Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

School of Social Sciences and Philosophy

Department of Economics

Environmental Behaviour and Decision-Making - Evidence from Laboratory, Online and Natural Experiments

Stefano Ceolotto

Supervisor: Prof. Eleanor Denny

Head of School: Prof. Carol Newman

2022

A thesis submitted to the University of Dublin, Trinity College

In application for the degree of Doctor of Philosophy

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and is my own work.

I agree to submit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgment.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Stefano Ceolotto

17th January 2022

Abstract

Tackling climate change and keeping global temperatures from rising by more than 1.5°C above pre-industrial levels is one of the major and most challenging issues that our societies are facing. Households are among the prime contributors to annual greenhouse gas emissions, with twenty percent of emissions being generated by residential energy consumption. Therefore, understanding individual decision-making in relation to environmental matters is key to designing effective climate change policies.

A vast body of literature studies how to encourage pro-environmental behaviours, from the adoption of price interventions and regulations, to the use of less invasive measures like information campaigns and other nudges. The research indicates a significant variation in the effectiveness of such policies and interventions, depending on the targeted behaviour, the selected measure and the context in which it was applied.

This thesis aims to provide a clearer picture of the link between the environment and individual decision-making by exploring it from three different perspectives. It studies how performing a pro-environmental behaviour affects people's decisions to undertake subsequent environmentally-friendly actions; how the framing of energy information impacts household consumption choices; and how environmental factors affect individuals' voting decisions. Considering such a cohesive framework is important because environmental behaviour can affect policy design

through the influence of environmental factors on voting, and, in turn, policy design determines the performance of pro-environmental behaviour.

Acknowledgements

I am immensely grateful to Declan Sheehan, my Grattan Scholarship sponsor. None of this would have been possible without you. My deepest gratitude goes to the Department of Economics for accepting me into the PhD programme and providing me with all the resources to successfully complete it. The research presented in this thesis was also funded by the TRiSS Academic Research Fellowship under the TRiSS Academic Research Projects call 2017/18; and the Irish Research Council under the New Horizons programme, Research Project REPRO/2015/28 entitled NEEPD.

I am indebted to my supervisor Prof. Eleanor Denny (Trinity College Dublin) for believing in me more than I did. You always had encouraging words to lift me up when I put myself down. And your feedback and mentorship over the past four and a half years made me become a better researcher than I ever hoped to be.

I am grateful to my Dr. Benjamin Elsner (University College Dublin) for his invaluable help with econometrics questions and, above all, for trusting me and treating me as a collegue since the very beginning of our collaboration. A huge thanks goes also to my co-authors Dr. Luna Bellani (University of Konstanz) and Dr. Nico Pestel (Maastricht University): working with you has been an honour and a pleasure.

I would like to thank Dr. James Carroll for the feedback on early versions of this dissertation. And Dr. Bryan Coyne for the constant support in the daily PhD life.

There are no words that truly reflect my gratitude for my mother, Renata, and my father, Paolo. Thank you for all the sacrifices you made, for your unconditioned love, for the education you gave me, and for constantly pushing me to be the best

version of myself. I would not be even a fraction of the man I am today without you.

A huge thanks to all my family. In particular to my grandmother Elena, who has always been my number one supporter. And to my grandparents Guido and Franca, who sadly left us two years ago. I miss you.

Finally, I would like to thank Serena for remaining at my side through the moments of frustration, when I could not see the light at the end of the tunnel, despite the difficulties and the distance. I know this journey has been as hard for you as it has been for me. Without your endless love, I would never have been able to get to the end of it. Although a meager compensation for the hardship I put you through, let me say that everything I accomplished during my PhD, and that culminates in this thesis, I dedicate to you.

Contents

List of Figures

List of Tables

Nomenclature and List of Abbreviations

CCA	Climate change attitude
CEM	Coarsened exact matching
DA	Donation-after experimental version
DF	Donation-first experimental version
Don	Having made the donation ($Donation = 1$) in the donation-first experimental version
EA	Environmental attitude
EUT	Expected utility theory
DCE	Discrete choice experiment
NoDon	Not having made the donation ($Donation = 0$) in the donation-first experimental version
PEB	Pro-environmental behaviour
PG	Public good
PM10	Particulate matter with an aerodynamic diameter smaller ten micrometers (μm)
RDCE	Randomised discrete choice experiment
SEU	Subjective expected utility
SOEP	German socio-economic panel survey
Total EECA	Total environmental and climate change attitude
WARP	Weak axiom of revealed preference
$v(x)$	Subjective value function (prospect theory)
μg	Microgram, $\mu g = 1 \times 10^{-6} g$
μm	Micrometer (or micron), $\mu m = 1 \times 10^{-6} m$
$\pi(p)$	Probability weighting function (prospect theory)

1 Introduction

The world's climate is changing, with increased concentrations of greenhouse gases, higher atmosphere and oceans temperatures, diminished snow and ice coverage and more extreme and frequent severe weather events (IPCC, 2014; WMO, 2020). The United Nations Intergovernmental Panel on Climate Change reports that, in large part, the causes of this change are anthropogenic, with human activities being responsible for approximately 1°C of global warming with respect to pre-industrial levels (IPCC, 2014, 2018). Tackling climate change and keeping global temperatures from rising by more than 1.5°C above pre-industrial levels is one of the major challenges that our societies are facing. Unfortunately, it is an extremely difficult one, due to the unique combination of global causes and long-term, uncertain and widespread consequences that climate change presents (Stern, 2008).

Already at the 1992 Rio Earth Summit it was stated that "altering consumption patterns is one of humanity's greatest challenges in the quest for environmentally sound and sustainable development" (Sitarz, 1993). Twenty percent of annual emissions is generated by residential energy consumption (International Energy Agency, 2019), with this phenomenon being particularly pronounced in developed countries (European Environmental Agency, 2020; U.S. Energy Information Administration, 2020). Therefore, scholars and researchers have highlighted the importance of individuals' decisions on environmental matters (DEFRA, 2008;

Dietz et al., 2009; Steg and Vlek, 2009). However, in the context of climate change, people often perceive their actions to have negligible impacts, either harmful or beneficial. In addition, environmental decisions typically involve a trade-off between *individual* costs versus *collective* benefits. Such a combination hinders the individuals' propensity to act, thus posing severe limitations to effectiveness of measures aimed at combating climate change.

This thesis is motivated by the belief that individual action is crucial to solve environmental problems. The economic literature has investigated individual decision-making processes since its earliest days. Throughout the following chapters, various mechanisms and factors influencing peoples' choices are analysed to improve the understanding of the link between the environment and individual decisions. The outcomes highlight the importance to incentivise the uptake of environmentally-friendly actions since this is likely to induce households to perform more of them in the future. But they also call for policymakers to carefully evaluate the national context and population's characteristics to design effective policies. Finally, they prove the mutual nature of the link showing the impact of environmental factors on peoples' decisions.

The remained of this chapter is thus structured. Section 1.1 outlines the evolution of economic theories on decision-making. Section 1.2 presents the applications of such theories in the context of environmental behaviour. Finally, Section 1.3 details the aspects of decision-making investigated in this thesis and its contributions to the current debate.

1.1 Economics and human behaviour

Economics is often characterized as the study of how agents make decisions (Mankiw and Taylor, 2014). The classic theory lies on the premise that individuals will choose the option that maximizes their utility. However, since utility is an unobservable entity, how can researchers be sure that individuals' choices are consistent with utility maximization? This fundamental question and the answer to it were proposed by Samuelson (1938) with the formalization of the Weak Axiom of Revealed Preference (WARP), whereby individuals behave as if they are maximizing utility if and only if they do not choose option B over option A after having previously chosen option A over option B.

