about one-third of the annual global greenhouse gas emissions (Crippa et al. 2021). It is one of the European Union's targeted areas to achieve climate neutrality by 2050, and changing food consumption is one of the key ways to reduce emissions (IPCC 2019).

To empower consumers to make more sustainable food choices, the European Commission plans to implement a unified labeling framework that will inform them about food products' carbon emissions and other environmental dimensions (European Commision 2020). Public support is one key advantage of information policies. Survey evidence indicates that more than 70% of European citizens favor a mandatory requirement to inform about products' carbon emissions (Gallup 2009). Information disclosure can also arise naturally in the market from the demand of consumers that utilize the information in their purchases. Some food brands of meat and dairy substitutes (Oatly 2022, Quorn 2022) already use carbon labels to inform consumers about the impact of their products on the climate, and Unilever (2022), one of the world's largest consumer goods producers, recently communicated their ambition to label all of its products.[1] If carbon labels cause consumers to reduce their emissions, they could offer a mitigation channel that is both feasible and scalable to several important consumption domains (Taufique et al. 2022).

In this paper, I provide evidence from a unique natural experiment to study whether carbon labels cause individuals to reduce their carbon emissions. Carbon labels were simultaneously implemented on 3,000 products (80 percent of the sales volume) in a supermarket on 12 November 2019 without prior announcement. Labels were implemented on products in all food categories, including low impact products (e.g. those produced by Oatly 2022 and Quorn 2022) and high-impact products such as beef or dairy products. To evaluate this natural experiment, I have access to a rich and unique data set of individual transactions. The data consist of pre- and post-treatment sales from the treated store and a similar control store. I

[1] Other recent examples include the multinational manufacturer Logitech International S.A. and footwear company Allbirds, Inc. (Allbirds 2022, Logitech 2022).

create a panel data set that tracks individual consumers' consumption and carbon emissions over time and use a difference-in-difference design that controls for consumers sorting themselves into (or out of) the treated store to study how consumers respond to carbon labels. My sample includes over 400,000 orders made by more than 30,000 consumers that I follow up to one year after carbon labels were implemented, allowing me to study both the short- and long-term response. Having access to individual transaction (scanner) data is unique and allows me study both the aggregate response and heterogeneity across several dimensions.

To the best of my knowledge, this paper provides evidence from the first supermarket in the world to label their assortment with carbon labels[2] and is the largest evaluation of carbon labels in any setting, particularly within the food domain. This paper's main contribution is to provide causal evidence of how consumers respond to carbon labels in their real-world consumption decisions on a substantially larger scale and time horizon than previously studied.

This paper is not the first to study how consumers respond to carbon labels and according to a recent review by Taufique et al. (2022), the existing evidence looks promising.[3] However, the authors also point out that the previous literature focuses primarily on a single-choice setting, such as how consumers respond to labeling on one product or a restaurant menu. Suppose, for example, that all products within supermarkets carry carbon labels. Consumers could then reduce their emissions through several product choices, and the salience of carbon labels could also increase compared with when only one or a few products are labeled.[4] Most studies further observe consumers up to a few weeks after carbon labels were implemented. Ex-ante, one could imagine the short-term effects of carbon labels on consumption behavior being either larger or smaller than over longer time hori-

[2] Tesco had plans to do so in 2011 but later abandoned these plans (Potter et al. 2021, Taufique et al. 2022).

[3] Food emissions are complex and arise throughout a product's life cycle, and gathering the necessary information for an individual consumer would require a substantial time investment. One argument is, therefore, that carbon labels and simple choice architecture could help consumers make informed choices and place salience on climate impact in their daily purchase decisions.

[4] However, choices with large differences in their carbon emissions could mute the aggregate response by crowding out attention from product choices with less significant differences.

zons. The salience of carbon labels in consumers' purchase decisions could be particularly strong in the short term. However, as time passes, consumers will have more time to incorporate the information from carbon labels into their consumption behavior. Bringing interventions to scale is essential to understand the generalizability and policy implications of previous findings to real-world settings when carbon labels are implemented on a large scale (DellaVigna & Linos 2022, Al-Ubaydli et al. 2017).

I begin by studying the short-term consumer response from November 2019 until the outbreak of the Covid pandemic in March 2020. Consumers' carbon footprint was reduced by around 3 percent, controlling for spending. In the long term, effects fade but remain statistically significant at around 1–2 percent one year after the introduction of carbon labels. The overall response looking at the entire one year post treatment period amounts to about 1.4 kg CO_2e less per order for consumers in the treated store. These effects are less than found in most previous studies based on lab or field experiments but are still economically sizable. Relating the emissions reductions from carbon labels found in this paper to previous estimates in the literature of implementing a carbon tax on meat and dairy products in the EU27 corresponds to a tax of about 20$ per tonne CO_2 (Wirsenius et al. 2011). The main driver was that consumers in the treated store responded to carbon labels by reducing their beef consumption and instead buying low-impact products (such as vegetarian or poultry) and middle-impact products such as pork or products that contain a mix of pork and beef. Consumers in the treated store became 4–6 percent less likely to include any beef in their order (around one-third of orders contain beef) and overall reduced their beef consumption by around 8–16 percent.

I also found evidence that carbon labels affected spending and shopping frequency at the treated store. Carbon labels caused existing consumers to reduce their spending at the treated store by 3.1 percent, which amounts to reduced spending by 7.5 euros per 45 days. Reduced spending when making an order (intensive margin) is consistent with consumers reducing their purchases of beef and increasing their consumption of less emission-intensive products such as pork or poultry, as those products are also cheaper. However, the main driver was reduced shopping frequency (extensive margin) at the treated store by around 2.1 percent, suggesting that some consumers prefer to avoid being informed about their carbon footprint. However, carbon labels also attract new consumers, and there is no statistically significant difference in shopping frequency or spending when considering both new and old consumers.

I then move on to explore heterogeneity. It is unclear for which consumers one should expect the largest responses. Consumers who already consider their carbon footprint when making consumption decisions could be well informed and have a limited possibility to lower their impact further. On the other hand, consumers with high meat and dairy consumption may have more potential, but less motivation, to reduce their carbon footprint.[5]

To avoid primarily capturing the household size of consumers, I calculate consumers' carbon emissions per kg of food in their consumption basket. This measures the emissions intensity for one kg of food in a consumer's basket. I then calculate the median carbon intensity per kg of food during the pre-treatment period among all consumers and divided the sample into the low-impact group, those below the median, and the high-impact group, those above the median. For the low-impact group, the effect is sizable in both the short and long term. These consumers reduce their carbon footprint by around 4 percent and their response remain stable and robust over the whole year after carbon labels were introduced. However, consumers above the median respond significantly in the first months, but their response appears to fade in the subsequent months. Instead, I find evidence that the high-impact group reduced their shopping frequency at the treated store by about 3 percent. This heterogeneity analysis reveals that carbon labels effectively caused consumers to reduce their carbon emissions in the short term among all consumers. However, in the long term, primarily, those consumers that already had a low carbon footprint remain affected by carbon labels.

The remainder of the paper is organized as follows. In the following subsection, 1.1.1, I discuss the contribution of my paper and its relation to previous literature. Section 1.2 provides background details on the Swedish food market, Section 1.3 describes the natural experiment, and Section

[5]Even for consumers who are not well-informed but respond to carbon labels, the treatment effect's direction is unclear. Some consumers would learn they were underestimating the carbon footprint of their diet; others would learn they were overestimating it.

1.4 discusses the data. Section 1.5 details the empirical methodology, and Section 1.6 presents the main results. Section 1.7 concludes with a short discussion of the results and avenues for further research.

1.1.1 Relation to the literature

My paper primarily contributes to four strands of literature. First, it is related to the fast-growing body of literature specifically targeting individuals' own carbon footprint and their consumption decisions. Evidence from the lab (Pace & van der Weele 2020, Kanay et al. 2021, Panzone, Ulph, Zizzo, Hilton & Clear 2021, Panzone, Ulph, Hilton, Gortemaker & Tajudeen 2021, Camilleri et al. 2019) and field experiments in restaurants (Brunner et al. 2018, Lohmann et al. 2022) in general find that consumers choose more climate-friendly products when provided with information on the products' carbon footprint.[6] There are also studies that found null effects of carbon labels. Hornibrook et al. (2015) used loyalty card data from Tesco and did not find that demand increased for products with the Tesco Carbon Trust's carbon reduction label, which appeared on laundry detergent, orange juice, light bulbs, and potatoes. Kortelainen et al. (2016) find no evidence that demand or price increased for Carbon Trust labeled detergent in Denmark. Lohmann et al. (2022) conducted a large-scale field experiment in which they implemented carbon footprint labels in university cafeterias and observed 2,228 college students' cafeteria choices nine weeks before and seven weeks after the intervention. They divide meals into low, middle, and high carbon footprint and find that high-emissions meals were replaced by middle-emissions meals—leading to an overall reduction of 4 percent in GHG emissions. Another set of studies has recently started exploring whether carbon calculators can help consumers reduce their emissions (Fosgaard et al. 2021, Enlund et al. 2022). Fosgaard et al. (2021) follow 258 consumers using a carbon calculator that shows them the GHG emissions associated with their food purchases over 38 weeks. The study finds that consumers reduce their grocery GHG emissions by 27 percent during the first month they use the app, with most of this reduction

[6] There are also stated preference surveys showing that consumers are interested and have a positive WTP to being guided by labeling schemes to help them consume more environmentally sustainable products (Carlsson et al. 2020).

coming from a 45 percent decrease in beef consumption. However, the effects fade over time, and the study finds no statistically significant change over the whole 19-week post-period.

My paper contributes to this literature by providing evidence from real-world setting where about 80 percent of the assortment in a supermarket is labeled with a carbon label. I follow more than 30,000 consumers making 400,000 orders up to one year post-treatment. An important dimension when studying food consumption is that habits drive food choices in the short run. My paper differs from Lohmann et al. (2022) in that consumers make choices involving products from all categories as opposed to a repeated single choice between restaurant meals. The individual consumer data also enable the study of heterogeneity in responses. The high-impact consumers drove the response in Lohmann et al. (2022) by switching to middle-impact meals. My findings differ from those of Lohmann et al. (2022), as consumers with the lowest pre-treatment carbon footprint are the ones driving the response in the long term. My findings are in line with Fosgaard et al. (2021) in that I find a large reduction in consumers beef consumption but also that the response fades over time. Carbon calculators do differ from carbon labels on products as they primarily provide information to the customer after making a purchase.

Second, this paper adds to the literature studying how consumers respond to environmental and ethical food labels (Elofsson et al. 2016 and Hainmueller et al. 2015). This literature generally finds positive and statistically significant effects from implementing food labels on a single product. Hainmueller et al. (2015) performed a multistore field experiment in which they introduced Fairtrade labels on a subset of coffee brands. They find an increase of around 8 percentage points in the demand for Fairtrade-labeled coffee brands. Elofsson et al. (2016) performed a similar study in Sweden in which they labeled milk packages with an environmentally friendly certification. They find that sales of certified milk packages increased by 11 percentage points. This paper contributes to this literature by studying a different setting where consumers do not make a binary choice between being and not being environmentally friendly. Instead, products vary significantly in their carbon footprint. Choosing 1 kg of pork over 1 kg of beef would reduce a consumer's carbon footprint by about the same amount as choosing 25 liters of soy milk over 25 liters of cow milk. It is therefore not evident that effect sizes from a single product (e.g., milk) would carry over when products from all food categories are labeled.

Third, this paper also relates to the literature on information and food labels in general. Large-scale implementation of information labels where researchers have access to a suitable control group is rare. Two related papers are Bollinger et al. (2011) and Dubois et al. (2021), which evaluate calorie postings and front pack nutritional information, respectively. Dubois et al. (2021) implemented four different front-of-pack nutrition labels on 1,266 food products during ten weeks in 2016 and compared purchase behavior to the corresponding ten weeks in 2015. They find an increase in nutrition quality by 2.5 percent from the most effective label. Bollinger et al. (2011) studied the effect of mandatory calorie postings in New York on the calorie consumption of Starbucks customers. The authors find that calories were reduced by 6 percent and food products drove the effect with no change in calories from beverages. To my knowledge, this is the first paper to provide causal evidence on a large-scale implementation of environmental labels. Carbon labels differ from nutritional labels in that they provide information about an externality, and there is no direct benefit to the individual from reducing their emissions.

Fourth, this paper relates to the literature on individuals' choice to actively sort themselves out of prosocial settings (Dana et al. 2006 Andreoni et al. 2017, Knutsson et al. 2013). Knutsson et al. (2013) studies a related natural experiment in Sweden in which grocery stores implemented the possibility of donating returned recycling deposits to a charity organization and finds that recycling rates are reduced in the treated stores. My results in this paper are aligned with their findings, as I find that carbon labels caused a slight reduction of 2 percent in the shopping frequency among existing consumers, indicating that some consumers actively avoid being exposed to prosocial choices.

1.2 Institutional background

This section begins by providing institutional details on the two supermarkets studied in this paper and then gives background details on why the treated supermarket implemented carbon labels on its products.

1.2.1 Stores

The Swedish food market is concentrated in a few large parent organizations that operate grocery stores or own subsidiary stores. Most sales occur in physical stores, and online sales are primarily supplied in urban areas, especially in Sweden's two largest cities, Stockholm and Gothenburg. Online sales had a market share of 3.4 percent in the first quarter of 2020, with a substantial increase during the Covid pandemic to 6 percent in the fourth quarter of 2020 (Swedish Food Retailers Federation 2020).

The treated store only operates online and was one of Sweden's first online supermarkets when it made its entry in the Gothenburg area in 2012. The control store started as a physical store but later adapted to the rising demand for online shopping and has offered both online and physical stores since 2016.

Table 1.1 contains institutional details on the two stores in this paper, which are subsidiaries of one of the largest grocery stores in Sweden. Each store is responsible for its pricing, and there is one notable difference in their pricing schemes. The treated store offers free order delivery above 700 SEK, whereas the control store always charges a delivery fee of 49–99 SEK (and an additional fee of 49 SEK if the order exceeds 700 SEK). However, the treated store has slightly higher prices on beef, pork, vegetarian, and poultry products, at around 20 SEK per kg, making the total order cost similar between stores.

The parent organization is primarily responsible for deciding what products each store should offer. At the end of 2018, a decision was made by the parent organization to merge the stores' supply channels. This shift affected the assortment in the treated store by replacing certain brands with those used in the control store. This process happened gradually during the spring of 2019 and was finalized in the late spring. From June 2019 onward, these two stores shared distribution and logistics channels, including having the same delivery truck.

Table 1.1: Institutional details on the two supermarkets

	Treated store	Control store
Operation		
Platform	Online	Online and Physical
Supply chain	Joint	
Delivery		
Gothenburg area	Yes	Yes
Stockholm area	Yes	Yes
Malmö area	Yes (until April 2020)	Yes
Cost (SEK)	29 (free above 700)	49-99 (+49 below 700)
Car	Shared	

Notes: The exchange rate was about 10.5 SEK for 1 Euro during 2019-2020.

1.2.2 Sustainability goals

In recent years, corporate social responsibility has become a common standard among business organizations (Bénabou & Tirole 2010 and Fioretti 2022).[7] This has taken the form of reducing the store's environmental impact in the grocery market. The largest grocery stores in Sweden all have sustainability goals targeted at reducing their environmental impact (Axfood AB 2020, ICA Gruppen 2022, COOP 2020). Sustainability goals targeting the store's carbon footprint are not unique to the Swedish or Nordic market; for example, the five largest supermarkets in the UK recently committed to reducing their carbon footprint by half by 2030 (WWF 2022).

As one means to achieve these goals, the stores have implemented various tools to assist their customers in consuming more sustainably. Grocery stores in France (Collibri Foundation 2021), Germany (COBIOM 2021), and the UK (Lidl 2021) have plans or started to implement environmental labels.[8] Unilever (2022) have also announced their plans to label their assortment. Other examples from the Swedish market include ICA Gruppen 2022, which offers climate-friendly meals and recently also started providing its consumers a personal carbon footprint target. COOP (2020) have implemented a sustainability declaration that gives consumers consumption

[7] It is beyond this paper's scope to study why stores implement these programs (see, e.g., Bénabou & Tirole 2010 and Fioretti 2022).

[8] Tesco had plans to do so in 2011 but later abandoned these plans (Potter et al. 2021, Taufique et al. 2022).

scores in 10 environmental areas. As another example, the control store of this paper introduced a bonus program in October 2020 that gave additional bonus points for "climate-friendly" purchases. Carbon labels are not exclusive to the food sector, for example Logitech (2022) have also labeled their assortment.

1.3 Natural experiment

Carbon labels were introduced on 12 November 2019 and have been in place since this date. This section provides details on the labels and how they were implemented.

1.3.1 Carbon labels

The term "carbon labels" in this paper refers to carbon footprint labels (CFLs). CFLs focus exclusively on a product's GHG emissions without considering other environmental dimensions, such as biodiversity. Figure 1.1 shows the implemented carbon labels that display estimates of kg carbon emissions equivalents (CO_2e) per kg of the product (analogous to the price per kg). Carbon emissions equivalents consider the global warming potential of all GHG emissions and convert them into a common metric.[9]

The carbon footprint is estimated by the Swedish research organization RISE (2020)[10] using a life-cycle analysis considering all production and transportation emissions. Consumers are provided with information on the supermarket's website that all estimates have been produced by RISE (2020). The website also contains information on the underlying assumptions behind the life-cycle estimates, that estimates will have uncertainties, and that the interested reader can access further details on the website of RISE (2020). Consumers are also provided with reference information, such as how emissions from 1 kg of beef relate to emissions from driving a car for 10 kilometers.

Why were all products not labeled? Store management explained that

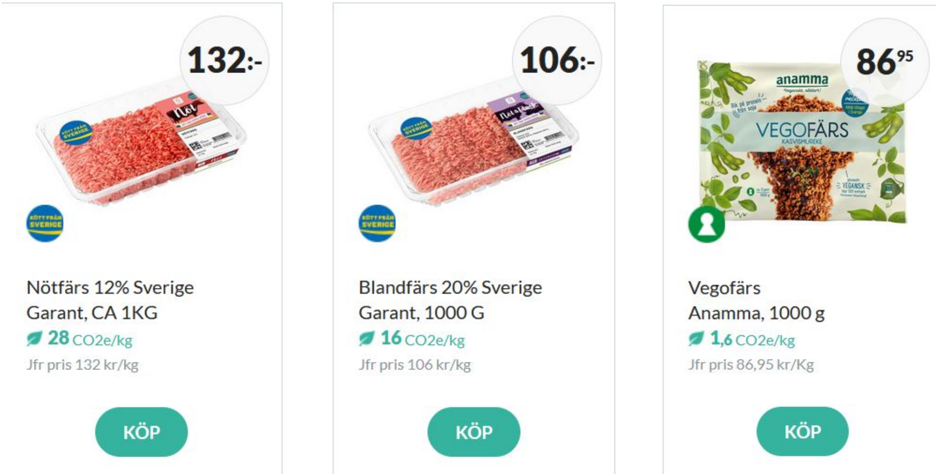
[9] For example, methane has around 28 times the global warming potential of carbon dioxide.

[10] RISE (2020) produces life-cycle estimates for many Swedish institutes, companies, and policy organizations.

they based their choice of products on sales volume and accuracy of measurement. Life-cycle analysis is complex and requires assumptions alongside some generalizations based on production standards. For products with multiple ingredients, such as charcuterie, it is challenging to estimate and generalize measurements from one brand to another. However, most emissions from meat products arise while the animals are alive and not in processing or transportation.

Figure 1.1: Examples of Carbon footprint labels

(a) Beef (High impact) (b) Beef and pork (Middle impact) (c) Vegan (Low impact)



Notes: This figure shows the carbon footprint labels implemented in the treated store on 12 November 2019. Labels display kg carbon emissions equivalents (CO_2e) per kg of the product. For example, the label in panel (a) should be interpreted as 28 kg CO_2e per kg beef. Carbon emissions equivalents consider the global warming potential of all greenhouse gases and convert them into a common metric. Emissions are estimated using life-cycle analysis, which considers all emissions from production to transportation.

1.3.2 Treatment

The supermarket implemented four treatment arms related to CFLs: (1) carbon footprint on products, (2) aggregate carbon footprint per order, (3) comparison with other consumers' carbon footprint, and (4) climate-smart product suggestions. Table 1.2 provides an overview of treatments implemented in the treated store.

1. Figure 1.1 shows the implemented carbon labels, where the carbon footprint is accompanied by a green leaf and placed next to other food labels displaying nutritional ingredients and other certificates (e.g., ecological, country of origin, and that the product is free of certain ingredients).[11]

2. Figure A1 displays the checkout page, where consumers can see the total carbon footprint of their order. This is calculated by considering the weight, quantity and carbon footprint of each product in the consumption basket. Consumers can also go back to their historical baskets on their member page.

3. On their member page, consumers can compare the average carbon footprint of their historical orders with that of all other consumers. The comparison is based on each basket's carbon footprint divided by the basket's weight.

4. In addition to CFLs, the store provides "climate-smart suggestions," placing a set of products with a lower carbon footprint at the top of the page when consumers search for a set of pre-selected keywords. Figure A2 shows how suggestions of similar vegan options were displayed for a search on ground meat.

This feature was pre-launched in a pilot study from June to November 2019. Consumers were asked in a pop-up window if they wanted to try out a new feature that would help their shopping be more "climate-smart." Consumers could opt in by clicking on "I want to try" or opt not take part by clicking on "Not now." Around 25 percent of consumers chose to opt in. During this pilot study, the interface was slightly different, as shown in Figure A3, where there is no green leaf next to the carbon footprint label.

From 12 November 2019 onward, climate-smart suggestions were activated by default for all consumers. Consumers could, however, opt out if they preferred not to get climate-smart suggestions. Around 99 percent of consumers decided to keep the default of having climate-smart suggestions turned on, and 1 percent actively turned off this feature.

[11] The green leaf provides a neutral frame for visualizing the numerical carbon footprint. Previous research has sometimes found that colorful schemes can increase the salience and effectiveness of food labels, whereas our treatment provides a neutral frame for this.

Table 1.2: Timeline of natural experiment

	Shown to consumers			
	On products		At checkout	
	Carbon footprint	Climate-smart suggestions	Carbon footprint	Comparison module
Time period				
-31/05/2019	No	No	No	No
01/06/2019-11/11/2019	No	Option to opt in (25% did)	No	No
12/11/2019-01/04/2022	Yes (100%)	Default with option to opt out (99% stayed)	Yes (100%)	Yes (100%)

Notes: This table provides the timeline of the natural experiment at the treated supermarket. The share of consumers affected by each treatment is given within parentheses.

1.3.3 Anticipation of treatment

During the pilot project on climate-smart suggestions described in Section 1.3.2, customers could read about the project on an information page on the store's website under the tab labeled "Food, environment, and climate/Climate-smart search." The information page described "climate-smart suggestions" as a new feature showing products with a lower climate impact when customers search on a set of keywords and explained that this feature was part of the store's broader work on climate. This page was available for all consumers to read, but they had to actively access it. The store informed its customers that RISE (2020) was responsible for the carbon footprint estimates and was working with the store to expand the products available in the feature.

As outlined in Section 1.2.2, it is common for Swedish supermarkets to communicate to their customers that they care about consumers' environmental impact (and climate in particular). However, one concern is that the pilot project might have attracted customers curious about this feature and that those preferences are correlated with their consumption's carbon footprint. In Section 1.6.5, I provide a robustness check where I rerun the main analysis, excluding those who opted to participate in the pilot project.

1.3.4 How can consumers reduce their carbon footprint?

Table 1.3 shows the carbon footprint of several products. Consumers can reduce their emissions in the treated store through two channels. First, they can instead purchase products with a lower carbon footprint. For example, if consumers would replace 1 kg of beef with 1 kg of poultry, that would reduce their carbon footprint by 25 kg CO_2e. However, consumers can also replace the beef with a mixture of beef and pork, which would also provide a significant reduction in their carbon footprint.

The second channel is through the volume of consumption at the treated store. I observe consumers' purchases of food within a single store and do not observe the same consumers' food purchases in other stores. We should therefore be worried about consumers reducing their carbon footprint in the treated store but increasing their carbon footprint in other stores. For example, if consumers reduce their beef purchases in the treated store without increasing their purchases of lower-impact products. In Section 1.6.2, I study consumers' purchasing patterns within the meat category to assess whether high-impact products are replaced by lower-impact products. In the analysis, I further control for the order size in spending or weight.

Table 1.3: Carbon footprint of food products

	kg CO2e	Country of origin	Calculation based on one kg
Meat			
Beef	41	Brazil	Boneless meat, not cooked
Beef	28	Sweden	Boneless meat, not cooked
Lamb	21	Sweden	Boneless meat, not cooked
Beef and pork mix	16	Sweden	Boneless meat, not cooked
Pork	4.2	Sweden	Boneless meat, not cooked
Poultry	2.6	Sweden	Boneless meat, not cooked
Quorn	1.7	Great Britain	Quorn mince
Peas, Lentils, and Beans	0.2–0.4	Sweden	Dried Peas, Lentils, or Beans
Carbohydrate sources			
Jasmine Rice	3.1	Thailand	Uncooked Rice
Pasta	0.8	Sweden	Uncooked Pasta
Potatoes	0.1	Sweden	Uncooked Potatoes
Dairy			
Butter	8	Sweden	Butter
Cheese (fat 31%)	5.3	Sweden	Cheese
Milk (fat 1.5%)	0.9	Sweden	Milk

Notes: This table shows kg carbon emissions equivalents from 1 kg of a food product from the publicly available list on the RISE (2020) website. RISE (2020) have estimated all carbon emissions equivalents displayed on the carbon labels in the treated store.

1.4 Consumer data

The data set consists of all online food transactions from 2018 to 2022 from both supermarkets.

1.4.1 Member account

To shop online, consumers need to have a registered member account. For privacy reasons, the stores do not share consumer backgrounds, so I do not observe whether an individual or household is simultaneously a member of both stores. In Section 1.6.5, I discuss how this may impact the results and present robustness checks to the main results.

For each account, I obtain an indicator for a private consumer or a business organization. In the analysis, I restrict the sample to private consumers.[12] I also observe members' postal codes on a three-digit level. I map postal codes with register data from Statistics Sweden to acquire aggregate socioeconomic background information on income, education, and household composition.[13]

1.4.2 Transactions

I obtained the universe of individual transactions from both supermarkets for January 2018 to February 2022. Each transaction includes the date, time, and postal code of delivery; furthermore, it includes member IDs, enabling me to follow individual members' consumption over time. For each product in the order, I observe the quantity, price, and any applicable discounts. I also have background information on a rich vector of characteristics for each product, including other food labels, such as ecological and Fairtrade labels; country of origin; and if the product is free of certain ingredients, such as milk or meat. To maintain consumers' privacy, the data set only includes food products.

1.4.3 Carbon footprint

Each product carries a unique product ID, and I obtained the implemented CFLs for all products that have been labeled in the treated store. Food consumption varies with the season, and products are regularly added and removed from the assortment in both the treated and control stores. The treated store updated its assortment by labeling newly added products and removing previously labeled (and unlabeled) products. Carbon-labeled products make up around 80 percent of the sales volume in the treated store, and this share has remained stable over time. I observed these labels ex post, which means the data set contains CFLs for all products that have

[12] I also obtain variables that unfortunately are not shared between stores, including age data in 10-year intervals from both stores. The control store collects the ages of all members, whereas the treated store only collects age data when customers pay by invoice. The treated store also has access to information regarding whether the customer lives in a house or an apartment.

[13] Covid affected the consumer base, with a large inflow of consumers in both stores. It also affected how many deliveries could occur, with the treated store being closer to its limit. This resulted in delayed delivery times, causing some consumers to switch to other stores.

been labeled at some point, but I did not observe the date a specific product was labeled.

To cover important emissions in the control store for products that are not shared between stores, I additionally measure emissions on meat and dairy products in the control store using the publicly available carbon footprint estimates from RISE (2020). The methodology is outlined in Appendix 1.9.1, and Figure A4 displays the volume share on which I can measure emissions. This study covers 80 percent of the sales volume in the treated store and 70 percent in the control store, and these shares have remained stable over time.

I create a data set that measures each order's aggregate carbon footprint to track a consumer's carbon footprint over time. Carbon labels (L^j) express product *j's* kg carbon emissions equivalents (kg CO_2e) per kg of the product. Hence, we need to multiply the label (L^j) by its weight in kg (w^j) to obtain product *j's* carbon footprint, $E^j = L^j * w^j$ in kg CO_2e. This is analogous to how one could multiply a product's price per kilo to obtain the product price. The aggregate carbon footprint of order k from all the j products in the order is given by $E_k = \sum_j E_k^j$.

1.4.4 Summary statistics

Table 1.4 provides summary statistics on orders for those consumers who are observed both in the pre-and post-treatment period. I observe around 450,000 orders purchased by 32,000 consumers. As online consumption generally contains a fixed delivery cost, transactions are sizable, and the average order cost is around 1,100 SEK. On average, the carbon footprint before carbon labels were implemented was around 33 kg CO_2e and 34 kg CO_2e per order in the treated and control store, respectively. After the introduction of carbon labels, the carbon footprint decreased slightly in the treated store to 32.9 kg CO_2e and rose in the control store to 34.5 kg CO_2e. After the outbreak of the Covid pandemic, the order size grew slightly in both cost and carbon footprint. During the full post-treatment period (from 12 November 2019 to October 2020), the average carbon footprint was 34.5 kg CO_2e in the treated store and 36 kg CO_2e in the control store. In Sweden, these numbers represent around 2.5 percent of the average annual carbon footprint from food per person.[14]

Table 1.4: Summary statistics on orders

	Short term (01/07/2019-01/03/2020)		Long term (01/07/2019-01/10/2020)	
	Treated	Control	Treated	Control
Orders (N)	202,628	67,786	316,410	119,650
Consumers (N)	21,053	7,214	24,030	8,275
Order per consumer	10	9	13	14
	(8)	(8)	(13)	(14)
Order amount (SEK/Order)	1,056	1,106	1,102	1,140
	(503)	(490)	(541)	(506)
kg CO_2e per order				
Before treatment period	33,1	33,8	33,1	33,6
	(23,4)	(27,9)	(23,5)	(27,7)
After treatment period	32,9	35,9	34,5	36,0
	(23,0)	(30,8)	(24,0)	(29,5)

Notes: This table reports the mean consumption order characteristics made by the sample observed before and after carbon labels were introduced in the treated store on 12 November 2019. Standard deviations in parenthesis. The short term is from July 2019 to March 2020, and the long term is from July 2019 to October 2020.

Table 1.5 presents summary statistics on aggregate food categories. Column 1 contains raw emissions and shows that around 70 percent of emissions arise from meat and dairy products. This figure is striking, but it is known from previous research and is representative of the carbon footprint for the general Swedish population (Naturvardsverket 2020).

Column 2 contains overall revenue shares and shows that meat and dairy contribute to around 40 percent of the store's revenue, and their emission share is at close to 70 percent. In column 3, I divide the emissions share by the revenue share. A number greater than 1 should be interpreted as the emission share being larger than its revenue share. The emissions share of meat is responsible for more than twice its revenue share. Dairy products are also overrepresented in their emissions share, but less so than meat. To summarize, meat and dairy products contribute to the large majority of emissions and more so than their revenue share.

[14]The average annual GHG emissions from food are around 1,400 kg CO_2e per capita in Sweden (Naturvardsverket 2020).

Table 1.5: Summary statistics on emissions and revenue from food categories

	Emissions share (%)	Revenue share (%)	$\frac{\text{Emissions share}}{\text{Revenue share}}$
Meat and charcuterie	38.0%	17.7%	2.147
Dairy and egg	31.1%	26.7%	1.165
Storage	8.7%	12.4%	0.702
Ready to eat	8.3%	5.4%	1.537
Fruit and vegetables	5.7%	17.0%	0.335

Notes: This table contains summary statistics on aggregate emissions from the five largest categories. Column 1 contains the raw emissions share per category, column 2 is the revenue share, and column 3 is the emissions share divided by the revenue share. A number greater than 1 indicates a higher emissions share than its revenue share.

1.5 Empirical methodology

This section first describes the main outcome measures I use to study whether carbon labels caused consumers to reduce their emissions. Next, I present the regression models and the assumptions that underlie the analysis to interpret the estimates as causal. I then provide graphical evidence of average emissions per order and details on whether there were any simultaneous changes in the prices of important product categories around the treatment date.

1.5.1 Main outcome measures

To evaluate if carbon footprint labels (CFLs) cause consumers to reduce their emissions, I compare consumers' carbon footprint before and after CFLs were introduced. My unit of measurement is a customer order, and I use four measures of the order's carbon footprint. The first measure is the sum of kg carbon emissions equivalents (kg CO_2e) per order. The sum of kg CO_2e (E) generated from the j products in order k is given by,

$$E_k = \sum_j E_k^j \tag{1.1}$$

As this is a partial equilibrium analysis, where individuals' consumption is observed within a single store, a concern is that consumers could reduce their carbon footprint by reducing the quantity of food purchased at the treated store (and buying those products in another store). I therefore also measure consumers' carbon footprint by the emissions intensity by the cost and weight of the order. I do this by dividing the order's kg CO_2e by the order cost, which is calculated by summing the cost in SEK of j products in order k ($C_k = \sum_j c_k^j$). Similarly, I divide the order's kg CO_2e by the order weight in kg, where the order weight of j products in order k is given by $W_k = \sum_j w_k^j$.

From a climate mitigation standpoint the absolute value of kg CO_2e (E_k) per order should be the most interesting outcome measure. Otherwise I will use the natural logarithm for each measure as our outcome measures. This allows us to interpret the consumer response proportional to each consumers own carbon footprint,

$$\widetilde{E}_k = \log(E_k) \tag{1.2}$$

$$\widetilde{E}_k^c = \log(\frac{E_k}{C_k}) \tag{1.3}$$

$$\widetilde{E}_k^w = \log(\frac{E_k}{W_k}). \tag{1.4}$$

$\widetilde{E}_k^c$ measures the logarithm of the order's carbon footprint, $\widetilde{E}_k^c$ measures the logarithm of the order's carbon footprint per order cost, and $\widetilde{E}_k^w$ measures the logarithm of the order's carbon footprint per kg of food in the order k. For small changes in the dependent variable, we can interpret the effect of a marginal change in the independent variable as a percentage change in the dependent variable. The independent variable in our case is the carbon label implementation, which is "switched on" from 0, where no products have a carbon label, to 1, where the assortment is labeled with each product's carbon footprint.

1.5.2 Main specification

To evaluate whether carbon labels cause consumers to reduce their emissions, I compare consumers' carbon footprint per order in the two stores before and after CFLs were implemented in the treated store. To this end, I estimate a series of difference-in-difference models of the following form:

$$\mathrm{Y}_{istk} = \beta_0 + \gamma_i + \gamma_t + \mathit{Treated} + \mathit{Post} + \beta_1(\mathit{Post} * \mathit{Treated}) + e_{istk}. \quad (1.5)$$

The dependent variable Y refers to one of our four outcome measures specified in Section 1.5.1, such as how much kg carbon emissions equivalents (kg CO_2e) an order contains (E). Y_{istk} measures the outcome of interest per order for consumer i when shopping in store s during time period t in order k. The store subscript s is either the treated store (s = *Treated*) or the control store (s = *Control*).

Individual fixed effects, γ_i, capture time-invariant individual characteristics correlated with a consumer's carbon footprint, such as allergies or other time-invariant dietary preferences. It also captures household composition for most households and other variables that remain constant during the study period. *Treated* is a dummy variable equal to 1 for consumers in the treated store and 0 for consumers in the control store. This dummy variable captures time-invariant differences between the stores, such as specific characteristics of the assortment or differences in the consumer base that remain constant over time. Time fixed effects, γ_t, capture variation in food consumption over time that is not correlated with a specific store. For example, online food purchases and meat consumption generally rise during the winter.

Post is a dummy variable taking the value of 1 from the date of treatment onward. The interaction term $\mathit{Post} * \mathit{Treated}$ captures variation in CFLs across stores over time. It is equal to 0 for both stores in the pre-treatment periods (as *Post* is equal to 0) but switches to 1 for all consumers in the treated store in the post-treatment period. As there is no variation in the individual's choice between the two supermarkets over time, *Treated* will be absorbed by all specifications that contain individual fixed effects, γ_i. This would also occur for *Post*, which will be absorbed by γ_t, as the time fixed effect level is on a lower level than the *Post* dummy. A more condensed version of equation (1.5), where I have omitted the absorbed store fixed effects and the *Post* dummy, is given by

$$\mathrm{Y}_{istk} = \beta_0 + \gamma_i + \gamma_t + \beta_1(\mathit{Post} * \mathit{Treated}) + e_{istk}. \tag{1.6}$$

I also estimate the short-term and long-term responses separately by splitting the post-treatment period into before and after the Covid pandemic, as illustrated in Table 1.6. I do this by splitting the *Post* dummy into two time periods.

Table 1.6: Treatment dummy variables

Variable name	*Post*	*Before Covid*	*After Covid*
Date			
01/07/2019-11/11/2019	0	0	0
12/11/2019-28/02/2020	1	1	0
01/03/2020-30/09/2020	1	0	1

Before Covid is a dummy variable taking the value of 1 from the date of treatment (12 November 2019) until the beginning of March 2020, and 0 otherwise. *After Covid* is a dummy variable taking the value of 1 from the beginning of March 2020 and onward, and 0 otherwise.

$$\begin{aligned} \mathrm{Y}_{istk} = \beta_0 + \gamma_i + \gamma_t &+ \beta_1(\mathit{Before\ Covid} * \mathit{Treated}) \\ &+ \beta_2(\mathit{After\ Covid} * \mathit{Treated}) + e_{istk}. \end{aligned} \tag{1.7}$$

The parameter of interest is β_1 (and β_2 in equation 1.7), which measures the causal average effect of carbon labels on consumers' carbon footprint per order for those consumers in the treated store who shopped before and after carbon labels were implemented.[15] For $\hat{\beta}_1$ ($\hat{\beta}_2$) to consistently estimate β_1 (β_2), I rely on the assumption that consumers in both stores would otherwise follow a parallel trend in the dependent variable. The parallel trend assumption would be invalid if the two stores followed different trends in emissions before introducing carbon labels or if the two stores have different trends post-treatment for some other factor that is not related to

[15] It is important to again emphasize that the data do not contain information on whether consumers were members in both stores simultaneously. The identification of β_1 (β_2) emerges from variation in CFLs across stores over time.

the treatment (for example, if one store gradually changed its assortment to become more environmentally friendly or suddenly changed its pricing structure on some environmentally friendly products).

To assess whether the two stores followed a parallel trend before carbon labels were implemented, I present event plots by estimating versions of model (1.8) where j refers to time periods before (leads) and l to time periods after (lags) carbon labels were implemented. The reference period for these events is the time period right before (t = -1) carbon labels were implemented and is therefore omitted from the model.

$$\mathrm{Y}_{istk} = \beta_0 + \gamma_i + \gamma_t + \sum_{j=1}^{J-1} \mathit{Treated} * \gamma_j + \sum_{l=0}^{L} \mathit{Treated} * \gamma_l + e_{istk}. \tag{1.8}$$

I cluster standard errors at the consumer level (member account level) in all specifications, allowing errors to be correlated within the same consumer (member account) over time but not across different consumers (member accounts).

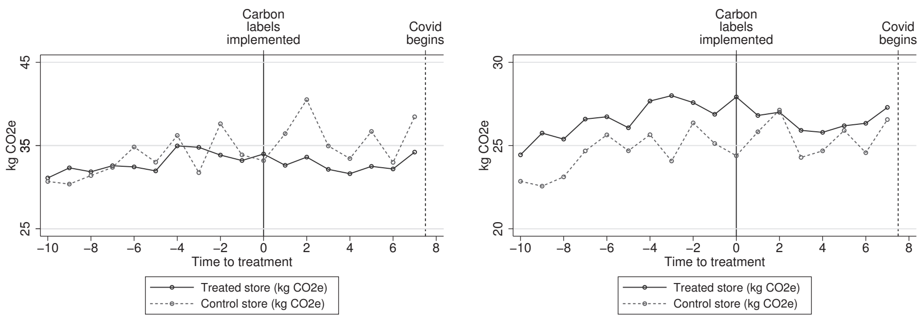
1.5.3 Graphical evidence and prices

Figure 1.2 presents average kg CO_2e per order for each store until the Covid pandemic began, and the vertical line indicates when carbon labels were introduced. Panel (a) contains the full sample, and panel (b) is restricted to those consumers that did not purchased an order above the 95th percentile of kg CO_2e per order (80 kg CO_2e). Food consumption varies depending on seasonal patterns, and there is a trend toward orders having a higher carbon footprint in winter than in summer. This is primarily explained by higher meat consumption during the winter. There is a slight drop in the carbon footprint per order in the treated store after the introduction of carbon labels. In contrast, the control store's carbon footprint does not display this decline. There is more variation in the control store, and one reason is that the sample size is around one-third of that in the treated store. There is notably a sharp increase in the control store during period 2 (in late December 2019). This is the result of a sale on beef implemented in the control during that period where some consumers purchased large quantities of beef. Looking at the bottom 95 percent of emitters in panel (b), the carbon footprint declines in the treated store and remains more stable in the control store and less responsive to the beef sale during period 2.

Figure 1.2: Emissions per order

(a) Full sample

(b) Excluding top 5% of emitters



Notes: This figure shows average emissions per order in each store from July 2019 to March 2020. Each diamond represents a time period of 14 days, and the vertical line indicates when carbon labels were implemented in the treated store (12 November 2019 and onward)

However, simultaneous price changes around the treatment could be an issue. For example, the treated store could have lowered the price if they wanted their customers to buy more low-impact products. To assess whether there were systematic price changes around the treatment date, I have plotted the average prices of certain meat products in each store over time in Figure A5. Meat is the main emissions source, in particular, beef products. I have therefore plotted the average prices in 14-day intervals of beef as well as pork or pork mixed with beef, poultry, and vegetarian products.[16] Baseline food prices (those prices that a store charges when there is no sale) are not adjusted regularly during the year, but some products go on sale from time to time. Baseline food prices differ slightly between stores, with prices of beef, pork, vegetarian products, and poultry generally being around 20 SEK higher per kg in the treated store. As mentioned in Section 1.2, this is explained by the treated store offering free delivery, making the

[16] I focus on three product categories with high, middle, and low carbon-intensive products, which are also analyzed in Section 1.6.3.

total order cost similar between stores.

The average prices of vegetarian products and poultry remain stable over time in both stores, without any notable sales from one 14-day interval to the next. One explanation is likely that these products are generally frozen, and the stores do not have the same incentive to sell them before the upcoming expiration date. However, beef and pork products do go on sale on a regular basis. This could be an issue if either the control or treated store changed its frequency of sales (or adjusted its baseline prices) around the treatment. For example, it would be a concern if one store permanently lowered its price on beef around the treatment date while the other store did not. In Table A1, I present the regression results of the price difference between the stores[17] on the *Post* dummy. I do this analysis for all three product categories: high-impact products (beef), middle-impact products (pork or pork mixed with beef), and low-impact products (vegetarian and poultry). I do not find any statistically significant difference between stores when I compare the price difference between stores before carbon labels were introduced with the price difference after carbon labels were introduced. This is reassuring and means I do not find statistical significant evidence that either store made any change in its sales or baseline prices in any meat category around the treatment date.

One might still be concerned that prices may have had an impact on the results. Price is an endogenous variable set by the store, so it is not apparent that one should want to control for it in the regression analysis. However, in Section 1.6.5, I rerun the main analysis by controlling for prices.

1.6 Results

I begin by considering the main research question of the paper: Do carbon labels cause consumers to reduce their carbon footprint? Individual transaction data enable us to track individuals' carbon footprint over time and study this question at the individual consumer level. Carbon labels could also affect a consumer's decision to shop at the treated store, and in Section 1.6.3, I examine how CFLs affected spending and supermarket

[17] I take the average price in the treated and subtract the average price in control stores for each 14-day interval.

choice.[18] In Section 1.6.4, I provide heterogeneity analysis based on consumers' carbon footprint prior to the introduction of carbon labels, and in Section 1.6.5, I conduct robustness checks to the main results. Section 1.6.6 presents some survey evidence on the representativeness of my sample and Swedish consumers in general compared with European citizens.

1.6.1 Main results

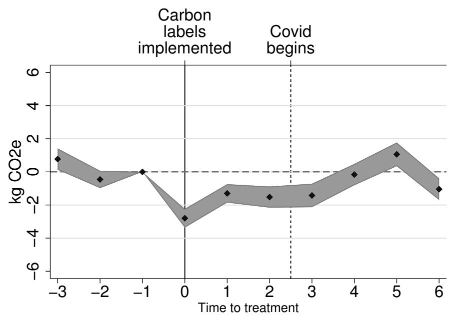
Figure 1.3 presents event plots from estimating equation (1.8) using as the dependent variable emissions per order in panel (a), the log of emissions per order in panel (b), the log of emissions per order cost in panel (c), and the log of emissions per kg of food in panel (d). Emissions per order refers to kg CO_2e per order. All models include fixed individual effects that control for time-invariant consumer characteristics. To reduce the impact of short-term sales, the event plots are presented with a 45-day interval. Each diamond represents the difference in emissions between stores, compared with the difference in carbon emissions in the period before CFLs were introduced. A negative value should be interpreted as the dependent variable being lower in the treated store than in the control store relative to the difference between stores in the reference period (-1).

All four figures show a decline in emissions per order in the treated store immediately after introducing carbon labels. Looking at the panel (a), consumers in the treated store reduced their carbon footprint by around 2 kg CO_2e per order and close to 3 kg CO_2e during the first 45-day interval after carbon labels were introduced. The effect of carbon labels is statistically significantly lower in the first four 45-day periods for all measures, but then rises somewhat in the last three 45-day periods.

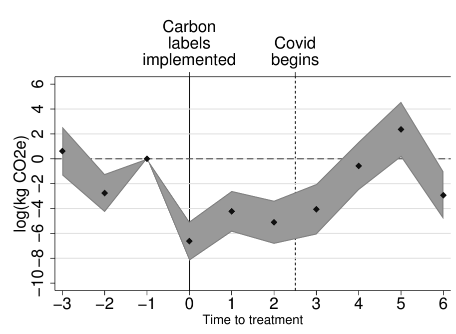
[18] With aggregate data, it would not have been possible to disentangle changes in the consumer base from individuals' choices of food products.

Figure 1.3: Event plots of effects of Carbon labels on emissions per order

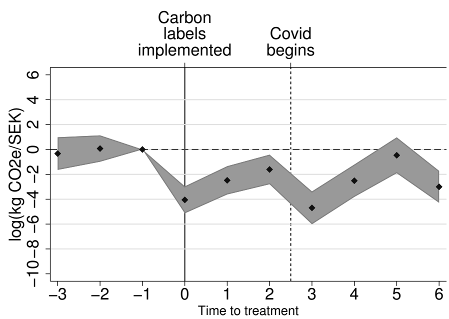
(a) Emissions



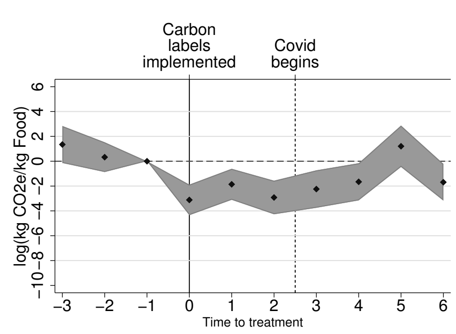
(b) log(Emissions)



(c) $\log(\frac{\text{Emissions}}{\text{Cost}})$



(d) $\log(\frac{\text{Emissions}}{\text{Weight}})$



Notes: This figure shows the long-term response (July 2019 to October 2020) of carbon labels on emissions per order from estimating equation (1.8). Emissions per order are measured as kg CO_2e per order. Cost and weight refer to the order cost and kg of food in the order, respectively. A time period refers to 45 days, and the solid vertical line indicates when carbon labels were implemented in the treated store (12 November 2019 and onward). The excluded period (time period -1) is the reference period and refers to the 45 days before the treatment was implemented. Each diamond indicates the difference-in-difference point estimate, γ_k, compared with the reference period, with a 95 percent confidence interval around it indicated by the grey area. A negative value should be interpreted as the dependent variable being lower in the treated store than in the control store relative to the difference between stores in the reference period (-1). All estimates where the dependent variable has been log-transformed have been multiplied by 100.

To assess the average impact of carbon labels on consumers' carbon footprint per order, I estimate equation (1.6) and (1.7) and look at three time periods: the average overall effect until October 2020, the average short-run effect before Covid, and the average effect after the breakout of the Covid pandemic.

Table 1.7 shows the difference-in-difference estimates, where column (1) contains the emissions per order as the dependent variable, column (2) is the log of emissions per order, and columns (3) and (4) show the log of emissions divided by the order cost and kg of food in the order, respectively. Emissions per order is (as before) defined as kg CO_2e per order.

All estimates where the dependent variable has been log-transformed have been multiplied by 100. For small changes in the dependent variable, we can therefore interpret the point estimates in column 2–4 as a percentage change in the dependent variable from implementing carbon labels in the treated store. For example, the point estimate of -2.693 in column (2) should be interpreted as kg CO_2e per order was reduced by about -2.7% after carbon labels were introduced in the treated store compared to the control store.

We can begin to look at the first row in Panel B, that shows how carbon labels affected consumers emissions in the treated store in short term. Column (1) contains kg CO_2e per order and in the short term, before Covid, consumers' carbon footprint per order was reduced by around 2 kg CO_2e in the treated store. Corresponding to around 6 percent of the average carbon footprint per order (the mean being 34 kg CO_2e per order). The reduction was also statistically significant for the other outcome measures in column 2–4. The log of emissions was reduced by about 5 percent, and the log of emissions divided by the order cost were reduced by 3 percent and the log of emissions per kg of food by close to 4 percent.

As indicated by the event plots in Figure 1.3, the effect of carbon labels on consumers' emissions per order does appear to decline over time. Looking at the period after the Covid pandemic began in the second row in Panel B, from March to October 2020, consumers in the treated store reduced emissions per order by around 0.8 kg CO_2e per order, corresponding to 2 percent of the average carbon footprint per order. The effect is not statistically significant, however, when looking at the log of emissions per order. The other two outcome measures show a similar pattern, where the effect after the onset of Covid is less than the short-term effect. However, the effect of carbon labels is statistically significantly lowering in column 3–4, where I use the log of kg CO_2e per order divided by the cost or the weight of the order as the dependent variable. Using these outcome measures consumers in the treated stores lowered their emissions per order by about 1–2 percent.

Turning our attention to the overall response in Panel A, using the whole post-treatment period as the comparison period, the overall effect lies between the short- and long-term effects. Overall, consumers' emissions per order was reduced by 1.4 kg CO_2e per order, corresponding to around 4 percent of the average kg CO_2e per order and around 2.5 percent when instead using the log of kg CO_2e per order divided by the cost or weight of the order as the dependent variable.

To summarize, I find that carbon labels reduce consumers' carbon footprint by around 3 to 5 percent from 12 November until the onset of the Covid pandemic at the beginning of March 2020. The consumer response declined over time to 1–2 percent after the onset of Covid, and the point estimates are here less precise and the reduction is not statistically significant for the log of emissions per order. The overall response over the whole data period from July 2019 to October 2020 is between the short- and long-term responses, at around 2.5 percent in reduced kg CO_2 per order for consumers in the treated store.

Table 1.7: Effects of Carbon labels on emissions per order

	(1) Emissions	(2) log(Emissions)	(3) $\log(\frac{\text{Emissions}}{\text{Cost}})$	(4) $\log(\frac{\text{Emissions}}{\text{Weight}})$
Panel A. Overall				
Carbon Labels	-1.418***	-2.693***	-2.715***	-2.423***
	(0.211)	(0.546)	(0.377)	(0.429)
Panel B. Post period split into before and after Covid				
Carbon Labels (Before Covid)	-2.150***	-4.880***	-3.225***	-3.788***
	(0.221)	(0.565)	(0.394)	(0.447)
Carbon Labels (After Covid)	-0.804**	-0.858	-2.287***	-1.278*
	(0.282)	(0.679)	(0.453)	(0.524)
Individual and time FE	Yes	Yes	Yes	Yes
Observations	443,131	443,131	443,131	443,131
Clusters	32,339	32,339	32,339	32,339

Standard errors clustered on member account level in parentheses

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Notes: This table shows the overall response (July 2019 to October 2020), before Covid response (July 2019 to March 2020) and after Covid results (March to October 2020) from estimating equation (1.7). Emissions per order are measured as kg CO_2e per order. Cost and weight refer to the order cost and kg of food in the order, respectively. The sample is restricted to those consumers who are observed before and after carbon labels were introduced in the treated store. The DiD estimates should be interpreted as the change in the dependent variable caused by introducing carbon labels in the treated store. Emissions refers to kg CO_2e per order. All regressions include individual and time fixed effects, and standard errors are clustered at the consumer level. All estimates where the dependent variable has been log-transformed have been multiplied by 100.

1.6.2 Mechanisms behind the carbon emissions reductions

The previous section documented that consumers responded to CFLs by reducing their carbon footprint. In this section, I focus on the mechanisms behind that reduction. We should expect consumers to increase their demand for low-impact products and consume fewer high-impact products, or at least not increase their demand for high-impact products. It is a priori less clear in what direction demand should shift for middle-impact products. Demand for such products may increase if consumers shift from high carbon-intensive products. However, demand could decrease if consumers shift from middle-impact products such as pork to less carbon-intensive products such as poultry or vegetarian options.

I create a data set of high-, middle-, and low-impact products for this analysis. I focus on the main driver behind carbon emissions: meat consumption. For meat products, the high-impact products are beef, the middle-impact products are pork and pork mixed with beef, and the low-impact products are vegetarian or vegan and poultry.

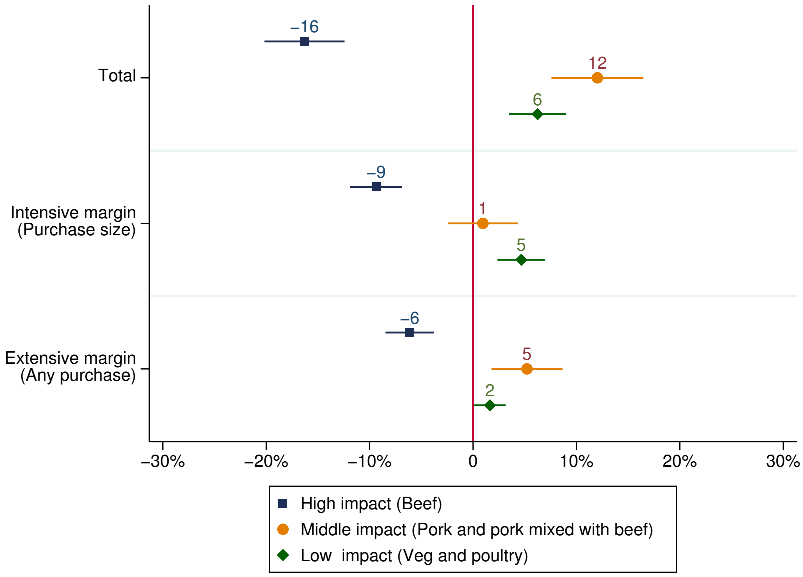
Figure 1.4 contains the short-term difference-in-difference estimates for the meat category, and the results are presented for the overall response and the extensive and intensive margins. I estimate this model separately for high-, middle-, and low-impact products compared with the average consumption of these products during the pre-treatment period. The plotted coefficient should be interpreted as the percentage change in high, middle or low impact products in the treated store compared with the control store after carbon labels were implemented until the onset of the Covid pandemic.

In the short term, until March 2020, the first finding is that beef consumption was reduced by 16 percent, corresponding to a 52-gram reduction per order. Middle- and low-impact products increased by 12 and 6 percent, or 23 and 42 grams, respectively—providing evidence that carbon labels caused consumers to replace beef with low- and middle-impact products. In emissions terms, this translates to a reduction of around 1.2 kg CO_2e, or 3.4 percent of the average carbon footprint, per order.

The lower beef consumption arises from a 10 percent decrease in how often consumers purchased beef (extensive margin) and a 6 percent decrease in the volume of beef consumption (intensive margin). For low-impact products, the intensive margin represents the most significant increase, and for middle-impact products, the extensive margin explains most of the response.

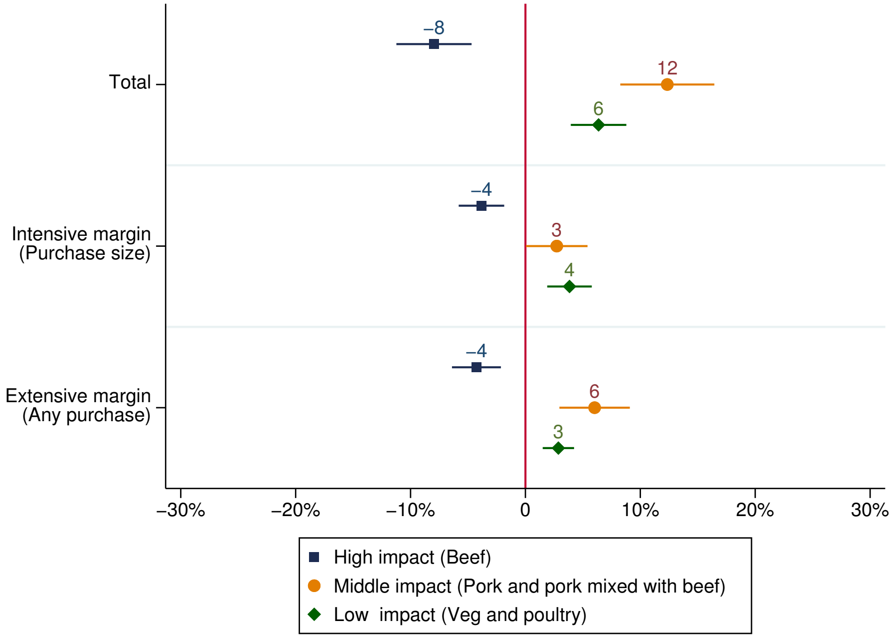
Figure 1.5 extends the time period until October 2020. Beef consumption is reduced by 8 percent, corresponding to 25 grams of beef. However, the increase in middle- and low-impact products remains at 12 and 6 percent, corresponding to increased consumption of about 24 and 43 grams of low- and middle-impact products, respectively. Replacing 25 grams of beef with middle- or low-impact products corresponds to a reduction of around 0.3 and 0.6 CO_2e, respectively. The extensive margin drives most of the response for middle-impact products, whereas the extensive and intensive margins are similar for low-impact products.

Figure 1.4: Short-term effects of carbon labels on meat product choice



Notes: This figure shows the short-term response (July 2019 to March 2020) of carbon labels on meat products for each product category. High-impact products contain beef, middle-impact products are either pork or a mixture of pork and beef, and low-impact products are vegetarian, vegan, or poultry products. The symbol indicates the DiD estimates for each category, and the bar indicates a 95 percent confidence interval. Estimates should be interpreted as the percentage change in the dependent variable relative to the average consumption in the pre-treatment period. All regressions include individual and time fixed effects, and standard errors are clustered at the consumer level.

Figure 1.5: Long-term effects of carbon labels on meat product choice



Notes: This figure shows the long-term response (July 2019 to October 2020) of carbon labels on meat products for each product category. High-impact products contain beef, middle-impact products are either pork or a mixture of pork and beef, and low-impact products are vegetarian, vegan, or poultry products. The symbol indicates the DiD estimates, and the bar indicates a 95 percent confidence interval. Estimates should be interpreted as the percentage change in the dependent variable relative to the average consumption in the pre-treatment period. All regressions include individual and time fixed effects, and standard errors are clustered at the consumer level.

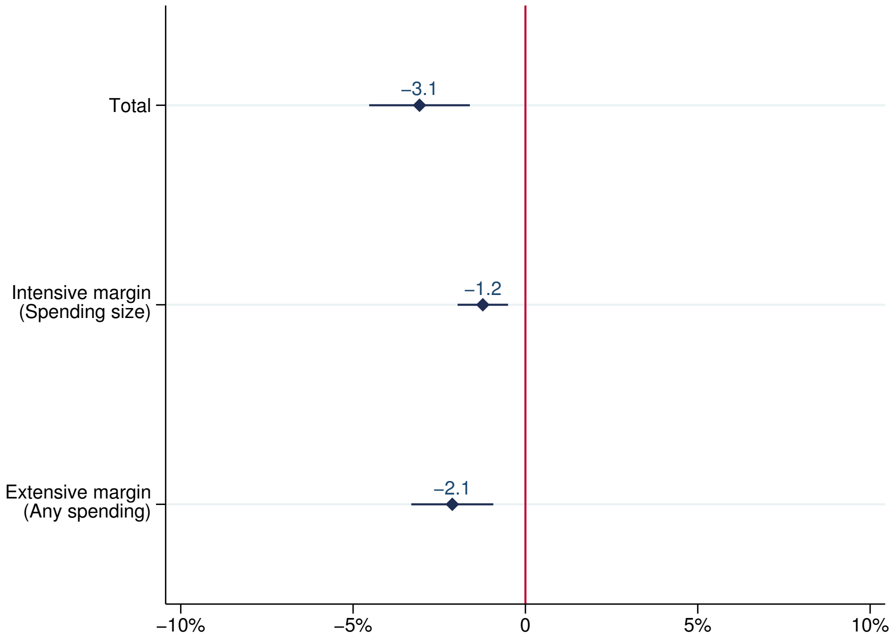
1.6.3 Do carbon labels affect spending?

In this section, I study how carbon labels affect (i) how often consumers shop at the treated store and (ii) how much they spend per order.

I begin to look at how carbon labels affected spending and supermarket choice, considering the shopping frequency and spending per order. In Figure 1.6, I have plotted total spending, average spending per order (intensive margin), and shopping frequency (extensive margin) for all consumers who were members prior to the treatment. I find an overall reduction in spending at the treated store of 3.1 percent, which amounts to about 7.5 euros per 45-day period. The reduction is driven primarily by the extensive margin, where I find that consumers in the treated store reduced their shopping frequency by around 2 percent.

I also find a reduction in the intensive margin of around 1.2 percent, which amounts to about 3 euros per 45-day interval. It should be emphasized that reduced spending per order of this magnitude is consistent with a consumer replacing beef with products such as pork or vegetarian products. Those products, in general, cost about 2–4 euros per kg less than beef.[19]

Figure 1.6: Effects of Carbon labels on spending per order



Notes: This figure shows average spending per consumer during each 45-day period in each store from July 2019 to March 2020. The DiD estimates should be interpreted as the percentage change in the dependent variable relative the average spending during pre-treatment period. The intensive margin refers to spending per order and the extensive margin the shopping frequency. All regressions include individual and time fixed effects, and standard errors are clustered at the consumer level.

I now use all the data to look at overall spending by aggregating spending

[19] This does not imply that the store's profit is reduced, as we need to consider the markup on those products and not only the revenue.

on the store level and every two weeks, comparing average spending before and after carbon labels were introduced. This means taking into account spending by consumers who were members before the treatment and those who became members afterward.

In Figure A6, I have plotted the coefficient estimates from this exercise and do not find any statistically significant change in overall spending. One explanation could be that even if previous customers reduced their shopping frequency at the treated store, carbon labels also attract new customers who appreciate knowing their carbon footprint. It should also be noted that one observation refers to two weeks for each store when aggregating store-level spending. Hence, the power to detect any change is likely small.

1.6.4 Consumer heterogeneity

In this section, I study consumer heterogeneity based on consumers' carbon footprint prior to the introduction of carbon labels. The background to this exercise is that consumers who already have a low carbon footprint may have a limited possibility to reduce their impact further, while consumers with a relatively high carbon footprint would have more potential to reduce their carbon footprint, but that could be the result of them caring less (on average) about their carbon footprint.

So as not to primarily capture the consumer's household size, I instead focus on the consumer's emissions intensity per kg of food in their order, where E_k refers to the kg CO_2e in order k and W_k to the weight (in kg) of order k. The emissions intensity per kg of food in order k is then given by

$$E_k^w = \frac{E_k}{W_k}. \tag{1.9}$$

I begin by calculating E^w for every order and sum over all of individual i's N_i orders during the pre-treatment period. I then divide by N_i to calculate that consumer's average E^w during the pre-treatment period:

$$\bar{E}_i^w = \frac{1}{N_i} \sum_{k=1}^{N_i} E_{ik}^w. \tag{1.10}$$

Table 1.8 contains the descriptive summary on this exercise. The second column in the first row shows the median value of E^w among all consumers during the pre-treatment period ($\bar{E}_i^w = 1.806$). The value $\bar{E}_i^w = 1.806$ refers to the median kg carbon emissions from 1 kilo of food among all consumers. The second row shows the sample of consumers who, on average, had orders below the median ($\bar{E}_i^w < 1.806$), and the fifth row shows the sample who, on average, had orders above the median ($\bar{E}_i^w > 1.806$). There are 4,215 (4,076) consumers below (above) the median in the control store and 11,955 (12,094) consumers below (above) the median in the treated store.

Table 1.8: Descriptive statistics on consumers emission intensity per kg food during the pre-treatment period

	Obs	Median	Mean	Min	Max	SD
All	32,340	1.806	2.009	0.100	23.481	1.060
Below median ($\bar{E}_i^w < 1.806$)	16,170	1.324	1.284	0.100	1.806	0.339
Below median: Control store	4,215	1.310	1.267	0.100	1.806	0.342
Below median: Treated store	11,955	1.330	1.290	0.100	1.806	0.338
Above median ($\bar{E}_i^w > 1.806$)	16,170	2.444	2.734	1.806	23.481	1.041
Above median: Control store	4,076	2.478	2.831	1.806	23.481	1.218
Above median: Treated store	12,094	2.431	2.701	1.806	21.147	0.972

Notes: This table contains summary statistics on $\bar{E}^w$ during the pre-treatment period. $\bar{E}^w$ measures each consumer's average emissions intensity per kg of food during the pre-treatment period. The emissions intensity per kg of food is calculated by dividing an order's kg CO_2e by the weight of the order.

In the following two subsections, I divide the sample into those above and those below the median ($\bar{E}_i^w = 1.806$) to study heterogeneity in how these two groups with high and low carbon intensity in their food consumption responded to carbon labels.

Carbon emissions per order

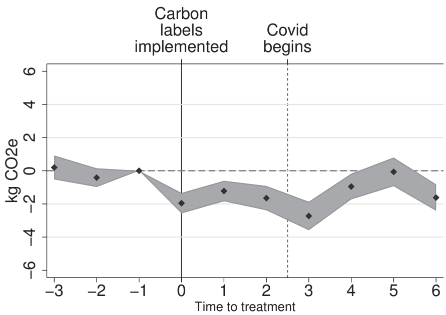
Figure A7 shows average emissions per order separately for those below ($\bar{E}_i^w < 1.806$) and those above ($\bar{E}_i^w < 1.806$) the median of E^w (1.806) during the pre-treatment period. Across both groups, there is a decline in the average carbon footprint per order for consumers in the treated store compared with consumers in the control store.

Figure 1.7 shows the event plot analysis, where I have estimated equation (1.8) separately for each subsample. Panel (a) plots the effect of carbon labels on kg emissions per order for consumers below the median and panel (b) plots the same for those above the median. The low-impact group has a similar pattern to that of the full sample, where the treatment effect stays around 2 kg CO_2e per order, with the most significant reduction in the first period. Consumers above the median in the treated store reduce their carbon footprint by almost 4 kg CO_2e during the first 45 days, but their response seems to vanish in the subsequent months. The same pattern is found when looking at the log of emissions and the log of the emissions intensity per kg in Figure A8 and Figure A9.

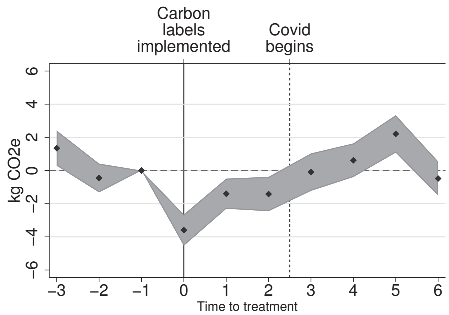
Tables 1.9 and 1.10 confirm these findings. The low-emissions groups reduce their carbon footprint by around 4 percent in the short term, and this effect remains one year afterward. For the group above the median, there is a significant reduction in the short term, up until Covid, by around 3.5 percent. However, it fades in the long term, and there is no statistically significant response one year afterward.

Figure 1.7: Heterogeneity based on consumers' pre-treatment emission intensity per kg food: Effects of carbon labels on emissions per order

(a) Consumers below the median ($\bar{E}_i^w < 1.806$)



(b) Consumers above the median ($\bar{E}_i^w > 1.806$)



Notes: This figure shows the long-term response (July 2019 to October 2020) of emissions per order as the dependent variable from estimating equation (1.8). The sample is separated based on consumers' average $\frac{\text{kg } CO_2e}{\text{kg food}}$ per order during the pre-treatment period. A time period refers to 45 days, and the solid vertical line indicates when carbon labels were implemented in the treated store. The excluded period (time period -1) is the reference period and refers to the 45 days before the treatment was implemented. Each diamond indicates the difference-in-difference point estimate, γ_k, compared with the reference period, with a 95 percent confidence interval around it indicated by the grey area. A negative value should be interpreted as average emissions per order being lower in the treated store than in the control store relative to the difference between stores in the reference period (-1).

Table 1.9: Effects of carbon labels on emissions per order for consumers below the median ($\bar{E}_i^w < 1.806$)

	(1) Emissions	(2) log(Emissions)	(3) $\log(\frac{\text{Emissions}}{\text{Cost}})$	(4) $\log(\frac{\text{Emissions}}{\text{Weight}})$
Panel A. Overall				
Carbon Labels	-1.509***	-3.798***	-3.697***	-4.124***
	(0.243)	(0.792)	(0.548)	(0.616)
Panel B. Post period split into before and after Covid				
Carbon Labels (Before Covid)	-1.465***	-4.707***	-3.019***	-4.051***
	(0.226)	(0.783)	(0.549)	(0.607)
Carbon Labels (After Covid)	-1.546***	-3.045**	-4.259***	-4.185***
	(0.323)	(1.006)	(0.672)	(0.769)
Individual FE	Yes	Yes	Yes	Yes
Observations	219,477	219,477	219,477	219,477
Clusters	16,162	16,162	16,162	16,162

Standard errors clustered on member account level in parentheses.

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Notes: This table shows the overall response (July 2019 to October 2020), before Covid response (July 2019 to March 2020), and after Covid results (March to October 2020) from estimating equation (1.7). Emissions per order are measured as kg CO_2e per order. Cost and weight refer to the order cost and kg of food in the order, respectively. The sample is restricted to those consumers who are observed before and after carbon labels were introduced in the treated store. The DiD estimates should be interpreted as the change in the dependent variable caused by introducing carbon labels in the treated store. All regressions include individual and time fixed effects, and standard errors are clustered at the consumer level. All estimates where the dependent variable has been log-transformed have been multiplied by 100.

Table 1.10: Effects of carbon labels on emissions per order for consumers above the median ($\bar{E}_i^w > 1.806$)

	(1) Emissions	(2) log(Emissions)	(3) $\log(\frac{\text{Emissions}}{\text{Cost}})$	(4) $\log(\frac{\text{Emissions}}{\text{Weight}})$
Panel A. Overall				
Carbon Labels	-1.337***	-1.659*	-1.799***	-0.824
	(0.343)	(0.736)	(0.493)	(0.551)
Panel B. Post period split into before and after Covid				
Carbon Labels (Before Covid)	-2.823***	-5.120***	-3.485***	-3.628***
	(0.375)	(0.805)	(0.553)	(0.630)
Carbon Labels (After Covid)	-0.074	1.283	-0.364	1.560*
	(0.455)	(0.889)	(0.576)	(0.651)
Individual FE	Yes	Yes	Yes	Yes
Observations	223,599	223,599	223,599	223,599
Cluster	16,163	16,163	16,163	16,163

Standard errors clustered on member account level in parentheses.

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Notes: This table shows the overall response (July 2019 to October 2020), before Covid response (July 2019 to March 2020), and after Covid results (March to October 2020) from estimating equation (1.7). Emissions per order are measured as kg CO_2e per order. Cost and weight refer to the order cost and kg of food in the order, respectively. The sample is restricted to those consumers who are observed before and after carbon labels were introduced in the treated store. The DiD estimates should be interpreted as the change in the dependent variable caused by introducing carbon labels in the treated store. All regressions include individual and time fixed effects, and standard errors are clustered at the consumer level. All estimates where the dependent variable has been log-transformed have been multiplied by 100.

Carbon labels effect on spending

Figure A10 shows how carbon labels affected spending and supermarket choice, considering the shopping frequency and spending per order. In Figure A10, I have plotted total spending, average spending per order (intensive margin), and shopping frequency (extensive margin) for all consumers who were members prior to the treatment.

I find no overall reduction in spending at the treated store for those below the median in either the intensive or extensive margin. Consumers above the median reduce both their shopping frequency and their spending on the intensive margin. I find that consumers above the median reduce their shopping frequency by around 3 percent and their intensive margin by 2 percent, and overall their spending is reduced by 5 percent, highlighting that this group is driving the aggregate response found earlier in Section 1.6.3. To summarize, high emitters became less likely to shop at the treated store after carbon labels were introduced. There is, however, no statistically significant change for low emitters.

1.6.5 Robustness

This section provides robustness checks to the main results.

Members of both stores simultaneously

If consumers are members of both stores simultaneously, this would cause measurement error in the treatment variable. For example, carbon labels may cause consumers to reduce their carbon footprint in both stores, and the response from those consumers in the treated store would thus be underestimated.

To assess how common this is and whether it affects the results, I first conducted a follow-up survey where I asked consumers how large of a share they consumed at the treated store, at the control store, or at other stores. On average, consumers stated that they purchased around 65 percent of their food at the treated store and around 1.5 percent at the control store (see Table A2). I then perform a formal test, where I rerun the main analysis but only include (odd) even zip codes from the treated store and (even) odd zip codes from the control store. This exercise has some caveats: first, it will naturally reduce the sample size, and second, it could introduce differences in socioeconomic background. However, it allows for the creation of a sample of consumers who should not simultaneously be present in both the treated and control groups. The event plot from this exercise is presented in Figure A11.[20] The patterns in both graphs are almost identical to the main results in Figure 1.3.

To summarize, the results from this exercise provide reassuring evidence that it is uncommon for consumers to be members of both stores and that

[20]Standard errors are slightly larger in panel (b), explained by odd zip codes having a larger population than even zip codes.

this does not seem to impact my estimates.

Length of time periods

We have previously seen in Section 1.5.3 that a sale of beef can impact the average carbon footprint per order. I have previously aggregated purchases into 45-day intervals to reduce the noise from sale offers and have a more balanced sample over time.

However, the length of periods is part of the researcher's degrees of freedom, with no definite correct choice. To assess how large an impact this choice has, I also present the results using a 14-day window. In Figure A12, I have plotted emissions for consumers below and above the median ($\bar{E}_i^w = 1.806$) during the pre-treatment period. This figure follows the same pattern as the findings in earlier sections, with consumers below the median reducing their kg carbon emissions per order in the short term, and the response is also stable for most periods over the longer term. Consumers above the median are more volatile around the treatment date, and the response fades with time.

There is notably high volatility around the treatment date for the high-impact group, mainly explained by sales promotions on beef not implemented in both stores simultaneously during each 14-day interval. The pattern is more stable for consumers below the median ($\bar{E}_i^w < 1.806$) than for those above the median ($\bar{E}_i^w > 1.806$), which is mainly explained by less elastic responses to beef promotions.

Prices

Price is an endogenous variable set by the store, so it is not apparent that one should want control for it in the regression analysis. However, even though I did not find any systematic changes in the pricing around the treatment date in Section 1.5.3, there may still be a concern that there is low power to detect such systematic changes.

To assess how prices affects the main results, I rerun the main analysis by controlling for prices of high-, middle-, and low-impact products in Table 1.11. Qualitatively, the results are robust, but in general, the point estimates are larger when also controlling for prices. In particular the point estimates for the period after Covid, where the effect size now indicates a 3–4 percent reduction in consumers' kg carbon emissions per order for all measures. To summarize, the results are robust to also controlling for important product prices, and the results would then indicate that the reduction in consumers' carbon footprint remained more stable over time, with larger reductions than when not including prices in the models.

Table 1.11: Effects of carbon labels on emissions per order, controlling for prices

	(1) Emissions	(2) log(Emissions)	(3) $\log(\frac{\text{Emissions}}{\text{Cost}})$	(4) $\log(\frac{\text{Emissions}}{\text{Weight}})$
Panel A. Overall				
Carbon Labels	-1.936***	-4.222***	-4.151***	-3.640***
	(0.211)	(0.590)	(0.404)	(0.460)
Panel B. Post period split into before and after Covid				
Carbon Labels (Before Covid)	-2.510***	-5.626***	-3.804***	-4.092***
	(0.255)	(0.631)	(0.437)	(0.492)
Carbon Labels (After Covid)	-1.946***	-3.166***	-4.412***	-3.300***
	(0.302)	(0.709)	(0.471)	(0.546)
Individual FE	Yes	Yes	Yes	Yes
Observations	443,131	443,131	443,131	443,131
Clusters	32,339	32,339	32,339	32,339

Standard errors clustered on member account level in parentheses.

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Notes: This table shows the overall response (July 2019 to October 2020), before Covid response (July 2019 to March 2020), and after Covid results (March to October 2020) from estimating equation (1.7). Emissions per order are measured as kg CO_2e per order. Cost and weight refer to the order cost and kg of food in the order, respectively. The sample is restricted to consumers observed before and after carbon labels were introduced in the treated store. The DiD estimates should be interpreted as the change in the dependent variable caused by introducing carbon labels in the treated store. All regressions include individual and time fixed effects, and the average price of high-impact (beef), middle-impact (pork or pork mixed with beef), and low-impact (vegetarian or poultry) products per 14-day interval. Standard errors are clustered at the consumer level. All estimates where the dependent variable has been log-transformed have been multiplied by 100.