The theory on decision-making has then evolved to explain decisions under risk and uncertainty. von Neumann and Morgenstern (1944) showed that under a certain set of assumptions it is possible to rank individuals' preferences over lotteries, the economic prototype of choices between risky alternatives. Based on the axioms of completeness, transitivity, continuity and independence, the Expected Utility Theory (EUT) posits that, when faced with the choice between risky prospects, agents will choose the alternative that provides the highest expected utility — i.e. that maximizes the value weighed by the probability of realization. In such instances where probabilities cannot be objectively determined, Savage (1954) suggested the adoption of subjective probabilities based on personal or social beliefs. This alternative conceptualization led to the reformulation of expected utility theory into a Subjective Expected Utility (SEU) model.

The economic and behavioural science literature has demonstrated that there are several instances in which individuals' behaviour does not align with the axioms of expected utility theory. The so-called Allais Paradox (Allais, 1953) shows that people tend to overvalue certain prospects (what Kahneman and Tversky (1979) call the certainty effect). Loss aversion (Kahneman and Tversky, 1979) implies that individuals behave differently when prospects entail gains than when they entail losses: risk aversion in the positive domain is coupled with risk seeking in the negative one. More generally, the way in which prospects are framed impacts individuals' decisions and can lead to preference reversal (Tversky and Kahneman, 1981). The Ellsberg Paradox (Ellsberg, 1961) posits that agents behave according to (subjective) expected utility theory when facing situations involving risk but not in situations involving uncertainty.

To provide a coherent explanation to these inconsistencies, Kahneman and Tversky (1979) developed the Prospect Theory. This approach implies that the agents' decision-making process is divided into two subsequent phases. Firstly, an editing phase in which a preliminary analysis of prospects yields a simple representation of them. Then, the edited prospects are evaluated and the one presenting the highest value is chosen. The revolution with respect to the classic expected utility approach lies in the editing phase, in which agents adopt several operations to reorganize, reformulate and transform the prospect's outcomes and probabilities, generating the inconsistencies highlighted by the previous research. The result is an evaluation defined in terms of gains and losses with respect to a reference point, which shows diminishing sensitivity in both domains but with a greater responsiveness to losses with respect to equally sized gains. Later, Tversky and Kahneman (1992) proposed a modification of prospect theory using a cumulative formulation of the probability weighting function (Cumulative Prospect Theory).

A related strand of the decision making literature examines how agents acquire and process the available information to derive their subjective evaluations. People may have limited computing ability or lack the tools to retrieve all the relevant information required to carry out the maximization exercise implied by the expected utility theory. Tversky and Kahneman (1974) argued that when they have to evaluate the probability of uncertain events individuals tend to rely on the degree to which these events are representative of a certain underlying parent population (*representativeness heuristic*); or they assess them based on the ease with which they can recall similar events (*availability heuristic*); or that they anchor their estimations on a starting point, which sometimes is the result of partial computations, subsequently failing to properly adjust them (*anchoring heuristic*). More recently, Stewart et al. (2006) conceptualized the decision-making process as an ordinal comparison based on prior experience. The Decision-by-Sampling approach posits that people retrieve from their memories an evaluation of attributes, use it to compare and rank attributes of a current prospect, and then choose the one with the highest rank instead of properly deriving expected values. In all these instances, people disregard actual probabilities, ignore relevant information or fail to adopt appropriate computational exercises, but rather employ (mental) shortcuts in their decision processes.

Many of the critiques and violations of the expected utility theory mentioned above are rooted in the cognitive dimension. Towards the end of the 1990s, a number of studies started to depart from the purely cognitive aspect of decision-making and incorporate emotions as crucial components. The literature on affective decision-making (see Lerner et al., 2015; Lowenstein and Lerner, 2003, for a review) investigates the influence of emotions on decision processes, highlighting that both integral emotions (those arising from the decision and its consequences) and incidental emotions (those generated by unrelated events, circumstances or decisions) play a role in shaping individuals' choices. In addition, emotions can also affect depth of processing, hence the degree to which people rely on some of the aforementioned heuristics.

Further developments have studied how to influence or help individuals in their decisions. Research on the framing of messages and information (Kahneman and Tversky, 1984; Lowenstein, 1988; Lowenstein and Prelec, 1992; Lowenstein and Thaler, 1989; Tversky and Kahneman, 1981) lies on the predication that "decision-makers respond differently to different but objectively equivalent descriptions of the same problem" (Levin et al., 1998). *Nudging* (Thaler and Sunstein, 2009) attempts to alter or correct people's behaviour by changing the choice architecture without altering the available options or the economic incentives, thus generating benefits for those individuals who are more prone to make errors without being harmful for those who are fully rational (Camerer et al., 2003). Somewhat differently, the literature on behavioural spillover investigates how the performance of an initial behaviour affects subsequent behaviours or decisions (Thøgersen, 1999).

1.2 Environmental decision-making

One area in which individuals' choices play a crucial role, and which has become ever more prominent in recent years, is that of environment-related actions and decisions. Pro-environmental behaviours (PEBs) are intended to cause the smallest possible harm to the environment or benefit it (Steg and Vlek, 2009). Examples include, among others, energy, water and fuel conservation, energy- or fuel-efficiency improvements, recycling, green consumerism and use of public transport or cycling.

Price interventions and legal regulations are obvious determinants of individuals' decision to engage in pro-environmental behaviours. However, a considerable interest has also been dedicated to non-price interventions aimed at incentivising green and sustainable choices. These can include making information more readily available or easy to understand, relying on social conformity and comparison, and using other nudges to promote the uptake of environmentally-friendly actions.

General information campaigns have proven to be not very effective in leading to behaviour changes (Koop et al., 2019; Steg and Vlek, 2009). A vast literature has studied how to better frame and present information or interventions, so as to enhance their effectiveness. While gain frames have been shown to lead to improved attitudes about recycling, pollution and climate change (Feinberg and Willer, 2011; Loroz, 2007), loss framing is more effective at leading to the target outcome or produces more sizable changes (Gonzales et al., 1988; Grazzini et al., 2018; Lord, 1994; Nabi et al., 2018; Poortinga and Whitaker, 2018; White et al., 2011).

In general, it appears that "framing the consequence[s] of environmentally-relevant decisions as a loss that must be prevented is more likely to lead to behavioural change than framing [them] as a gain to be achieved" (Ropert Homar and Cvelbar, 2021). In the context of energy- and fuel-efficiency, various studies have analysed the effect of reframing energy and fuel consumption in a variety of ways, from short-run and long-run monetary information to personalised information (Allcott and Knittel, 2019; Andor et al., 2020; Carroll et al., 2021, 2016a,b; Davis and Metcalf, 2016; Department of Energy and Climate Change, 2014; Heinzel, 2012; Heinzle and Wüstenhagen, 2012; Jain et al., 1994; Kallbekken et al., 2013; Newell and Siikamäki, 2014; Shen and Saijo, 2009). Results have been mixed, with some papers finding the intervention to lead to efficiency improvements, while others failing to detect a significant effect.