Climate-smart suggestions

In this section, I study the role of Climate-smart suggestions in consumers' responses. To this aim, I exploit the fact that around 25 percent of consumers had self-selected to try out the nudge from June to November 2019. In Figure A13, I add an interaction term for opting in to the pilot project in June, when about 25 percent of consumers opted to participate in the pilot project. The estimated difference between these groups is insignificant for all periods, providing suggestive evidence that the search nudge did not impact consumers' shopping significantly. To summarize, I find no significant impact of the climate-smart suggestions on consumers' carbon footprint.

There may still be a concern that consumers who took part in the pilot project were already informed about some of the carbon footprint estimates and that the treatment is therefore a staggered treatment design that could cause problems when using two-way fixed effects (de Chaisemartin & D'Haultfœuille 2020). Therefore, I rerun the main analysis without those who took part in the pilot project and present the results of this analysis in Table 1.12. The point estimates are very similar to the main results, except that the treatment period after Covid is now statistically significant which was not the case for the whole sample in Table 1.7.

To summarize, this section provides suggestive evidence that the climate-smart suggestions, at least during the pilot study, were ineffective in changing consumer behavior. Second, the main results are robust to using only the sample that did not participate in the pilot project.

Table 1.12: Effects of carbon labels on emissions per order on consumers who did not take part in the pilot study

	(1) Emissions	(2) log(Emissions)	(3) $\log(\frac{\text{Emissions}}{\text{Cost}})$	(4) $\log(\frac{\text{Emissions}}{\text{Weight}})$
Panel A. Overall				
Carbon Labels	-1.543***	-3.107***	-2.831***	-2.586***
	(0.218)	(0.571)	(0.391)	(0.449)
Panel B. Post period split into before and after Covid				
Carbon Labels (Before Covid)	-2.225***	-5.086***	-3.407***	-3.991***
	(0.228)	(0.591)	(0.409)	(0.467)
Carbon Labels (After Covid)	-0.955**	-1.401*	-2.334***	-1.375*
	(0.291)	(0.714)	(0.472)	(0.553)
Individual FE	Yes	Yes	Yes	Yes
Observations	349,450	349,450	349,450	349,450
Clusters	25,638	25,638	25,638	25,638

Standard errors clustered on member account level in parentheses.
* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Notes: This table shows the overall response (July 2019 to October 2020), before Covid response (July 2019 to March 2020), and after Covid results (March to October 2020) from estimating equation (1.7). Emissions per order are measured as kg CO_2e per order. Cost and weight refer to the order cost and kg of food in the order, respectively. The sample is restricted to those consumers who were observed before and after carbon labels were introduced in the treated store. The DiD estimates should be interpreted as the change in the dependent variable caused by introducing carbon labels in the treated store. All regressions include individual and time fixed effects, and standard errors are clustered at the consumer level. All estimates where the dependent variable has been log-transformed have been multiplied by 100.

1.6.6 External validity

During spring 2022, I distributed a follow-up survey by email to about 8,000 customers of the treated supermarket. Among the group that opened the email, 259 respondents chose to participate (they were incentivized by around 7 euros). Table A2 shows the responses from this survey.

One objective of the survey was to measure the sample's representativeness. For this purpose, the following question was asked: How seriously do you view climate change as a problem today, on a scale from 1 to 10, where 10 is extremely important and 1 is not important? This question was chosen because it was previously used in a special edition on climate change in the Eurobarometer (European Commission 2021) and sent out to a representative sample in several European countries, including Sweden. I find that my sample respond almost exactly to the representative sample of Swedish citizens from the Eurobarometer.[21] Survey evidence with low response frequency has to be interpreted cautiously, but it is not apparent that this should be correlated with respondents' preferences toward the climate. Survey evidence with low response frequency has to be interpreted cautiously, but it is not apparent that this should be correlated with their preferences towards the climate.

If we look at data on actual consumption patterns, Swedes have higher meat consumption and eat more beef and less poultry compared with the average European. Swedish household emissions are also higher than those in other European countries, at around 9 tonnes CO_2e per capita compared with around 6 tonnes CO_2e.

However, another pattern emerges when we look at survey data on attitudes toward climate change. Table A3 presents survey responses from the Eurobarometer on Europeans' attitudes toward climate change (European Commission 2021). The left column contains the Swedish responses, the second column includes Germany as a reference country, and the third column is the average response across all European countries.

The Swedish population has similar views to those of people in most other European countries regarding how seriously they view climate change as a problem today. However, Swedes perceive themselves as having more personal responsibility than Europeans in general, and are more likely to have individual action to fight climate change in the last six months. This pattern also emerges in questions about whether respondents have used an alternative to their car or bought and ate less meat in the last six months. Another interesting and highly relevant question for this paper is whether respondents consider their carbon footprint and sometimes adapt their behavior concerning food (or transport). Here Swedes are more inclined to answer that they have done so, at 34 percent compared with 22 percent for Germany and 16 percent for the whole sample.

[21] On average, respondents in the treated store rated the seriousness of climate change as a problem today at 7.83, and 7.86 by the representative sample of Swedish citizens in the Eurobarometer.

1.7 Conclusion

This paper exploited a unique large-scale implementation of Carbon labels in a supermarket from one day to the next to study if carbon labels cause consumers to reduce their carbon emissions. To the best of my knowledge, this paper provides the largest evaluation of carbon labels in any setting, particularly within the food domain.

Carbon labels were implemented in the treated store on 12 November 2019 and I find that carbon labels caused consumers to reduce their carbon emissions per order by around 3–5 percent in the first four months before the Covid pandemic. The effect of carbon labels does fade over time to about 1–2 percent measured over the period from March 2020 until October 2020. The overall effect measured over the entire year after labels were introduced is in between the short- and long-term response. Carbon labels caused consumers in the treated store to reduce carbon emissions by around 2.5 percent per order, measured over the entire year after labels were introduced. This estimate remains stable for all measures of emissions I use, including dividing emissions by the weight or cost of the order.

I then explored heterogeneous responses by calculating consumers' emissions intensity per kg of food they purchased before the carbon labels were introduced and divided the sample into those above or below the median. Among consumers above the median, carbon labels caused them to reduce their carbon footprint by 4 percent in the first four months. However, there was no statistically significant effect in the last five months. I also find evidence that those consumers reduce their shopping frequency at the treated store. Suggesting that some consumers may prefer to hide their visible carbon footprint in the treated store in favor of other stores where their carbon footprint may not be as salient. The pattern is different for consumers below the median. Those consumers showed a stable response over the entire post-period by reducing their carbon footprint per order by 4% over the entire post-evaluation year, with no statistically significant change in their shopping frequency.

The overall emission reduction among all consumers amounts to about 2.5 percent, corresponding on average to around 27 (35) kg CO_2e per year and person in EU27 (Sweden).[22] In conversations with the developers of the carbon footprint estimates at RISE (2020), the cost to develop and maintain the database has been around 300 000 euros in fixed cost and 100 000 euros in annual maintenance. Carbon labels would cover its fixed cost in four years for a consumer base of around 100,000 consumers and valuation of 50$ per kg CO_2e and disregarding discounting. My results thereby indicate that carbon labels implemented on a sufficiently large scale could be a cost-effective complement to economic policy instruments. The rise in online shopping could provide a particularly suitable opportunity for carbon labels as each product package does not need to be labeled, and the online interface allows consumers to get a clear overview of potential substitutes.

It can be interesting to relate our estimated effect size of carbon labels to studies that have estimated how a carbon tax on food products would affect GHG emissions. Säll & Gren (2015) found that GHG emissions would be reduced by about 12 percent from implementing a tax of 120$ per tonne CO_2e on meat and dairy products in the Swedish market. Our estimated effect size of carbon labels would then be about one-fifth of a carbon tax of 120$ per tonne CO_2e in the Swedish market. If the response is linear in the tax, it corresponds to a tax of 24$ per tonne CO_2e. A similar result is found in Wirsenius et al. (2011) that looked at implementing a carbon tax on meat and dairy products in the EU27. Wirsenius et al. (2011) found that a tax of about 20$ per tonne CO_2e would reduce GHG emissions by around 2.5 percent.

This paper speaks directly to the recent discussion on whether behavioral scientists should shift their focus from individual actions to studying how individuals can support large-scale systematic changes (Chater & Loewenstein 2022). By evaluating carbon labels outside the laboratory on a large scale, this paper provides evidence of what the upside of a scalable and relevant i-frame policy can bring. Having such estimates should be of interest for both those favoring i-frame policy's and those that oppose them. Scaling our estimated effect size to the 447 million individuals in the EU27

[22] The average carbon footprint from food consumption in the EU27 were 1070 kg CO_2e per year and person in 2018 according to Sandström et al. (2018). The annual carbon footprint per capita in Sweden was about 1,400 kg CO_2e according to Naturvardsverket 2020.

would reduce emissions by 12 million tonnes CO2e and year, or about 0.35% of the EU27 annual GHG emissions [23]. However, one might expect the effectiveness of carbon labels to be lower in physical stores where the salience and overview of carbon labels could be reduced. Our estimate is further based on Swedish urban households that already cares more about their carbon footprint than individuals in most other European countries. On the other hand, the social dimension would likely be increased in physical stores compared to the online setting.

As Taufique et al. (2022) notes in their review, carbon labels are not a panacea, but this paper provides evidence that they could serve as a cost-effective complement to other economic policies. If carbon labels are more or less effective in physical stores is an interesting question that is left for future research. Another question for future work is how carbon labels interact with a carbon tax. One possibility is, of course, that a carbon tax would crowd out the voluntary response if consumers feel they are already paying for their emissions. Another highly relevant question is, of course, the general equilibrium response when labels are implemented not only on food products within a single store but in other grocery stores and outside of the food domain.

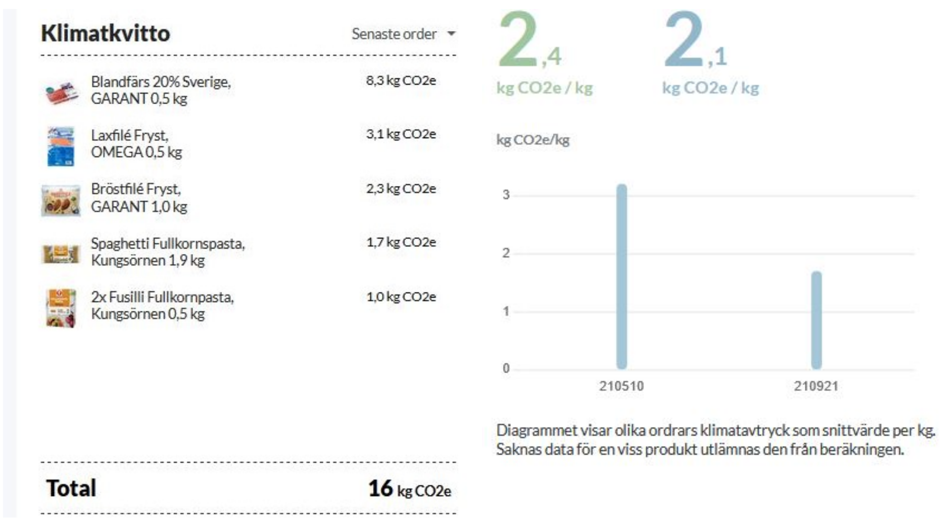
[23] GHG emissions in the EU27 were 3,457 million tonnes CO_2e per year during 2018 (Crippa et al. 2022).

1.8 Appendix A

1.8.1 Figures

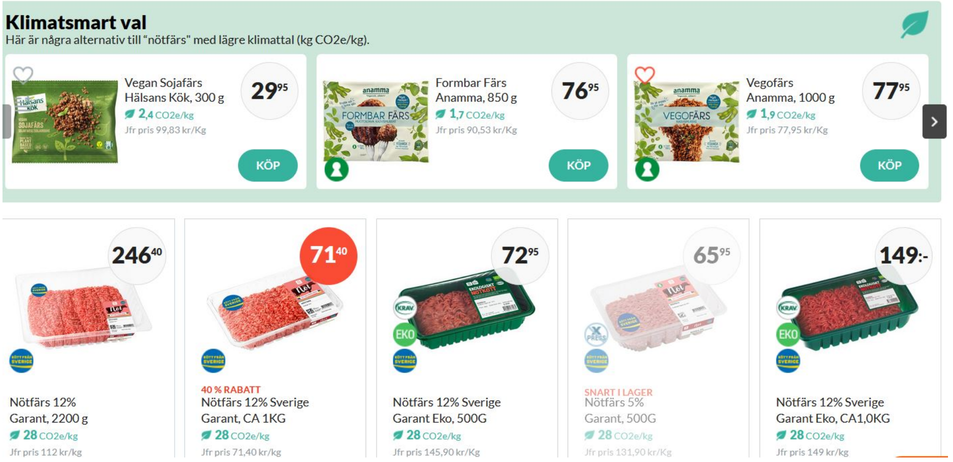
Figure A1: Aggregate carbon footprint information at checkout

(a) Climate receipt (b) Social comparison module



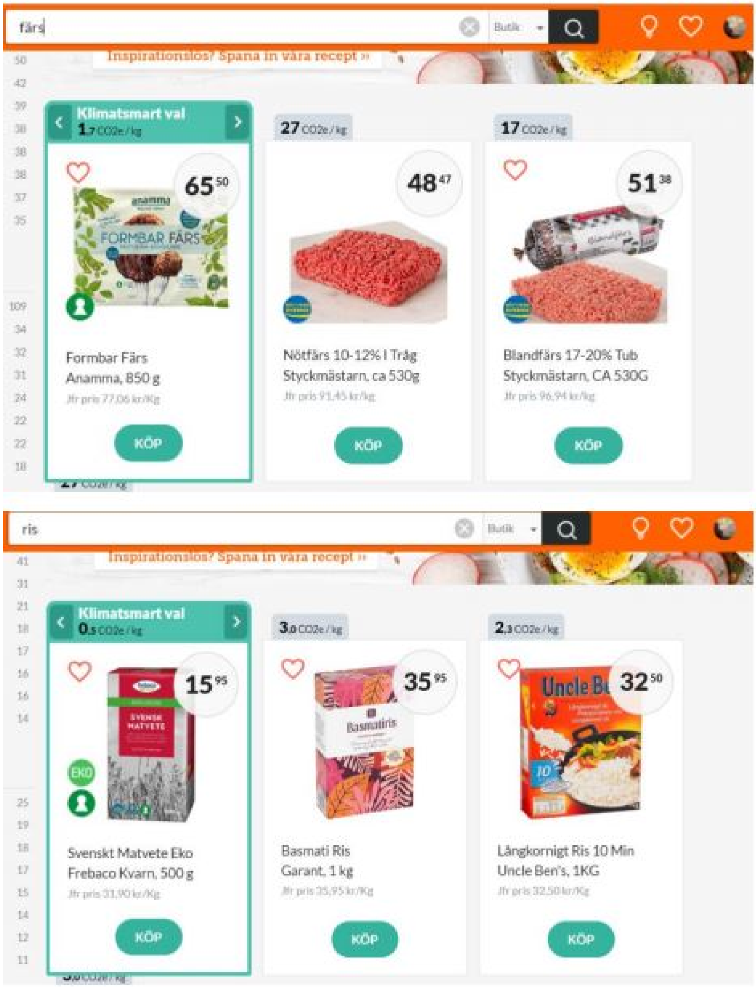
Notes: Consumers can access their order history on their member page. Panel (a) shows the total carbon footprint for that order. Panel (b) shows the consumers' total carbon footprint divided by the weight of the order. Each line indicates one specific order, and the green-colored number shows the customer's average across all orders; the blue number shows the average across all consumers at the store.

Figure A2: Climate smart suggestions during the Natural experiment



Notes: This figure shows the climate-smart suggestions activated by default for all consumers from 12 November 2019 onward. The picture displays the results of a search on "ground meat." Consumers could opt out if they preferred not to get climate-smart suggestions. Around 99 percent of consumers decided to keep the default of having climate-smart suggestions turned on, while 1 percent actively turned off this feature.

Figure A3: Climate-smart suggestions during the pilot study



Notes: This figure shows the climate-smart suggestions in the pilot study from June to November 2019. The top picture displays the results of a search on “ground meat,” and the bottom picture a search on “rice.”

Figure A4: The volume share covered by CFLs

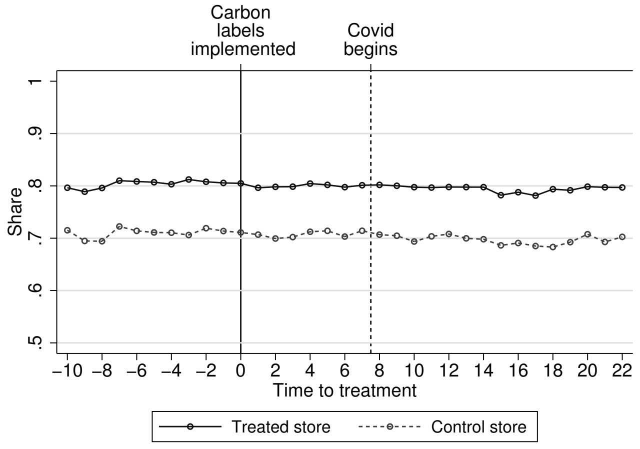
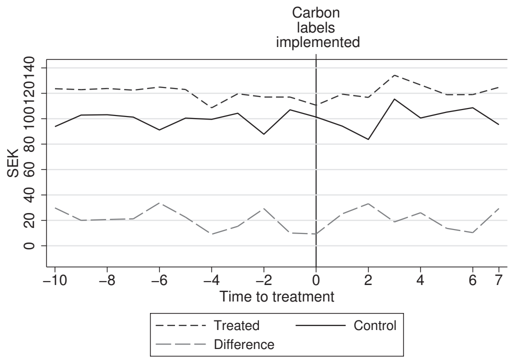
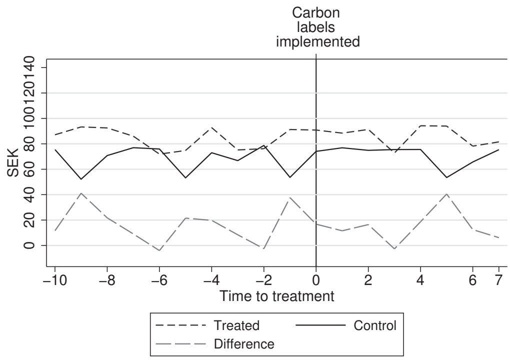


Figure A5: Average prices over time

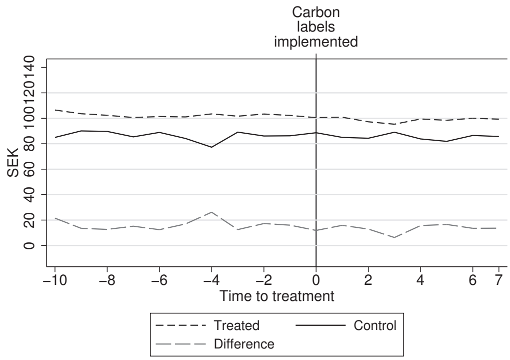
(a) High impact



(b) Middle impact

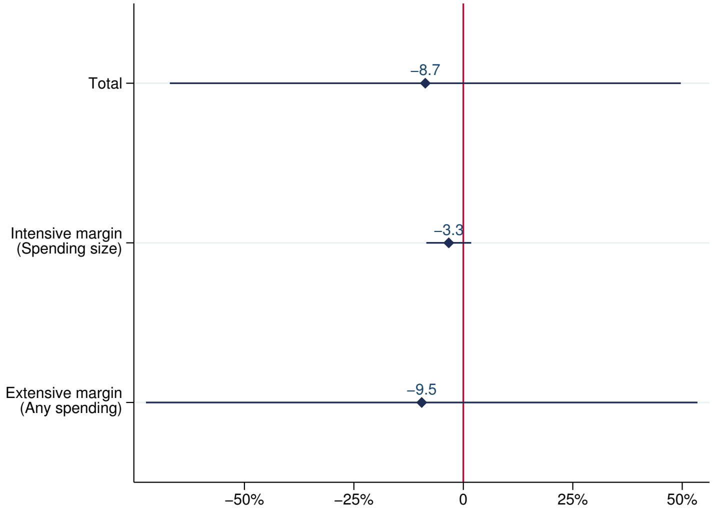


(c) Low impact



Notes: This graph contains the average high-impact (beef), middle-impact (pork and pork mixed with beef), and low-impact (vegetarian and poultry) average prices at 14-day intervals from June 2019 to March 2020.

Figure A6: Aggregate store-level spending

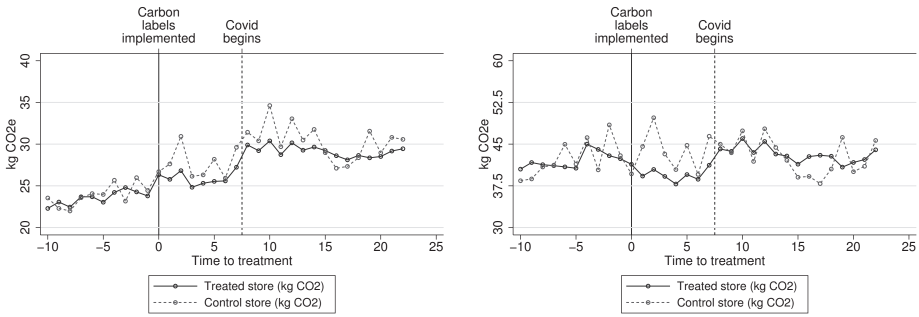


Notes: This figure shows aggregate spending per 45 days in each store from July 2019 to March 2020. The DiD estimates should be interpreted as the percentage change in the dependent variable relative to the average spending during the pre-treatment period.

Figure A7: Heterogeneity based on consumers pre-treatment carbon emission intensity per kg food: Average emissions per order

(a) Consumers below the median ($\bar{E}_i^w$) < 1.806)

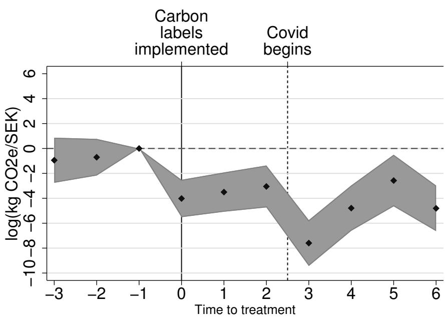
(b) Consumers above the median ($\bar{E}_i^w$) > 1.806)



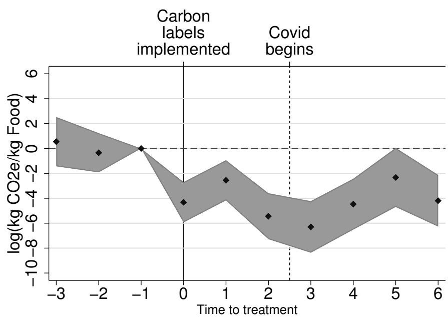
Notes: This figure shows average emissions per order in each store from July 2019 until October 2020. Each diamond represents 14 days, and the straight line indicates that carbon labels were from period 0 onwards (12th November 2019).

Figure A8: Effects of carbon labels on emissions per order for consumers below the median

(a) $\log(\frac{\text{Emissions}}{\text{Cost}})$

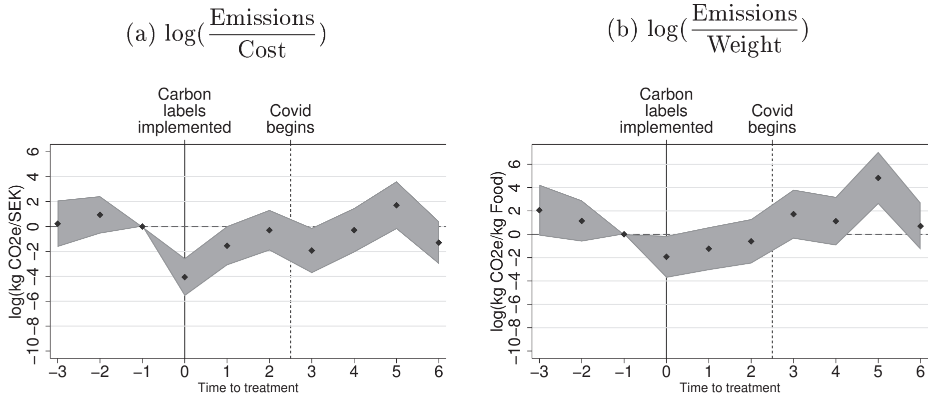


(b) $\log(\frac{\text{Emissions}}{\text{Weight}})$



Notes: This figure shows the long-run response (July 2019 to October 2020) for consumers who were on average below the median (1.806 $\frac{\text{kg } CO_2e}{\text{kg food}}$) during the pre-treatment period. A time period refers to 45 days, and the solid vertical line indicates when carbon labels were implemented in the treated store. The excluded period (time period -1) is the reference period and refers to the 45 days before the treatment was implemented. Each diamond indicates the difference-in-difference point estimate, γ_k, compared with the reference period, with a 95 percent confidence interval around it indicated by the grey area. A negative value should be interpreted as average emissions per order being lower in the treated store than in the control store relative to the difference between stores in the reference period (-1). All estimates where the dependent variable has been log-transformed have been multiplied by 100.

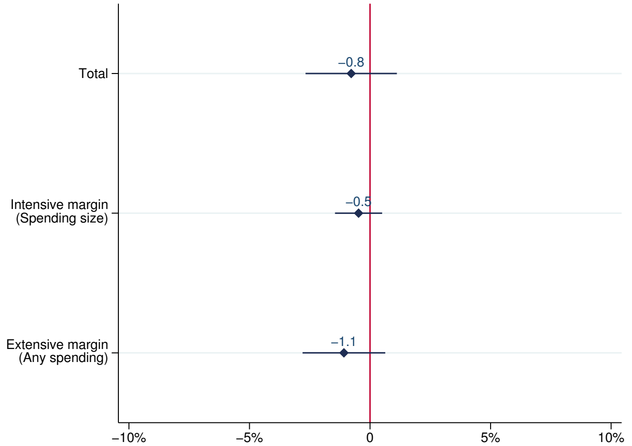
Figure A9: Effects of carbon labels on emissions per order for consumers above the median



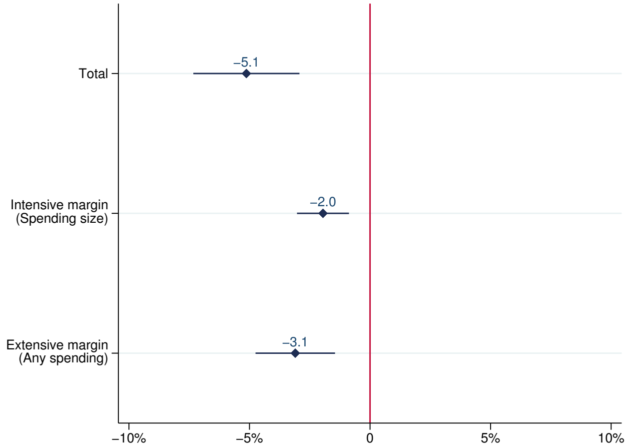
Notes: This figure shows the long-run response (July 2019 to October 2020) for consumers who were on average above the median (1.806 $\frac{\text{kg } CO_2e}{\text{kg food}}$) during the pre-treatment period. A time period refers to 45 days, and the solid vertical line indicates when carbon labels were implemented in the treated store. Emissions per order are measured as kg CO_2e per order. Cost and weight refer to the order cost and kg of food in the order, respectively. The excluded period (time period -1) is the reference period and refers to the 45 days before the treatment was implemented. Each diamond indicates the difference-in-difference point estimate, γ_k, compared with the reference period, with a 95 percent confidence interval around it indicated by the grey area. A negative value should be interpreted as average emissions per order being lower in the treated store than in the control store relative to the difference between stores in the reference period (-1). All estimates where the dependent variable has been log-transformed have been multiplied by 100.

Figure A10: Heterogeneity in short-term effects of carbon labels on spending

(a) Consumers below the median ($\bar{E}_i^w < 1.806$):



(b) Consumers above the median ($\bar{E}_i^w > 1.806$):

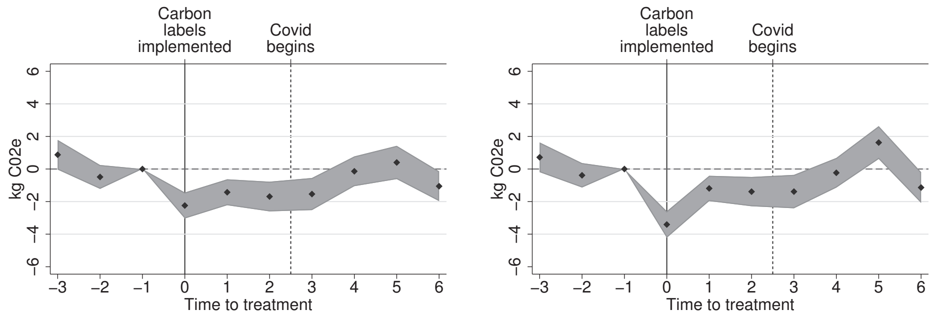


Notes: This figure shows average spending per consumer during 45-day intervals in each store from July 2019 to March 2020. The DiD estimates should be interpreted as the percentage change in the dependent variable relative to average spending during the pre-treatment period. The intensive margin refers to spending per order, and the extensive margin is shopping frequency. All regressions include individual and time fixed effects, and standard errors are clustered at the consumer level. Panel (a) and panel (b) shows consumers below and above the median emission intensity per kg food and order during the pre-treatment period.

Figure A11: Event plot of emissions per order when excluding odd (even) zip codes in the treated (control) store

(a) Treatment (control) store even (odd) zip codes.

(b) Treatment (control) store odd (even) zip codes.

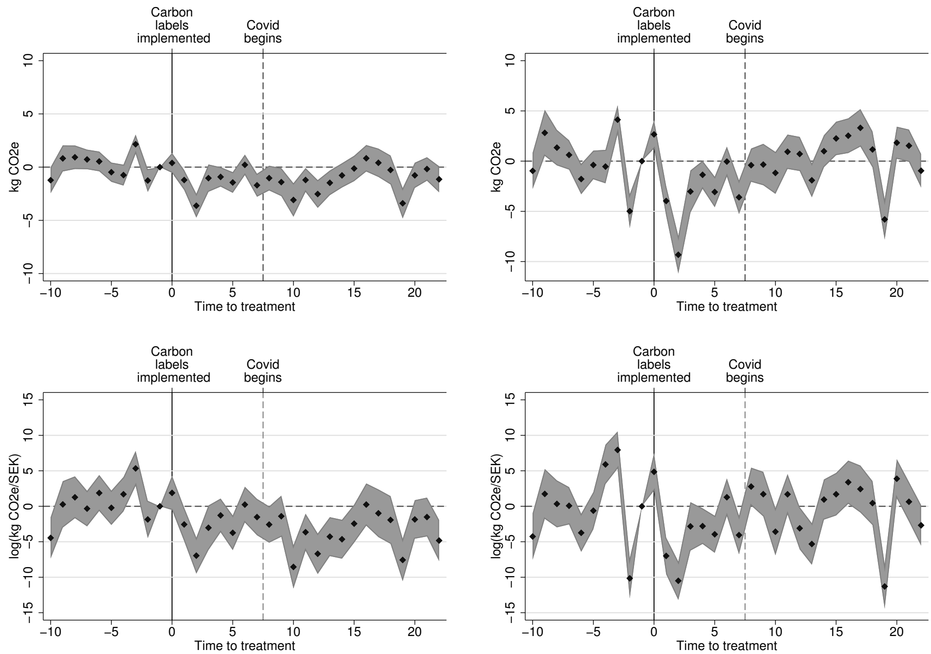


Notes: This figure shows emissions per order as the dependent variable from estimating equation (1.8). The excluded time period is the reference period (-1) and the solid vertical line refers to when carbon labels were implemented. Each diamond refers to a 45-day interval. Emissions per order are measured as kg CO_2e per order. Cost and weight refer to the order cost and kg of food in the order, respectively. A negative value should be interpreted as emissions per order being lower in the treated store compared to the control store relative to the reference period. In panel (a), only even zip codes are used in the treated store and only odd ones in the control store. In panel (b), only odd zip codes are used in the treated store and only even ones in the control store. .

Figure A12: Heterogeneity based on consumers' pre-treatment carbon emissions intensity per kg of food: Effects of carbon labels on carbon emissions per order

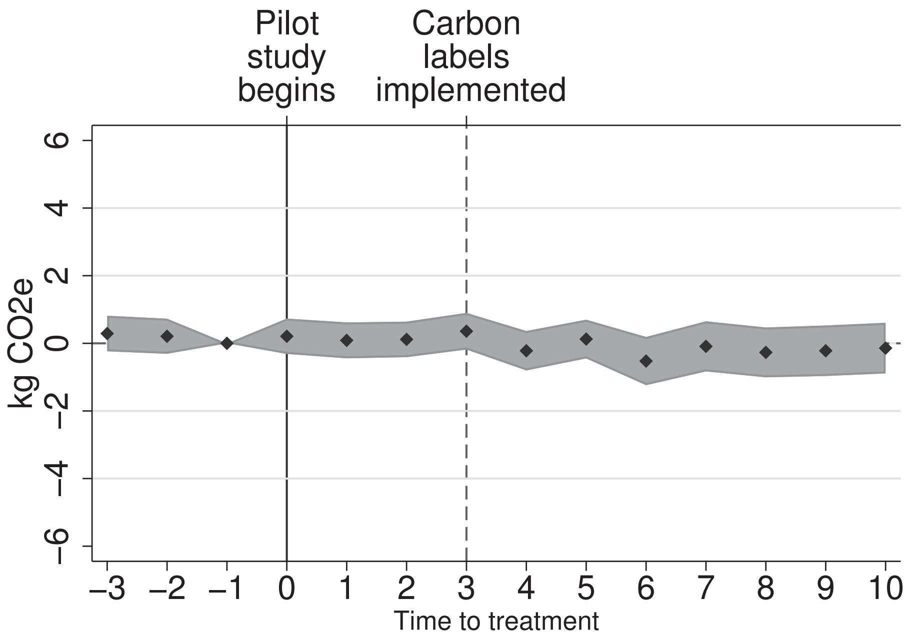
(a) Consumers below the median ($\bar{E}_i^w < 1.806$)

(b) Consumers above the median ($\bar{E}_i^w > 1.806$)



Notes: This figure shows the long-term response (July 2019 to October 2020) of carbon labels on consumers' kg carbon emissions per order. A time period refers to 14 days, and the solid vertical line indicates the date the carbon label was implemented (12 November 2019). The excluded time period is the reference category and refers to the 14 days before the treatment was implemented.

Figure A13: Heterogeneous response among those who took part in the pilot study



Notes: This figure shows the DiD estimates between those who opted to try out climate-friendly suggestions and those who did not take part. The dependent variable is kg CO_2e per order. A negative value should be interpreted as consumers who did not take part in the pilot project having lower kg CO_2e per order than consumers who did. A time period refers to 45 days, and the solid vertical line indicates when the pilot project started, on 1 June 2019. The dashed line refers to when carbon labels were implemented, which took place in the middle of that 45-day interval. The excluded time period is the reference category and refers to periods before the pilot project started.

1.8.2 Tables

Table A1: Price differences after carbon labels were introduced

	(1)	(2)	(3)
	High	Middle	Low
Post*Treated store	-0.447	-1.458	-3.134
	(4.058)	(6.628)	(1.880)
Constant	21.16***	16.45**	16.43***
	(2.705)	(4.419)	(1.254)
N	18	18	18

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Notes: This table contains the regression of the price difference (in SEK/kg) on the post-treatment dummy interacted with the treated store from July 2019 to March 2020. Column (1) contains beef prices, column (2) pork and pork mixed with beef prices, and column (3) vegetarian products and poultry prices. One observation is a 14-day interval, and the dependent variable is the price difference in each product category between the stores. A negative coefficient should be interpreted as the price in the treated store being lower compared with the control store during the post-treatment period relative to the price difference during the period before carbon labels were introduced.

Table A2: Descriptive statistics from the follow-up survey during spring 2022

	count	mean	sd	min	max
How serious of a problem is climate change? (0-100)	259	78.309	21.459	0	100
Share of grocery purchase (0-100)					
Treated store	259	66.734	25.727	7	100
Control storel	259	1.432	6.848	0	45
Other online store	259	7.718	15.709	0	85
Physical store	259	24.116	22.800	0	89
CFLs (1/0)					
Know they exist	259	0.761	0.428	0	1
I sometimes use CFLs when I shop in (1/0)					
Treated store (1/0)	259	0.552	0.498	0	1
Tot_use_online (1/0)	259	0.359	0.481	0	1
Tot_use_phy (1/0)	259	0.355	0.480	0	1
Easy to use	197	0.604	0.490	0	1
Sometimes prefer not to see carbon footprint	197	0.157	0.365	0	1
Important reason that I shop at the treated store	197	0.269	0.445	0	1
What is your age?					
Age_1 (<30)	29	1.000	0.000	1	1
Age_2 (31-40)	49	1.000	0.000	1	1
Age_3 (41-50)	59	1.000	0.000	1	1
Age_4 (51-65)	60	1.000	0.000	1	1
Age_5 (>65)	61	1.000	0.000	1	1
Where do you live?					
Gothenburg	148	1.000	0.000	1	1
Stockholm	98	1.000	0.000	1	1
Other	10	1.000	0.000	1	1
N	259				
Duration in sec.	259	1,936.305	25,077.089	41	403,043

Table A3: Attitudes toward climate change and personal action from the European Commission (2021)

	Sweden	Germany	EU
How serious a problem do you think (1-10)			
Climate change is at this moment	7,86	7,96	7,93
Personal			
Responsibility	56%	56%	41%
Action	74%	79%	64%
You regularly			
Use alternatives to your private car	56%	56%	41%
Buy and eat less meat	46%	51%	31%
Consider carbon footprint			
and sometimes adapt accordingly			
Food	34%	22%	16%
Transport (long distance)	27%	21%	11%

1.9 Appendix B

1.9.1 Measure

To measure carbon footprint from consumers in the control store, I employ the following strategy. (i) I conduct one-to-one matches on unique product bar codes with products that have CFLs in the treated store. (ii) For those products that are not identical in the stores, I use carbon footprint estimates from RISE, which is the company responsible for all carbon footprints used in the treated store, and multiply this number by the meat percentage of meat products in the control store.