Social norms are shared rules of conduct sustained by approval and disapproval (Elster, 1989) and are characterised by being implicit, conditionally followed, and externally motivated (Farrow et al., 2017). A basic distinction is drawn between descriptive (what people do) and injunctive (what people approve of doing) norms, with further categorisations dividing between perceived (subjective beliefs) and actual (objective beliefs) norms, and between prescriptive (what to do) and proscriptive (what not to do) norms. Social norms have been found to significantly affect pro-environmental behaviours like water conservation (Bernedo et al., 2014; Ferraro et al., 2011; Jaeger and Schultz, 2017; Landon et al., 2018), energy conservation (Allcott, 2011b; Costa and Kahn, 2013), recycling (Nigbur et al., 2010; Viscusi et al., 2014) or soil conservation practices (Willy and Holm-Müller, 2013). In general, descriptive norms appear more effective than injunctive ones at incentivising the target behaviour (Farrow et al., 2017).

Social comparison, on the other hand, refers to the process of giving households information about their performance on a certain behaviour in relation to that of comparable households (Andor and Fels, 2018). Evidence suggests that social comparison positively affects energy use (Alberts et al., 2016; Allcott, 2011b; Allcott and Rogers, 2014; Ayres et al., 2013; Costa and Kahn, 2013; Ferraro and Price, 2013; Mizobuchi and Takeuchi, 2013; Peschiera et al., 2010; Schultz et al., 2015; Seyranian et al., 2015; Tiefenbeck et al., 2013) and that it can also improve the efficacy of other treatments (Ferraro and Price, 2013; Mizobuchi and Takeuchi, 2013; Schultz et al., 2015; Tiefenbeck et al., 2013). However, long-run effects, or effects on other pro-environmental behaviours are more ambiguous (Andor and Fels, 2018; Koop et al., 2019).

Commitment devices are measures that help individuals binding themselves to the performance of a certain future behavior in an attempt to reduce the emergence of time inconsistencies and procrastination. In this sense, goal setting entails the commitment to a concrete reference point and/or deadline rather than a more vaguely defined intention. Self-set goals appear to be an effective tool to allow people to reduce energy (Harding and Hsiaw, 2014; McCalley and Midden, 2002; Winett et al., 1979) and water (Jaeger and Schultz, 2017; Tijs et al., 2017) consumption. Conversely, externally-set goals are generally less effective, albeit with some exceptions (Abrahamse et al., 2007; Winett et al., 1982).

To achieve the lifestyle changes required to tackle the environmental problems our societies are and will be facing, however, it is necessary to ensure that people do not engage in PEBs as one-off occurrences. In addition to the long-term effects of the aforementioned interventions, another crucial aspect to be considered is how the decision to engage in a PEB depends on the performance of past PEBs. Behavioural spillover is the phenomenon by which performing a behaviour either increases (positive spillover) or reduces (negative spillover) the likelihood of performing subsequent (un)related behaviours (Thøgersen, 1999). In the context of PEBs, the spillover literature has found mixed results, with some studies showing positive effects (Alacevich et al., 2021; Lacasse, 2017; Lanzini and Thøgersen, 2014; Margetts and Kashima, 2017; Thøgersen and Noblet, 2012; Thomas et al., 2016), and others negative effects (Geng et al., 2016; Tiefenbeck et al., 2013; Truelove et al., 2016). In addition, surprisingly little attention has been given to long-term dynamics.

Finally, not only do individuals' decisions impact the environment, but also environmental factors affect people's decision-making processes (Lerner et al., 2015; Lowenstein and Lerner, 2003). Meteorologic conditions such as temperature, precipitations or solar radiation, have been shown to affect turnout and voting preferences (Eisinga et al., 2012a,b; Meier et al., 2019; Sforza, 2014), stock markets (Hirshleifer and Shumway, 2003; Kamstra et al., 2003) and sentence decisions (Heyes and Saberian, 2019). In addition, poor air quality conditions have been linked to reduced productivity (Chang et al., 2016, 2019; Graff Zivin and Neidell, 2012; Heyes et al., 2016; Huang et al., 2020; Klingen and van Jos N. Ommeren, 2020; Künn et al., 2021; Meyer and Pagel, 2017), stock returns (Levy and Yagil, 2011), criminal activity (Bondy et al., 2020; Burkhardt et al., 2019; Herrnstadt et al., 2020) and cognitive performance (Ebenstein et al., 2016).

1.3 Aspects of decision-making examined in this thesis

This thesis explores the link between the environment and individual decision-making by investigating it from three different perspectives. Specifically, it studies how engaging in a pro-environmental behaviour affects people's decisions to perform subsequent environmentally-friendly actions; how the framing of energy information impacts consumption choices of household appliances; and how environmental factors affect individuals' decision-making through their influences on emotions.

Chapter 2: Investigating the presence and direction of behavioural spillovers between PEBs

The efficacy of policies aimed at incentivising a more environmentally-friendly lifestyle lies in part on individuals' ability to perform multiple pro-environmental behaviours without deviating from this course of action. The previous literature analysing behavioural spillovers in the environmental context found mixed evidence, with some studies highlighting the presence of a positive spillover (Alacevich et al., 2021; Lacasse, 2017; Lanzini and Thøgersen, 2014; Margetts and Kashima, 2017; Thøgersen and Noblet, 2012; Thomas et al., 2016), others of a negative one (Geng et al., 2016; Tiefenbeck et al., 2013; Truelove et al., 2016); and others yet reporting mixed or inconclusive results (Carrico et al., 2018; Thøgersen, 1999; Thøgersen and Ölander, 2003). However, many of these paper rely on stated rather than actual behaviour, they do not consider the persistence of the effect, nor whether it is limited to the environmental domain.

This study investigates the presence of behavioural spillovers through a lab experiment that links the voluntary performance of an initial PEB (making a donation to a pro-environmental organization) to the engagement in a series of subsequent environmentally-friendly actions (contributing for the provision of environmental public goods). Such a framework, by considering actual behaviour, allows for a more accurate representation of spillover effects and it controls for potential endogeneity problems with the use of a matching estimation. In addition, the design of the public goods game allows the testing of the persistence of the effect as well as whether it can spread across different domains (since the game includes different types of public goods).

The findings highlight the presence of a strong, positive behavioural spillover, with participants who performed the initial PEB contributing on average 11 experimental units more for the provision of public goods, a result which is confirmed in the matched sample. The analysis also shows that the effect does not disappear after a few rounds but remains at a persistent level throughout the entire game, and that it is stronger for environmental public goods than for generic ones. Overall, this suggests that if individuals engage in PEBs they are more likely to perform further PEBs in the future and to sustain a more environmentally-friendly lifestyle.

Chapter 3: Studying the effects of energy cost information on consumption choices

Energy efficiency is seen as a fundamental tool to reduce greenhouse gas emissions and combat climate change (McKinsey, 2010) and it represents a key policy focus for many Governments and organizations (European Parliament, 2018). More energy efficient products typically present a trade-off between a higher upfront purchasing price and lower lifetime operating costs. The individuals' inability to properly evaluate such a trade-off generates an *energy efficiency gap* (Jaffe and Stavins, 1994), whereby investments in energy-efficiency that would be beneficial from a net present value standpoint are not undertaken.

Energy labels have been widely adopted (Collaborative Labeling and Appliance Standards Program, 2005) to make energy information more readily available to consumers and facilitate the comparison among different products as well as between purchasing price and operating costs. Several studies have investigated whether framing the energy information provided by the labels in monetary terms can improve their effectiveness (Allcott and Knittel, 2019; Andor et al., 2020; Carroll et al., 2021, 2016a; Davis and Metcalf, 2016; Department of Energy and Climate Change, 2014; Heinzel, 2012; Jain et al., 1994; Kallbekken et al., 2013; Newell and Siikamäki, 2014; Shen and Saijo, 2009). These papers have found mixed results, and it is not entirely clear whether they are due to differences in products, contexts or methodologies.