For each product I have access to a unique EAN bar code that allows me to map exact products between stores. For identical articles I can directly implement CFLs onto articles in the control store. To label articles that are not shared between stores I use the open source access "Open List" from RISE (2020). Labeling is done by taking into account these characteristics of an article:

- Meat percentage
- Country of origin
- Meat
- Carbon footprint according to RISE (2020)

The carbon footprint label L_j on article j, contains the carbon emission from consuming one kilogram of article j. By multiplying the carbon footprint label on product j with the weight of product j, gives us emission from product j. Hence, emissions from product j, with carbon label L_j and weight w_j , is given by, $E_j = L_j * w_j$. Total carbon emissions for consumer i, from j article and in time period t, is then given by,

$$E_{it} = \sum_j E_{ijt}. \tag{1.11}$$

Products that are shared between stores account for about 60 percent of the sales volume (measured in kg). Labels are added to milk and eggs, beef, pork, salmon, chicken, turkey, and lamb. S_{kj} = share of meat j in article k, P_k = meat percentage of article k (80–100 percent), and L_{j*} = carbon footprint of meat j with 100 percent meat. I then calculate the carbon label for article k by

$$L_k = \frac{P_k}{100} \sum_j L_{kj*} * S_{kj} \tag{1.12}$$

Example: Ground beef and pork (100 percent meat) that contains 50 percent beef, with a carbon footprint of 28 kg CO_2e, and 50 percent pork, with a carbon footprint of 4.1 CO_2e:

$$CFL_{ground\ beef} = 28 * 0.5 + 4.1 * 0.5 = 16.05 \tag{1.13}$$

Chapter 2

Combining monetary incentives and nudges to promote green consumption: Evidence from a large-scale natural experiment

Abstract: In this paper, I study a large-scale natural experiment where a supermarket, without prior announcement, increased the bonus points on fruit and vegetables and labeled them as "good deeds." The additional bonus points function as an equivalent price reduction of 0.6–2%, and the labeling provides a normative nudge on fruit and vegetables as environmentally friendly products. I use panel data that cover more than 40,000 consumers who place over 800,000 orders several months prior to the implementation of the bonus, throughout the entire year when the bonus is in place, and for several months after the bonus has been removed. The results indicate a larger consumer response than expected solely on the basis of the monetary incentive. The bonus program increased overall fruit and vegetable consumption by, on average, 6–9% per order. I further find no evidence that increasing the monetary incentive from an equivalent price reduction of 0.6% to 2% had any additional impact on consumption.

2.1 Introduction

Increasing the consumption of fruits and vegetables is an important global challenge due to its impact on both health and the environment (Albani et al. 2017, IPCC 2019, Griffith 2022). According to the Lancet Commission on Healthy Diets from Sustainable Food, unhealthy diets pose a more substantial risk of morbidity and mortality than the combined impact of unsafe sex, alcohol, drug, and tobacco use (Willett et al. 2019). Accordingly, the Commission has recommended increasing the intake of fruits and vegetables as one of the key changes required to address this issue. Shifting towards a plant-based diet can also help to reduce greenhouse gas emissions (IPCC 2019). Despite the numerous advantages of fruit and vegetable consumption, many households worldwide, including in developed regions like Europe, have lower intake than recommended (WHO 2011).[1]

This paper investigates the impact of combining a normative appeal with a small monetary incentive to promote fruit and vegetable consumption. I exploited a unique natural experiment where a supermarket increased their bonus points on fresh fruit and vegetable products and labeled them "good deeds" that are sustainable for the environment and individual health.[2] The increased bonus points lowered the price of (fresh) fruit and vegetables by 0.6–2% and the normative appeal places salience on these products as sustainable good deeds. The supermarket made no prior announcement of these changes. This is important since it prevents consumers from sorting themselves into or out of the treated store in anticipation of the changes.[3]

To evaluate the natural experiment, I have access to a unique data set of individual transactions from the treated supermarket and a similar control store.[4] This enables me to track the purchases of individual consumers over time, create a panel data set of consumers that are members before and

[1] Dietary risks is the second most prominent source of mortality in Sweden (Murray et al. 2020).

[2] In recent years, food retailers have begun implementing various programs to help their customers consume more environmentally friendly products. The background to this is provided in more detail in subsection 2.2.2.

[3] For example, one worry would be that consumers who consume a relatively high volume of fruit and vegetables become members in anticipation of the bonus, changing the consumer composition in the treated store compared to the control store.

[4] See Dubois et al. (2022) for an overview of the pros and cons of scanner data in economic research.

after the intervention is implemented, and employ a difference-in-difference design to analyze the causal effect of combining a monetary incentive with a normative appeal to promote fruit and vegetable consumption.

Several studies have attempted to promote plant-based consumption by different methods, such as changing the menu order (Kurz 2018), using cues (Cornelissen et al. 2008) or different informational messages (Palomo-Vélez et al. 2018, Shreedhar & Galizzi 2021, Perino & Schwirplies 2022, Pace & van der Weele 2020, Kanay et al. 2021, Panzone, Ulph, Hilton, Gortemaker & Tajudeen 2021, Lohmann et al. 2022). Most studies in this area are lab or small-scale field experiments, and we have less evidence from large-scale interventions (see Broers et al. 2017, Cadario & Chandon 2020 and Carlsson et al. 2021 for recent overviews). Understanding if interventions are effective when implemented at scale is essential for policy implications (DellaVigna & Linos 2022, Al-Ubaydli et al. 2017).

There are several reasons to expect that treatment effects could differ in a lab or smaller field setting compared to if the intervention is implemented in a supermarket. First, in real-world supermarkets there will be thousands of products and other promotional campaigns, which could alter the treatment's salience compared to a lab or smaller field setting. Second, interventions could have different short versus long terms impacts. For example, a normative appeal could be particularly salient in the first days, weeks, or months but fade over time. On the other hand, over longer time horizons, consumers would have more time to adapt and incorporate a monetary incentive into their budget. That effect may therefore grow or remain more stable as time passes.

The main contribution of this paper is to provide large-scale and real-world evidence on the causal effects of combining a small monetary incentive with a normative appeal to promote fruit and vegetable consumption. In this paper I follow more than 40,000 consumers in two supermarkets who place over 800,000 orders several months before the intervention, throughout the entire year when the intervention is in place, and for several months after the intervention has been removed. This allows me to examine how consumers in the treated store responded in the short and long term, and to study whether the effects persist after removing the bonus.

I begin by studying the average response over the year the bonus program is in place. Overall, fresh fruit and vegetable purchases increased by 390 grams per order, corresponding to a 9% increase per order. This effect is substantial in relation to the small monetary incentive. For example, in their meta-analysis, Afshin et al. (2017) found that fruit and vegetables would increase by 11-17% for a 10% price reduction. Furthermore, while the consumption of fruit and vegetables began to decline after the bonus program was removed, the effect remained statistically significant for up to two months after the program ended.

The increase in fresh fruit and vegetable purchases was driven by the (intensive margin) volume of fresh fruit and vegetables per order rather than an increase in the (extensive margin) share of orders that contained fresh fruit and vegetables. Furthermore, the increase in fresh fruit and vegetables were not caused by consumers substituting away from frozen fruit and vegetable products. Instead, the consumers' overall consumption of fruit and vegetables increased.

The bonus program consists of four different bonus levels. The additional bonus points for (fresh) fruit and vegetables correspond to different equivalent price reductions depending on the customers' bonus level. The additional bonus points implied a 0.6% equivalent price reduction for the lowest bonus level and increased to 2% for the highest bonus level. I attempt to isolate the monetary incentive from the normative appeal by studying the heterogeneous response among consumers in the different bonus groups. However, contrary to what one might expect, I do not find that consumers who faced a larger price incentive responded differently from the other groups.

The remainder of the paper is organized as follows. In the next subsection, I describe how my paper is related to the literature and discuss the main contributions of my paper. Sections 2.2 and 2.3 provide background details on the treated and control store and the natural experiment. Section 2.4 describes the data and Section 2.5 the empirical methodology. Section 2.6 presents the results and Section 2.7 concludes with a short discussion of the results.

2.1.1 Relation to the literature

This paper is connected to several strands of literature. First, it relates to studies on promoting fruit and vegetable consumption through various nudging techniques. Several studies have explored how to increase fruit and vegetable consumption by manipulating salience or placement (Thorndike et al. 2017, Adams et al. 2016, Hanks et al. 2012, Cohen et al. 2015, Van Kleef et al. 2012, Adams et al. 2005, Kurz 2018, List & Samek 2015). Other studies have employed cues and social norm messages (Cornelissen et al. 2008, Gonçalves et al. 2021), food labels and information campaigns (Levy et al. 2012, Capacci & Mazzocchi 2011) or environmental moral appeals (Shreedhar & Galizzi 2021). Although most studies on nudges or information interventions are based on laboratory or small-scale field experiments in lunch cafeterias, some investigate nationwide policies.[5] Capacci & Mazzocchi (2011), e.g., examined the nationwide 5-a-day information program in the UK.[6]

The main contribution of this paper compared to previous lab or field experiments is to provide large-scale evidence along several important dimensions. To begin with, the intervention in this paper is implemented in a real-world supermarket. From a policy perspective, this is particularly important to understand the external validity if, for example, the intervention would be implemented as a national policy in all supermarkets. Supermarkets have several thousands of products and many other food categories than fruit and vegetables, which could impact the salience of an intervention compared to a lab or smaller field setting.

This paper's sample size and time horizon are substantially larger than previous lab or field experiments. I follow over 40,000 consumers in two supermarkets who place over 800,000 orders several months before the intervention, throughout the entire year when the intervention is in place,

[5] See Broers et al. (2017) and Carlsson et al. (2021) for recent reviews.

[6] The 5-a-day program supports various initiatives, including the National School Fruit Scheme, which provides free fruit or vegetables to children aged between 4 and 6 in schools. The program also involves a multimedia communications campaign and collaboration with private and public partners. The program's goal, implemented in numerous countries worldwide, is to encourage fruit and vegetable consumption and reach the recommended daily intake of five portions (equivalent to 400 grams) as recommended by WHO (2011). Stables et al. (2002) studied the program in the US.

and several months after the intervention has been removed. This allows me to examine how consumers in the treated store responded in the short and long term and to study whether the effects persist after removing the bonus.

Furthermore, compared to the studies investigating nationwide policies, this paper uses individual transaction data and a suitable control group that does not receive the treatment. This allows me to use a difference-in-differences design on the individual consumer level to study the causal effects of the intervention, compared to Capacci & Mazzocchi (2011), which uses aggregate data at the national level for the UK.

Second, this study is also related to the literature focusing on monetary incentives to promote fruit and vegetable consumption (An 2013). The majority have looked at incentives targeted to low-income families (Herman et al. 2008, Klerman et al. 2014, Berkowitz et al. 2021, Griffith et al. 2018).[7] Two closely related papers are Panzone et al. (2022) and Polacsek et al. (2018). Panzone et al. (2022) studied a promotion campaign implemented by a retailer in Croatia, where stickers were affixed to fruit and vegetable products targeted to children. Polacsek et al. (2018) offered low-income families a 5% discount on fruit and vegetable purchases.

Compared to Panzone et al. (2022), this paper has access to a suitable control store and tracks consumers in the treated and control store several months before, during the one year the intervention is implemented, and several months after the intervention was removed. This allows me to use a difference-in-difference strategy to provide causal evidence of the program's overall effectiveness, and study its heterogeneous impacts along consumer characteristics and different time horizons. Compared to Polacsek et al. (2018), this paper's sample size is substantially larger. Polacsek et al. (2018) follows about 400 participants, whereas this paper follows more than 40,000 consumers.

The monetary incentive in this paper (0.6-2%) is smaller than in most

[7] Klerman et al. (2014) examined the effects of offering a 30% incentive for purchases of targeted fruits and vegetables (TFVs) for participants in the HIP program. Klerman et al. (2014) found that after four to six months of implementing the incentive program, the mean daily intake of TFVs for HIP participants was 24% higher than that of the control group. Griffith et al. (2018) found that the UK Healthy Start scheme, which introduced vouchers for fruit, vegetables, and milk, was more effective than an equivalent-value cash benefit.

other studies employing monetary incentives. However, the bonus program analyzed in this paper also incorporates a normative appeal, labeling the affected products as "good deeds", and I observe larger effect sizes in this paper than expected from the monetary incentive alone.[8] The intervention in this paper is further not limited to targeting young children, as in the program analyzed by Panzone et al. (2022), or low-income households, as in a number of other studies (Griffith et al. 2018, Herman et al. 2008, Klerman et al. 2014, Berkowitz et al. 2021).

Third, this paper also relates to the literature on using salience nudges and monetary incentives in isolation or combination (List & Samek 2015, Stuber et al. 2021, Ni Mhurchu et al. 2010). Several papers show that consumers do not pay full attention to sales taxes, and that making them more salient can increase awareness and their effect on consumption (Chetty et al. 2009, Gabaix 2019, Tiefenbeck et al. 2018). Other studies have investigated if intrinsic motivation, peer comparison (Allcott & Rogers 2014, Allcott & Taubinsky 2015, Kotchen & Moore 2007, Ghesla et al. 2020), moral appeals (Egebark & Ekström 2016) or climate labels (Bilén 2022) affect behavior.

It is, however, not evident that normative or moral appeals should go hand in hand with monetary incentives (Gneezy & Rustichini 2000, Gneezy et al. 2011, Bowles & Polania-Reyes 2012, Bénabou & Tirole 2006, Titmuss et al. 1970).[9] For example, Sudarshan (2017) found peer comparison reports on electricity consumption were as effective in reducing electricity consumption, but combining peer comparison and monetary incentives to reduce consumption was ineffective in changing behavior. Mellström & Johannesson (2008) found that providing a monetary incentive to donate blood crowded out donations for women but not for males.

This paper provides evidence that combining a monetary incentive with a normative appeal effectively promotes fruit and vegetable consumption. I find larger effect sizes in this paper than in most other large-scale field or natural experiments outside of lab environments (Carlsson et al. 2021). The setting in this paper differs from, for example, blood donations since fruit

[8] Afshin et al. (2017) estimate in their meta-analysis that a 10% subsidy on fruit and vegetable consumption would increase purchases by 11-17%.

[9] The extrinsic reward could lower the signal of the individual's intrinsic motivation and may therefore crowd out some behavioral change through the reputation channel of the individual's choices.

and vegetable consumption carry an individually beneficial health component in addition to the prosocial motive of consuming those products for the environment.

2.2 Institutional background

This section begins by providing institutional details on the two supermarkets studied in this paper and then gives background details on why the treated supermarket implemented the bonus on fruits and vegetables.

2.2.1 Treated and control store

In Sweden, the food market is dominated by a few large parent organizations (retail food groups) which own grocery stores or subsidiary stores. Most sales occur in physical stores, while online sales are primarily concentrated in urban areas, particularly in Stockholm and Gothenburg. To my knowledge, there is no data on the share of online sales in Stockholm and Gothenburg. On the national level, online purchases accounted for about 6 percent of all sales at the end of 2020 (Swedish Food Retailers Federation 2020). An important feature of the Swedish market during 2020 and 2021 is that Sweden did not implement strict lockdown policies in response to the pandemic as did many other countries. Shopping for food in physical stores was still available, but the pandemic did increase the market share for online sales.

Table 2.1 contains institutional details on the two stores in this paper, which are subsidiaries of one of Sweden's largest food retail groups. Each store is responsible for its pricing, whereas the retail group is primarily responsible for assortment decisions. There is one notable difference in their pricing schemes. The control store offers free order delivery above 700 SEK. In contrast, the treated store always charges a delivery fee of 49–99 SEK (and an additional fee of 49 SEK if the order exceeds 700 SEK). However, the control store has slightly higher product prices, making the total order cost similar between stores.

The control store only operate online. They were one of the first online supermarkets to enter the Swedish online market when they started offering online food purchases in the Gothenburg area in 2012 and have a relatively large market share in the Gothenburg area compared to other areas, such as Stockholm. The treated store initially operated exclusively as a physical store but later adapted to the growing demand for online shopping by offering both online and physical stores from 2016 onward. They also offer their customers a bonus (loyalty) program that rewards customers with bonus points when they spend money at the treated store. Section 2.3 will describe this bonus program in more detail.

In section 2.4, I will provide an overview of aggregate purchase data on fruit and vegetables in each store, and subsection 2.5.2 will provide data on the monthly purchase patterns before and after the natural experiment.

Table 2.1: Institutional details on the two supermarkets

	Treated store	Control store
Loyalty program		
Bonus points	Yes	No
Operation		
Store	Online and Physical	Online
Supply chain	Joint	
Delivery		
Gothenburg area	Yes	Yes
Stockholm area	Yes	Yes
Malmö area	Yes	Yes (until April 2020)
Pick-up in store	Yes	No
Cost (SEK)	49-99 (+49 below 700)	29 (free above 700)
Time	Mon - Sun	Mon - Sun
Car	Shared	

Notes: The exchange rate was about 10 SEK to 1 Euro during 2020-2021.

2.2.2 Sustainability goals

Corporate social responsibility has become a prevalent standard among business organizations in recent years (Bénabou & Tirole 2010 and Fioretti 2022). Grocery stores are well aware that the food system is one of the largest emissions sources, and the main grocery stores in Sweden have all established sustainability goals aimed at reducing their business carbon footprint (Axfood AB 2020, ICA Gruppen 2022, COOP 2020). This trend is not unique to Sweden or the Nordic market.[10] For example, the five largest supermarkets in the UK have also committed to reducing their carbon footprint by half by 2030, and Unilever, which is one the world's largest producer of consumer goods, have committed to reaching net zero in emissions by 2030 (WWF 2022, Unilever 2022).

To achieve their sustainability goals, grocery stores have implemented various tools to help customers consume more sustainably. The treated store motivated its additional bonus on fresh fruits and vegetables by citing the lower climate impact and additional health benefits, referring to the Lancelot commission's recommendations on increased fruit and vegetable consumption (Willett et al. 2019). Other supermarkets in Sweden have tried to help customers by offering climate-friendly meals and personal carbon footprint targets (ICA Gruppen 2022, COOP 2020). Grocery stores in France (Collibri Foundation 2021), Germany (COBIOM 2021), and the UK (Lidl 2021) have plans or have begun to implement environmental labels.[11]

Examples outside of the food industry include Allbirds, Inc. and Logitech International S.A. that provide estimates of their products carbon footprint (Allbirds 2022, Logitech 2022).[12]

2.3 Natural experiment

This paper analyzes a natural experiment in the treated supermarket between October 5th, 2020, and September 30th, 2021. During this period, the supermarket made two changes to purchases of fresh fruit and vegetable products: (i) offering double points and (ii) labeling those purchases as good deeds. Importantly, the supermarket made no prior announcement and took no other steps to inform customers about the upcoming changes.

In subsection 2.3.1, I describe the changes to the bonus program that constitutes the natural experiment's treatment, and how the bonus points are calculated and awarded at the treated store. In subsection 2.3.2, I

[10] Ytreberg et al. (2023) provides a recent review of the interventions implemented by Nordic retailers.

[11] Tesco had plans to do so in 2011 but later abandoned them (Potter et al. 2021, Taufique et al. 2022).

[12] Logitech International S.A. is a multinational manufacturer of computer peripherals and software, and Allbirds, Inc. is an American company that sells footwear and apparel.

explain the monetary incentive for consumers to collect bonus points.

2.3.1 Treatment

The bonus program incentivizes customers to spend more at the treated store by offering bonus points for purchases. The bonus points are awarded based on the amount of SEK spent, including online and physical purchases, and customers use the same member account for both transaction sources. Table 2.2 provides an overview of the bonus points awarded for products belonging to specific categories.

The first column of the table displays the bonus points awarded prior to the natural experiment, in which every SEK spent was rewarded with one bonus point. The second column illustrates the bonus points awarded during the natural experiment from October 5, 2020, to September 30, 2021. Customers earned two bonus points for each SEK spent on fresh fruits and vegetables throughout this period. All other food categories, including frozen fruit and vegetable products, were rewarded with one bonus point per SEK spent.

The bonus points awarded after the natural experiment are displayed in the third column. From October 1, 2021, the additional bonus points for fresh fruits and vegetables and the good deeds label were removed. Subsequently, the store shifted the good deeds label and double-point rewards to its line of ecological products, denoted as Store's Own Brand in Table 2.2. This paper studies the effects on fruit and vegetable consumption and does not analyze the effects on the store's own brand products after October 2021.

Table 2.2: Overview of bonus program in the treated store

Period	Before	Natural experiment	After
Time	-10/04/2020	10/05/2020-09/30/2021	10/01/2021-01/01/2022
Treated category			
Fruit and vegetables (Fresh)			
Bonus points per SEK	1	2	1
Category label	No	Good deeds	No
Untreated categories			
Fruit and vegetables (Frozen)			
Bonus points per SEK	1	1	1
Category label	No	No	No
Store's own brand (Ecological)			
Bonus points per SEK	1	1	2
Category label	No	No	Good deeds
Other			
Bonus points per SEK	1	1	1
Category label	No	No	No

Note: This table provides the timeline of the natural experiment at the treated supermarket. For each spent SEK, customers collect bonus points, and the number of points per SEK is displayed in the table.

2.3.2 Bonus levels

Consumers collect bonus points during three bonus periods each year: from February 1st to May 31st, from June 1st to September 30th, and from October 1st to January 31st. The number of bonus points the customer collected during the previous period will determine to which bonus level the member will be assigned for the next bonus period. Collecting bonus points is valuable because customers receive a bonus check for every 2,500 bonus points they accumulate, and the value of the bonus check increases the higher the customer's bonus level status is.

Table 2.3 displays an overview of the bonus program and the monetary reward for collecting bonus points. For instance, customers who collected less than 5,000 bonus points in the previous period will be placed in bonus level 1 for the upcoming period. They will receive a bonus check of 15 SEK for every 2,500 bonus points they collect in the current period. Customers who collected at least 15,000 points during the previous period will be placed in bonus level 4 and will now receive a bonus check of 50 SEK for every 2,500 points they collect during the current period.

Customers would receive double bonus points on fresh fruit and vegetables during the natural experiment. Since the value of the bonus check increases depending on the customer's bonus level, the additional bonus point on fresh fruit and vegetables will cause a larger equivalent price reduction for customers in higher bonus levels than those in lower bonus levels. The additional bonus point corresponds to an equivalent price reduction of 0.6% for customers in the lowest bonus stage, 0.8% for customers in level two, 1.2% in level three, and a 2% equivalent price reduction for customers in the highest bonus level.

Table B1 displays the number of consumers in each bonus level during the natural experiment. Consumers can move between bonus levels depending on their accumulated bonus points during the last four-month period. Figure B1 displays the movement between bonus levels from one period to the next. About half of the members in a current bonus level will stay at the same level in the next period. About 15% will move one level up (down), 5% will move two levels up (down), and 3% will move three levels up (down).

Table 2.3: Bonus levels during the natural experiment

Bonus level	1	2	3	4
Bonus point details				
Required points (previous period)	< 5000	5000-9999	10,000-14,999	15,000-
Value of bonus check for every 2,500 points collected	15 SEK	20 SEK	30 SEK	50 SEK
Equivalent price reduction during the natural experiment				
Fresh fruit and vegetables	0.6%	0.8%	1.2%	2%
Frozen fruit and vegetables	0%	0%	0%	0%
Other categories	0%	0%	0%	0%

Notes: This table overviews the design of the bonus program during the natural experiment. For every 2,500 bonus points, customers at the treated store will receive a bonus check. The value of the bonus check is based on the member's current bonus level. Four bonus levels exist and are based on how many points the customer collected during the previous period (one period comprises four months). During the natural experiment, the bonus points were increased from 1 to 2 for fresh fruit and vegetables. All other categories remained at one bonus point per SEK. The equivalent price reduction is calculated as $\frac{Value\ of\ bonus\ check}{2{,}500}$.

2.4 Data

2.4.1 Transactions

Consumers are required to register a member account to shop online. For each account, I obtain an indicator to distinguish between private consumers and business organizations, and I restrict the sample to private consumers. The transaction data cover all online food purchases for all private consumer member accounts from 2020 to 2022.[13] For each order transaction, the data includes all the products in the order, the date, time, postal code of delivery, and the consumer's member ID. This allows me to follow individual members' purchases over time.

For each product in the order, the data includes the quantity, price, weight, and applicable discounts. I also have background information on a rich vector of characteristics for each product, including other food labels, such as Ecological and Fairtrade labels; country of origin; and if the product is free of certain ingredients, such as milk or meat.

2.4.2 Categories

Each product in the dataset contains a unique product identifier, which is linked to a hierarchical structure of product categories. The product hierarchy includes both aggregate categories, such as fruit and vegetables or meat, and subcategories, such as fresh or frozen fruit and vegetables.[14] The monetary bonus is implemented on fresh fruit and vegetables but is not implemented on frozen fruit and vegetables.

Table 2.4 presents summary statistics on overall fruit and vegetable consumption, fresh fruit and vegetable consumption, and frozen fruit and vegetable consumption. The pre-period spans from January 2020 until 4 October 2020, and the treatment period spans from 5 October 2020 until 30 September 2021. Column (1) and (2) contain average kg of fruit and

[13] I also obtained all order amounts from the treated store in physical stores, which allows me to track the members' bonus level over time, since the bonus system considers orders in online and physical stores.

[14] It should be noted that while the product categories are similar across stores, they are not completely overlapping. By observing scanner codes, unique product identifiers can be linked between stores, enabling the mapping of the category structure across stores.

vegetables per order with standard deviations in parenthesis for consumers in the treated store before and after the bonus change was implemented, and column (3) and (4) contains corresponding data for the control store. Column (5) contains statistics for the full sample and the entire data period.

The sample includes 857,028 orders, and the data is based on the sample of consumers observed at least once both in the pre-period and the treatment period. About 94% of the orders contain at least some fresh fruit and vegetables, compared to 44% that include frozen fruit and vegetables (looking at the whole sample in both stores, which is also similar to the sub-samples in each store separately). The average volume of consumed fruit and vegetables per order is about 4 kg, while the average volume of consumed frozen fruit and vegetables per order is 0.5 kg. The average consumption among orders that do contain at least some fresh fruit and vegetables is 4.4 kg, and 1.1 kg for frozen fruit and vegetables.

Table 2.4: Summary statistics on fruit and vegetable consumption per order

	Treated store		**Control store**		**Full sample**
	Before	During	Before	During	
Total					
All	4.46	4.99	4.46	4.57	4.62
	(3.57)	(3.84)	(3.39)	(3.36)	(3.53)
Fresh	3.97	4.52	3.93	4.06	4.12
	(3.30)	(3.55)	(3.08)	(3.05)	(3.23)
Frozen	0.48	0.46	0.53	0.51	0.50
	(0.91)	(0.90)	(0.98)	(0.97)	(0.95)
Intensive margin					
All	4.72	5.25	4.67	4.76	4.85
	(3.50)	(3.76)	(3.32)	(3.29)	(3.46)
Fresh	4.28	4.82	4.18	4.28	4.38
	(3.23)	(3.47)	(3.01)	(2.98)	(3.16)
Frozen	1.14	1.13	1.12	1.15	1.14
	(1.09)	(1.12)	(1.18)	(1.18)	(1.15)
Extensive margin					
All	0.94	0.95	0.95	0.96	0.95
	(0.23)	(0.22)	(0.21)	(0.20)	(0.21)
Fresh	0.93	0.94	0.94	0.95	0.94
	(0.26)	(0.24)	(0.24)	(0.22)	(0.24)
Frozen	0.42	0.41	0.47	0.45	0.44
	(0.49)	(0.49)	(0.50)	(0.50)	(0.50)
Observations	132,986	201,618	220,351	302,073	857,028

Note: This table contains summary statistics on the consumption of average kg of vegetables per order. The sample contains private consumers observed before and after the natural experiment. "Before" refers to the period before the treatment was implemented 01/01/2020-10/04/2020, and "During" to the period when the treatment was in place 10/05/2020-09/30/2021. The intensive margin refers to the volume per order when the customers purchase at least one product from the category. The extensive margin refers to the share of orders containing at least some products from the category.

2.4.3 Demographics

To shop online, customers must have a registered member account containing personal information, including the member's home address.[15] In this section, I present a statistical summary of the demographic background of households based on their reported home addresses.

To ensure the anonymity of the members, I obtained the postal code of the member account at a three-digit level from both stores. I linked the postal codes with register data from Statistics Sweden to obtain aggregate socioeconomic background information on income, education, and household composition. The treated store collect age data on all its members in ten-year intervals, while the control store only collected age data when customers paid by invoice. [16]

Table 2.5 summarizes the key demographic characteristics of the sample of consumers from both stores. Online food delivery is primarily available in urban areas, and our sample of consumers is concentrated in the two largest cities in Sweden, Stockholm and Gothenburg. In the treated store, the share of consumers in Stockholm and Gothenburg corresponds to the population size of each city, with Stockholm having almost twice the population of Gothenburg. However, the control store has relatively more consumers in Gothenburg, which is explained by the store's initial market entry into Gothenburg in 2012.

I divide the sample into three age brackets: below 40, 40-60, and above 60. In the treated store, there is an even representation of consumers across all three age brackets. However, before the pandemic, the age distribution was more skewed towards the middle-aged and those over 60 were under-represented. The control store collected age data when consumers paid by invoice, which explains the large share of middle-aged members.

[15] Customers also have the opportunity to have a delivery made to another address for a specific order.

[16] The control store also collected information on the house type of the customers (apartment or house), while the treated store did not.

Table 2.5: Demographic summary statistics

	Treated store			Control store		
	count	mean	sd	count	mean	sd
Location						
Stockholm	9,111			13,033		
Gothenburg	3,795			16,092		
Other	3,286			370		
Age						
< 41	6,223			1,121		
41-60	4,866			1,981		
60+	5,103			325		
No info	0			26,068		
Income						
Median	16,039	588.825	268.902	29,495	581.255	233.359
No info	137			0		
Members	16,192			29,495		

Notes: This table presents the demographic summary statistics of treated and control stores. The columns include count, mean, and standard deviation (sd) for each demographic variable, separated by treated and control stores. The demographic data that is collected from the stores include the location (Stockholm, Gothenburg, and Other) and age groups (less than 41, 41-60, 60+, and No info). The treated store collects age information on all customers, whereas the control store only collects age when customers pay by invoice. Data on income have been collected from Statistics Sweden based on the postal codes of customers on a three-digit level.

2.5 Empirical methodology

In this section, I will first describe the regression models utilized and the assumption that ensures a causal interpretation of the estimates. I will then provide graphical evidence of how the average consumption of fresh fruit and vegetables developed over time in the two stores and discuss the parallel trend assumption.

2.5.1 Main specification

To study how the bonus and labeling (fresh) fruit and vegetables as good deeds affect consumption among consumers in the treated store, I estimate the model in equation (2.1). The dependent variable Y measures the outcome of interest per order for consumer i when shopping in store s during time period t in order k. The store subscript s refers to either the treated store (s = *Treated*) or the control store (s = *Control*). Our main outcome of interest is kg of fresh fruit and vegetables per order, but I will also estimate this model for frozen fruit and vegetables.

$$\mathrm{Y}_{i,s,t,k} = \beta_0 + \gamma_i + \gamma_t + \beta_1(Post * Treated) + e_{i,s,t,k}. \tag{2.1}$$

To control for individual characteristics that might be correlated with a consumer's consumption of fruit and vegetables, such as allergies or other time-invariant dietary preferences, I include individual fixed effects, γ_i, in the model. Individual fixed effects will also control for other variables that remain constant during the study period, which for example could be household composition. The dummy variable *Treated* equals 1 for consumers in the treated store and 0 for consumers in the control store. It is important to note that the data does not provide information on whether consumers were members of both stores simultaneously. Therefore, the individual fixed effects will also absorb store fixed effects that control for time-invariant store characteristics that remain constant over time but may differ between stores, such as larger consumption of fruit and vegetables in one of the stores.

Month dummies, γ_t, capture variation in fruit and vegetable consumption over time that is not correlated with a specific store. For example, more locally produced fresh fruit and vegetables are consumed during the summer period in Sweden. The dummy variable *Post* takes the value of 1 from the date of the natural experiment and onward. The interaction term $Post * Treated$ captures variation in the bonus program over time. It equals 0 for both stores in the pre-treatment periods but switches to 1 for all consumers in the treated store in the post-treatment period. Both the dummy variables *Treated* and *Post* will be absorbed by the model specifications containing individual fixed effects and month dummies, respectively.

The main parameter of interest is β_1, which measures the average causal effect of the bonus program on (fresh) fruit and vegetables per order for consumers in the treated store who shopped before and after the bonus change. The identification of β_1 is based on the variation in the bonus across stores over time.

To consistently estimate β_1, I rely on the assumption that consumers in both stores would have followed parallel trends in their (fresh) fruit and vegetable consumption if the bonus program had not changed. This assumption would be invalid if consumers in the two stores had different trends in fruit and vegetable consumption before the natural experiment, or if customers in the two stores had different trends afterward due to other factors unrelated to the treatment.

To assess if the customers had different pre-trends, I will provide graphical evidence of the average purchase volume of (fresh) fruit and vegetable purchases in the two stores over time. To assess if the other simultaneous changes affected the consumption patterns, I will present graphical evidence of the prices of fruit and vegetables. I will also provide robustness checks to the main results in subsection 2.6.2 where I include the average price per kg fruit and vegetables in the regression models. It is important to note that the parallel trend assumption would not be invalid if the customer recruitment changed during the natural experiment, for example, if the treated store recruited those consumers with particularly high consumption of fruit and vegetables, as I have access to the individual transaction data that allows me to disentangle consumer change choices from changes in the consumer base. Standard errors are clustered at the consumer level (member account level) across all specifications, permitting errors to be correlated within an individual consumer (member account) over time, but not across distinct consumers (member accounts).

I will also conduct an analysis using the (event study) model described in equation (2.2), where I interact the dummy variable for the treated store with each time-month dummy. In this model, j refers to the time periods before (leads) the implementation of the bonus, and l represents the time periods after (lags) its introduction. The excluded reference period is the time period before the bonus was implemented (t = -1) and represents the difference between the consumers' fruit and vegetable consumption during this period.

$$\mathrm{Y}_{i,s,t,k} = \beta_0 + \gamma_i + \gamma_t + \sum_{j=1}^{J-1} \textit{Treated} * \gamma_j + \sum_{l=0}^{L} \textit{Treated} * \gamma_l + e_{i,s,t,k}. \quad (2.2)$$

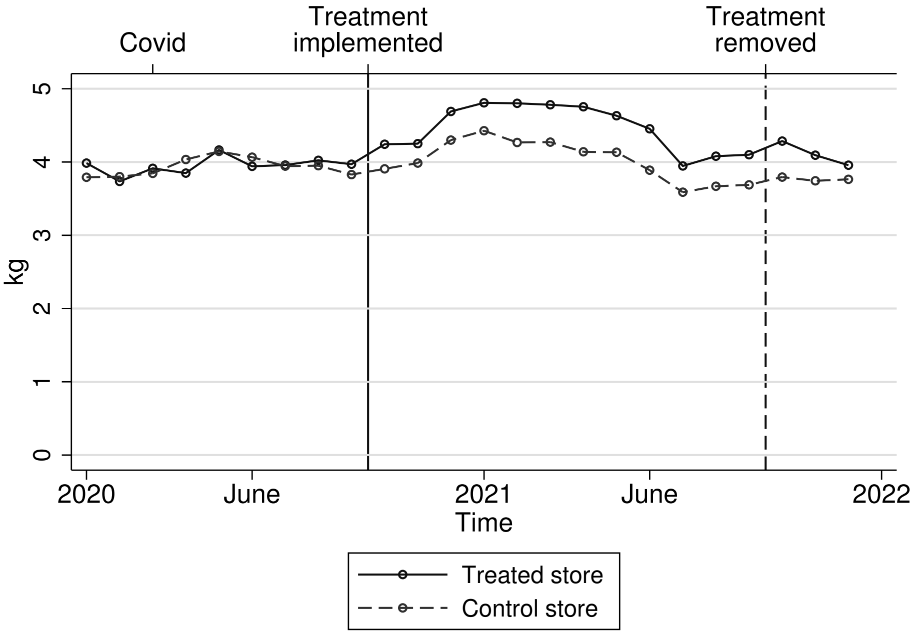
If both stores followed a parallel trend in fruit and vegetable consumption before the treatment was introduced, we should not expect any statistically significant estimates for the estimated parameters γ_j. In contrast, we would expect γ_l to be positive and statistically significant if the bonus program and the normative appeal caused consumers in the treated store to increase their fruit and vegetable consumption.

2.5.2 Graphical evidence

Figure 2.1 shows how average (kg) fresh fruit and vegetables per order developed over time in 30-day intervals. The first period began in January 2020. The vertical line indicates the start date of the natural experiment, 5 October 2020, when the treated store implemented the additional bonus point and labeled (fresh) fruit and vegetables as "good deeds."

There is a stable and parallel trend in the consumption of fresh fruit and vegetables from January 2020 until October 2020, with close to no difference in the average order volume between the stores. There is a clear rise in consumption of fresh fruit and vegetables for consumers in the treated store following the treatment. The increased consumption amounts to, on average, about 200-400 grams of fresh fruit and vegetables per order for consumers in the treated store, compared to consumers in the control store evaluated over the entire year when the treatment was in place.

Figure 2.1: Average (kg) fresh fruit and vegetables per order



Notes: This figure shows the average consumption of fresh fruit and vegetables per order in each store from January 2020 to December 2021. Each diamond represents a time period of 30 days. The vertical line indicates when the natural experiment begins (5 October 2020 and onward), and the shaded line indicates when the natural experiment ends (1 October 2021 and onward).

2.5.3 Price and other simultaneous changes

The validity of the difference-in-difference estimate as a measure of the average causal effect of the bonus program on consumers' fruit and vegetable consumption relies on a parallel trend in consumption among shoppers in the two stores that would otherwise continue in the absence of the change in the bonus program. Hence, no other significant changes should occur simultaneously with the treatment that impacts the consumption of fruit and vegetables.

I begin by looking at how the price of fruit and vegetables evolved around the timing of the bonus program—for example, if the treated store also changed its pricing of fruit and vegetables to further stimulate demand. To assess whether there was a systematic change in either store's price around the treatment date, I have plotted the average prices of one kilo of fresh and frozen fruit and vegetables in Figure B2 with a 30-day interval. The average kilo price remained stable two months prior and two months after the bonus program was implemented, at around 40 SEK per kilo. The control store charges a kilo price of 5 SEK higher than the treated store, which aligns with their pricing for other food categories whereas the control store instead offers free delivery as described in Section 2.2.[17] One explanation is that even if sales occur regularly, they are not implemented on all fruit and vegetable products, but instead on single products that could impact the substitution within the category but do not have much impact on the average price across the whole category. One might still be concerned that prices may have impacted the results. Price is an endogenous variable set by the store, so it is not apparent that one should want to control for it in the regression analysis. However, in subsection 2.6.2, I rerun the main analysis by controlling for prices.

2.6 Results

In this section, I first present the main results in subsection 2.6.1, on how the bonus program impacted purchases of (fresh) fruit and vegetables in the treated store. The treatment effect from the bonus program consists of the combined effect of the monetary incentive from the additional bonus points and the normative appeal.

In subsection 2.6.2, I assess if the main results are robust to controlling for prices. In subsection 2.6.3, I first try to disentangle the role of the monetary incentive by looking into how the different bonus-level groups were affected as the bonus point produced different price reduction equivalents for each bonus level. I then look at heterogeneity based on the age and income levels of the consumers. Lastly, in subsection 2.6.4, I study if the changes in consumers' fruit and vegetable consumption are large enough to also translate into an environmental impact.

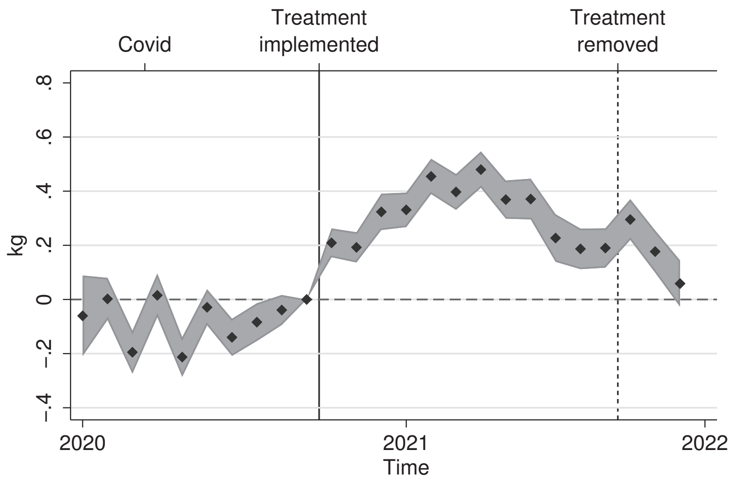
[17] In Bilén (2022), I have plotted the price of meat and meat substitutes, which display a similar pattern.

2.6.1 Main results

Figure 2.2 displays the event plot estimates from estimating equation (2.2) of fruit and vegetable consumption for each 30-day interval. The period before the bonus implementation is omitted from the analysis and serves as the reference period. Each point estimate shows the difference-in-difference estimate compared to the omitted period, controlling for time invariant individual consumer characteristics and month dummies. A positive coefficient should be interpreted as the consumption of fresh fruit and vegetables per order being larger in the treated store compared to the control store than before the bonus was implemented.

There was an immediate increase in (fresh) fruit and vegetable consumption per order, and the effect increased over the first five months. Fresh fruit and vegetable consumption was about 200-400 grams larger in the treated store, and the difference was statistically significant for the entire year that the additional bonus was in place. The effect started to decline when the bonus was removed in October 2021. There was no statistically significant difference during the last period in December 2021, even though the point estimate remained positive.

Figure 2.2: Event plot of effects of the bonus program on fresh fruit and vegetables per order



Note: This figure shows an event plot from estimating equation (2.2), and the dependent variable is kg of fresh fruit and vegetables per order. A time period refers to 30 days, and the solid vertical line indicates when the bonus program was implemented in the treated store (5 October 2020). The excluded period (time period -1) is the reference period and refers to the 30 days before the treatment was implemented. Each diamond indicates the difference-in-difference point estimate, γ_k, compared with the reference period, with a 95 percent confidence interval around it indicated by the grey area. A positive value should be interpreted as the dependent variable being larger in the treated store than in the control store, relative to the difference between stores in the reference period (-1).

The bonus was implemented on fresh fruit and vegetables, while frozen fruit and vegetables remained unchanged. To assess the average impact of the treatment, I estimate equation (2.1) on purchases of fresh and frozen fruit and vegetables. To simplify the notation, I will label the difference-in-difference term *Post*Treated* as *Treatment.*

Table 2.6 shows the difference-in-difference estimates, with columns (1) and (2) containing fresh fruit and vegetables per order and columns (3) and (4) containing frozen fruit and vegetables per order. The treatment increased fresh fruit and vegetable consumption by 386 grams per order during the entire year the treatment was implemented. Column (2) expands the study period to encompass the three months after it was removed. The dummy variable *Treatment*Removed* should be interpreted as the impact of the treatment on purchases of (fresh) fruit and vegetables were 131 grams lower during the three months after the bonus was removed compared to when it was in place. However, the effect of the treatment was positive also after it was removed. The sum of *Treatment + Treatment*Removed* shows that, on average, fresh fruit and vegetable consumption was 254 grams larger in the treated store during the three months when the treatment had been removed compared to the period before the treatment was implemented.

If we now turn our attention to what happened with frozen fruit and vegetables in column (3), we see no statistically significant decline, and the point estimate is narrowly estimated around 0. This implies that the increase in fresh products was not driven by substitutions from frozen products but instead consisted of increased fruit and vegetable consumption per order. In column (4), I find a statistically significant reduction. However, that point estimate is 1 gram per order, which is close to zero in economic impact compared to the increase in fresh products.

Table 2.6: Effects of bonus program on fruit and vegetable consumption per order

	Affected by Bonus		**Not affected by Bonus**	
	(1)	(2)	(3)	(4)
	Fresh	Fresh	Frozen	Frozen
Treatment	0.386***	0.385***	-0.009	-0.010*
	(0.018)	(0.018)	(0.005)	(0.005)
Treatment*Removed		-0.131***		0.011
		(0.024)		(0.007)
Constant	4.033***	3.997***	0.502***	0.495***
	(0.004)	(0.007)	(0.001)	(0.002)
Individual FE	Yes	Yes	Yes	Yes
Treatment + Treatment*Removed		0.254***		0.001
		(0.029)		(0.008)
Observations	857,028	942,205	857,028	942,205
Clusters	44,781	44,781	44,781	44,781

Standard errors clustered on the consumer level in parentheses

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

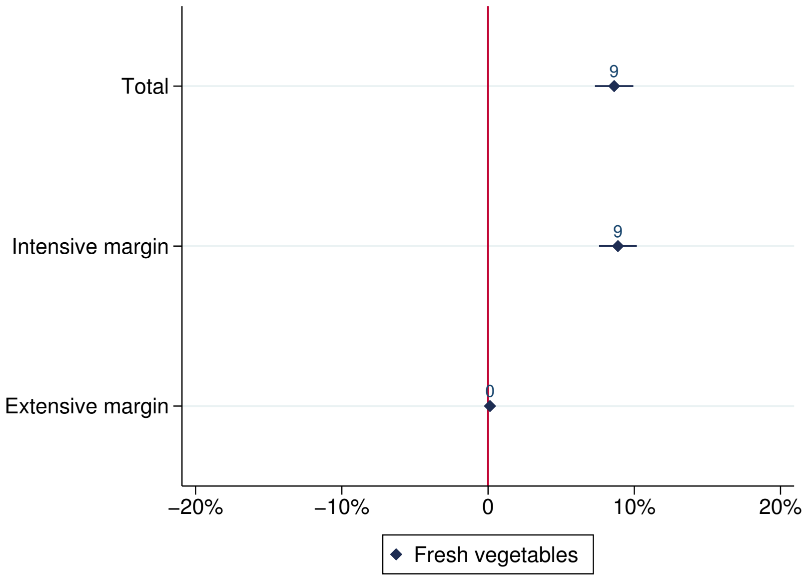
Note: This table presents the overall impact of the bonus program on orders containing fresh and frozen fruits and vegetables by the estimating equation (2.1). The difference-in-difference term *Post*Treated* is labeled as *Treatment*. The dependent variable represents the weight (in kg) of fresh or frozen fruits and vegetables per order. The analysis is limited to consumers observed before and after implementing the bonus program in the treated store. All regressions include individual and monthly fixed effects, with standard errors clustered at the consumer level. *Treatment + Treatment*Removed* refers to the difference-in-difference estimate comparing the period before the bonus program to the period after the program was removed.

Intensive and extensive margin

Figure 2.3 displays the treatment effect on fresh fruit and vegetables consumption separated by the total effect and the intensive and extensive margins. The intensive margin refers to the volume of fresh fruit and vegetables consumed per order for orders containing fresh fruit and vegetables, and the extensive margin refers to the share of orders containing fresh fruit and vegetables.

Before the treatment was implemented, about 94% of orders contained fresh fruit and vegetables, with an average volume of about 4 kg per order. The intensive margin drives the treatment effect, increasing the average order size from 400 to 440 grams. This effect size corresponds to an increase in the total consumption of fresh fruit and vegetable by about 9% per order. On the other hand, the extensive margin does not exhibit any statistically significant change due to the intervention. The null effect on the extensive margin is precisely estimated with a tight confidence interval of around 0.

Figure 2.3: Effects of the bonus program on fresh fruit and vegetable consumption per order: Intensive and extensive margin



Note: This figure illustrates the consumption of fresh fruits and vegetables per order. The difference-in-difference estimates should be interpreted as the percentage change in the dependent variable in the treated store compared to the control store relative to the average consumption before the natural experiment. The intensive margin denotes the volume of fresh vegetables per order, while the extensive margin represents the proportion of orders containing fresh fruits and vegetables. All regressions incorporate individual and month-fixed effects, with standard errors clustered at the consumer level.

2.6.2 Robustness

Price

Price is a choice variable for each store, both before and during the natural experiment. One worry would be that the treated store may respond by changing its prices as an endogenous response to the increased demand from the bonus program. It is not evident that one should control for it in the regression analysis, but it is also important to know if price changes or the bonus program drives the results. As shown in Section 2.5.3, the average kilo price of fruits and vegetables remained stable without any clear sales or price changes surrounding the treatment, but the price in the treated store was lowered about two months before the natural experiment. In this section, I will rerun the main analysis while controlling for prices and run the analysis restricted to the two-month pre-period when prices remained unchanged.

Table B2 presents the regression estimates when controlling for the average kilo price of fruits and vegetables for each 30-day interval at each store. I find that the consumption of fresh fruit and vegetables, on average, increases by 220-240 grams per order. The estimate for frozen vegetables displayed a decrease of close to 30 grams per order, and that change is statistically significantly different from zero. However, the qualitative results remained unaffected, with an overall increase in fruit and vegetable purchases per order at about 200 grams or 6% per order.

Furthermore, in Table B3, I run the analysis by restricting the time period to two months before and after the treatment implementation, during which prices remained steady without significant sales or price changes. Encouragingly, the effect sizes were almost identical to those for the entire time period, with an increase of 245 grams in fresh products and a reduction of 26 grams in frozen products. This is nearly equivalent to the estimates derived from the complete data period. During this specific time period, the price remained unchanged.

2.6.3 Heterogeneity

Monetary incentive

In this subsection, I will attempt to isolate the role of the monetary incentive. As described in Section 2.3, the monetary incentive differed between the bonus levels. The additional bonus point corresponds to an equivalent price reduction of 0.6% for bonus level one and increases to an equivalent price reduction of 2% for bonus level four. If the monetary incentive is the primary driver, we should expect consumers in higher bonus levels to respond more than consumers in lower bonus level groups.

I begin by estimating the model in equation (2.1) when I restrict the treated sample to one of the bonus levels, respectively, compared to all consumers in the control store. Column (1) in Table 2.7 contains consumers in bonus level group 1, and column 2-4 contains those in bonus levels two, three, and respectively.[18] I find that customers in all four bonus levels increased their fresh fruit and vegetable consumption compared to customers in the control store. The largest point estimate is found in the lowest bonus level group, which increased their consumption by 368 grams per order. The other bonus level groups' estimates are similar, from 289 grams for level two and 335 and 319 grams per order for levels three and four, respectively.

In column (5), I rerun the analysis but restricted to consumers in the treated store and include an indicator for the customer's bonus level and interacted with the treatment. The interaction term captures a linear relationship between the treatment and the monetary incentive that the different bonus levels face. Suppose the monetary incentive is the primary driver. In that case, we should expect a positive relationship as the equivalent price reduction increases with the bonus level (from 0.6% for level one up to 2% for level four). However, I find no difference between the bonus levels groups' response to the treatment with a narrow confidence interval around the null estimate.

The increased bonus points were implemented without a prior announcement and affected consumers in all bonus levels. As I showed in Section 2.3.2, in a given bonus period (every bonus period covers four months), about 50% of the consumers stay in their bonus level, while the other half move to another bonus level. To ensure that the sample of consumers in a given bonus level remains fixed, I conducted the analysis by restricting the time period from June 2020 until February 2021. During this time period, the before-treatment period comprises the bonus point period from 1 June 2020 until 1 October 2020, and consumers will stay at that bonus level during the post-treatment period from 5 October 2020 until 31 January.

Table B4 shows the results from this exercise. I find a slightly higher

[18] A consumer's bonus level is based on the previous four shopping months, and this means consumers in bonus level 1, when the bonus was implemented, would face the same price incentive for the next four months.

point estimate in all bonus levels. However, the qualitative pattern remains the same, with a higher point estimate of a 430-gram increase for bonus level one and an increase between 264 and 370 grams for the other bonus levels. The interaction term captures the linear relationship between the bonus level and the treatment and is not statistically significant with a narrow 95% confidence interval around 0.

Table 2.7: Heterogeneity by bonus level: Effects of bonus program on (fresh) fruit and vegetable consumption per order

	(1) Level 1	(2) Level 2	(3) Level 3	(4) Level 4	(5) Comparing bonus levels
Treatment	0.368***	0.289***	0.335***	0.319***	
	(0.032)	(0.039)	(0.048)	(0.066)	
Bonus level					0.100*
					(0.041)
Bonus level*Treatment					-0.011
					(0.025)
Constant	3.969***	3.976***	4.038***	4.220***	4.165***
	(0.002)	(0.004)	(0.005)	(0.010)	(0.078)
Individual FE	Yes	Yes	Yes	Yes	Yes
Observations	459,980	474,786	463,301	491,452	271,531
Clusters	34,518	33,182	30,924	30,123	15,098

Standard errors clustered on the consumer level in parentheses

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Note: This table displays the overall effect of the bonus program on orders containing fresh fruits and vegetables by estimating equation (2.1). The dependent variable indicates the weight (in kg) of fresh fruits and vegetables per order. In columns 1-4, I compared consumers in the control store with consumers belonging to one of the bonus levels in the treated store. In column (5), the sample is restricted to only consumers in the treated store. All regressions include individual and monthly fixed effects, with standard errors clustered at the consumer level.

Demographics

This subsection examines heterogeneity in response to the bonus program related to the demographics of consumers in the treated store. Specifically, I analyze the impact of age and income on the program's effectiveness in increasing fruit and vegetable consumption. The treated store collects data on the members' age for all members and the control store for those members that pay by invoice. I divide the sample into three age groups: below 40, 40 to 60, and above 60. Table B5 presents the results of the treatment for each age group separately in columns 1-3, and then in column 4, I compare the effects across age groups. The results indicate that the treatment increases fruit and vegetable consumption by 400 grams per order among consumers below 40. In contrast, the effect is slightly smaller for the other two age groups, at about 355 and 370 grams per order. However, the differences in response across age groups are not statistically significant.

Next, I examined heterogeneity in the response for different income levels. To do this, I divided the sample into income terciles based on the median income of the customers' postal code of delivery. Table B6 reports the results separately for each income group. Column 1 refers to the lowest-income group, while columns 2 and 3 refer to the middle and high-income groups. The results show that the treatment increased fruit and vegetable consumption in all income terciles, with the lowest income group increasing their consumption by around 250 grams per order. The response among the two highest income terciles was substantially larger, with the middle-income group increasing their consumption by 460 grams on average and the high-income group increasing their consumption by 370 grams per order. The middle-income group increases their consumption by 270 grams per order more than the low-income group. The high-income group increases their consumption by about 200 grams more per order than the low-income group. These differences between income groups are statistically significant and economically significant in magnitude, and correspond to almost twice the effect compared to the response among the low-income group. However, there was no significant difference between the middle and high-income groups.

2.6.4 Environment

This section examines the environmental impact of the treatment by investigating the effect on emissions per order. To quantify emissions per order, I utilize the dataset employed in Bilén (2022), which explored the environmental consequences of carbon labeling on emissions per order in the two stores also studied in this paper.

I compute emissions per order by aggregating the carbon footprint of each product to determine the total carbon footprint per order.[19] Table B7 displays the regression estimates. Column (1) presents emissions per order, column (2) emissions intensity per kilogram of food per order, and in columns (3) and (4), I include controls for the monthly average kg prices of beef, beef mixed with pork, poultry, and chicken, as well as the prices of fresh and frozen fruits and vegetables.

Ex ante, one hypothesis would be that fruit and vegetables have, on average, a lower carbon footprint than meat products and that increasing the consumption volume of those products would lower the consumer's carbon footprint. However, I do not observe a statistically significant change in the emissions per order of consumers in the treated stores compared to those in the control store for either specification. While accounting for prices marginally modifies the point estimates, I cannot reject the null hypothesis in both cases.

One potential explanation for this null result could be attributed to a limitation of the dataset, which is the comparatively lower share of carbon footprint measurements for fruits and vegetables relative to, for example, meat products.

2.7 Conclusion

This paper studies the impact of a large-scale intervention that combines monetary incentives with normative appeals to encourage fruit and vegetable consumption. The monetary incentive corresponds to a price reduction of 0.6% to 2% on fresh fruit and vegetables. The normative appeal labeled the purchase of fruit and vegetable products as "good deeds" to nudge consumers towards making healthier and more environmentally-friendly choices.

The combined monetary incentive and normative appeal increased overall fresh vegetable consumption by, on average, 6-9%, looking at the entire year after the bonus program was implemented. This increase translates to about 390 grams of fresh fruit and vegetables per order. I did not find that this increase resulted from consumers substituting frozen fruit and vegeta-

[19] More information on the measures used, the dataset, and background data, can be found in Bilén (2022).

bles for fresh products. Instead, I find that the increase in fresh fruit and vegetables raised consumers' overall fruit and vegetable consumption. Furthermore, although vegetable consumption began to decline after the bonus program was removed, the effect remained statistically significant for up to two months after the program ended.

The effect is large given the small monetary incentive, compared to previous research findings on nudges or other interventions to promote environmentally friendly behavior (Carlsson et al. 2021). The effect size is also economically meaningful and indicates substantial health impacts. The effect sizes presented in this paper suggest an increased intake of 200-400 grams of fruits and vegetables per order transaction and household. For a two-person household on a weekly basis, this translates to an additional daily intake of 15-30 grams of fruits and vegetables per person. According to estimates in the epidemiology literature (Aune et al. 2017), this increased consumption implies a 1-3% reduction in the relative risk of coronary heart disease, stroke, and cardiovascular disease.[20]

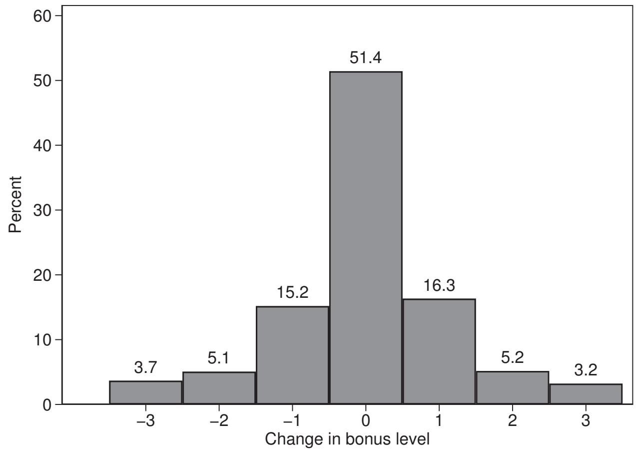
This study's findings have important implications for policymakers and retailers promoting sustainable consumption patterns. The results in this paper provide evidence that combined monetary incentives and normative appeals to promote fruit and vegetable consumption do not crowd each other out, contrary to what has previously been found for daycare fines (Gneezy & Rustichini 2000) and blood donation (Mellström & Johannesson 2008). While not explicitly investigated in this paper, future research could explore whether the monetary incentive provided by the bonus points is potentially more salient and, consequently, more effective than if the monetary incentive were applied directly through a reduction in product prices.

[20] The effect sizes are still economically sizable even if we consider that some fruit and vegetables are not consumed (Schott & Andersson 2015).

2.8 Appendix

2.8.1 Figures

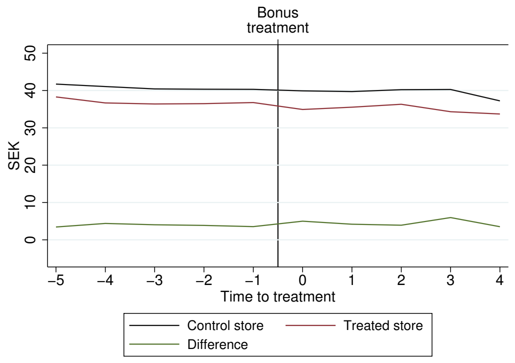
Figure B1: Customer change in bonus levels from period to period



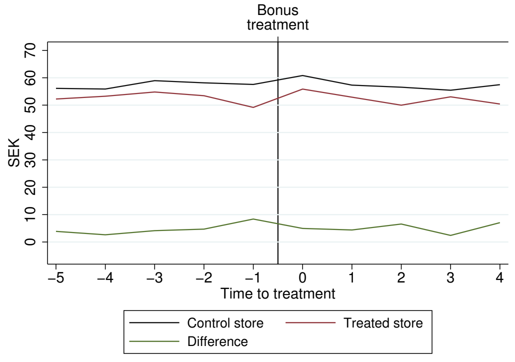
Note: This figure illustrates the transitions of consumers between various bonus levels. A positive value indicates that the consumer was in a lower bonus level during the previous period compared to the current period, while a negative value means that the consumer was in a higher bonus level. Zero means that the consumer remained at the same bonus level as in the previous period.

Figure B2: Average fruit and vegetable prices over time

(a) Fresh



(b) Frozen



Note: This graph contains the average fruit and vegetable price per kg in 30-day intervals from August 2020 to December 2020.

2.8.2 Tables

Table B1: Members in each bonus level over time

Bonus level	1	2	3	4	All
Period					
06-01-20 - 09-30-20	5,855	3,733	1,974	1,855	13,417
10-01-20 - 01-31-21	5,867	4,013	2,221	2,128	14,229
02-01-21 - 05-31-21	3,525	3,506	2,293	2,781	12,105
06-01-21 - 09-30-21	2,423	2,499	1,857	2,528	9,307
10-01-21 - 12-31-21	2,325	2,206	1,381	1,444	7,356
Total	19,995	15,957	9,726	10,736	56,414

Note: This table displays the number of customers in each bonus level in the treated store during every bonus level period.

Table B2: Controlling for prices: Effects of bonus program on fresh fruit and vegetable consumption per order

	Affected by Bonus		**Not affected by Bonus**	
	(1)	(2)	(3)	(4)
	Fresh	Fresh	Frozen	Frozen
Treatment	0.242***	0.220***	-0.027***	-0.029***
	(0.024)	(0.023)	(0.007)	(0.007)
Treatment*Removed		-0.146***		0.016*
		(0.024)		(0.007)
Constant	7.062***	6.933***	1.431***	1.416***
	(0.298)	(0.298)	(0.099)	(0.099)
Treatment + Treatment*Removed		0.074*		-0.014
		(0.034)		(0.010)
Individual FE	Yes	Yes	Yes	Yes
Observations	857,028	942,205	857,028	942,205
Clusters	44,781	44,781	44,781	44,781

Standard errors clustered on the consumer level in parentheses

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Note: This table presents the overall impact of the bonus program on orders containing fresh and frozen fruits and vegetables by the estimating equation (2.1). The difference-in-difference term *Post*Treated* is labeled as *Treatment*. The dependent variable represents the weight (in kg) of fresh or frozen fruits and vegetables per order. The analysis is limited to consumers observed before and after implementing the bonus program in the treated store. All regressions include individual and monthly fixed effects and controls for the average monthly price of fresh and frozen fruit and vegetables per kg, with standard errors clustered at the consumer level. *Treatment + Treatment*Removed* refers to the difference-in-difference estimate comparing the period before the bonus program to the period after the program was removed.

Table B3: Short term: Effects of bonus program on fresh fruit and vegetable consumption per order

	(1) Fresh	(2) Frozen
Treatment	0.245***	-0.026*
	(0.032)	(0.010)
Constant	4.117*	1.119*
	(1.717)	(0.548)
Individual FE	Yes	Yes
Observations	193,416	193,416
Clusters	29,970	29,970

Standard errors clustered on the consumer level in parentheses

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Note: This table presents the overall short-term impact of the bonus program on orders containing fresh and frozen fruits and vegetables by the estimating equation (2.1) from August until December 2020. The dependent variable represents the weight (in kg) of fresh or frozen fruits and vegetables per order. The analysis is limited to consumers observed before and after implementing the bonus program in the treated store. All regressions include individual and monthly fixed effects, with standard errors clustered at the consumer level.

Table B4: Heterogeneity by bonus level: Short term effects of bonus program on (fresh) fruit and vegetable consumption per order

	(1) Level 1	(2) Level 2	(3) Level 3	(4) Level 4	(5) Comparing bonus levels
Treatment	0.432***	0.340***	0.373***	0.264***	
	(0.034)	(0.032)	(0.059)	(0.075)	
Bonus level					-0.041
					(0.036)
Bonus level * Treatment					-0.038
					(0.034)
Constant	3.991***	3.979***	4.053***	4.225***	4.444***
	(0.003)	(0.003)	(0.004)	(0.006)	(0.068)
Individual FE	Yes	Yes	Yes	Yes	Yes
Observations	247,150	259,014	251,467	260,719	148,562
Clusters	28,067	26,726	25,003	24,924	13,752

Standard errors clustered on the consumer level in parentheses

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Note: This table displays the overall effect of the bonus program on fresh fruits and vegetables during the initial four months of the bonus program from estimating equation (2.1) for different bonus levels. The dependent variable indicates the weight (in kg) of fresh fruits and vegetables per order. The treated sample is limited to one bonus level in columns 1-4, compared to all consumers in the control store. In column (5), the sample is restricted to only consumers in the treated store. All regressions include individual and monthly fixed effects, with standard errors clustered at the consumer level.

Table B5: Heterogeneity by age: Effects of bonus program on (fresh) fruit and vegetable consumption per order

	(1)	(2)	(3)	(4)
Age group	< 40	40-60	60+	All
Treatment	0.411***	0.355***	0.370***	0.422***
	(0.0613)	(0.0712)	(0.0753)	(0.0420)
Treatment *(40-60)				-0.0741
				(0.0716)
Treatment *(60+)				-0.00904
				(0.0235)
Individual FE	Yes	Yes	Yes	Yes
Observations	129,740	153,367	131,151	414,258
Consumers	7,167	6,721	5,241	19,129

Standard errors clustered on the consumer level in parentheses

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Note: This table shows the overall response of the bonus program on fresh fruit and vegetables per order from estimating equation (2.1) for different age groups. The dependent variable represents the weight (in kg) of fresh or frozen fruits and vegetables per order. The analysis is limited to consumers observed before and after implementing the bonus program in the treated store. All regressions include individual and monthly fixed effects, with standard errors clustered at the consumer level.

Table B6: Heterogeneity by income: Effects of bonus program on (fresh) fruit and vegetable consumption per order

	(1)	(2)	(3)	(4)
Income group	Low income	Middle income	High income	All
Treatment	0.250***	0.464***	0.367***	0.206***
	(0.035)	(0.054)	(0.051)	(0.031)
Treatment *Middle income				0.267***
				(0.060)
Treatment *High income				0.200***
				(0.057)
Constant	3.874***	4.041***	4.260***	4.052***
	(0.008)	(0.012)	(0.013)	(0.006)
Individual FE	Yes	Yes	Yes	Yes
Observations	293,478	276,748	282,641	856,091
Clusters	14,889	14,996	14,678	44,714

Standard errors clustered on the consumer level in parentheses

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Note: This table shows the overall response of the bonus program on fresh fruit and vegetables per order from estimating equation (2.1) for different income groups. The dependent variable represents the weight (in kg) of fresh or frozen fruits and vegetables per order. The analysis is limited to consumers observed before and after implementing the bonus program in the treated store. All regressions include individual and monthly fixed effects, with standard errors clustered at the consumer level.

Table B7: Effects of bonus on the emissions per order

	(1) Emissions	(2) $\frac{\text{Emissions}}{\text{Weight}}$	(3) Emissions	(4) $\frac{\text{Emissions}}{\text{Weight}}$
Treatment	0.247	-0.005	0.418	0.016
	(0.139)	(0.006)	(0.218)	(0.012)
Constant	35.522***	1.911***	58.396***	2.832***
	(0.033)	(0.001)	(4.151)	(0.236)
Individual FE	Yes	Yes	Yes	Yes
Price dummies	No	No	Yes	Yes
Observations	851,729	851,729	851,729	851,729
Clusters	44,689	44,689	44,689	44,689

Standard errors clustered on the consumer level in parentheses

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Note: This table presents the overall impact of the bonus program on emissions per order by estimating equation (2.1). Emissions per order are measured as kg CO_2e per order, and weight refers to kg of food in order. The analysis is limited to consumers observed before and after implementing the bonus program in the treated store. All regressions include individual and monthly fixed effects, with standard errors clustered at the consumer level. Price dummies refer to the monthly kg price of fresh and frozen fruit and vegetables, beef, vegan and poultry, and pork or pork mixed with beef.

Chapter 3

Are women more generous than men? A meta-analysis

This article is published in Journal of the Economic Science Association (2021)*

Abstract: We perform a meta analysis of gender differences in the standard windfall gains dictator game (DG) by collecting raw data from 53 studies with 117 conditions, giving us 15,016 unique individual observations. We find that women on average give 4 percentage points more than men (Cohen's d=0.16), and that this difference decreases to 3.1% points (Cohen's d=0.13) if we exclude studies where dictators can only give all or nothing. The gender difference is larger if the recipient in the DG is a charity, compared to the standard DG with an anonymous individual as the recipient (a 10.9 versus a 2.3% points gender difference). These effect sizes imply that many individual studies on gender differences are underpowered; the median power in our sample of standard DG studies is only 9% to detect the meta-analytic gender difference at the 5% significance level. Moving forward on this topic, sample sizes should thus be substantially larger than what has been the norm in the past.

*Please cite as Bilén, D., Dreber, A. & Johannesson, M. Are women more generous than men? A meta-analysis. J Econ Sci Assoc 7, 1–18 (2021).

3.1 Introduction

To what extent are there gender differences in altruism or prosocial behavior? The answer to this question could be important for understanding for instance gender differences on the labor market, voting, volunteer work, and charitable giving. In this paper we do a meta-analysis of gender differences in the dictator game (DG; Kahneman et al. 1986, Forsythe et al. 1994). While giving in this game is not necessarily due to altruistic concerns - for example the results of List (2007), Krupka & Weber (2013), Bardsley (2008) and Dana et al. (2006) suggest that DG giving is influenced by the strategy space, reference points and expectations of social norms - this is the most commonly studied game in order to understand non-strategic prosocial behavior. In the standard DG, one individual - the dictator - anonymously decides how to split a windfall endowment with another individual. In an alternative version of the DG, the dictator decides how much of the endowment to give to a charity (charity DG) (Eckel & Grossman 1996).

We include both of these standard windfall gains versions of the DG and collect raw data from 53 papers, both published work and working papers, with a total of 117 conditions and 15,016 unique individual observations. We only include experiments where monetary endowments are windfalls, the most selfish option is to give nothing, the price of giving is equal to 1, both men and women are represented, and where no reciprocity is involved.

There are several previous papers studying gender differences in DG giving (with early papers including e.g. Eckel & Grossman (1998), Bolton & Katok (1995), Andreoni & Vesterlund (2001), and with more recent reviews including e.g. Croson & Gneezy (2009) and Bertrand et al. (2010)), with individual studies typically finding that women on average give more or that there is no statistically significant gender difference. Most related to our study is the extensive meta-analysis of DG giving by Engel (2011). Engel, using reported coefficients rather than raw data, found that women on average gave 5.8 percentage points more than men.

Analyzing raw data allows us to include studies that collected gender information but either did not include gender in their analysis, or included gender but did not explicitly publish any results related to gender in the paper. Including studies where gender is not the main variable of interest could reduce publication bias; as statistically significant gender differences may be more likely to be published. Compared to Engel (2011), we have a substantially larger sample size for estimating the gender difference and we also explicitly compare DGs where the recipient is a person or a charity. The paper is organized in the following way. Section 3.2 describes the inclusion criteria and data. Section 3.3 describes the meta-analysis methods and Section 3.4 the results. Section 3.5 concludes.

3.2 Inclusion criteria and data

3.2.1 Inclusion criteria

Our inclusion criteria are summarized in Table 3.1. We restrict our analysis to the original windfall version of the dictator game (see Forsythe et al. 1994), where the experimenter unconditionally transfers an endowment to dictators and dictators decide how much of the endowment to give to recipient players. This restriction excludes experiments where participants first earned their endowment from performing a task. We also exclude versions of the Take Game, where the dictators also have the option to take money from recipients. Both double-blind (where neither the recipient or the experimenter can identify individual dictator decisions) and single-blind studies (where only the recipient is blind to individual dictator decisions) are included. In the initial stages of the project we had planned to only include double-blind studies, but this was revised when we realized that our sample would be too small (around 70% of our sample consists of single-blind studies). We exclude studies without a monetary endowment, but have no further restriction with respect to the size of the monetary endowment and we also allow conditions where only a randomly drawn share of participants are paid. We exclude studies where participants are matched and known to each other (for example spouses).

There is no participant age restriction, but as we require monetary incentives we exclude studies on young children. Only studies where the price of giving is equal to 1 are included (excluding studies with multipliers). We also limit the inclusion to conditions where there is no reciprocity involved, while we allow individuals to play the DG in both roles.

There is also variation between studies in the choice set of dictators. To give an example, endowments may be 4$ and dictators can give x$ to the recipient, where $x \in \{0, 1, 2, 3, 4\}$. We include all studies where dictators are allowed to give or keep the full endowment, making no further restriction on the choice set. Thus, in the extreme case dictators play an all or nothing game and decide whether to either give or keep the endowment, which is the case in one of our included papers (Tinghög et al. 2016). Finally, we only include studies that have data on gender and where both genders participate (excluding single-gender studies). The search for studies started on the database Econlit with a search for the keyword "Dictator game". This gave us 513 hits and if the studies fitted our inclusion criteria we sent out an email to request the raw data from the corresponding authors. In May of 2018 we also sent out a request to the experimental email group (ESA) describing our sample of included studies and our inclusion criteria, and we closed the inclusion of studies in September 2018.

Defining inclusion and exclusion criteria involves a certain degree of arbitrariness and there is a tradeoff between including studies that use as similar experimental design as possible, which allows for internal validity, and expanding the inclusion criteria to more heterogeneous designs that may increase the statistical power and the generalizability of results. We potentially lose external validity when for instance not including studies varying the price of giving and studies where the dictator earns the endowment (which excludes experiments such as the all or nothing experiment in Bekkers (2007), where the dictator earns the endowment in the experiment). However, we think the standard DG version is a reasonable starting point for analyzing gender differences in the dictator game. Generosity is unconditional, with no involvement of reciprocity, and involves no efficiency gains by changing the size of the endowment by giving and no party has done more (or less) to earn the endowment. We do include charity organizations as recipients which may increase external validity, as donating directly to anonymous individuals as in the standard DG is rare outside of the laboratory. We did not restrict our inclusion criteria to studies specifically designed to study gender differences, but included all studies meeting our inclusion criteria that had collected data on gender. It could be argued that ideally only studies designed to study gender differences should be included

Table 3.1: Exclusion criteria used in the meta-analysis.

Domain	Inclusion criteria
Source of Endowment	Windfall
Anonymity	Single-blind or Double-blind
Recipient	Anonymous individual or Charity organisation
Price of giving	1
Experimental setting	No restriction
	(Lab, Field, Online etc)
Reciprocity	No
Endowment	Monetary
Payment scheme	Deterministic or random
Choice set	The dictator can keep (give 0) or give the full endowment
Age of participant	No restriction
Genders represented in experiment	Both

as they may be designed to eliminate confounding experimental designs that could influence the gender gap. However, it is not straightforward to define which studies were ex ante designed to study gender differences, and the observable "gender studies" may be published because they found gender differences and not due to their superior designs leading to publication bias. In an attempt to test if observable gender studies differ, we compare results for studies having gender in the title of the paper to the other studies in our data. An additional potential limitation of our data collection is that we did not explicitly search for discussion papers, which may induce publication bias. However, we test for publication bias to assess the importance of this limitation. The final sample consists of 15,016 unique observations where we also have gender data.

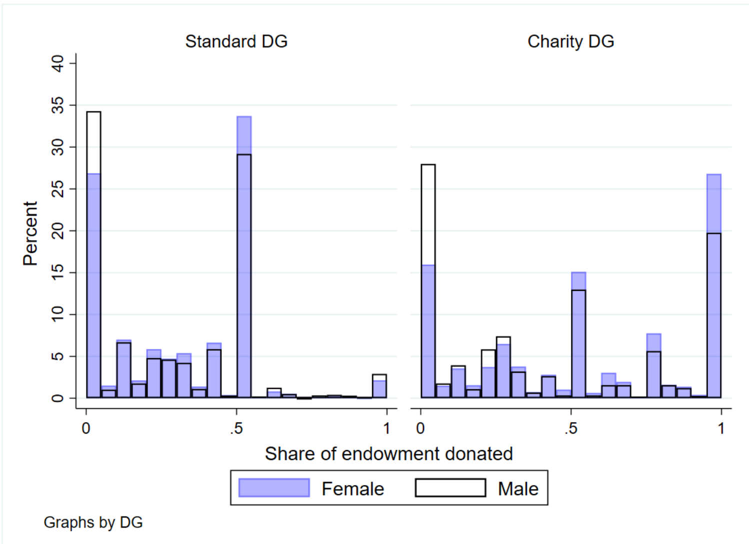
3.2.2 Data

In Table 3.2 we present summary statistics of the included data. The full sample consists of 15,122 observations but for some observations either gender or the dictator's decision is missing or have been incorrectly coded. Excluding those observations gives us a sample of 15,016 unique individual observations where we have both gender and the donation decision by the subject.[2] In Figure 3.1 we plot the distributions of DG giving for each

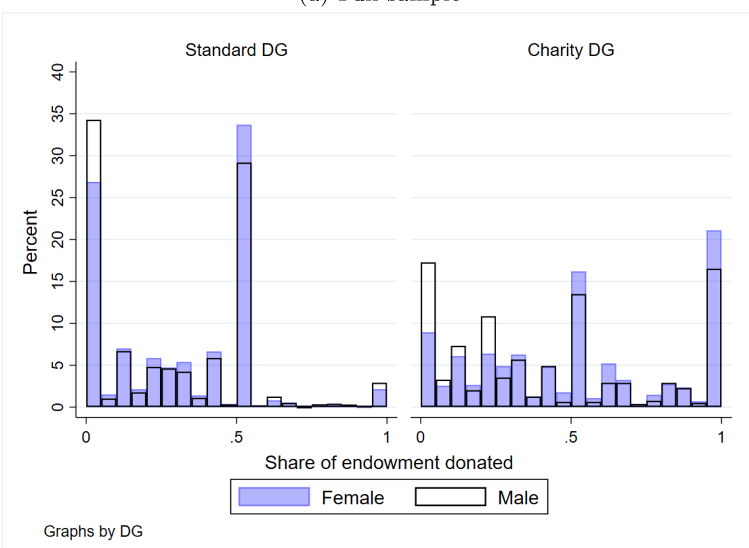
[2]If an experiment consists of several independent rounds we calculate the average across all rounds for each participant so that we only have one unique observation per individual in the data set.

gender and by recipient type. As previously shown by e.g. Engel (2011), the distribution of DG giving in the standard DG is concentrated at giving nothing or half. In the charity DG a substantial fraction of subjects also give the entire endowment. When we exclude the 'all or nothing' study the density on the two extreme points of the distribution decreases somewhat for the charity DG.