This study sheds light on such an ambiguity through an online randomised discrete choice experiment (RDCE) that adopts the same methodology in four countries. Specifically, it asks respondents from Canada, the Republic of Ireland, the United Kingdom and the United States to express their preferences for tumble dryers which vary over a number of attributes. Energy information is reported in three ways: (i) as the standard energy label which is customary in each country; (ii) as the 10-years energy cost based on average usage; and (iii) as the "personalised" 10-years energy cost based on self-reported individual usage.

The key insight is that the effectiveness of monetary energy information varies across countries, with no significant effects in Ireland and the United States, a negative effect in Canada and a positive one in the United Kingdom. Decomposing the analysis by individual characteristics highlights that the negative effect of personalised information in Canada comes primarily from households with a lower usage and that providing monetary information seems to crowd out individuals who would buy a more energy efficient product for environmental motivations. This suggests that there is no silver bullet to improve the efficacy of energy efficiency information and policymakers should carefully tailor interventions to their specific national and social context.

Chapter 3: Evaluating the affective impact of environmental factors on decision-making

Not only do people's decisions have consequences on the environment, but also environmental factors affect decision-making (Lerner et al., 2015; Lowenstein and Lerner, 2003). A poor air quality has been shown to impair cognitive functions and generate negative emotions (Chen, 2019; Graff Zivin and Neidell, 2018; Trushna et al., 2020), thus having a knock-on effect on individuals' choices (Chen, 2019).

The economic literature has documented the effects of air pollution in a variety of contexts, from criminal activity (Bondy et al., 2020; Burkhardt et al., 2019;

Herrnstadt et al., 2020), to investors' behaviour (Heyes et al., 2016; Levy and Yagil, 2011), to school attainments (Heissel et al., 2021; Lavy et al., 2014), to productivity (Archsmith et al., 2018; Chang et al., 2019; Graff Zivin and Neidell, 2012; Heyes et al., 2019; Künn et al., 2021). These studies, however, focus on specific subgroups of the population and therefore it is challenging to ascertain their external validity.

The analysis presented in this chapter uses national and state election in Germany as a large-scale real-world experiment to investigate the affective impact of PM10 on decision-making. The main measure of interest is represented by the vote share of incumbent parties, since it informs about voters' support for the political status quo as well as their risk preferences. This is then integrated with the information provided by two large-scale representative surveys: the *Politbarometer* and the German Socio-Economic Panel (SOEP).

The results evidence that a higher concentration of PM10 in the air reduces the support for incumbent parties at the expenses of the established opposition. This trend is confirmed by the survey data, which additionally highlights how such an effect is driven by an increase in negative emotions and not by a change in people's perception of the socio-economic situation. These findings provide a compelling evidence of the affective impacts of environmental factors on the decisions of the general population.

Table 1.1: Summary of research studies

	Donation to a pro-environmental organization and behavioural spillover: Evidence from the lab	**Putting a new 'spin' on energy labels: measuring the impact of reframing energy efficiency on tumble dryer choices**	**Air Pollution Affects Decision-Making: Evidence from the Ballot Box**
Research questions	Does performing a PEB generate a positive or negative behavioural spillover? Is the behavioural spillover persistent? Does the behavioural spillover spreads across different domains?	Does providing long-term average energy consumption information in monetary terms increase uptake of more efficient technologies? Does providing personalised long-term energy consumption information in monetary terms increase uptake of more efficient technologies? Does the effect differ across countries?	Does ambient air pollution affect voting decisions? What are the mechanisms driving the effect?
Methodology	Laboratory experiment with multi-round public goods game. Econometric analysis: Fixed effects panel regressions; Coarsened exact matching estimation.	Online randomised discrete choice experiment. Econometric analysis: Mixed multinomial logit estimations; Willingness-to-pay analysis.	Econometric analysis: Fixed effects regressions on panel and survey data; Instrumental variable estimation.
Authors	Stefano Ceolotto	Stefano Ceolotto, Eleanor Denny	Luna Bellani, Stefano Ceolotto, Benjamin Elsner, Nico Pestel

Table 1.1 — continued

	Study 1	**Study 2**	**Study 3**
Own Contribution	Stefano Ceolotto designed and ran the experiment, conducted the data analysis and drafted the manuscript.	Stefano Ceolotto led the data analysis and manuscript drafting. He received valuable contributions and feedback from his co-author throughout.	Stefano Ceolotto led the data analysis, contributed to the conceptualisation and to the manuscript drafting. He received valuable contributions and feedback from his co-authors throughout.
Publication status (as of November 2021)	Expected to be submitted to *Ecological Economics* in 2022.	TCD Working Paper (TEP1521). Expected to be submitted to *The Energy Journal* in 2022.	IZA Discussion Paper (DP14718). Submitted to *American Economic Review*.

Chapter 5: Conclusions

The final chapter of the thesis presents the concluding remarks, discussing the implications of the findings of the previous three chapters, the limitations of those analyses, and the possibilities for future research avenues.

2 Donation to a pro-environmental organization and behavioural spillover: Evidence from the lab

2.1 Introduction

It is widely recognized that the climate system is undergoing profound changes (IPCC, 2014). The concentration of greenhouse gases has increased, especially in the past twenty years; the atmosphere and oceans have warmed; the amounts of snow and ice have diminished; the sea level has risen and the oceans' heat storage and acidification have increased, reducing their ability to moderate climatic changes (IPCC, 2014; WMO, 2020). Extreme weather events have become more frequent and severe (IPCC, 2014). In the past couple of years there have been intense precipitation events and flooding in Africa, Asia and Europe; increases in warm temperature extremes that exacerbate the risk and frequency of wildfires; and extremely active North Atlantic cyclone seasons (IPCC, 2014; WMO, 2020). These changes are having widespread impacts on natural and human systems, with an estimated 9.8 million people displaced in 2020 due to hydrometeorological hazards and disasters, disruptions to the agricultural sector which have elevated the levels of food insecurity, causing tens of billions of dollars in economic losses and numerous casualties (Agarwal et al., 2021; IPCC, 2014; Wing et al., 2021; WMO, 2020).

These changes are largely due to anthropogenic influences (IPCC, 2014). To limit climate change and prevent global average temperatures from rising to more than 1.5°C above pre-industrial levels, interventions to achieve a substantial and sustained reduction of greenhouse gas emissions are required (IPCC, 2014). Although governmental and international interventions are deemed more effective (Hale, 2010; Stavins, 2008), research has suggested that there is scope for individual actions (DEFRA, 2008; Dietz et al., 2009; Steg and Vlek, 2009).

When it comes to incentivising pro-environmental behaviours (PEBs) a key concept is that of behavioural spillover — i.e. when performing a behaviour either increases (positive spillover) or reduces (negative spillover) the likelihood of performing subsequent (un)related behaviours (Dolan and Galizzi, 2015; Thøgersen, 1999; Thøgersen and Crompton, 2009). However, there is no consensus in the literature as to whether positive or negative spillovers are to be expected in the context of PEBs. Some studies have found that engaging in a PEB increases the likelihood of performing other PEBs (Alacevich et al., 2021; Lacasse, 2017; Lanzini and Thøgersen, 2014; Margetts and Kashima, 2017; Thøgersen and Noblet, 2012; Thomas et al., 2016). Others have detected negative spillovers between environmentally-friendly actions (Geng et al., 2016; Tiefenbeck et al., 2013; Truelove et al., 2016). While others yet have evidenced the presence of mixed effects (Carrico et al., 2018; Thøgersen, 1999; Thøgersen and Ölander, 2003).