Figure 3.1: The full sample contains 15,016 unique individual dictator decisions. There are 11,802 observations in the standard DG and 3,214 observations in the charity DG. Excluding the 'all or nothing' study reduces the sample size in charity DG to 1,812 observations.



(a) Full sample



(b) Excluding 'all or nothing' study

Table 3.2: Descriptive statistics of the data included in the meta-analysis.

	N	mean	sd	min	max
Share donated (Full sample)	15,085	.3245847	.2927535	0	1
Share donated (Standard DG)	11,829	.2818374	.2487024	0	1
Share donated (Charity DG)	3,256	.4798851	.3763469	0	1
1 if Female	15,050	.4942193	.4999832	0	1
Obs with gender and donation data	15,016				
Papers	53				
Conditions	117				
Year of paper		2014	3.8	1998	2018
Partition (the smallest share a dictator can donate)	15,122	.156878	.2751402	.0000667	1
1 if Student sample	15,122	.4427986	.4967336	0	1
1 if Econ student	15,122	.0675837	.2510384	0	1
1 if Random payment scheme	15,122	.2526782	.4345623	0	1
1 if Double-blind	15,122	.3166248	.4651751	0	1
Experiment setting					
Field	968				
Lab	7,890				
Online	5,375				
Mturk	3,211				
Phone	889				
Age					
No info	4,521				
<15	229				
15-30	5911				
31-40	1540				
41-50	1166				
50-60	973				
60+	782				
Region					
Europe	7,844				
Nordic	4,759				
N America	4,847				
US	4,831				
Asia	882				
Africa	934				
S America	151				
Oceania	454				

3.3 Meta-analysis methods

By collecting the individual participation data, a meta-analysis can either be done by the traditional approach of pooling effect sizes in a random effects model or with individual regression models. As Burke et al. (2017) note, these methods in general produce similar results and differences largely occur when researchers use different modelling assumptions. We begin by performing traditional random-effects meta-analysis which allow us to estimate the heterogeneity in the gender difference across the conditions included in the meta-analysis (the heterogeneity is captured by the estimate of Tau, which is the standard deviation in the true effect size across the conditions).[3] We then estimate one stage individual regression models where we also provide several robustness checks of our results. Several experiments contain multiple conditions. These conditions create a natural clustering of the individual observations in our data. In total we have 117 conditions from 53 papers, which allows us to treat each condition within an experiment as a separate cluster. We estimate a separate effect size for each condition in the random-effects model, and we cluster the standard errors on the condition level in all individual regression models.[4] Our definition of a condition follows the definition within each paper. If the same condition within an experiment is conducted in different countries (except if the study is done online (MTurk)), we define these as separate conditions to account for the country level clustering of the experiments. We estimate equation (3.1) where S_{ij} denotes the share of the endowment donated by participant i in condition j, X is a vector of individual covariates and Z is a vector of treatment condition controls. We also replace Z with a condition fixed effect using dummy variables for each condition.[5] The gender coefficient β_1 and the interaction between gender and charity recipient β_3 are the coefficients of interest,

$$\begin{aligned} S_{ij} = \beta_0 + \beta_1 Female + \beta_2 Charity + \beta_3 (Female \star Charity) \\ + \beta_4 X_{ij} + \beta_5 Z_j + \epsilon_{ij}. \end{aligned} \tag{3.1}$$

3.4 Results

We use a significance threshold of $p < 0.005$ for "statistically significant evidence" and a threshold of $p < 0.05$ for "suggestive evidence" in our results below in line with the recent recommendation of Benjamin et al. (2018).

[3]The estimation is done by the Ipdmetan-command by Fischer in Stata, with the estimation of the between study variance τ^2 by DerSimonian and Laird (Fisher (2015).

[4]An alternative strategy would be to cluster on the paper level. However, there are only 12 papers where a charity is the recipient, making this clustering unfeasible for us.

[5]As the recipient type does not vary within a condition, the binary variable for the charity DG is already captured by the condition fixed effects. However, as gender varies within conditions we can still estimate the interaction effect between gender and charity DG.

All our tests are two-sided.

3.4.1 Random-effects meta-analysis

Figure 3.2 shows a forest plot of the estimated gender gap for each of the 117 conditions in our sample; and the random effects results are also reported in Appendix Table C1. We show the results both separately for the standard DG and the charity DG, and pooled for both DG versions. Women give on average 4 percentage points more than men and the gender gap is statistically significant. The average donation in our data is 32 percent of the endowment (see Table 3.2) and hence the gender gap is more than 10 percent of the average donation, and women on average give 13 percent more than men (Cohen's d=0.16)[6]. The standard deviation in the true effect size - the variation between studies over and above sampling variation - is slightly higher than the average effect size at $\hat{\tau} = 4.6$ percentage points. To further assess heterogeneity in the gender gap, we estimate the gender gap in the standard DG and the charity DG respectively.

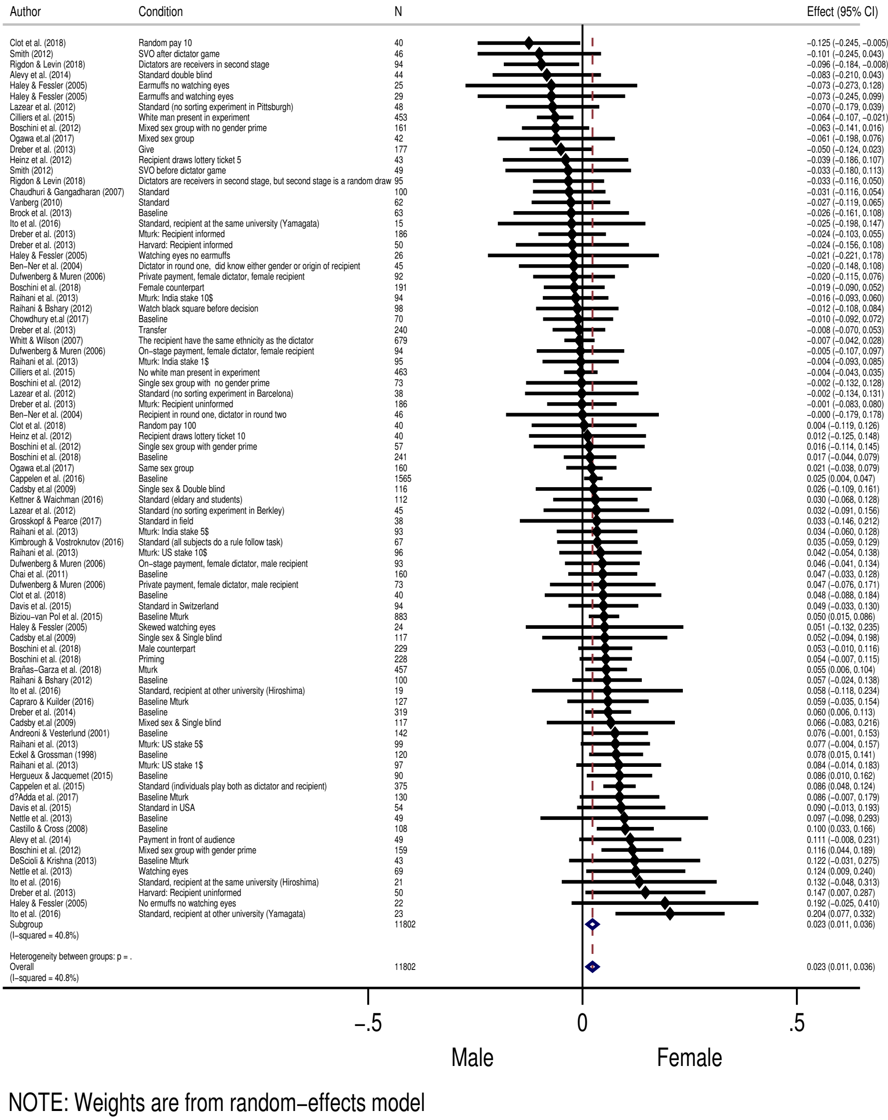
In the standard DG where the recipient is another participant, women on average donate 2.3 percentage points more than men, with a standard deviation in the true effect size of $\hat{\tau}=$ 3.3 percentage points. In the charity DG the gender gap is larger, with women on average giving 10.9 percentage points more than men with a standard deviation in the true effect size of $\hat{\tau}=$ 6.4 percentage points. A meta-regression in Table C2 confirms that there is a statistically significant difference in the gender gap between the standard DG and the charity DG. The gender gaps of 2.3 percentage units in the standard DG and 10.9 percentage units in the charity DG imply that women give 9 percent more than men in the standard DG (Cohen's d=0.10) and 26 percent more than men in the charity DG (Cohen's d=0.35).

Excluding the 'all or nothing' conditions reduces the gender gap found in the overall sample from 4 to 3.1 percentage points (Cohen's d=0.13), but

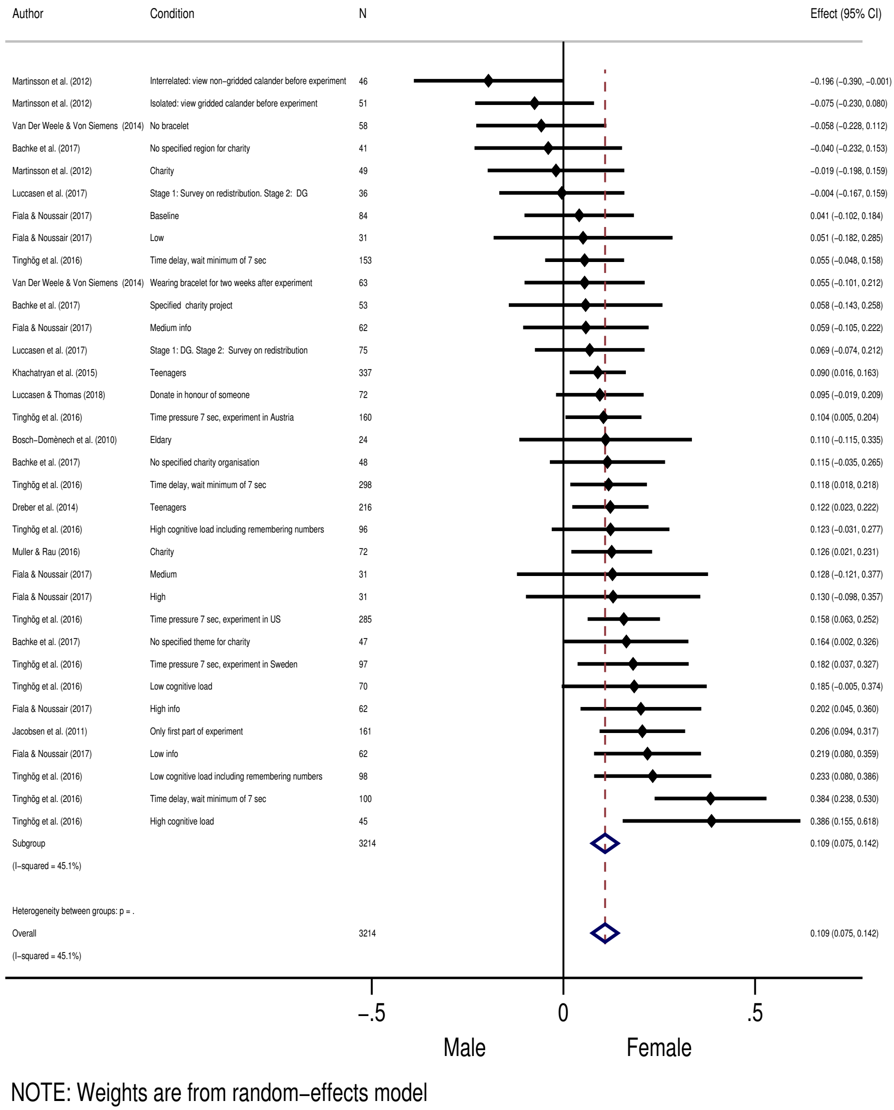
[6]The average donation for men in the whole sample is 30 percent of the endowment, 27 percent in the standard DG and 42 percent in the charity DG. We compare the gender gap in percentage terms with respect to the average donation for men. To convert effect sizes to Cohen's d (the effect size as a fraction of the STD), we calculate the STD for each paper and take the average STD across papers. The STD in the whole sample is 0.250 (0.247 if we exclude 'all or nothing' conditions), while in the standard DG it is 0.233 and in the charity DG it is 0.310 (0.300 if we exclude 'all or nothing' conditions).

it remains statistically significant with heterogeneity in the true effect size of $\hat{\tau} = 3.8$ percentage points. In the charity DG the gender gap is reduced from 10.9 to 8 percentage points (Cohen's d=0.27), with heterogeneity $\hat{\tau} = 4.7$ percentage points. The smaller gender gap in the charity DG decreases the meta-regression estimate of the difference in the gender gap between the standard DG and the charity DG from 8.7 to 5.9 percentage points.

Figure 3.2: Random effects model (estimated with the Ipdmetan command in Stata). Figure 3.2a contains experiments with the standard DG and Figure 3.2b contains experiments with the charity DG. The blue diamonds indicate the estimated effect size (and the CI) for each DG.



(a) Standard DG



(b) Charity DG

3.4.2 Individual level regression analysis

In Table 3.3 we report the results of the individual level regression analysis. In column 1, where we only include a binary variable for the gender of the dictator, women give on average 4.8 percentage points more than men, which is similar to the gender gap of 5.8 percentage points reported in Engel (2011). Controlling for the charity DG in column 2 gives an overall gender gap of 4 percentage points, which is identical to the gender gap found with the Random-effects meta-analysis.[7] In column 3-6 we add an interaction between the female variable and the charity DG. The gender gap is statistically significantly larger in the charity DG compared to the standard DG in all four specifications, with an interaction coefficient of between 9.3 and 9.8 percentage points. Women give on average around 2 percentage points more than men in the standard DG and 11-12 percentage points more in the charity DG. These gender gaps are statistically significant for both types of DG in all the four regression models, except for the standard DG in column 3 where there is suggestive evidence of a gender difference (p=.0058). If we exclude the 'all or nothing' conditions, the results are similar (see Table 3.4)

[7] The gender gap in our data is around 1 percentage point lower in Double blind studies but this difference is not statistically significant (p =565).

Table 3.3: OLS results of the estimated gender gap in the DG. Standard errors clustered on the condition level in parentheses.

	(1) Share	(2) Share	(3) Share	(4) Share	(5) Share	(6) Share
Female	0.048***	0.040***	0.020*	0.020***	0.023***	0.021***
	(0.010)	(0.008)	(0.007)	(0.006)	(0.006)	(0.006)
Charity DG		0.194***	0.144***	0.141***		
		(0.029)	(0.029)	(0.035)		
Charity DG * Female			0.094***	0.093***	0.097***	0.098***
			(0.018)	(0.017)	(0.016)	(0.016)
Constant	0.300***	0.263***	0.272***	0.463***	0.302***	0.358***
	(0.012)	(0.012)	(0.011)	(0.050)	(0.007)	(0.024)
Condition fixed effects	No	No	No	No	Yes	Yes
Individual controls[a]	No	No	No	Yes	No	Yes
Treatment controls[b]	No	No	No	Yes	No	No
Female +(Charity DG * Female)			0.113***	0.114***	0.120***	0.119***
			(0.016)	(0.015)	(0.015)	(0.014)
Observations	15,016	15,016	15,016	15,016	15,016	15,016
Number of conditions	117	117	117	117	117	117

[a] Individual controls: Student characteristics, age and region.
[b] Treatment controls: Double-blind, setting characteristics, random payment and partitioning of endowment.
* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Table 3.4: OLS results of the gender difference in the DG, excluding the "all or nothing" DG study. Standard errors clustered on the condition level in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)
	Share	Share	Share	Share	Share	Share
Female	0.035***	0.028***	0.020*	0.021***	0.023***	0.021***
	(0.009)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Charity DG		0.184***	0.147***	0.137***		
		(0.036)	(0.034)	(0.035)		
Charity DG * Female			0.066**	0.067**	0.065***	0.066***
			(0.022)	(0.020)	(0.018)	(0.017)
Constant	0.289***	0.268***	0.272***	0.496***	0.317***	0.370***
	(0.012)	(0.011)	(0.011)	(0.046)	(0.007)	(0.025)
Condition fixed effects	No	No	No	No	Yes	Yes
Individual controls[a]	No	No	No	Yes	No	Yes
Treatment controls[b]	No	No	No	Yes	No	No
Female + (Charity DG * Female)			0.086***	0.087***	0.088***	0.087***
			(0.021)	(0.019)	(0.017)	(0.016)
Observations	13,614	13,614	13,614	13,614	13,614	13,614
Number of conditions	107	107	107	107	107	107

Individual controls: Student characteristics, age and region.
Treatment controls: Double-blind, setting characteristics, random payment and partitioning of endowment.
* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Robustness checks

In a robustness test, we estimate a mixed random effects model where we allow for both the intercept and the gender gap to have random effects on the condition level.[8] These results are reported in tables C3 and C4. We find very similar results in these estimations, although the pooled gender gap of 4.6 percentage points is slightly higher. The gender gap in the standard DG is around 2 percentage points and the gender gap in the charity DG is 11 percentage points and significantly higher than in the standard DG. The gender gap is statistically significant in both the standard DG and the

[8]We thank the anonymous reviewers for suggesting this robustness check and the robustness check below using the tobit model.

charity DG, both with and without the ‘all or nothing’ conditions included.

In a second robustness test, we re-estimate our results using a tobit model. The action space is limited to a donation between 0 and 1 even though some subjects may possibly prefer to take from the recipient or give more than the endowment. In the tobit model we allow censoring to occur at both 0 and 1. We report these results in tables C5 and C6, and the coefficients in these tables should be interpreted as the gender gap with respect to the latent (that in theory can take on both negative values and values above 1) dependent variable. The tobit model yields higher estimates of the gender gap, with a gender gap of 6.6 percentage points in the pooled sample. The gender gap is around 4 percentage points in the standard DG and 17 percentage points in the charity DG, and this difference is statistically significant. The gender gap is statistically significant in both the standard DG and the charity DG, both with and without the ‘all or nothing’ conditions included. As can be seen from Figure 3.1, men are more likely to donate zero in both the standard DG and the charity DG and women are more likely to donate the full endowment in the charity DG. These differences at the censoring points of 0 and 1 result in a higher estimated gender gap in the tobit model when these observations are interpreted as being censored.

So far we have measured the gender gap as the difference in the share of the endowment donated. As both men and women donate more in the charity DG it is possible that the gender gap is larger in percentage points but not in terms of percentage of the average donation. We test this in an additional robustness check where we instead use the measure from equation (3.2) below, where we divide the individual share donated in the DG (s_ij) by individual i that took part in condition j by the average donation in condition j,

$$\hat{s}_{ij} = \frac{s_{ij}}{\overline{s_j}}. \tag{3.2}$$

Multiplying this measure with 100 allows us to interpret each observation in terms of percent of the average donation within the condition that the participant took part.

Our results are confirmed by this standardization when we include all the DG studies in Table C7. Women on average give around 9 percent more than men in the standard DG and 25 percent more in the charity DG; and this difference is statistically significant. In Table C8 we carry out this analysis excluding the "all or nothing " DG study. This reduces the gender gap in the charity DG with one fifth to around 20 percent, and the gender gap is not statistically significantly larger in the charity DG any more (but there is suggestive evidence for a larger gender gap in the charity DG in all models). When we measure donations in relative terms the evidence of a larger gender difference in the charity DG is thus less strong, as the donations are larger on average in the charity DG compared to the standard DG.

3.4.3 Gender in the title of the studies

Of the 53 (117) papers (conditions) included in our study 16 (31) have gender in the title of the paper, and we test if the gender difference differs between papers with and without gender in the title.[9] If we observe such a difference we cannot tell if this is due to that studies explicitly designed to study gender differences lead to different results, or if the difference is due to that studies with gender in the title were framed as studies of gender differences and published because they found a significant gender difference. We return to the issue of publication bias in that section below.

A meta-regression in Table C9 provides no evidence that the gender gap differs between papers with gender in the title and the other DG studies. We also estimate equation (3.1) in an OLS model in Table C10, where we include a dummy for "gender in the title", that is interacted with the female variable to test if the gender gap is significantly larger in papers with gender in the title. The null results from the meta-regression are confirmed in the OLS model. In Table C10 we also report the estimated gender difference among the subset of papers with gender in the title (it is the sum of the gender coefficient and the interaction coefficient) and it is between 3.8 and 3.9 percentage points when we include both types of DGs, and between 2.6 and 2.9 percentage points for the standard DG and between 10 and 10.8 percentage points for the charity DG. The gender difference is statistically significant in all models except in model 3 where there is suggestive evidence.

[9] With "gender in the title", we do not explicitly mean the word "gender", so we include Eckel & Grossman (1998) that mention "women" and "men" in the title.

3.4.4 Statistical power

In Table C11 we summarize the statistical power to detect the gender gap found in this meta-analysis. The power estimates are based on estimating the mean difference between males and females using two-sided hypothesis testing, and they are based on the average standard deviation of the standard DG papers (STD= 0.233) and the charity DG papers (STD= 0.310). We use the random effects results of a gender difference of 2.3 percentage units in the standard DG and 10.9 percentage units in the charity DG to estimate power in each DG type; but we also report power for the overall gender difference of 4 percentage units for both types of DGs. We do the power calculations for tests at the 5% level, as that is most commonly used in the literature. But we also report results for the more stringent 0.5% threshold used in this paper.

We calculate the sample size as the total number of observations in a paper, which means that we sum over all conditions within the paper. The median sample size in the standard DG papers is 130 observations, which yields a statistical power of only 9% (16%) to detect an effect size of 2.3 (4) percentage points. To reach 80 percent power a paper would need around 3,224 (1,068) observations to detect an effect size of 2.3 (4) percentage points in the standard DG. The median sample size for the charity DG papers is 192 observations, yielding 68% (14%) power to detect an effect size of 10.9 (4) percentage points. All the above estimates are based on tests at the 5% level, and using the more stringent 0.5% threshold leads to even lower power (see Table C11).

Some of the included DG studies were not designed to study gender differences, which may explain the inadequate power. However, the power for studies that have gender in the title of the paper are only slightly higher. There are 13 papers in the standard DG and three papers in the charity DG that have gender in the title of the paper. The median sample size of the 13 standard DG studies is 191, which gives 10% (22%) statistical power to detect a 2.3 (4) percentage units gender difference. The median sample size of the three charity DG studies is 216, which gives 73% (16%) statistical power to detect a 10.9 (4) percentage units gender difference. Our analysis here mainly points to the sample sizes needed for high-powered future research on this topic.

3.4.5 Publication bias

Figure C1 shows funnel plots for the full sample, the standard DG sample, and the charity DG sample with the estimated effect sizes on the x-axis and the corresponding standard errors on the y-axis. An asymmetric plot could be evidence of publication bias, where only significant studies are published. The outliers at the far right in Figure C1 (a) and C1 (c) are two of the 'all or nothing' conditions. The funnel plots in Figure C1 do not provide any clear visual evidence of publication bias. In Figure C2 we restrict the funnel plots to papers with gender in the paper's title. There is no clear visual evidence of any asymmetry in these plots either.[10]

To statistically test if there is evidence of publication bias we have employed Egger's and Begg's tests of publication bias in Table C12. We find no evidence of publication bias in either the pooled sample or when looking at each dictator game separately. We furthermore carry out these tests including only papers with gender in the title, but we do not find a statistically significant publication bias in these tests either.

3.5 Discussion

Our results suggest that women give more than men on average in both the standard and the charity DG, but the gender gap is modest in size (4 percentage points in the pooled data and a Cohen's d of 0.16). This is similar to the gender gap in the Engel meta-analysis Engel (2011). Looking at the standard DG and the charity DG separately, we find that the gender gap is 2.3 percentage points (Cohen's d=0.10) in the former and 10.9 percentage points (Cohen's d=0.35) in the latter (and if we exclude the "all or nothing study" this gender gap decreases to 8 percentage units; Cohen's d=0.27). It is interesting to compare these results to the recent study by Falk et al. (2018), measuring economic preferences in a global preference survey with a sample size of about 80,000 individuals. They measured al-

[10]For the charity DG there are only three papers with gender in the title making it impossible to test for publication bias, but for completeness we show a funnel plot of these three observations as well in Figure C2.

truism by combining the answers to two survey questions. One of these was a hypothetical donation question similar to the charity DG and the other question measured the willingness to give to good causes on an 11-point scale. They found significantly higher altruism for women than for men, with an estimated gender difference of 0.10 Cohen's d units. Our findings are consistent with those of Falk et al. (2018), with a similar effect size if we pool our data for the standard and charity DG. For the charity DG we find a larger effect size than in Falk et al. (2018), but this effect size is also less precisely estimated in our study.[11]

The estimated gender differences in our study implies that the typical DG study in the literature is underpowered to test for gender differences. With the median sample size being 130 observations in the standard DG in our sample, this leads to a median power of only 9% to detect a 2.3 percentage point difference, while for the charity DG the equivalent numbers are 192 observations and 68% power to detect an 10.9 percentage point difference (for tests at the commonly used 5% level; using the more stringent 0.5% level would decrease power further). Power problems have previously been reported in economics in general (Ioannidis et al. (2017)) as well as for the DG (Ortmann & Zhang (2015)). However, for some of the datasets included in the meta-analysis, the researchers may never have had the intention to study gender differences and may have been well powered to study their main research question. Our power results should thus mainly guide future research that aims to explore gender differences in DG giving and moving forward researchers may need substantially larger sample sizes than what has previously been the norm.

There are also several additional caveats to our conclusions. First, it is not clear whether important datasets are missing from our analysis, and whether the inclusion of these would change any of our conclusions. Publication bias may lead to inflated effect sizes in meta-analysis, which was also observed in a recent study Kvarven et al. (2020) comparing meta-analyses to pre-registered multiple-laboratory replication projects. In testing for publi-

[11] Falk et al. (2018) do not report gender differences separately for the hypothetical donation question that is similar to the charity DG and the 11-point scale donation question. We therefore cannot compare the results of the real and hypothetical charity DG questions across our studies (but only compare our results to their combined altruism measure).

cation bias we found no strong evidence of publication bias, but there may still be publication bias. We find substantial heterogeneity in the gender gap between conditions. In our paper we explore heterogeneity with respect to if the recipient is a charity organization or another participant, but there is much more work that can be done in this regard. One should also be careful to generalize our findings to also hold in other designs such as when the price of giving varies or when the dictator earns the endowment. For example, Andreoni & Vesterlund (2001) report gender differences to be conditional on the price of giving.

Our results suggest a larger gender difference in the charity DG compared to the standard DG, although the strength of this evidence depends on if the "all or nothing" charity DG study is included or not and if the difference is measured in absolute or relative terms (as the average donations are higher in the charity DG). To draw strong conclusions about if the gender difference is larger in the charity DG than the standard DG, it would be interesting to conduct a well-powered study to directly compare the gender difference in these two versions of the DG. A possible explanation for a larger gender difference in the charity DG could be that the charity DG is more closely related to empathy and altruism, whereas the standard DG is more related to fairness preferences (deviating from the 50/50 norm). In the standard DG it is unusual to observe donations over 50% of the endowment, whereas donating 100% of the endowment is relatively common in the charity DG. Altruism as a motivation for donations is consistent with a stronger tendency for such corner solutions of donating all or nothing. Further work is needed to better understand if the two types of DGs measure different forms of social preferences.

3.6 Appendix

Online Appendix: Tables and Figures

This document contains additional tables and figures to **Are women more generous than men? A meta-analysis** by David Bilén, Anna Dreber and Magnus Johannesson.

Table C1: Gender differences in the DG estimated by the random effects model. Results are shown both for the data pooled across all DG studies and separately for the standard DG and the charity DG. Standard errors in parentheses.

	(1)	(2)	(3)	(4)	(5)
				All or nothing DG study excluded	
	Pooled	Standard DG	Charity DG	Pooled	Charity DG
Female	0.04***	0.023***	0.109***	0.031***	0.080***
	(0.007)	(0.006)	(0.017)	(0.006)	(0.018)
$\hat{\tau}$	0.046	0.033	0.064	0.038	0.047
Conditions	117	83	34	107	24

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Table C2: Meta-regression results of the difference in the gender gap between the charity DG and the standard DG (the between study variance is estimated by method of moments and without Knapp-Hartung modifications). Standard errors in parentheses.

	(1) Full sample	(2) All or nothing DG study excluded
Charity DG	0.087***	0.059**
	(0.016)	(0.018)
Constant	0.023***	0.023***
	(0.007)	(0.006)
Observations	117	107
$\hat{\tau}$	0.038	0.034
Conditions	117	107

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Table C3: Mixed random effects results of gender differences in the DG. Each model includes a random intercept for each condition and a random slope for the gender gap in each condition. Standard errors clustered on the condition level in parentheses and the co-variance between random effects is unstructured.

	(1) Share	(2) Share	(3) Share	(4) Share
Female	0.047***	0.046***	0.022***	0.020***
	(0.008)	(0.008)	(0.006)	(0.006)
Charity DG		0.168***	0.137***	0.088**
		(0.025)	(0.025)	(0.028)
Charity DG ∗ Female			0.094***	0.095***
			(0.018)	(0.018)
Constant	0.300***	0.252***	0.259***	0.404***
	(0.011)	(0.009)	(0.008)	(0.052)
Condition random effects	Yes	Yes	Yes	Yes
Individual controls[a]	No	No	No	Yes
Treatment controls[b]	No	No	No	Yes
Female + (Charity DG ∗ Female)			0.115***	0.115***
			(0.016)	(0.016)
Condition	117	117	117	117
Observations	15016	15016	15016	15016

[a] Individual controls: Student characteristics, age and region.
[b] Treatment controls: Double-blind, setting characteristics, random payment and partitioning of endowment.
* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Table C4: Mixed random effects results of gender differences in the DG, excluding the all or nothing study. Each model includes a random intercept for each condition and a random slope for the gender gap in each condition. Standard errors clustered on the condition level in parentheses and the co-variance between random effects is unstructured.

	(1) Share	(2) Share	(3) Share	(4) Share
Female	0.033***	0.033***	0.022***	0.020***
	(0.007)	(0.007)	(0.006)	(0.006)
Charity DG		0.159***	0.127***	0.107***
		(0.028)	(0.029)	(0.027)
Charity DG * Female			0.060**	0.062***
			(0.019)	(0.019)
Constant	0.288***	0.254***	0.259***	0.406***
	(0.010)	(0.008)	(0.008)	(0.053)
Condition random effects	Yes	Yes	Yes	Yes
Individual controls	No	No	No	Yes
Treatment controls	No	No	No	Yes
Female + (Charity DG * Female)			0.083***	0.082***
			(0.018)	(0.018)
Conditions	107	107	107	107
Observations	13614	13614	13614	13614

[a] Individual controls: Student characteristics, age and region.
[b] Treatment controls: Double-blind, setting characteristics, random payment and partitioning of endowment.
* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Table C5: Tobit results of the estimated gender gap in the DG. We model censoring of the donated share donated both from below at 0 and above at 1. Standard errors clustered on the condition level in parentheses.

	(1) Share	(2) Share	(3) Share	(4) Share	(5) Share	(6) Share
Female	0.078***	0.066***	0.039**	0.040***	0.044***	0.041***
	(0.017)	(0.014)	(0.012)	(0.011)	(0.011)	(0.011)
Charity DG		0.267***	0.201***	0.214***		
		(0.043)	(0.045)	(0.051)		
Charity DG * Female			0.124***	0.123***	0.130***	0.130***
			(0.027)	(0.026)	(0.025)	(0.024)
Constant	0.227***	0.179***	0.192***	0.454***	0.239***	0.294***
	(0.020)	(0.024)	(0.023)	(0.077)	(0.012)	(0.036)
Condition fixed effects	No	No	No	No	Yes	Yes
Individual controls[a]	No	No	No	Yes	No	Yes
Treatment controls[b]	No	No	No	Yes	No	No
Female +(Charity DG * Female)			0.164***	0.164***	0.174***	0.172***
			(0.025)	(0.024)	(0.023)	(0.022)
Observations	15016	15016	15016	15016	15016	15016
Number of conditions	117	117	117	117	117	117

[a] Individual controls: Student characteristics, age and region.
[b] Treatment controls: Double-blind, setting characteristics, random payment and partitioning of endowment.
* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Table C6: Tobit results of the estimated gender gap in the DG, excluding the all or nothing study. We model censoring of the donated share donated both from below at 0 and above at 1. Standard errors clustered on the condition level in parentheses.

	(1) Share	(2) Share	(3) Share	(4) Share	(5) Share	(6) Share
Female	0.058***	0.047***	0.037**	0.039***	0.042***	0.039***
	(0.014)	(0.011)	(0.012)	(0.010)	(0.011)	(0.010)
Charity DG		0.256***	0.213***	0.191***		
		(0.048)	(0.047)	(0.047)		
Charity DG*Female			0.077*	0.079**	0.076**	0.077***
			(0.030)	(0.027)	(0.025)	(0.023)
Constant	0.224***	0.197***	0.202***	0.545***	0.269***	0.320***
	(0.020)	(0.021)	(0.021)	(0.064)	(0.011)	(0.036)
Condition fixed effects	No	No	No	No	Yes	Yes
Individual controls[a]	No	No	No	Yes	No	Yes
Treatment controls[b]	No	No	No	Yes	No	No
Female +(Charity DG * Female)			0.114***	0.118***	0.118***	0.116***
			(0.028)	(0.025)	(0.023)	(0.021)
Observations	13614	13614	13614	13614	13614	13614
Number of conditions	107	107	107	107	107	107

[a] Individual controls: Student characteristics, age and region.
[b] Treatment controls: Double-blind, setting characteristics, random payment and partitioning of endowment.
* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Table C7: OLS results of the estimated gender gap in the DG. The dependent variable is the share of the endowment donated in the DG divided by the average donation within the condition the participant took part. We multiply this measure with 100 to interpret the results in terms of percentage of the average donation within a study condition. Standard errors clustered on the condition level in parentheses.

	(1) Percent	(2) Percent	(3) Percent	(4) Percent	(5) Percent	(6) Percent
Female	12.279***	12.307***	8.878***	8.147**	9.089***	8.380**
	(2.245)	(2.253)	(2.547)	(2.461)	(2.626)	(2.517)
Charity DG		-0.689*	-9.203***	-9.257***		
		(0.272)	(1.961)	(2.600)		
Charity DG * Female			16.094***	15.930***	16.697***	16.916***
			(3.822)	(3.847)	(3.950)	(3.850)
Constant	93.895***	94.029***	95.679***	97.735***	88.642***	108.237***
	(1.114)	(1.075)	(1.220)	(2.541)	(1.300)	(6.711)
Condition fixed effects	No	No	No	No	Yes	Yes
Individual controls[a]	No	No	No	Yes	No	Yes
Treatment controls[b]	No	No	No	Yes	No	No
Female + (Charity DG * Female)			24.972***	24.076***	25.786***	25.296***
			(2.849)	(2.881)	(2.951)	(2.853)
Observations	15016	15016	15016	15016	15016	15016
Number of conditions	117	117	117	117	117	117

[a] Individual controls: Student characteristics, age and region.
[b] Treatment controls: Double-blind, setting characteristics, random payment and partitioning of endowment.
* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Table C8: OLS results of the estimated gender gap in the DG, excluding the "all or nothing" DG study. The dependent variable is the share of the endowment donated in the DG divided by the average donation within the condition the participant took part. We multiply this measure with 100 to interpret the results in terms of percentage of the average donation within a study condition. Standard errors clustered on the condition level in parentheses.

	(1) Percent	(2) Percent	(3) Percent	(4) Percent	(5) Percent	(6) Percent
Female	10.259***	10.289***	8.878***	8.248**	9.089***	8.391**
	(2.298)	(2.307)	(2.548)	(2.466)	(2.627)	(2.513)
Charity DG		-0.795*	-6.715*	-6.175*		
		(0.284)	(2.368)	(2.611)		
Charity DG * Female			10.732*	11.159*	11.066*	11.662*
			(4.350)	(4.274)	(4.465)	(4.288)
Constant	94.909***	95.000***	95.679***	104.847***	91.122***	109.688***
	(1.129)	(1.102)	(1.220)	(4.293)	(1.591)	(6.783)
Condition fixed effects	No	No	No	No	Yes	Yes
Individual controls[a]	No	No	No	Yes	No	Yes
Treatment controls[b]	No	No	No	Yes	No	No
Female + (Charity DG * Female)			19.610***	19.407***	20.155***	20.053***
			(3.525)	3.473)	(3.611)	(3.417)
Observations	13614	13614	13614	13614	13614	13614
Number of conditions	107	107	107	107	107	107

[a] Individual controls: Student characteristics, age and region.
[b] Treatment controls: Double-blind, setting characteristics, random payment and partitioning of endowment.
* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Table C9: Meta-regression results of the difference in the gender gap between conditions that had gender in the title of the paper and those that did not (the between study variance is estimated by method of moments and without Knapp-Hartung modifcations). Standard errors in parentheses.

	(1) Pooled	(2) Standard DG	(3) Charity DG
Gender in title	-0.013	0.002	0.003
	(0.014)	0.013)	(0.051)
Constant	0.045***	0.023**	0.108***
	(0.008)	(0.008)	(0.019)
Observations	117	83	34

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Table C10: OLS results of differences in the estimated gender gap in conditions that either had or did not have gender in the title of the paper. Standard errors clustered on the condition level in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled	Pooled	Standard DG	Standard DG	Charity DG	Charity DG
Female	0.041***	0.041***	0.019*	0.018*	0.118***	0.114***
	(0.010)	(0.009)	(0.009)	(0.008)	(0.019)	(0.018)
Gender in title	-0.017	-0.008	-0.037	-0.050*	0.067	0.264
	(0.020)	(0.023)	(0.019)	(0.019)	(0.052)	(0.134)
Gender in title*Female	-0.004	-0.001	0.007	0.011	-0.018	-0.006
	(0.015)	(0.014)	(0.013)	(0.012)	(0.022)	(0.020)
Constant	0.268***	0.453***	0.283***	0.467***	0.403***	0.864***
	(0.015)	(0.049)	(0.014)	(0.042)	(0.029)	(0.212)
Individual controls[a]	No	Yes	No	Yes	No	Yes
Treatment controls[b]	No	Yes	No	Yes	No	Yes
Charity DG dummy	Yes	Yes	No	No	No	No
Female + Gender in title*Female	0.038***	0.040***	0.026*	0.029**	0.100***	0.108***
	(0.010)	(0.010)	(0.009)	(0.009)	(0.010)	(0.007)
Conditions	117	117	83	83	34	34
Observations	15016	15016	11802	11802	3214	3214

[a] Individual controls: Student characteristics, age and region.
[b] Treatment controls: Double-blind, setting characteristics, random payment and partitioning of endowment.
* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Table C11: Power to detect the effect sizes estimated in the random effects model for the mean and median sample sizes in the standard DG (median $N = 130$, mean $N = 288$) and the charity DG (median $N = 192$, mean $N = 271$). The power is shown for the game specific effect sizes (0.023 and 0.109), with results for the pooled effect size (0.04) in parentheses.[a]

Effect size	DG	α	Power (median N)	Power (mean N)	N for 80% power	% papers with at least 80% power
0.023 (0.04)	Standard	0.05	0.086 (0.163)	0.148 (0.306)	3,224 (1,068)	0 (2)
0.023 (0.04)	Standard	0.005	0.013 (0.033)	0.024 (0.087)	5,470 (1,812)	0 (0)
0.109 (0.04)	Charity	0.05	0.679 (0.144)	0.821 (0.184)	256 (1,888)	25 (0)
0.109 (0.04)	Charity	0.005	0.346 (0.027)	0.524 (0.04)	436 (3,204)	8 (0)

[a] The power estimations are based on the average STD in the standard DG studies (0.233) and the average STD in the charity DG studies (0.310).

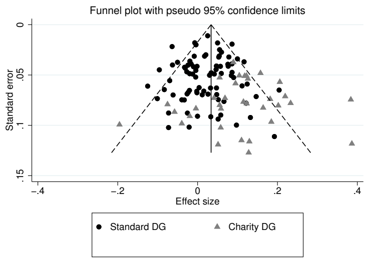
Table C12: Egger's and Begg's test of publication bias[a]. The tests are carried out both based on all DG studies pooled and separately for the standard DG and charity DG studies. Column 4-5 only includes studies with gender in the title of the paper[b]. Standard errors in parentheses.

	(1) Pooled	(2) Standard DG	(3) Charity DG	(4) Pooled	(5) Standard DG
Egger's test					
slope	0.015	0.026*	0.138*	0.074*	0.079*
	(0.011)	(0.011)	(0.054)	(0.032)	(0.031)
bias	0.502	-0.069	-0.411	-0.955	-1.233
	(0.263)	(0.291)	(0.781)	(0.763)	(0.733)
Begg's test (continuity corrected)					
z-score	1.17	0.37	0.95	1.12	1.13
p-value	0.244	0.712	0.343	0.262	0.260
Observations	117	83	34	31	28

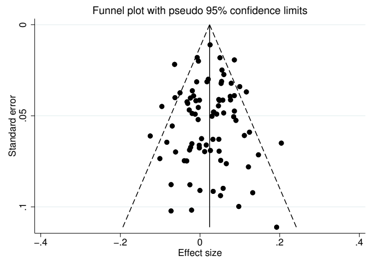
[a] The Egger's test estimates $\frac{ES_j}{SE_j} = \_0 + \beta_1 \frac{1}{SE_j} + \epsilon_j$ and if the intercept is different from zero this could be evidence of publication bias. A statistically significant result does not necessarily imply evidence of publication bias, we could also have true heterogeneity in the data that is not due to publication bias.

[b] There are only three studies with gender in the title of the paper for the charity DG, and it is therefore not meaningful to test for publication bias for charity DG studies with gender in the title. The tests of publication bias for papers with gender in the title are therefore only done for the pooled sample and standard DG studies.

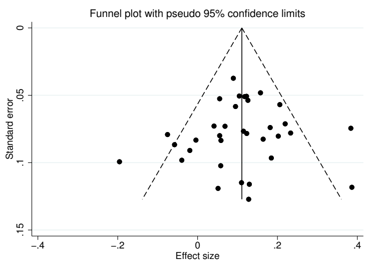
* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$



(a) Pooled

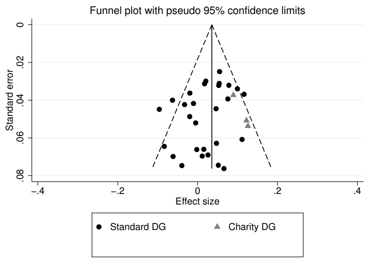


(b) Standard DG

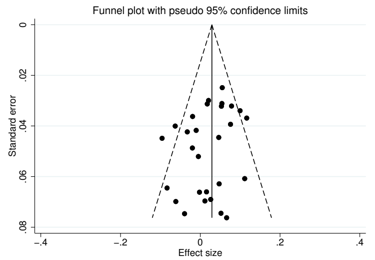


(c) Charity DG

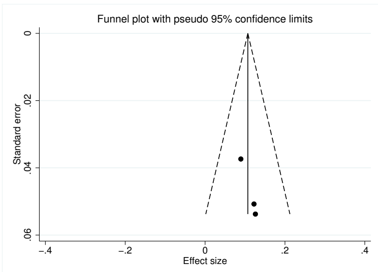
Figure C1: Funnel plots using all studies. The lines represent the pooled effect size in each sample.



(a) Pooled



(b) Standard DG



(c) Charity DG

Figure C2: Funnel plots restricted to studies that had gender in the title of the paper. The lines represent the pooled effect size in each sample.

Supplemental Online Material B: papers included in the meta-analysis

References

Alevy, Jonathan E., Francis L. Jeffries, and Yonggang Lu (2014). "Gender- and frame-specific audience effects in dictator games". *Economics Letters* 122.1, pp. 50–54.

Andreoni, James and Lise Vesterlund (2001). "Which is the fair sex? Gender differences in altruism". *The Quarterly Journal of Economics* 116.1, pp. 293–312.

Bachke, Maren Elise, Frode Alfnes, and Mette Wik (2017). "Information and donations to development aid projects". *Journal of Behavioral and Experimental Economics* 66. Experiments in Charitable Giving, pp. 23–28.

Ben-Ner, Avner et al. (2004). "Reciprocity in a two-part dictator game". *Journal of Economic Behavior & Organization* 53.3, pp. 333–352.

Biziou-van-Pol, Laura et al. (2015). "Does telling white lies signal pro-social preferences?" *Judgment and Decision Making* 10.6, pp. 538–548.

Bosch-Domènech, Antoni, Rosemarie Nagel, and Juan V. Sánchez-Andrés (2010). "Prosocial Capabilities in Alzheimer's Patients". *The Journals of Gerontology: Series B* 65B.1, pp. 119–128.

Boschini, Anne, Anna Dreber, et al. (2018). "Gender and altruism in a random sample". *Journal of Behavioral and Experimental Economics* 77, pp. 72–77.

Boschini, Anne, Astri Muren, and Mats Persson (2012). "Constructing gender differences in the economics lab". *Journal of Economic Behavior & Organization* 84.3, pp. 741–752.

Brañas-Garza, Pablo, Valerio Capraro, and Ericka Rascón-Ramírez (2018). "Gender differences in altruism on Mechanical Turk: Expectations and actual behaviour". *Economics Letters* 170, pp. 19–23.

Brock, J. Michelle, Andreas Lange, and Erkut Y. Ozbay (Feb. 2013). "Dictating the Risk: Experimental Evidence on Giving in Risky Environments". *American Economic Review* 103.1, pp. 415–37.

Cadsby, C. Bram, Maroš Servátka, and Fei Song (Sept. 2010). "Gender and generosity: does degree of anonymity or group gender composition matter?" *Experimental Economics* 13.3, pp. 299–308.

Cappelen, Alexander W., Ulrik H. Nielsen, et al. (Dec. 2016). "Fairness is intuitive". *Experimental Economics* 19.4, pp. 727–740.

Cappelen, Alexander W., Knut Nygaard, et al. (2015). "Social Preferences in the Lab: A Comparison of Students and a Representative Population." *Scandinavian Journal of Economics* 117.4, pp. 1306–1326.
Capraro, Valerio and Jotte Kuilder (2016). "To know or not to know? Looking at payoffs signals selfish behavior, but it does not actually mean so". *Journal of Behavioral and Experimental Economics* 65, pp. 79–84.
Castillo, Marco E. and Philip J. Cross (2008). "Of mice and men: Within gender variation in strategic behavior". *Games and Economic Behavior* 64.2. Special Issue in Honor of Michael B. Maschler, pp. 421–432.
Chai, Sun-Ki et al. (2011). "Cultural values and behavior in dictator, ultimatum, trust games: an experimental study". *Working Paper*.
Chaudhuri, Ananish and Lata Gangadharan (2007). "An Experimental Analysis of Trust and Trustworthiness". *Southern Economic Journal* 73.4, pp. 959–985.
Chowdhury, Subhasish M., Jeon Joo Young, and Saha Bibhas (2017). "Gender Differences in the Giving and Taking Variants of the Dictator Game". *Southern Economic Journal* 84, pp. 474–483.
Cilliers, Jacobus, Oeindrila Dube, and Bilal Siddiqi (2015). "The white-man effect: How foreigner presence affects behavior in experiments". *Journal of Economic Behavior and Organization* 118. Economic Experiments in Developing Countries, pp. 397–414.
Clot, Sophie, Gilles Grolleau, and Lisette Ibanez (2018). "Shall we pay all? An experimental test of Random Incentivized Systems". *Journal of Behavioral and Experimental Economics* 73, pp. 93–98.
d'Adda, Giovanna, Valerio Capraro, and Massimo Tavoni (2017). "Push, don't nudge: Behavioral spillovers and policy instruments". *Economics Letters* 154, pp. 92–95.
Davis, Alexander L. et al. (2015). "Generosity Across Contexts". *Working Paper*.
DeScioli, Peter and Siddhi Krishna (2013). "Giving to whom? Altruism in different types of relationships". *Journal of Economic Psychology* 34, pp. 218–228.
Dreber, Anna, Tore Ellingsen, et al. (2013). "Do people care about social context? Framing effects in dictator games". *Experimental Economics* 16.3, pp. 349–371.
Dreber, Anna, Emma von Essen, and Eva Ranehill (2014). "Gender and competition in adolescence: task matters". *Experimental Economics* 17.1, pp. 154–172.
Dreber, Anna, Drew Fudenberg, and David G. Rand (2014). "Who cooperates in repeated games: The role of altruism, inequity aversion, and demo-

graphics”. *Journal of Economic Behavior and Organization* 98, pp. 41–55.
Dufwenberg, Martin and Astri Muren (2006). “Generosity, anonymity, gender”. *Journal of Economic Behavior & Organization* 61.1, pp. 42–49.
Eckel, Catherine C. and Philip J. Grossman (1998). “Are Women Less Selfish Than Men?: Evidence From Dictator Experiments”. *The Economic Journal* 108.448, pp. 726–735.
Fiala, Lenka and Charles N. Noussair (2017). “Charitable giving, emotions, and the default effect”. *Economic Inquiry* 55.4, pp. 1792–1812.
Grosskopf, Brit and Graeme Pearce (2017). “Discrimination in a deprived neighbourhood: An artefactual field experiment”. *Journal of Economic Behavior & Organization* 141, pp. 29–42.
Haley, Kevin J. and Daniel M.T. Fessler (2005). “Nobody’s watching?: Subtle cues affect generosity in an anonymous economic game”. *Evolution and Human Behavior* 26.3, pp. 245–256.
Heinz, Matthias, Steffen Juranek, and Holger A. Rau (2012). “Do women behave more reciprocally than men? Gender differences in real effort dictator games”. *Journal of Economic Behavior & Organization* 83.1, pp. 105–110.
Hergueux, Jérôme and Nicolas Jacquemet (June 2015). “Social preferences in the online laboratory: a randomized experiment”. *Experimental Economics* 18.2, pp. 251–283.
Ito, Takehiro et al. (2016). “Contagion of Self-Interested Behavior: Evidence from Group Dictator Game Experiments”. *German Economic Review* 17.4, pp. 425–437.
Jacobsen, Karin J. et al. (2011). “Are nurses more altruistic than real estate brokers?” *Journal of Economic Psychology* 32.5, pp. 818–831.
Kettner, Sara Elisa and Israel Waichman (2016). “Old age and prosocial behavior: Social preferences or experimental confounds?” *Journal of Economic Psychology* 53, pp. 118–130.
Khachatryan, Karen et al. (2015). “Gender and preferences at a young age: Evidence from Armenia”. *Journal of Economic Behavior & Organization* 118, pp. 318–332.
Kimbrough, Erik O. and Alexander Vostroknutov (2016). “Norms Make Preferences Social”. *Journal of the European Economic Association* 14.3, pp. 608–638.
Lazear, Edward P., Ulrike Malmendier, and Roberto A. Weber (Jan. 2012). “Sorting in Experiments with Application to Social Preferences”. *American Economic Journal: Applied Economics* 4.1, pp. 136–63.
Luccasen, R. A., M. K. Thomas, and Philip J. Grossman (Aug. 2017). “Giving to poverty relief charities: the impact of beliefs and misperceptions toward

income redistribution in a real donation experiment". *Social Choice and Welfare* 49.2, pp. 387–409.

Luccasen, R. Andrew and M. Kathleen Thomas (Jan. 2020). "Voluntary taxation and the arts". *Journal of Cultural Economics*.

Martinsson, Peter, Kristian Ove R. Myrseth, and Conny Wollbrant (2012). "Reconciling pro-social vs. selfish behavior: On the role of self-control." *Judgment & Decision Making* 7.3, pp. 1–20.

Müller, Stephan and Holger A. Rau (2016). "How Gender and Risk Preferences Influence Charitable Giving: Experimental Evidence". *Working Paper*.

Nettle, Daniel et al. (2013). "The watching eyes effect in the Dictator Game: it's not how much you give, it's being seen to give something". *Evolution and Human Behavior* 34.1, pp. 35–40.

Ogawa, Kazuhito et al. (2017). "Reconsidering Whether Women are Less Selfish than Men: Group Gender Composition Matters in Dictator Games". *Working Paper*.

Raihani, Nichola J. and Redouan Bshary (2012). "A positive effect of flowers rather than eye images in a large-scale, cross-cultural dictator game". *Proceedings: Biological Sciences* 279.1742, pp. 3556–3564.

Raihani, Nichola J., Ruth Mace, and Shakti Lamba (2013). "The Effect of $1, $5 and $10 Stakes in an Online Dictator Game". *PLoS ONE* 8.

Rigdon, Mary L. and Adam Seth Levine (2018). "Gender, Expectations, and the Price of Giving". *Review of Behavioral Economics* 5.1, pp. 39–59.

Smith, John (Dec. 2012). "The endogenous nature of the measurement of social preferences". *Mind & Society* 11.2, pp. 235–256.

Tinghög, Gustav et al. (2016). "Intuition and Moral Decision-Making – The Effect of Time Pressure and Cognitive Load on Moral Judgment and Altruistic Behavior." *PLoS ONE* 11.10, pp. 1–19.

van der Weele, Joël J. and Ferdinand A. von Siemens (2020). "Bracelets of pride and guilt? An experimental test of self-signaling". *Journal of Economic Behavior and Organization* 172, pp. 280–291.

Vanberg, Christoph (2010). "Voting on a sharing norm in a dictator game". *Journal of Economic Psychology* 31.3, pp. 285–292.

Whitt, Sam and Rick K. Wilson (2007). "The Dictator Game, Fairness and Ethnicity in Postwar Bosnia". *American Journal of Political Science* 51.3, pp. 655–668.

Bibliography

Adams, M. A., Bruening, M., Ohri-Vachaspati, P. & Hurley, J. C. (2016), 'Location of school lunch salad bars and fruit and vegetable consumption in middle schools: A cross-sectional plate waste study', Journal of the Academy of Nutrition and Dietetics **116**(3), 407–416.

Adams, M. A., Pelletier, R. L., Zive, M. M. & Sallis, J. F. (2005), 'Salad bars and fruit and vegetable consumption in elementary schools: a plate waste study', Journal of the American Dietetic Association **105**(11), 1789–1792.

Afshin, A., Penalvo, J. L., Del Gobbo, L., Silva, J., Michaelson, M., O'Flaherty, M., Capewell, S., Spiegelman, D., Danaei, G. & Mozaffarian, D. (2017), 'The prospective impact of food pricing on improving dietary consumption: a systematic review and meta-analysis', PloS one **12**(3), e0172277.

Al-Ubaydli, O., List, J. A. & Suskind, D. L. (2017), 'What Can We Learn from Experiments? Understanding the Threats to the Scalability of Experimental Results', American Economic Review **107**(5), 282–86.

Albani, V., Butler, L. T., Traill, W. B. & Kennedy, O. B. (2017), 'Fruit and vegetable intake: change with age across childhood and adolescence', British Journal of Nutrition **117**(5), 759–765.

Alevy, J. E., Jeffries, F. L. & Lu, Y. (2014), 'Gender- and frame-specific audience effects in dictator games', Economics Letters **122**(1), 50 – 54.

Allbirds (2022), 'Sustainability Guide & Practices'.
URL: *https://www.allbirds.com/pages/sustainable-practices*

Allcott, H. & Rogers, T. (2014), 'The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation', American Economic Review **104**(10), 3003–3037.

Allcott, H. & Taubinsky, D. (2015), 'Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market', American Economic Review **105**(8), 2501–38.

An, R. (2013), 'Effectiveness of subsidies in promoting healthy food purchases and consumption: a review of field experiments', Public health nutrition **16**(7), 1215–1228.

Andreoni, J., Rao, J. M. & Trachtman, H. (2017), 'Avoiding the ask: A field experiment on altruism, empathy, and charitable giving', Journal of Political Economy **125**(3), 625–653.

Andreoni, J. & Vesterlund, L. (2001), 'Which is the fair sex? gender differences in altruism', The Quarterly Journal of Economics **116**(1), 293–312.

Aune, D., Giovannucci, E., Boffetta, P., Fadnes, L. T., Keum, N., Norat, T., Greenwood, D. C., Riboli, E., Vatten, L. J. & Tonstad, S. (2017), 'Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality—a systematic review and dose-response meta-analysis of prospective studies', International journal of epidemiology **46**(3), 1029–1056.

Axfood AB (2020), 'Environmental targets'.
URL: *https://www.axfood.com/about-axfood/goals-and-strategy*

Bachke, M. E., Alfnes, F. & Wik, M. (2017), 'Information and donations to development aid projects', Journal of Behavioral and Experimental Economics **66**, 23 – 28. Experiments in Charitable Giving.

Bardsley, N. (2008), 'Dictator game giving: altruism or artefact?', Experimental Economics **11**(2), 122–133.

Bekkers, R. (2007), 'Measuring altruistic behavior in surveys: The all-or-nothing dictator game', Survey Research Methods **1**(3), 139–144.

Ben-Ner, A., Putterman, L., Kong, F. & Magan, D. (2004), 'Reciprocity in a two-part dictator game', Journal of Economic Behavior & Organization **53**(3), 333 – 352.

Bénabou, R. & Tirole, J. (2010), 'Individual and corporate social responsibility', Economica **77**(305), 1–19.

Berkowitz, S. A., Curran, N., Hoeffler, S., Henderson, R., Price, A. & Ng, S. W. (2021), 'Association of a fruit and vegetable subsidy program with food purchases by individuals with low income in the us', JAMA Network Open **4**(8), e2120377–e2120377.

Bertrand, M., Goldin, C. & Katz, L. F. (2010), 'Dynamics of the gender gap for young professionals in the financial and corporate sectors', American Economic Journal: Applied Economics **2**(3), 228–55.

Bilén, D. (2022), ‘Do carbon labels cause consumers to reduce their emissions? Evidence from a large-scale natural experiment’.

Biziou-van Pol, L., Haenen, J., Novaro, A., Liberman, A. O. & Capraro, V. (2015), ‘Does telling white lies signal pro-social preferences?’, Judgment & Decision Making **10**(6), 538 – 548.

Bénabou, R. & Tirole, J. (2006), ‘Incentives and prosocial behavior’, American Economic Review **96**(5), 1652–1678.

Bollinger, B., Leslie, P. & Sorensen, A. (2011), ‘Calorie posting in chain restaurants’, American Economic Journal: Economic Policy **3**(1), 91–128.

Bolton, G. E. & Katok, E. (1995), ‘An experimental test for gender differences in beneficent behavior’, Economics Letters **48**(3), 287–292.

Bosch-Domènech, A., Nagel, R. & Sánchez-Andrés, J. V. (2010), ‘Prosocial capabilities in alzheimer’s patients’, The Journals of Gerontology: Series B **65B**(1), 119–128.

Boschini, A., Dreber, A., von Essen, E., Muren, A. & Ranehill, E. (2018), ‘Gender and altruism in a random sample’, Journal of Behavioral and Experimental Economics **77**, 72 – 77.

Boschini, A., Muren, A. & Persson, M. (2012), ‘Constructing gender differences in the economics lab’, Journal of Economic Behavior & Organization **84**(3), 741–752.

Bowles, S. & Polania-Reyes, S. (2012), ‘Economic incentives and social preferences: Substitutes or complements?’, Journal of Economic Literature **50**(2), 368–425.

Brañas-Garza, P., Capraro, V. & Rascón-Ramírez, E. (2018), ‘Gender differences in altruism on mechanical turk: Expectations and actual behaviour’, Economics Letters **170**, 19 – 23.

Brock, J. M., Lange, A. & Ozbay, E. Y. (2013), ‘Dictating the risk: Experimental evidence on giving in risky environments’, American Economic Review **103**(1), 415–37.

Broers, V. J., De Breucker, C., Van den Broucke, S. & Luminet, O. (2017), ‘A systematic review and meta-analysis of the effectiveness of nudging to increase fruit and vegetable choice’, The European Journal of Public Health **27**(5), 912–920.

Brunner, F., Kurz, V., Bryngelsson, D. & Hedenus, F. (2018), 'Carbon label at a university restaurant – label implementation and evaluation', Ecological Economics **146**, 658 – 667.

Burke, D. L., Ensor, J. & Riley, R. D. (2017), 'Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ', Statistics in Medicine **36**(5), 855–875.

Cadario, R. & Chandon, P. (2020), 'Which healthy eating nudges work best? a meta-analysis of field experiments', Marketing Science **39**(3), 465–486.

Cadsby, C. B., Servátka, M. & Song, F. (2010), 'Gender and generosity: does degree of anonymity or group gender composition matter?', Experimental Economics **13**(3), 299–308.

Camilleri, A. R., Larrick, R. P., Hossain, S. & Patino-Echeverri, D. (2019), 'Consumers underestimate the emissions associated with food but are aided by labels', Nature Climate Change **9**(1), 53–58.

Capacci, S. & Mazzocchi, M. (2011), 'Five-a-day, a price to pay: an evaluation of the uk program impact accounting for market forces', Journal of health economics **30**(1), 87–98.

Cappelen, A. W., Nielsen, U. H., Tungodden, B., Tyran, J.-R. & Wengström, E. (2016), 'Fairness is intuitive', Experimental Economics **19**(4), 727–740.

Cappelen, A. W., Nygaard, K., Sørensen, E. . & Tungodden, B. (2015), 'Social preferences in the lab: A comparison of students and a representative population.', Scandinavian Journal of Economics **117**(4), 1306 – 1326.

Capraro, V. & Kuilder, J. (2016), 'To know or not to know? looking at payoffs signals selfish behavior, but it does not actually mean so', Journal of Behavioral and Experimental Economics **65**, 79 – 84.

Carlsson, F., Gravert, C., Johansson-Stenman, O. & Kurz, V. (2021), 'The use of green nudges as an environmental policy instrument', Review of Environmental Economics and Policy **15**(2), 216–237.

Carlsson, F., Kataria, M., Lampi, E., Nyberg, E. & Sterner, T. (2020), Food labels: how consumers value moral, environmental, and health aspects of meat consumption, Working papers in economics, University of Gothenburg, Department of Economics.

Castillo, M. E. & Cross, P. J. (2008), 'Of mice and men: Within gender variation in strategic behavior', Games and Economic Behavior **64**(2), 421 – 432. Special Issue in Honor of Michael B. Maschler.

Chai, S.-K., Dorj, D., Kim, M. S., Liu, M. & Sherstyuk, K. (2011), 'Cultural values and behavior in dictator, ultimatum, trust games: an experimental study', Working Paper .

Chater, N. & Loewenstein, G. (2022), 'The i-Frame and the s-Frame: How Focusing on Individual-Level Solutions Has Led Behavioral Public Policy Astray'.

Chaudhuri, A. & Gangadharan, L. (2007), 'An experimental analysis of trust and trustworthiness', Southern Economic Journal **73**(4), 959–985.

Chetty, R., Looney, A. & Kroft, K. (2009), 'Salience and taxation: Theory and evidence', American Economic Review **99**(4), 1145–77.

Chowdhury, S. M., Young, J. J. & Bibhas, S. (2017), 'Gender differences in the giving and taking variants of the dictator game', Southern Economic Journal **84**, 474 – 483.

Cilliers, J., Dube, O. & Siddiqi, B. (2015), 'The white-man effect: How foreigner presence affects behavior in experiments', Journal of Economic Behavior & Organization **118**, 397 – 414. Economic Experiments in Developing Countries.

COBIOM (2021), 'Together for Carbon Labelling The Initiative for more CO2e-Transparency in the Food Industry'.
URL: *https://together-for-carbon-labelling.com*

Cohen, J. F., Richardson, S. A., Cluggish, S. A., Parker, E., Catalano, P. J. & Rimm, E. B. (2015), 'Effects of choice architecture and chef-enhanced meals on the selection and consumption of healthier school foods: A randomized clinical trial', JAMA pediatrics **169**(5), 431–437.

Collibri Foundation (2021), 'We're one of the first UK retailers to trial Eco-Score labelling'.
URL: *https://www.colruytgroup.com/en/conscious-consuming/eco-score*

COOP (2020), 'Coops klimatmål'.
URL: *https://www.coop.se/hallbarhet/hallbarhetsarbete/klimat–energi/*

Cornelissen, G., Pandelaere, M., Warlop, L. & Dewitte, S. (2008), 'Positive cueing: Promoting sustainable consumer behavior by cueing common environmental behaviors as environmental', International Journal of Research in Marketing **25**(1), 46–55.

Crippa, M., Guizzardi, D., Banja, M., Solazzo, E., Muntean, M., Schaaf, E., Pagani, F., Monforti-Ferrario, F., Olivier, J., Quadrelli, R., Risquez Martin, A., Taghavi-Moharamli, P., Grassi, G., Rossi, S., Jacome Felix Oom, D., Branco, A., San-Miguel-Ayanz, J. & Vignati, f. E. (2022), Co2 emissions of all world countries, Scientific analysis or review KJ-NA-31-182-EN-N (online),KJ-NA-31-182-EN-C (print), Luxembourg (Luxembourg).

Crippa, M., Solazzo, E., Guizzardi, D., Monforti-Ferrario, F., Tubiello, F. N. & Leip, A. (2021), 'Food systems are responsible for a third of global anthropogenic GHG emissions', **2**(3), 198–209. Number: 3 Publisher: Nature Publishing Group.

Croson, R. & Gneezy, U. (2009), 'Gender differences in preferences', Journal of Economic literature **47**(2), 448–474.

d'Adda, G., Capraro, V. & Tavoni, M. (2017), 'Push, don't nudge: Behavioral spillovers and policy instruments', Economics Letters **154**, 92 – 95.

Dana, J., Cain, D. M. & Dawes, R. M. (2006), 'What you don't know won't hurt me: Costly (but quiet) exit in dictator games', Organizational Behavior and Human Decision Processes **100**(2), 193–201.

Davis, A. L., Jehli, N. R., Miller, J. H. & Weber, R. A. (2015), 'Generosity across contexts', Working Paper .

de Chaisemartin, C. & D'Haultfœuille, X. (2020), 'Two-way fixed effects estimators with heterogeneous treatment effects', American Economic Review **110**(9), 2964–96.

DellaVigna, S. & Linos, E. (2022), 'RCTs to Scale: Comprehensive Evidence From Two Nudge Units', Econometrica **90**(1), 81–116.

DeScioli, P. & Krishna, S. (2013), 'Giving to whom? altruism in different types of relationships', Journal of Economic Psychology **34**, 218 – 228.

Dreber, A., Ellingsen, T., Johannesson, M. & Rand, D. G. (2013), 'Do people care about social context? framing effects in dictator games', Experimental Economics **16**(3), 349–371.

Dreber, A., Fudenberg, D. & Rand, D. G. (2014), 'Who cooperates in repeated games: The role of altruism, inequity aversion, and demographics', Journal of Economic Behavior & Organization **98**, 41 – 55.

Dreber, A., von Essen, E. & Ranehill, E. (2014), 'Gender and competition in adolescence: task matters', Experimental Economics **17**(1), 154–172.

Dubois, P., Albuquerque, P., Allais, O., Bonnet, C., Bertail, P., Combris, P., Lahlou, S., Rigal, N., Ruffieux, B. & Chandon, P. (2021), 'Effects of front-of-pack labels on the nutritional quality of supermarket food purchases: evidence from a large-scale randomized controlled trial', **49**(1), 119–138.

Dubois, P., Griffith, R. & O'Connell, M. (2022), 'The use of scanner data for economics research', Annual Review of Economics **14**, 723–745.

Dufwenberg, M. & Muren, A. (2006), 'Generosity, anonymity, gender', Journal of Economic Behavior & Organization **61**(1), 42–49.

Eckel, C. C. & Grossman, P. J. (1996), 'Altruism in anonymous dictator games', Games and Economic Behavior **16**(2), 181 – 191.

Eckel, C. C. & Grossman, P. J. (1998), 'Are women less selfish than men?: Evidence from dictator experiments', The Economic Journal **108**(448), 726–735.

Egebark, J. & Ekström, M. (2016), 'Can indifference make the world greener?', Journal of Environmental Economics and Management **76**, 1–13.

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

Engel, C. (2011), 'Dictator games: a meta study', Experimental Economics **14**(4), 583–610.

Enlund, J., Andersson, D. & Carlsson, F. (2022), 'Individual Carbon Footprint Reduction: Evidence from Pro-environmental Users of a Carbon Calculator'.

European Commision (2020), 'Farm to fork strategy'.
URL: *https://food.ec.europa.eu/horizontal-topics/farm-fork-strategy*

European Commission (2021), 'data.europa.eu'. Accessed: 2022-9-12.
URL: *http://data.europa.eu/88u/dataset/S2273_95_1_513_ENG*

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D. & Sunde, U. (2018), 'Global Evidence on Economic Preferences*', The Quarterly Journal of Economics **133**(4), 1645–1692.

Fioretti, M. (2022), 'Caring or Pretending to Care? Social Impact, Firms' Objectives, and Welfare', Journal of Political Economy **130**(11), 2898–2942.

Forsythe, R., Horowitz, J. L., Savin, N. & Sefton, M. (1994), 'Fairness in simple bargaining experiments', Games and Economic Behavior **6**(3), 347 – 369.

Fosgaard, T. R., Pizzo, A. & Sadoff, S. (2021), 'Do people respond to the climate impact of their behavior? the effect of carbon footprint information on grocery purchases', p. 29.

Gabaix, X. (2019), Chapter 4 - behavioral inattention, in B. D. Bernheim, S. DellaVigna & D. Laibson, eds, 'Handbook of Behavioral Economics - Foundations and Applications 2', Vol. 2 of Handbook of Behavioral Economics: Applications and Foundations 1, North-Holland, pp. 261–343.

Gallup (2009), 'Europeans' attitudes towards the issue of sustainable consumption and production'.

Ghesla, C., Grieder, M. & Schubert, R. (2020), 'Nudging the poor and the rich–a field study on the distributional effects of green electricity defaults', Energy Economics **86**, 104616.

Gneezy, U., Meier, S. & Rey-Biel, P. (2011), 'When and why incentives (don't) work to modify behavior', Journal of economic perspectives **25**(4), 191–210.

Gneezy, U. & Rustichini, A. (2000), 'A fine is a price', The Journal of Legal Studies **29**(1), 1–17.

Gonçalves, D., Coelho, P., Martinez, L. F. & Monteiro, P. (2021), 'Nudging consumers toward healthier food choices: A field study on the effect of social norms', Sustainability **13**(4), 1660.

Griffith, R. (2022), 'Obesity, poverty and public policy', The Economic Journal **132**(644), 1235–1258.

Griffith, R., von Hinke, S. & Smith, S. (2018), 'Getting a healthy start: The effectiveness of targeted benefits for improving dietary choices', Journal of health economics **58**, 176–187.

Grosskopf, B. & Pearce, G. (2017), 'Discrimination in a deprived neighbourhood: An artefactual field experiment', Journal of Economic Behavior & Organization **141**, 29 – 42.

Hainmueller, J., Hiscox, M. J. & Sequeira, S. (2015), 'Consumer demand for fair trade: Evidence from a multistore field experiment', The Review of Economics and Statistics **97**(2), 242–256.

Haley, K. J. & Fessler, D. M. (2005), 'Nobody's watching?: Subtle cues affect generosity in an anonymous economic game', Evolution and Human Behavior **26**(3), 245 – 256.

Hanks, A. S., Just, D. R., Smith, L. E. & Wansink, B. (2012), 'Healthy convenience: nudging students toward healthier choices in the lunchroom', Journal of public health **34**(3), 370–376.

Heinz, M., Juranek, S. & Rau, H. A. (2012), 'Do women behave more reciprocally than men? gender differences in real effort dictator games', Journal of Economic Behavior & Organization **83**(1), 105 – 110.

Hergueux, J. & Jacquemet, N. (2015), 'Social preferences in the online laboratory: a randomized experiment', Experimental Economics **18**(2), 251–283.

Herman, D. R., Harrison, G. G., Afifi, A. A. & Jenks, E. (2008), 'Effect of a targeted subsidy on intake of fruits and vegetables among low-income women in the special supplemental nutrition program for women, infants, and children', American journal of public health **98**(1), 98–105.

Hornibrook, S., May, C. & Fearne, A. (2015), 'Sustainable development and the consumer: Exploring the role of carbon labelling in retail supply chains', Business Strategy and the Environment **24**(4), 266–276.

ICA Gruppen (2022), 'ICA Gruppen: Beyond Climate Neutrality Sweden'. **URL:** *https://www.icagruppen.se/en/about-ica-gruppen/goals-and-strategy*

Ioannidis, J. P. A., Stanley, T. D. & Doucouliagos, H. (2017), 'The power of bias in economics research', The Economic Journal **127**(605), F236–F265.

IPCC (2019), IPCC, 2019: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems [P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacemi, J. Malley, (eds.)]. In press., Technical report, University of Zurich, Department of Informatics.

Ito, T., Ogawa, K., Suzuki, A., Takahashi, H. & Takemoto, T. (2016), 'Contagion of self-interested behavior: Evidence from group dictator game experiments', German Economic Review **17**.

Jacobsen, K. J., Eika, K. H., Helland, L., Lind, J. T. & Nyborg, K. (2011), 'Are nurses more altruistic than real estate brokers?', Journal of Economic Psychology **32**(5), 818 – 831.

Kahneman, D., Knetsch, J. L. & Thaler, R. H. (1986), 'Fairness and the assumptions of economics', Journal of business pp. S285–S300.

Kanay, A., Hilton, D., Charalambides, L., Corrégé, J.-B., Inaudi, E., Waroquier, L. & Cézéra, S. (2021), 'Making the carbon basket count: Goal setting promotes sustainable consumption in a simulated online supermarket', **83**, 102348.

Kettner, S. E. & Waichman, I. (2016), 'Old age and prosocial behavior: Social preferences or experimental confounds?', Journal of Economic Psychology **53**, 118 – 130.

Khachatryan, K., Dreber, A., von Essen, E. & Ranehill, E. (2015), 'Gender and preferences at a young age: Evidence from armenia', Journal of Economic Behavior & Organization **118**, 318 – 332. Economic Experiments in Developing Countries.
URL: *http://www.sciencedirect.com/science/article/pii/S0167268115000645*

Kimbrough, E. O. & Vostroknutov, A. (2016), 'Norms make preferences social', Journal of the European Economic Association **14**(3), 608–638.

Klerman, J. A., Bartlett, S., Wilde, P. & Olsho, L. (2014), 'The short-run impact of the healthy incentives pilot program on fruit and vegetable intake', American Journal of Agricultural Economics **96**(5), 1372–1382.

Knutsson, M., Martinsson, P. & Wollbrant, C. (2013), 'Do people avoid opportunities to donate?: A natural field experiment on recycling and charitable giving', Journal of Economic Behavior & Organization **93**, 71–77.

Kortelainen, M., Raychaudhuri, J. & Roussillon, B. (2016), 'Effects of carbon reduction labels: Evidence from scanner data', Economic Inquiry **54**(2), 1167–1187.

Kotchen, M. J. & Moore, M. R. (2007), 'Private provision of environmental public goods: Household participation in green-electricity programs', Journal of Environmental Economics and management **53**(1), 1–16.

Krupka, E. L. & Weber, R. A. (2013), 'Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?', Journal of the European Economic Association **11**(3), 495–524.

Kurz, V. (2018), 'Nudging to reduce meat consumption: Immediate and persistent effects of an intervention at a university restaurant', Journal of Environmental Economics and Management **90**, 317 – 341.

Kvarven, A., Strømland, E. & Johannesson, M. (2020), 'Comparing meta-analyses and preregistered multiple-laboratory replication projects', **4**(4), 423–434. Number: 4 Publisher: Nature Publishing Group.

Lazear, E. P., Malmendier, U. & Weber, R. A. (2012), 'Sorting in experiments with application to social preferences', American Economic Journal: Applied Economics **4**(1), 136–63.

Levy, D. E., Riis, J., Sonnenberg, L. M., Barraclough, S. J. & Thorndike, A. N. (2012), 'Food choices of minority and low-income employees: a cafeteria intervention', American journal of preventive medicine **43**(3), 240–248.

Lidl (2021), 'We're one of the first UK retailers to trial Eco-Score labelling.'.
URL: *https://www.lidl.co.uk/about-us/lidl-changes-for-the-better/eco-score*

List, J. A. (2007), 'On the interpretation of giving in dictator games', Journal of Political economy **115**(3), 482–493.

List, J. A. & Samek, A. S. (2015), 'The behavioralist as nutritionist: Leveraging behavioral economics to improve child food choice and consumption', Journal of health economics **39**, 135–146.

Logitech (2022), 'Carbon labeling and measuring carbon impact of products'.
URL: *https://www.logitech.com/en-us/sustainability/carbon-labeling-measuring.html*

Lohmann, P. M., Gsottbauer, E., Doherty, A. & Kontoleon, A. (2022), 'Do print labels promote climatarian diets? evidence from a large-scale field experiment', **114**, 102693.

Luccasen, R. A. & Thomas, M. K. (2020), 'Voluntary taxation and the arts', Journal of Cultural Economics .

Luccasen, R. A., Thomas, M. K. & Grossman, P. J. (2017), 'Giving to poverty relief charities: the impact of beliefs and misperceptions toward income redistribution in a real donation experiment', Social Choice and Welfare **49**(2), 387–409.

Martinsson, P., Myrseth, K. O. R. & Wollbrant, C. (2012), 'Reconciling pro-social vs. selfish behavior: On the role of self-control.', Judgment & Decision Making **7**(3), 1 – 20.