This paper tries to fill in some of the gaps in the literature on spillover effects from PEBs by answering the following questions: "Does purposefully engaging in a PEB lead to a positive or a negative spillover?"; "How persistent is the spillover effect?"; and "Is the behavioural spillover restricted to the same domain as the initial action or does it span to other domains?" This is done with a lab experiment where participants first decide whether to donate an amount from their final payoff to a well-known environmental organization — the World Wildlife Fund (WWF) — and then they play a multi-round public goods (PGs) game. In this context, a positive spillover would emerge if people who decided to donate have higher contribution levels for the provision of the PGs than those who did not donate. Conversely, a negative spillover would be represented by lower contribution levels from individuals who made the donation. This corresponds to testing the following set of hypotheses:

$$
\begin{aligned}
H0 &: \text{Av. Contribution}_{Don} = \text{Av. Contribution}_{NoDon} \\
H1a &: \text{Av. Contribution}_{Don} > \text{Av. Contribution}_{NoDon} \\
H1b &: \text{Av. Contribution}_{Don} < \text{Av. Contribution}_{NoDon},
\end{aligned}
$$

where $H1a$ represents a positive spillover and $H1b$ a negative spillover.

The results highlight the presence of a strong positive spillover. Participants who decided to donate contribute significantly more for the provision of PGs than participants who did not donate. The magnitude of the effect is noteworthy: contribution levels increase by roughly 11 experimental units[1], which corresponds to roughly 40% of the mean value at baseline. This is considerably greater than what had been found in the majority of previous studies that detected the presence of a positive spillover where the effect tended to be relatively small in magnitude (Maki et al., 2019). Such a discrepancy may be due to fundamental differences in the experimental design, as detailed below and later in the paper[2]. It also appears that the positive spillover tends to be rather stable and persistent. The effect of having decided to donate, on the contribution for the provision of PGs, is not limited to the first few rounds, but it remains at the same level throughout the entire game. Finally, the analysis suggests that the behavioural spillover is circumscribed primarily to the environmental domain. The effect on environmental PGs is

[1] More details on the experimental design are reported in Section 2.2.

[2] A more detailed discussion of the differences in experimental design and results are reported in Sections 2.5 and 2.5.

statistically greater than that on generic PGs, which, conversely, is not significant. Such results show that, in the context of this study, there are no cross-domain spillover effects. This is at odds with previous studies that detected behavioural spillover across different domains (Mazar and Zhong, 2010; Sachdeva et al., 2009). Once again, differences in terms of experimental design exist between those studies and the current one, which might explain the contradicting outcomes.

The design of my experiment overcomes several of the limitations present in the previous literature. Many former studies rely on stated performance for either the initial (Thøgersen and Noblet, 2012) or subsequent (Lacasse, 2017; Lanzini and Thøgersen, 2014) PEBs, or both (Thøgersen, 1999; Thøgersen and Ölander, 2003; Thomas et al., 2016). This type of data is inherently correlational, hence not well-suited to investigate the causal process of behavioural spillover (Carrico et al., 2018; Truelove et al., 2014). In addition, it introduces possible biases, like incorrect recollection bias, or an intentional misreporting to avoid a negative judgment in the eyes of the experimenters. Others focus on spillover effects following an experimental intervention which targets a specific PEB (Carrico et al., 2018; Lanzini and Thøgersen, 2014; Tiefenbeck et al., 2013). In other cases yet, the design of the experiments artificially induces participants to engage in PEBs (Geng et al., 2016). However, in the real life performing a PEB entails some sort of trade-off or opportunity cost[3]. It is the free and conscious choice to engage in an environmentally-friendly action that should signal an individual's environmental commitment and reinforce their propensity to act pro-environmentally in the future (Lacasse, 2017; van der Werff et al., 2014), or license people to "rest on their laurels" believing they have already done their share (Thøgersen and Crompton, 2009; Truelove et al., 2016). Hence, if spillover effects are to emerge, they should do so following the purposeful adoption of a PEB.

[3]People often have to chose whether to perform an easier or less expensive action, or a more difficult or more expensive PEB. For example: taking the car versus public transport to commute, leaving appliances in stand-by versus switching them off, eating meat versus following a vegetarian diet, and so on.

For this reason, in my experiment, the decision of whether to make the donation or not is left entirely to the discretion of participants. In addition, a donation entails a trade-off — individuals are foregoing some money to do an environmentally-friendly action — and it has been recognised as a valid measure of significant PEB (Clements et al., 2015). This better represents the purposeful engagement in a PEB that takes place in everyday life, thus it should offer a more accurate and robust test for the presence of spillover effects. The choice of incorporating a PGs game is also a sensible one, since many environmental behaviours and interventions present public goods characteristics (Clark et al., 2003; Waichman et al., 2021).

However, the decision to perform the initial PEB might depend on an individual's values and environmental attitude. Namely, it is potentially endogenous. To account for the concerns regarding the endogeneity of the initial PEB, a matching estimation is adopted, using coarsened exact matching (CEM) to compare participants who donated to others with a similar environmental attitude but who were not asked to make the donation, thus deriving the average treatment effect on the treated. In a recent study, Alacevich et al. (2021) cope with similar endogeneity issues adopting an instrumental variable approach where they use a policy implementation as an intention-to-treat variable to instrumentalize for the endogenous decision to engage in the initial PEB. Such an approach would not be particularly well-suited in the context of the current experiment (see Section 2.4 for a discussion of the motivations). Therefore, a matching strategy has been chosen instead.

Another aspect that is typically overlooked is the persistence of the effect. Experimental studies focus on the impact of an initial PEB on a targeted subsequent environmentally-friendly action (Carrico et al., 2018; Geng et al., 2016; Lanzini and Thøgersen, 2014; Thøgersen and Noblet, 2012; Truelove et al., 2016). Longitudinal studies, although more suited to investigate the effect over a longer period of time and multiple behaviours, make use of self-reported performance measures which, as previously said, are correlational in nature (Thøgersen, 1999; Thøgersen and Ölander, 2003; Thomas et al., 2016). To cope with the pressing concerns that climate change poses, structural lifestyle changes will be required (IPCC, 2014; MacKay, 2009; Truelove et al., 2016). So it is important to understand not only if a spillover effect will emerge, but also how long it is going to last. In their study, Alacevich et al. (2021) found that the waste reduction following a waste separation program tended to disappear after a few months. However, further investigation of the persistence of behavioural spillovers is warranted.

In this regard, the presence of multiple rounds in the PGs game allows for the evaluation of whether the spillover effect is only limited to the very first rounds following the initial PEB, or if it persists several rounds into the game. This helps understand if the consequences from acting pro-environmentally tend to disappear shortly after the initial behaviour was done, or if they are more profound and can potentially lead to fundamental behavioural changes.

However, climate change and environmental problems are far from being the sole challenges of our times. Poverty, healthcare, homelessness are all prominent issues that affect our societies and that have been exacerbated by the recent Covid-19 pandemic. Hence, it is important to test whether the effect of engaging in a PEB spills over exclusively to other environmentally-friendly actions or if there can be cross-domain effects. Some papers have evidenced the presence of spillover effects spanning across domains (Mazar and Zhong, 2010; Sachdeva et al., 2009). Still, the previous research has not simultaneously tested the effect on the same and different domains.

For this reason, my experiment includes environmental together with generic PGs in the public goods game. In such a way it is possible to investigate if spillovers take place exclusively in the same domain as the first behaviour or if they can spread to other, unrelated domains.

Finally, the use of the experimental method, which links an initial PEB to subsequent actions, overcomes the limits of survey and stated preference data and is better suited to explain the causal relation which inherently characterises behavioural spillover.

This paper provides important contributions to the literature on behavioural spillover and PEBs. By adopting a rigorous experimental approach and econometric techniques to link the purposeful engagement in an initial, effortful environmentally-friendly action to the performance of future PEBs, it overcomes several limitations affecting previous analyses and offers a more sensible and representative investigation of the causal relationship which characterises the spillover process. It also accounts for potential endogeneity concerns connected to the purposeful engagement in the initial PEB with the adoption of a matching approach. In addition, it is among the first studies to explicitly test the persistence of the spillover effect throughout multiple subsequent behaviours as well as its ramification across different domains.

The findings help to shed light on and advance the ongoing debate regarding environmental behaviour, suggesting that when people freely decide to engage in a PEB which requires to make a concrete effort they are more likely to act pro-environmentally in the future and to sustain this course of action. This outcome appears particularly sensible and relevant from a policy standpoint in light of the types of behaviours considered — donating to a pro-environmental organization and contributing to the provision of public goods. In fact, many environmental interventions for mitigation of and adaptation to climatic changes have public good characteristics and are provided by pro-environmental organizations.

One final consideration is needed. I do not test the potential mechanisms that could be guiding the observed effect. The literature has taken inspiration from various psychological theories to justify why positive and negative spillovers emerge, form goal activation (Dhar and Simonson, 1999) to self-perception (Bem, 1972), from cognitive dissonance (Festinger, 1957) to action learning (Nigg et al., 1999) and moral licensing (Merritt et al., 2010; Nisan and Horenczyk, 1990)[4]. While this paper does not specifically investigate any of the aforementioned mechanisms, it is still possible to advance hypotheses regarding which ones might drive the observed results.

The remainder of the paper is structured as follows. Section 2.2 details the experimental design. Section 2.3 describes the data and its characteristics. Section 2.4 outlines the empirical strategy. Section 2.5 presents the results of the analysis. Section 2.6 discusses the potential implications and concludes.

2.2 Experimental Design

A total of 111 participants took part in the experiment. They were recruited through fliers posted in all main buildings of Trinity College Dublin's campus, as well as advertisement in some large undergraduate classes in the School of Social Sciences and Philosophy.

The experimental sessions were conducted in the Psychology Lab at Trinity College Dublin from February 2019 to February 2020. Upon arrival, participants were assigned to a computer station and were advised to take their decisions independently and to avoid any sort of communication with other players. An experimenter monitored the room throughout the experiment to prevent illicit behaviours, to answer participants' questions and to fix any technical issue that might emerge. Sessions consisted of 10 or 15 people. At the beginning of each session, participants were presented with a brief introduction detailing the structure of the experiment. They were also told that their final payoff would be paid to the in the form of a voucher.

[4]See Thøgersen and Noblet (2012) for a discussion of the proposed theories and their applications in the context of PEBs.

The experiment was divided into two versions. One where participants were asked whether they want to donate part of their final payoff to the pro-environmental organization *before* the PGs game. The other in which they could decide to make the donation *after* the PGs game. Participants were randomly assigned to one or the other[5]. The first version represents the main experimental treatment which will be the focus of most of the analysis. Conversely, the second version serves mainly as a control, and it is not used to investigate the presence of behavioural spillover *per se*. For ease of exposition, in the remainder of the paper the former will be referred to as the 'donation-first' (DF) version, while the latter as the 'donation-after' (DA) version. In the end, 9 experimental sessions were conducted: 7 donation-first sessions, with 82 participants; and 2 donation-after sessions, with 29 participants[6]. Unfortunately, from March 2020, all experimental sessions had to be canceled due to the outbreak of the Covid-19 pandemic. This means the sample size is considerably smaller than what originally desired[7], which constitutes a limitation of the current analysis. Conducting the experiment online rather than in the lab would have certainly helped in collecting a wider sample, and would have allowed to continue the study also during the pandemic. However, the dynamics of the public goods game, which are explained in greater detail later in this section, prevents the experiment to be reliably conducted online. Hence, an in-person laboratory setting has been chosen.

The reason for having two experimental versions is twofold. First, it seems that

[5]See Appendix A.1 for more information on the randomization procedure.

[6]In one of the donation-after sessions, a person who had originally confirmed his presence did not show up. Since the PGs game can only be played in multiples of 5, the 14 remaining participants were give two options and asked which one they preferred. One was to play the game with 10 players, turning away 4 people at random, who would then be ensured the participation in a later session if they so desired. The other was to play it with 15 players, with the experimenter stepping in to fill-in the remaining slot. In this case, they were assured the experimenter would play a sequence of pre-determined random contributions, to avoid influencing the outcome of the game. The 14 participants unanimously decided to allow the experimenter to step in and play the game with 15 players and 3 groups. The sequence of randomly generated contributions was: 52, 88, 92, 58, 6, 48, 21, 42, 42, 44, 88, 8, 59, 69, 4, 89, 19, 71, 70, 92. The experimenter's observation is removed from the analysis.

[7]For example, Blanken et al. (2015) state that to have sufficient power to uncover a moral licensing effect, a sample size of at least 150 observations per cell would be required.

merely the fact of having seen the initial PEB could lead to different behaviours later on (Mazar and Zhong, 2010). So, by having two different versions it is possible to check the presence of a priming effect. Second, since participants are free to choose whether they want to perform the initial PEB, it could be argued that said choice is endogenous and depends on an individual's environmental attitudes and self-identity. That is, only environmentally-friendly people will decide to donate, while non-environmentally-friendly ones will not. If this is the case, simply comparing the contributions of those who donated with those who did not might not provide an appropriate representation of the spillover effect. To circumvent this problem, it is possible create a counterfactual scenario matching participants who made the initial PEB in the DF version with individuals from the DA version with a similar environmental attitude[8]. This makes it possible to deal with the potential endogeneity issues connected to the spontaneous decision to engage in the initial PEB, and to interpret the estimated effect as the average treatment effect on the treated.

As mentioned in the Section 2.1, the initial PEB in the donation-first version is represented by a donation to the World Wildlife Fund. Specifically, participants were asked whether they would like to donate €2 from their final payoff to the WWF[9]. They were informed the final earnings could be anything between €8 and €22. But the fact that they do not know the exact amount and that they have the power, with their decisions, to define said amount should prevent (or at least considerably limit) income effects to emerge. The choice to have a donation

[8]Merely comparing participants who decided to donate in both experimental versions would not be possible, since deciding to donate at the end of the PGs game is fundamentally different from deciding to donate at the beginning of it. In fact, even though participants have not seen their final payoff, they could still have formed an educated guess of what that could be, based on the feedback provided at the end of each round. It is possible that some decided to donate because they thought their payoff would be so high they could afford a small reduction. While others because they believed it was so low it would not represent a significant addition to their finances, and they might as well just donate to contribute to a good cause. These reasoning do not apply to the decision of donating before the PGs game, because participants cannot know nor infer the amount of their final payoff.