Mellström, C. & Johannesson, M. (2008), 'Crowding out in blood donation: was titmuss right?', Journal of the European Economic Association **6**(4), 845–863.

Müller, S. & Rau, H. A. (2016), 'How gender and risk preferences influence charitable giving: Experimental evidence', Working Paper .

Murray, C. J., Aravkin, A. Y., Zheng, P., Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., Abd-Allah, F., Abdelalim, A., Abdollahi, M., Abdollahpour, I. et al. (2020), 'Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019', The lancet **396**(10258), 1223–1249.

Naturvardsverket (2020), 'Konsumtionsbaserade vaxthusgasutslapp per person och år – naturvardsverket'.
URL: *https://www.naturvardsverket.se/data-och-statistik/konsumtion/vaxthusgaser-konsumtionsbaserade-utslapp-per-person/*

Ni Mhurchu, C., Blakely, T., Jiang, Y., Eyles, H. C. & Rodgers, A. (2010), 'Effects of price discounts and tailored nutrition education on supermarket purchases: a randomized controlled trial', The American journal of clinical nutrition **91**(3), 736–747.

Oatly (2022), 'Climate footprint'.
URL: *https://www.oatly.com/en-gb/stuff-we-make/climate-footprint*

Ogawa, K., Takemoto, T., Takahashi, H. & Suzuki, A. (2017), 'Reconsidering whether women are less selfish than men: Group gender composition matters in dictator games', Working Paper .

Ortmann, A. & Zhang, L. (2015), Sample size null hypothesis significance testing, Greenwood Press/ABC-CLIO, pp. 373–374.

Pace, D. & van der Weele, J. J. (2020), 'Curbing carbon: An experiment on uncertainty and information about CO_2 emissions'.

Palomo-Vélez, G., Tybur, J. M. & Van Vugt, M. (2018), 'Unsustainable, unhealthy, or disgusting? comparing different persuasive messages against meat consumption', Journal of Environmental Psychology **58**, 63–71.

Panzone, L. A., Tocco, B., Brecic, R. & Gorton, M. (2022), 'A frequency reward program with a costly reward: An evaluation of a campaign to motivate consumption of fruit and vegetables'.

Panzone, L. A., Ulph, A., Hilton, D., Gortemaker, I. & Tajudeen, I. A. (2021), 'Sustainable by design: Choice architecture and the carbon footprint of grocery shopping', Journal of Public Policy & Marketing **40**(4), 463–486.

Panzone, L. A., Ulph, A., Zizzo, D. J., Hilton, D. & Clear, A. (2021), 'The impact of environmental recall and carbon taxation on the carbon footprint of supermarket shopping', **109**, 102137.

Perino, G. & Schwirplies, C. (2022), 'Meaty arguments and fishy effects: Field experimental evidence on the impact of reasons to reduce meat consumption', Journal of Environmental Economics and Management **114**, 102667.

Polacsek, M., Moran, A., Thorndike, A. N., Boulos, R., Franckle, R. L., Greene, J. C., Blue, D. J., Block, J. P. & Rimm, E. B. (2018), 'A supermarket double-dollar incentive program increases purchases of fresh fruits and vegetables among low-income families with children: the healthy double study', Journal of nutrition education and behavior **50**(3), 217–228.

Potter, C., Bastounis, A., Hartmann-Boyce, J., Stewart, C., Frie, K., Tudor, K., Bianchi, F., Cartwright, E., Cook, B., Rayner, M. & Jebb, S. A. (2021), 'The effects of environmental sustainability labels on selection, purchase, and consumption of food and drink products: A systematic review', Environment and Behavior **53**(8), 891–925. PMID: 34456340.
URL: *https://doi.org/10.1177/0013916521995473*

Quorn (2022), 'Quorn unveils carbon footprint labelling of its products and calls on other brands to follow suit as it launches new campaign to help consumers battle climate change'.
URL: *https://www.quorn.co.uk/company/press/quorn-unveils-carbon-footprint-labelling-of-its-products-and-calls-on-other*

Raihani, N. J. & Bshary, R. (2012), 'A positive effect of flowers rather than eye images in a large-scale, cross-cultural dictator game', Proceedings: Biological Sciences **279**(1742), 3556–3564.

Raihani, N. J., Mace, R. & Lamba, S. (2013), 'The effect of $1, $5 and $10 stakes in an online dictator game', PLoS ONE **8**.

Rigdon, M. L. & Levine, A. S. (2018), 'Gender, expectations, and the price of giving', Review of Behavioral Economics **5**(1), 39–59.

RISE (2020), 'The open access list – an extract from the rise food climate database 1.7 (2020)'.

URL: *https://www.ri.se/en/what-we-do/expertises/rise-food-climate-database*

Sandström, V., Valin, H., Krisztin, T., Havlík, P., Herrero, M. & Kastner, T. (2018), 'The role of trade in the greenhouse gas footprints of eu diets', Global Food Security **19**, 48–55.

Schott, A. B. S. & Andersson, T. (2015), 'Food waste minimization from a life-cycle perspective', Journal of Environmental Management **147**, 219–226.

Shreedhar, G. & Galizzi, M. M. (2021), 'Personal or planetary health? direct, spillover and carryover effects of non-monetary benefits of vegetarian behaviour', Journal of Environmental Psychology **78**, 101710.

Säll, S. & Gren, I.-M. (2015), 'Effects of an environmental tax on meat and dairy consumption in sweden', Food Policy **55**, 41–53.

Smith, J. (2012), 'The endogenous nature of the measurement of social preferences', Mind & Society **11**(2), 235–256.

Stables, G. J., Subar, A. F., Patterson, B. H., Dodd, K., Heimendinger, J., Van Duyn, M. A. S. & Nebeling, L. (2002), 'Changes in vegetable and fruit consumption and awareness among us adults: results of the 1991 and 1997 5 a day for better health program surveys', Journal of the American Dietetic Association **102**(6), 809–817.

Stuber, J. M., Hoenink, J. C., Beulens, J. W., Mackenbach, J. D. & Lakerveld, J. (2021), 'Shifting toward a healthier dietary pattern through nudging and pricing strategies: A secondary analysis of a randomized virtual supermarket experiment', The American journal of clinical nutrition **114**(2), 628–637.

Sudarshan, A. (2017), 'Nudges in the marketplace: The response of household electricity consumption to information and monetary incentives', Journal of Economic Behavior & Organization **134**, 320–335.

Swedish Food Retailers Federation (2020), 'The Food Retail Index December and 2020'.
URL: *https://www.svenskdagligvaruhandel.se/wp-content/uploads/Dagligvaruindex-december-2020.pdf*

Taufique, K. M. R., Nielsen, K. S., Dietz, T., Shwom, R., Stern, P. C. & Vandenbergh, M. P. (2022), 'Revisiting the promise of carbon labelling', Nat. Clim. Chang. **12**(2), 132–140.

Thorndike, A. N., Bright, O.-J. M., Dimond, M. A., Fishman, R. & Levy, D. E. (2017), 'Choice architecture to promote fruit and vegetable purchases by families participating in the special supplemental program for women, infants, and children (wic): randomized corner store pilot study', Public health nutrition **20**(7), 1297–1305.

Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R. & Staake, T. (2018), 'Overcoming salience bias: How real-time feedback fosters resource conservation', Management Science **64**(3), 1458–1476.

Tinghög, G., Andersson, D., Bonn, C., Johannesson, M., Kirchler, M., Koppel, L. & Västfjäll, D. (2016), 'Intuition and moral decision-making – the effect of time pressure and cognitive load on moral judgment and altruistic behavior.', PLoS ONE **11**(10), 1 – 19.

Titmuss, R. et al. (1970), 'The gift relationship. from human blood to social policy.', The gift relationship. From human blood to social policy. .

Unilever (2022), 'Carbon footprints – now it's personal'.
URL: *https://www.unilever.com/news/news-search/2021/carbon-footprints-now-its-personal/*

Van Kleef, E., Otten, K. & van Trijp, H. C. (2012), 'Healthy snacks at the checkout counter: A lab and field study on the impact of shelf arrangement and assortment structure on consumer choices', BMC public health **12**, 1–10.

Vanberg, C. (2010), 'Voting on a sharing norm in a dictator game', Journal of Economic Psychology **31**(3), 285 – 292.

Whitt, S. & Wilson, R. K. (2007), 'The dictator game, fairness and ethnicity in postwar bosnia', American Journal of Political Science **51**(3), 655–668.

WHO (2011), Global atlas on cardiovascular disease prevention and control, Technical report, World Health Organization.
URL: *http://www.jstor.org/stable/resrep30133.13*

Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., Garnett, T., Tilman, D., DeClerck, F., Wood, A. et al. (2019), 'Food in the anthropocene: the eat–lancet commission on healthy diets from sustainable food systems', The lancet **393**(10170), 447–492.

Wirsenius, S., Hedenus, F. & Mohlin, K. (2011), 'Greenhouse gas taxes on animal food products: rationale, tax scheme and climate mitigation effects', Climatic Change **108**(1), 159–184.

WWF (2022), ‘Leading supermarkets join WWF’.
URL: *https://www.wwf.org.uk/press-release/leading-supermarkets-join-wwf*

Ytreberg, N. S., Alfnes, F. & Oort, B. v. (2023), ‘Mapping of the digital climate nudges in nordic online grocery stores’.

Zhang, L. & Ortmann, A. (2014), ‘The effects of the take-option in dictator-game experiments: a comment on engel’s (2011) meta-study’, Experimental Economics **17**(3), 414–420.

Previous doctoral theses in the Department of Economics, Gothenburg

Avhandlingar publicerade innan serien Ekonomiska Studier startades
(Theses published before the series Ekonomiska Studier was started):

Östman, Hugo (1911), Norrlands ekonomiska utveckling
Moritz, Marcus (1911), Den svenska tobaksindustrien
Sundbom, I. (1933), Prisbildning och ändamålsenlighet
Gerhard, I. (1948), Problem rörande Sveriges utrikeshandel 1936/38
Hegeland, Hugo (1951), The Quantity Theory of Money
Mattsson, Bengt (1970), Cost-Benefit analys
Rosengren, Björn (1975), Valutareglering och nationell ekonomisk politik
Hjalmarsson, Lennart (1975), Studies in a Dynamic Theory of Production and its Applications
Örtendahl, Per-Anders (1975), Substitutionsaspekter på produktionsprocessen vid massaframställning
Anderson, Arne M. (1976), Produktion, kapacitet och kostnader vid ett helautomatiskt emballageglasbruk
Ohlsson, Olle (1976), Substitution och odelbarheter i produktionsprocessen vid massaframställning
Gunnarsson, Jan (1976), Produktionssystem och tätortshierarki – om sambandet mellan rumslig och ekonomisk struktur
Köstner, Evert (1976), Optimal allokering av tid mellan utbildning och arbete
Wigren, Rune (1976), Analys av regionala effektivitetsskillnader inom industribranscher
Wästlund, Jan (1976), Skattning och analys av regionala effektivitetsskillnader inom industribranscher
Flöjstad, Gunnar (1976), Studies in Distortions, Trade and Allocation Problems
Sandelin, Bo (1977), Prisutveckling och kapitalvinster på bostadsfastigheter
Dahlberg, Lars (1977), Empirical Studies in Public Planning
Lönnroth, Johan (1977), Marxism som matematisk ekonomi
Johansson, Börje (1978), Contributions to Sequential Analysis of Oligopolistic Competition

Ekonomiska Studier, utgivna av Nationalekonomiska institutionen vid Göteborgs Universitet. Nr 1 och 4 var inte doktorsavhandlingar. (The contributions to the department series 'Ekonomiska Studier' where no. 1 and 4 were no doctoral theses):

2. **Ambjörn, Erik** (1959), Svenskt importberoende 1926-1956: en ekonomisk-statistisk kartläggning med kommentarer
3. **Landgren, K-G.** (1960), Den "Nya ekonomien" i Sverige: J.M. Keynes, E. Wigfors och utecklingen 1927-39
5. **Bigsten, Arne** (1979), Regional Inequality and Development: A Case Study of Kenya
6. **Andersson, Lars** (1979), Statens styrning av de kommunala budgetarnas struktur (Central Government Influence on the Structure of the Municipal Budget)
7. **Gustafsson, Björn** (1979), Inkomst- och uppväxtförhållanden (Income and Family Background)
8. **Granholm, Arne** (1981), Interregional Planning Models for the Allocation of Private and Public Investments
9. **Lundborg, Per** (1982), Trade Policy and Development: Income Distributional Effects in the Less Developed Countries of the US and EEC Policies for Agricultural

Commodities

10. **Juås, Birgitta** (1982), Värdering av risken för personskador. En jämförande studie av implicita och explicita värden. (Valuation of Personal Injuries. A comparison of Explicit and Implicit Values)
11. **Bergendahl, Per-Anders** (1982), Energi och ekonomi - tillämpningar av input-output analys (Energy and the Economy - Applications of Input-Output Analysis)
12. **Blomström, Magnus** (1983), Foreign Investment, Technical Efficiency and Structural Change - Evidence from the Mexican Manufacturing Industry
13. **Larsson, Lars-Göran** (1983), Comparative Statics on the Basis of Optimization Methods
14. **Persson, Håkan** (1983), Theory and Applications of Multisectoral Growth Models
15. **Sterner, Thomas** (1986), Energy Use in Mexican Industry.
16. **Flood, Lennart** (1986), On the Application of Time Use and Expenditure Allocation Models.
17. **Schuller, Bernd-Joachim** (1986), Ekonomi och kriminalitet - en empirisk undersökning av brottsligheten i Sverige (Economics of crime - an empirical analysis of crime in Sweden)
18. **Walfridson, Bo** (1987), Dynamic Models of Factor Demand. An Application to Swedish Industry.
19. **Stålhammar, Nils-Olov** (1987), Strukturomvandling, företagsbeteende och förväntningsbildning inom den svenska tillverkningsindustrin (Structural Change, Firm Behaviour and Expectation Formation in Swedish Manufactury)
20. **Anxo, Dominique** (1988), Sysselsättningseffekter av en allmän arbetstidsförkortning (Employment effects of a general shortage of the working time)
21. **Mbelle, Ammon** (1988), Foreign Exchange and Industrial Development: A Study of Tanzania.
22. **Ongaro, Wilfred** (1988), Adoption of New Farming Technology: A Case Study of Maize Production in Western Kenya.
23. **Zejan, Mario** (1988), Studies in the Behavior of Swedish Multinationals.
24. **Görling, Anders** (1988), Ekonomisk tillväxt och miljö. Förorenings-struktur och ekonomiska effekter av olika miljövårdsprogram. (Economic Growth and Environment. Pollution Structure and Economic Effects of Some Environmental Programs).
25. **Aguilar, Renato** (1988), Efficiency in Production: Theory and an Application on Kenyan Smallholders.
26. **Kayizzi-Mugerwa, Steve** (1988), External Shocks and Adjustment in Zambia.
27. **Bornmalm-Jardelöw, Gunilla** (1988), Högre utbildning och arbetsmarknad (Higher Education and the Labour Market)
28. **Tansini, Ruben** (1989), Technology Transfer: Dairy Industries in Sweden and Uruguay.
29. **Andersson, Irene** (1989), Familjebeskattning, konsumtion och arbetsutbud - En ekonometrisk analys av löne- och inkomstelasticiteter samt policysimuleringar för svenska hushåll (Family Taxation, Consumption and Labour Supply - An Econometric Analysis of Wage and Income Elasticities and Policy Simulations for Swedish Households)
30. **Henrekson, Magnus** (1990), An Economic Analysis of Swedish Government Expenditure
31. **Sjöö, Boo** (1990), Monetary Policy in a Continuous Time Dynamic Model for Sweden

32. **Rosén, Åsa** (1991), Contributions to the Theory of Labour Contracts.
33. **Loureiro, Joao M. de Matos** (1992), Foreign Exchange Intervention, Sterilization and Credibility in the EMS: An Empirical Study
34. **Irandoust, Manuchehr** (1993), Essays on the Behavior and Performance of the Car Industry
35. **Tasiran, Ali Cevat** (1993), Wage and Income Effects on the Timing and Spacing of Births in Sweden and the United States
36. **Milopoulos, Christos** (1993), Investment Behaviour under Uncertainty: An Econometric Analysis of Swedish Panel Data
37. **Andersson, Per-Åke** (1993), Labour Market Structure in a Controlled Economy: The Case of Zambia
38. **Storrie, Donald W.** (1993), The Anatomy of a Large Swedish Plant Closure
39. **Semboja, Haji Hatibu Haji** (1993), Energy and Development in Kenya
40. **Makonnen, Negatu** (1993), Labor Supply and the Distribution of Economic Well-Being: A Case Study of Lesotho
41. **Julin, Eva** (1993), Structural Change in Rural Kenya
42. **Durevall, Dick** (1993), Essays on Chronic Inflation: The Brazilian Experience
43. **Veiderpass, Ann** (1993), Swedish Retail Electricity Distribution: A Non-Parametric Approach to Efficiency and Productivity Change
44. **Odeck, James** (1993), Measuring Productivity Growth and Efficiency with Data Envelopment Analysis: An Application on the Norwegian Road Sector
45. **Mwenda, Abraham** (1993), Credit Rationing and Investment Behaviour under Market Imperfections: Evidence from Commercial Agriculture in Zambia
46. **Mlambo, Kupukile** (1993), Total Factor Productivity Growth: An Empirical Analysis of Zimbabwe's Manufacturing Sector Based on Factor Demand Modelling
47. **Ndung'u, Njuguna** (1993), Dynamics of the Inflationary Process in Kenya
48. **Modén, Karl-Markus** (1993), Tax Incentives of Corporate Mergers and Foreign Direct Investments
49. **Franzén, Mikael** (1994), Gasoline Demand - A Comparison of Models
50. **Heshmati, Almas** (1994), Estimating Technical Efficiency, Productivity Growth And Selectivity Bias Using Rotating Panel Data: An Application to Swedish Agriculture
51. **Salas, Osvaldo** (1994), Efficiency and Productivity Change: A Micro Data Case Study of the Colombian Cement Industry
52. **Bjurek, Hans** (1994), Essays on Efficiency and Productivity Change with Applications to Public Service Production
53. **Cabezas Vega, Luis** (1994), Factor Substitution, Capacity Utilization and Total Factor Productivity Growth in the Peruvian Manufacturing Industry
54. **Katz, Katarina** (1994), Gender Differentiation and Discrimination. A Study of Soviet Wages
55. **Asal, Maher** (1995), Real Exchange Rate Determination and the Adjustment Process: An Empirical Study in the Cases of Sweden and Egypt
56. **Kjulin, Urban** (1995), Economic Perspectives on Child Care
57. **Andersson, Göran** (1995), Volatility Forecasting and Efficiency of the Swedish Call Options Market
58. **Forteza, Alvaro** (1996), Credibility, Inflation and Incentive Distortions in the Welfare State
59. **Locking, Håkan** (1996), Essays on Swedish Wage Formation
60. **Välilä, Timo** (1996), Essays on the Credibility of Central Bank Independence

61. **Yilma, Mulugeta** (1996), Measuring Smallholder Efficiency: Ugandan Coffee and Food-Crop Production
62. **Mabugu, Ramos E.** (1996), Tax Policy Analysis in Zimbabwe Applying General Equilibrium Models
63. **Johansson, Olof** (1996), Welfare, Externalities, and Taxation; Theory and Some Road Transport Applications.
64. **Chitiga, Margaret** (1996), Computable General Equilibrium Analysis of Income Distribution Policies in Zimbabwe
65. **Leander, Per** (1996), Foreign Exchange Market Behavior Expectations and Chaos
66. **Hansen, Jörgen** (1997), Essays on Earnings and Labor Supply
67. **Cotfas, Mihai** (1997), Essays on Productivity and Efficiency in the Romanian Cement Industry
68. **Horgby, Per-Johan** (1997), Essays on Sharing, Management and Evaluation of Health Risks
69. **Nafar, Nosratollah** (1997), Efficiency and Productivity in Iranian Manufacturing Industries
70. **Zheng, Jinghai** (1997), Essays on Industrial Structure, Technical Change, Employment Adjustment, and Technical Efficiency
71. **Isaksson, Anders** (1997), Essays on Financial Liberalisation in Developing Countries: Capital mobility, price stability, and savings
72. **Gerdin, Anders** (1997), On Productivity and Growth in Kenya, 1964-94
73. **Sharifi, Alimorad** (1998), The Electricity Supply Industry in Iran: Organization, performance and future development
74. **Zamanian, Max** (1997), Methods for Mutual Fund Portfolio Evaluation: An application to the Swedish market
75. **Manda, Damiano Kulundu** (1997), Labour Supply, Returns to Education, and the Effect of Firm Size on Wages: The case of Kenya
76. **Holmén, Martin** (1998), Essays on Corporate Acquisitions and Stock Market Introductions
77. **Pan, Kelvin** (1998), Essays on Enforcement in Money and Banking
78. **Rogat, Jorge** (1998), The Value of Improved Air Quality in Santiago de Chile
79. **Peterson, Stefan** (1998), Essays on Large Shareholders and Corporate Control
80. **Belhaj, Mohammed** (1998), Energy, Transportation and Urban Environment in Africa: The Case of Rabat-Salé, Morocco
81. **Mekonnen, Alemu** (1998), Rural Energy and Afforestation: Case Studies from Ethiopia
82. **Johansson, Anders** (1998), Empirical Essays on Financial and Real Investment Behavior
83. **Köhlin, Gunnar** (1998), The Value of Social Forestry in Orissa, India
84. **Levin, Jörgen** (1998), Structural Adjustment and Poverty: The Case of Kenya
85. **Ncube, Mkhululi** (1998), Analysis of Employment Behaviour in Zimbabwe
86. **Mwansa, Ladslous** (1998), Determinants of Inflation in Zambia
87. **Agnarsson, Sveinn** (1998), Of Men and Machines: Essays in Applied Labour and Production Economics
88. **Kadenge, Phineas** (1998), Essays on Macroeconomic Adjustment in Zimbabwe: Inflation, Money Demand, and the Real Exchange Rate
89. **Nyman, Håkan** (1998), An Economic Analysis of Lone Motherhood in Sweden
90. **Carlsson, Fredrik** (1999), Essays on Externalities and Transport
91. **Johansson, Mats** (1999), Empirical Studies of Income Distribution

92. **Alemu, Tekie** (1999), Land Tenure and Soil Conservation: Evidence from Ethiopia
93. **Lundvall, Karl** (1999), Essays on Manufacturing Production in a Developing Economy: Kenya 1992-94
94. **Zhang, Jianhua** (1999), Essays on Emerging Market Finance
95. **Mlima, Aziz Ponary** (1999), Four Essays on Efficiency and Productivity in Swedish Banking
96. **Davidsen, Björn-Ivar** (2000), Bidrag til den økonomisk-metodologiske tenkningen (Contributions to the Economic Methodological Thinking)
97. **Ericson, Peter** (2000), Essays on Labor Supply
98. **Söderbom, Måns** (2000), Investment in African Manufacturing: A Microeconomic Analysis
99. **Höglund, Lena** (2000), Essays on Environmental Regulation with Applications to Sweden
100. **Olsson, Ola** (2000), Perspectives on Knowledge and Growth
101. **Meuller, Lars** (2000), Essays on Money and Credit
102. **Österberg, Torun** (2000), Economic Perspectives on Immigrants and Intergenerational Transmissions
103. **Kalinda Mkenda, Beatrice** (2001), Essays on Purchasing Power Parity, RealExchange Rate, and Optimum Currency Areas
104. **Nerhagen, Lena** (2001), Travel Demand and Value of Time - Towards an Understanding of Individuals Choice Behavior
105. **Mkenda, Adolf** (2001), Fishery Resources and Welfare in Rural Zanzibar
106. **Eggert, Håkan** (2001), Essays on Fisheries Economics
107. **Andrén, Daniela** (2001), Work, Sickness, Earnings, and Early Exits from the Labor Market. An Empirical Analysis Using Swedish Longitudinal Data
108. **Nivorozhkin, Eugene** (2001), Essays on Capital Structure
109. **Hammar, Henrik** (2001), Essays on Policy Instruments: Applications to Smoking and the Environment
110. **Nannyonjo, Justine** (2002), Financial Sector Reforms in Uganda (1990-2000): Interest Rate Spreads, Market Structure, Bank Performance and Monetary Policy
111. **Wu, Hong** (2002), Essays on Insurance Economics
112. **Linde-Rahr, Martin** (2002), Household Economics of Agriculture and Forestry in Rural Vienam
113. **Maneschiöld, Per-Ola** (2002), Essays on Exchange Rates and Central Bank Credibility
114. **Andrén, Thomas** (2002), Essays on Training, Welfare and Labor Supply
115. **Granér, Mats** (2002), Essays on Trade and Productivity: Case Studies of Manufacturing in Chile and Kenya
116. **Jaldell, Henrik** (2002), Essays on the Performance of Fire and Rescue Services
117. **Alpizar, Francisco, R.** (2002), Essays on Environmental Policy-Making in Developing Countries: Applications to Costa Rica
118. **Wahlberg, Roger** (2002), Essays on Discrimination, Welfare and Labor Supply
119. **Piculescu, Violeta** (2002), Studies on the Post-Communist Transition
120. **Pylkkänen, Elina** (2003), Studies on Household Labor Supply and Home Production
121. **Löfgren, Åsa** (2003), Environmental Taxation – Empirical and Theoretical Applications
122. **Ivaschenko, Oleksiy** (2003), Essays on Poverty, Income Inequality and Health in Transition Economies

123. **Lundström, Susanna** (2003), On Institutions, Economic Growth and the Environment
124. **Wambugu, Anthony** (2003), Essays on Earnings and Human Capital in Kenya
125. **Adler, Johan** (2003), Aspects of Macroeconomic Saving
126. **Erlandsson, Mattias** (2003), On Monetary Integration and Macroeconomic Policy
127. **Brink, Anna** (2003), On the Political Economy of Municipality Break-Ups
128. **Ljungwall, Christer** (2003), Essays on China's Economic Performance During the Reform Period
129. **Chifamba, Ronald** (2003), Analysis of Mining Investments in Zimbabwe
130. **Muchapondwa, Edwin** (2003), The Economics of Community-Based Wildlife Conservation in Zimbabwe
131. **Hammes, Klaus** (2003), Essays on Capital Structure and Trade Financing
132. **Abou-Ali, Hala** (2003), Water and Health in Egypt: An Empirical Analysis
133. **Simatele, Munacinga** (2004), Financial Sector Reforms and Monetary Policy in Zambia
134. **Tezic, Kerem** (2004), Essays on Immigrants' Economic Integration
135. INSTÄLLD
136. **Gjirja, Matilda** (2004), Efficiency and Productivity in Swedish Banking
137. **Andersson, Jessica** (2004), Welfare Environment and Tourism in Developing Countries
138. **Chen, Yinghong** (2004), Essays on Voting Power, Corporate Governance and Capital Structure
139. **Yesuf, Mahmud** (2004), Risk, Time and Land Management under Market Imperfections: Applications to Ethiopia
140. **Kateregga, Eseza** (2005), Essays on the Infestation of Lake Victoria by the Water Hyacinth
141. **Edvardsen, Dag Fjeld** (2004), Four Essays on the Measurement of Productive Efficiency
142. **Lidén, Erik** (2005), Essays on Information and Conflicts of Interest in Stock Recommendations
143. **Dieden, Sten** (2005), Income Generation in the African and Coloured Population – Three Essays on the Origins of Household Incomes in South Africa
144. **Eliasson, Marcus** (2005), Individual and Family Consequences of Involuntary Job Loss
145. **Mahmud, Minhaj** (2005), Measuring Trust and the Value of Statistical Lives: Evidence from Bangladesh
146. **Lokina, Razack Bakari** (2005), Efficiency, Risk and Regulation Compliance: Applications to Lake Victoria Fisheries in Tanzania
147. **Jussila Hammes, Johanna** (2005), Essays on the Political Economy of Land Use Change
148. **Nyangena, Wilfred** (2006), Essays on Soil Conservation, Social Capital and Technology Adoption
149. **Nivorozhkin, Anton** (2006), Essays on Unemployment Duration and Programme Evaluation
150. **Sandén, Klas** (2006), Essays on the Skill Premium
151. **Deng, Daniel** (2006), Three Essays on Electricity Spot and Financial Derivative Prices at the Nordic Power Exchange
152. **Gebreeyesus, Mulu** (2006), Essays on Firm Turnover, Growth, and Investment Behavior in Ethiopian Manufacturing

153. **Islam, Nizamul Md.** (2006), Essays on Labor Supply and Poverty: A Microeconometric Application
154. **Kjaer, Mats** (2006), Pricing of Some Path-Dependent Options on Equities and Commodities
155. **Shimeles, Abebe** (2006), Essays on Poverty, Risk and Consumption Dynamics in Ethiopia
156. **Larsson, Jan** (2006), Four Essays on Technology, Productivity and Environment
157. **Congdon Fors, Heather** (2006), Essays in Institutional and Development Economics
158. **Akpalu, Wisdom** (2006), Essays on Economics of Natural Resource Management and Experiments
159. **Daruvala, Dinky** (2006), Experimental Studies on Risk, Inequality and Relative Standing
160. **García, Jorge** (2007), Essays on Asymmetric Information and Environmental Regulation through Disclosure
161. **Bezabih, Mintewab** (2007), Essays on Land Lease Markets, Productivity, Biodiversity, and Environmental Variability
162. **Visser, Martine** (2007), Fairness, Reciprocity and Inequality: Experimental Evidence from South Africa
163. **Holm, Louise** (2007), A Non-Stationary Perspective on the European and Swedish Business Cycle
164. **Herbertsson, Alexander** (2007), Pricing Portfolio Credit Derivatives
165. **Johansson, Anders C.** (2007), Essays in Empirical Finance: Volatility, Interdependencies, and Risk in Emerging Markets
166. **Ibáñez Díaz, Marcela** (2007), Social Dilemmas: The Role of Incentives, Norms and Institutions
167. **Ekbom, Anders** (2007), Economic Analysis of Soil Capital, Land Use and Agricultural Production in Kenya
168. **Sjöberg, Pål** (2007), Essays on Performance and Growth in Swedish Banking
169. **Palma Aguirre, Grisha Alexis** (2008), Explaining Earnings and Income Inequality in Chile
170. **Akay, Alpaslan** (2008), Essays on Microeconometrics and Immigrant Assimilation
171. **Carlsson, Evert** (2008), After Work – Investing for Retirement
172. **Munshi, Farzana** (2008), Essays on Globalization and Occupational Wages
173. **Tsakas, Elias** (2008), Essays on Epistemology and Evolutionary Game Theory
174. **Erlandzon, Karl** (2008), Retirement Planning: Portfolio Choice for Long-Term Investors
175. **Lampi, Elina** (2008), Individual Preferences, Choices, and Risk Perceptions – Survey Based Evidence
176. **Mitrut, Andreea** (2008), Four Essays on Interhousehold Transfers and Institutions in Post-Communist Romania
177. **Hansson, Gustav** (2008), Essays on Social Distance, Institutions, and Economic Growth
178. **Zikhali, Precious** (2008), Land Reform, Trust and Natural Resource Management in Africa
179. **Tengstam, Sven** (2008), Essays on Smallholder Diversification, Industry Location, Debt Relief, and Disability and Utility
180. **Boman, Anders** (2009), Geographic Labour Mobility – Causes and Consequences
181. **Qin, Ping** (2009), Risk, Relative Standing and Property Rights: Rural Household Decision-Making in China

182. **Wei, Jiegen** (2009), Essays in Climate Change and Forest Management
183. **Belu, Constantin** (2009), Essays on Efficiency Measurement and Corporate Social Responsibility
184. **Ahlerup, Pelle** (2009), Essays on Conflict, Institutions, and Ethnic Diversity
185. **Quiroga, Miguel** (2009), Microeconomic Policy for Development: Essays on Trade and Environment, Poverty and Education
186. **Zerfu, Daniel** (2010), Essays on Institutions and Economic Outcomes
187. **Wollbrant, Conny** (2010), Self-Control and Altruism
188. **Villegas Palacio, Clara** (2010), Formal and Informal Regulations: Enforcement and Compliance
189. **Maican, Florin** (2010), Essays in Industry Dynamics on Imperfectly Competitive Markets
190. **Jakobsson, Niklas** (2010), Laws, Attitudes and Public Policy
191. **Manescu, Cristiana** (2010), Economic Implications of Corporate Social Responsibility and Responsible Investments
192. **He, Haoran** (2010), Environmental and Behavioral Economics – Applications to China
193. **Andersson, Fredrik W.** (2011), Essays on Social Comparison
194. **Isaksson, Ann-Sofie** (2011), Essays on Institutions, Inequality and Development
195. **Pham, Khanh Nam** (2011), Prosocial Behavior, Social Interaction and Development: Experimental Evidence from Vietnam
196. **Lindskog, Annika** (2011), Essays on Economic Behaviour: HIV/AIDS, Schooling, and Inequality
197. **Kotsadam, Andreas** (2011), Gender, Work, and Attitudes
198. **Alem, Yonas** (2011), Essays on Shocks, Welfare, and Poverty Dynamics: Microeconometric Evidence from Ethiopia
199. **Köksal-Ayhan, Miyase Yesim** (2011), Parallel Trade, Reference Pricing and Competition in the Pharmaceutical Market: Theory and Evidence
200. **Vondolia, Godwin Kofi** (2011), Essays on Natural Resource Economics
201. **Widerberg, Anna** (2011), Essays on Energy and Climate Policy – Green Certificates, Emissions Trading and Electricity Prices
202. **Siba, Eyerusalem** (2011), Essays on Industrial Development and Political Economy of Africa
203. **Orth, Matilda** (2012), Entry, Competition and Productivity in Retail
204. **Nerman, Måns** (2012), Essays on Development: Household Income, Education, and Female Participation and Representation
205. **Wicks, Rick** (2012), The Place of Conventional Economics in a World with Communities and Social Goods
206. **Sato, Yoshihiro** (2012), Dynamic Investment Models, Employment Generation and Productivity – Evidence from Swedish Data
207. **Valsecchi, Michele** (2012), Essays in Political Economy of Development
208. **Teklewold Belayneh, Hailemariam** (2012), Essays on the Economics of Sustainable Agricultural Technologies in Ethiopia
209. **Wagura Ndiritu, Simon** (2013), Essays on Gender Issues, Food Security, and Technology Adoption in East Africa
210. **Ruist, Joakim** (2013), Immigration, Work, and Welfare
211. **Nordén, Anna** (2013), Essays on Behavioral Economics and Policies for Provision of Ecosystem Services
212. **Yang, Xiaojun** (2013), Household Decision Making, Time Preferences, and

Positional Concern: Experimental Evidence from Rural China

213. **Bonilla Londoño, Jorge Alexander** (2013), Essays on the Economics of Air Quality Control
214. **Mohlin, Kristina** (2013), Essays on Environmental Taxation and Climate Policy
215. **Medhin, Haileselassie** (2013), The Poor and Their Neighbors: Essays on Behavioral and Experimental Economics
216. **Andersson, Lisa** (2013), Essays on Development and Experimental Economics: Migration, Discrimination and Positional Concerns
217. **Weng, Qian** (2014), Essays on Team Cooperation and Firm Performance
218. **Zhang, Xiao-Bing** (2015), Cooperation and Paradoxes in Climate Economics
219. **Jaime Torres, Monica Marcela** (2015) Essays on Behavioral Economics and Policy Design
220. **Bejenariu, Simona** (2015) Determinants of Health Capital at Birth: Evidence from Policy Interventions
221. **Nguyen, Van Diem** (2015) Essays on Takeovers and Executive Compensation
222. **Tolonen, Anja** (2015) Mining Booms in Africa and Local Welfare Effects: Labor Markets, Women's Empowerment and Criminality
223. **Hassen, Sied** (2015) On the Adoption and Dis-adoption of Household Energy and Farm Technologies
224. **Moursli, Mohamed-Reda** (2015) Corporate Governance and the Design of Board of Directors
225. **Borcan, Oana** (2015) Economic Determinants and Consequences of Political Institutions
226. **Ruhinduka, Remidius Denis** (2015) Essays on Field Experiments and Impact Evaluation
227. **Persson, Emil** (2016) Essays on Behavioral and Experimental Economics, Cooperation, Emotions and Health.
228. **Martinangeli, Andrea** (2017) Bitter divisions: inequality, identity and cooperation
229. **Björk, Lisa** (2017) Essays on Behavioral Economics and Fisheries: Coordination and Cooperation
230. **Chegere, Martin Julius** (2017) Post-Harvest Losses, Intimate Partner Violence and Food Security in Tanzania
231. **Villalobos-Fiatt, Laura** (2017) Essays on forest conservation policies, weather and school attendance
232. **Yi, Yuanyuan** (2017) Incentives and Forest Reform: Evidence from China
233. **Kurz, Verena** (2017) Essays on behavioral economics: Nudges, food consumption and procedural fairness
234. **Yashodha, Yashodha** (2017) Contract Choice and Trust in Informal Groundwater Markets
235. **Meles, Tensay Hadus** (2017) Power Outages, Increasing Block Tariffs and Billing Knowledge
236. **Mühlrad, Hanna** (2018) The Impact of Reproductive and Birth Technologies on Health, Fertility and Labor Outcomes
237. **Gakii Gatua, Josephine** (2018) Primary Health Care Interventions and Social Ties in Kenya
238. **Felgendreher, Simon** (2018) Essays on behavioral economics and the effects of the colonial rule on Java
239. **Demeke, Eyoual** (2019) Essays on Environmental and Behavioral Economics
240. **Kiss, Tamas** (2019) Predictability in Equity Markets: Estimation and Inference
241. **Khomenko, Maksym** (2019) Essays on the Design of Public Policies and .

Regulations
242. **Mukanjari, Samson** (2019) Climate Policy and Financial Markets
243. **Schürz, Simon** (2020) Economic and Intergenerational Decision-Making in Families
244. **Lau, Debbie** (2020) Empirical Essays on Education and Health Policy Evaluation
245. **Ho, Hoang-Anh** (2020) Essays on Culture, Institutions, and Economic Development
246. **Rubio Ramos, Melissa** (2020) The Economics of Coercive Institutions, Conflict, and Development
247. **Lindahl, Anna** (2021) Empirical tests of exchange rate and stock return models
248. **Sjöholm, Carolin** (2021) Essays on public policy in the informal sector context
249. **Tian, Ruijie** (2021) Impacts of Climate Policy and Natural Disasters: Evidence from China
250. **Larsson, Sebastian** (2021) Providers and Profiteers: Essays on Profits and Competition in the Provision of Public Services
251. **Norrgren, Lisa** (2022) Carpe Diem or Seize your Health? The Economics of Time Preferences, Health, and Education
252. **Andersson, Lina** (2022) Emotions in Game Theory: Fear, friendliness and hostility
253. **Enlund, Jakob** (2022) Three essays in applied economics
254. **Berggren, Andrea** (2023) Choices among Doctors, Students and Primary Care Providers: Empirical Evidence from Sweden
255. **Hansson, Magnus** (2023) Decentralized Finance and Central Bank Communication
256. **Bilén, David** (2023) Sustainable Consumption and Prosocial Actions