[9]The decision to have a fixed donation was taken to avoid concerns of participants adopting a certain contribution behaviour in the PGs game to affect the amount of the donation, which would confound the outcome of the analysis.

to the WWF stands on two reasons. On the one hand, making a donation is a concrete and effortful action which better mirrors the PEBs people make in their everyday life rather than simply reporting previous behaviour or imagining doing something. On the other hand, the WWF as a pro-environmental organization is widely known, with a vast range of activities and it does not entail political considerations[10].

The core of the experiment is represented by a linear public goods game[11] with a voluntary contribution mechanism (Ledyard, 1995). Participants play in groups of $i = 1, \ldots, N$ members and the game is divided into $t = 1, \ldots, T$ rounds. In each round, participants receive an endowment, which is constant for all individuals in all rounds ($W_{it} = W$), and they have to divide it between a 'private account' (i.e. the money they keep for themselves) and a 'public account' (i.e. the public good). The amount a player puts in the public account is called her/his contribution (c_{it}), while the rest remains in the private account ($x_{it} = W - c_{it}$). The money put in the private account is only available for the individual. Conversely, the amount placed in the public account generates earnings for all group members, based on the total group contribution ($C_t = \sum_{i=1}^{N} c_{it}$) and the marginal per-capita return of the public account (α, with $1/N < \alpha < 1$), such that the return from the public account is $Y_t = \alpha C_t$. The total earnings of individual i in round t is given by the sum of what she/he put in the private account plus the return from the public account ($\pi_{it} = x_{it} + Y_t$); and the final payoff is the sum of earnings across all rounds ($\Pi_i = \sum_{t=1}^{T} \pi_i$).

[10]The World Wildlife Fund is active in six main areas of intervention: Food, Climate, Wildlife Conservation, Forests, Fresh Water and Oceans. Participants were allowed to specify to which of these six areas they wanted their donation to be directed. Other widely known organizations like Greenpeace or PETA were not chosen since they could be viewed as more controversial or politically partial.

[11]The reason why a linear public good was chosen rather than a threshold public good is because the latter provides an incentive to contribute in order to meet the threshold. The objective of this research is to study the presence of behavioural spillover, and not mechanisms to promote contributions to PGs. Therefore, the experiment was designed to remove any incentive in this sense. Moreover, linear public goods are quite common in the environmental domain. Think about recycling, pollution reduction, oceans' clean-up, all these activities do not require a minimum level to be achieved in order to generate benefits.

The Nash equilibrium of the game is that every player contributes 0 to the public account and keeps the entire endowment of herself/himself, because α being smaller than 1 implies that every amount put in the public account generates a lower return than if it were put in the private account. However, since $\alpha > 1/N$, if all group members contribute their entire endowment the earnings for each one of them would be greater than if no one contributed anything, which corresponds to the social optimum. Previous studies on linear PGs find that both the Nash equilibrium and the social optimum are rarely achieved (Bernasconi et al., 2009; Blackwell and McKee, 2003; Fellner and Lünser, 2014; Isaac and Walker, 1998; Ledyard, 1995).

In this experiment, participants in each session were divided into groups of 5 players (including themselves, $N = 5$) and played a total of 20 rounds ($T = 20$), each featuring a single public good. This multi-round design allows the investigation of the persistence of the spillover effect. Since the main focus of the study is not seeing how individuals coordinate but how behavioural spillover affects future PEBs, participants are randomly re-grouped before every round[12]. Tokens are used as experimental currency, a common practice in the PGs literature, with an exchange rate of 1 token = 3.25 cents[13]. In each round, participants receive an endowment of 100 tokens ($W = 100$), and are asked how many of these they want to put in their private account and how many in the public account. The marginal per-capita return of the public account is $\alpha = 0.4$[14].

At the beginning of the PGs game, or after having made the donation decision in the DF version, participants were shown the instructions of the game. These were displayed on each participant's computer and read aloud by the experimenter.

[12]This implies that each round corresponds essentially to a prisoner's dilemma.
[13]Approximately 31 tokens = 1 Euro.
[14]This implies that if all group members contribute their whole endowment, each one of them would receive double that amount back: $100 \times 0.4 \times 5 = 200$. Although participants received all the relevant figures with the game's instructions, this piece of information was not explicitly reported and they had to infer it on their own.

They were also prompted to ask questions if anything was unclear.

Before the actual game started, participants would engage in 2 practice rounds to help them familiarize with the mechanism. It was made clear that these rounds would not count towards the final payoff. After the practice session was over, they had the chance to read the instructions once more. Then, the 20-rounds PGs game took place. Two unrelated tasks were presented after rounds 7 and 14 respectively, to give participants a bit of a break. At the end of the game, a survey elicited individual characteristics and attitudes towards the environment, climate change, risk and impatience. Figure 2.1 reports a flowchart of the experiment, highlighting the differences between the two experimental versions and the point of randomization. The full text of the Introduction and Instructions for each treatment can be found in Appendix A.1, while the survey is reported in Appendix A.2.

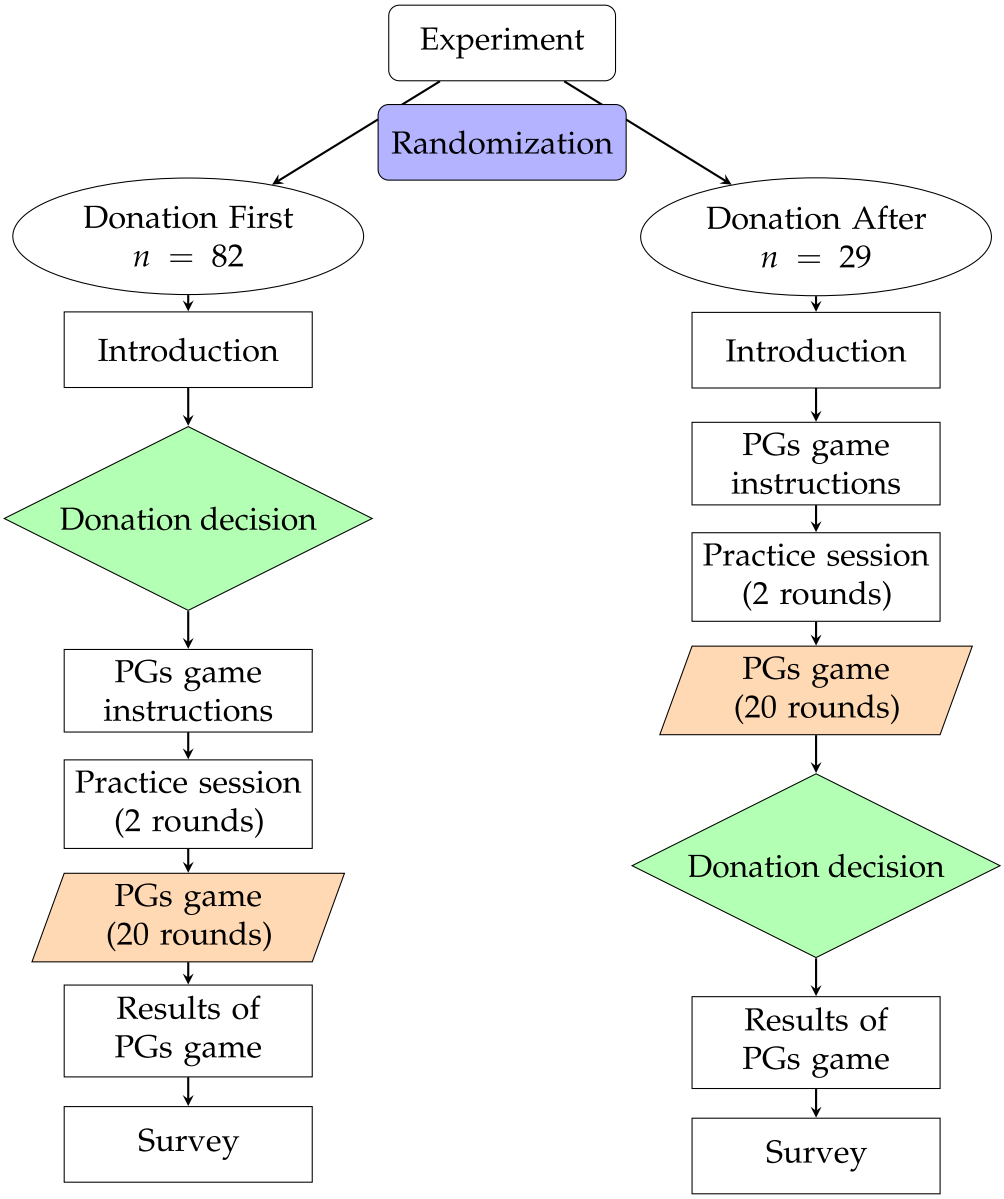
Since the objective of this study is not only to assess whether doing something good for the environment affects future PEBs, but also to test if the effect can span across different domains, the public accounts need to reflect both connotations. Therefore, the PGs game has two types of public accounts: environmental and generic[15]. Participants were told that, irrespective of their preferences, the two typologies produce the same returns, and that they only differ in terms of the "channels" through which these returns are generated. Specifically, the experiment's instructions said:

> Think about the Public Account's earnings ($0.4 \times M$)[16] as a service the

[15]The distinction between different types of public accounts is not new in the public goods literature. Blackwell and McKee (2003) and Fellner and Lünser (2014) distinguish between local and global public accounts; while Corazzini et al. (2015) present collective accounts with geometric names.

[16]M is the symbol used in the instructions provided to participants to represent the total group contribution, which has been called C_t in the previous discussion. This decision was taken to make the instructions as simple as possible and avoid confusion between individual contribution, c, and group contribution, C, since many participants did not not come from an economics or mathematics background and they could get lost in the technical terminology.

Figure 2.1: Structure of the experiment



Public Account can generate: the money (tokens) contributed to the Public Account are used to create a service which provides benefits (earnings) to the whole society (Group). So, even if the benefits (earnings) are the same for the two types, the services that generate them are different.

- **Environmental Public Account.** Goods and services provided may include, but are not limited to, national parks, oceans and shores' clean-up, waste recycling, afforestation, greenhouse gas emissions' reductions, renewable electricity power plants, and so on.

- **Generic Public Account.** Goods and services provided may include, but are not limited to, street lighting, public radio and television broadcasts, national defence, research on diseases' treatments, the Red Cross, and so on.

Participants were also told that the order in which the Public Accounts are presented was randomly determined by the software, but that it was the same for all groups in the same experimental session. However, they did not know how many environmental and how many generic public accounts there were. Out of the 20 rounds, 13 were environmental and 7 were generic, and this applied to all experimental sessions, only the order in which they were presented differed.

In the end, the average payoff was €12.63, inclusive of a €5 show-up fee. Thirty-one of the 82 participants in the donation-first version decided to donate, for a total amount of €62, which has been donated to the World Wildlife Fund by the experimenter on behalf of the participants.

The experiment was designed with the *oTree* software (Chen et al., 2016).

2.3 Data

Table 2.1 reports descriptive statistics of individual characteristics and attitudes towards the environment, climate change, impatience and risk over the whole sample, and a comparison of the two experimental versions. The list of demographic questions included in the experiment is reported in Appendix A.2.

From Table 2.1 it emerges that the sample consists primarily of young and educated people who do not have a stable employment[17], which is not surprising considering that the vast majority of participants were students. The gender ratio is close to 50%, with a slight prevalence of females. Most participants are not in a relationship, they are not living alone and they are coping on current

[17]More than 96% of participants have between 18 and 35 years of age and almost 65% have a degree, are enrolled at the university or have been.

Table 2.1: Descriptive statistics and comparison of experimental versions

	Descriptive statistics					Versions comparison		
	Obs.	Mean All	Std.Dev.	min	Max	Mean DA	Mean DF	Diff. DA–DF
Education	111	4.31	1.52	0	8	4.72	4.16	0.57*
Age	111	1.19	0.51	1	4	1.17	1.20	-0.02
Female	111	0.54	0.50	0	1	0.48	0.56	-0.08
Employed	111	0.14	0.34	0	1	0.28	0.09	0.19**
Single	111	0.93	0.26	0	1	0.93	0.93	0.00
Living alone	111	0.07	0.26	0	1	0.03	0.09	-0.05
Economic background	111	0.35	0.48	0	1	0.38	0.34	0.04
Experience	111	0.07	0.26	0	1	0.10	0.06	0.04
Income	111	2.85	1.30	0	5	3.14	2.74	0.39
EA	111	27.60	4.97	12	34	27.66	27.59	0.07
CCA	111	25.79	3.71	12	30	25.79	25.79	0.00
Total ECCA	111	53.40	7.38	29	64	53.45	53.38	0.07
Patience	111	5.91	2.15	2	10	5.66	6.00	-0.34
Risk	111	6.23	2.03	2	10	6.45	6.15	0.30

Notes: The left-hand side of the table displays descriptive statistics for the full sample. The right-hand side reports the mean values in the two experimental versions together with the results of *t*-tests on their equality. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

income[18].

A potential concern is that people's knowledge of the public goods theory may impact their contribution behaviour, thus confounding the results. Roughly one-third of the sample has a background in economics, but less than 10% has prior experience with economic experiments and only a subset of them had taken part in coordination or public goods games before.

Questions to elicit individuals' attitudes towards the environment and climate change are designed as 1-5 Likert scale questions, where respondents are asked how much they agree or disagree with certain statements. The Environmental Attitude (EA) variable is given by the sum of seven questions (two reverse scale,

[18]The income variable takes values from 1 to 5, with values increasing with income. It is based on the following question: "How would you describe your current income situation? (If you are married or in a domestic partnership consider your combined income)". Answer options were: "Finding it very difficult to live on current income; Finding it difficult to live on current income; Coping on current income; Living comfortably on current income; Living very comfortably on current income; Prefer not to state". The question and the available options are reported in Appendix A.2.

Cronbach's $\alpha = 0.79$); the total score can range from a minimum of 7 to a maximum of 35. The Climate Change Attitude (CCA) variable is given by the sum of six questions (three reverse scale, Cronbach's $\alpha = 0.68$), with the total score ranging from 6 to 30. Both sets of questions are adapted from previous research on pro-environmental behaviour and climate change (Della Giusta et al., 2012; Leiserowitz et al., 2013; Minton and Rose, 1997; Sparks and Shepherd, 1992). It can be seen that, on average, the experimental sample presents fairly high levels of environmental and climate change attitudes, with means close to the maximum of both scales. These two variables are then merged to form single comprehensive measure, labelled Total Environmental and Climate Change Attitude (Total ECCA). This can range from a minimum of 13 to a maximum of 65, and presents a Cronbach's α of 0.81. The sample mean is greater than 53, meaning that, on average, participants tend to be environmentally-friendly and sensitive to climate change problems.

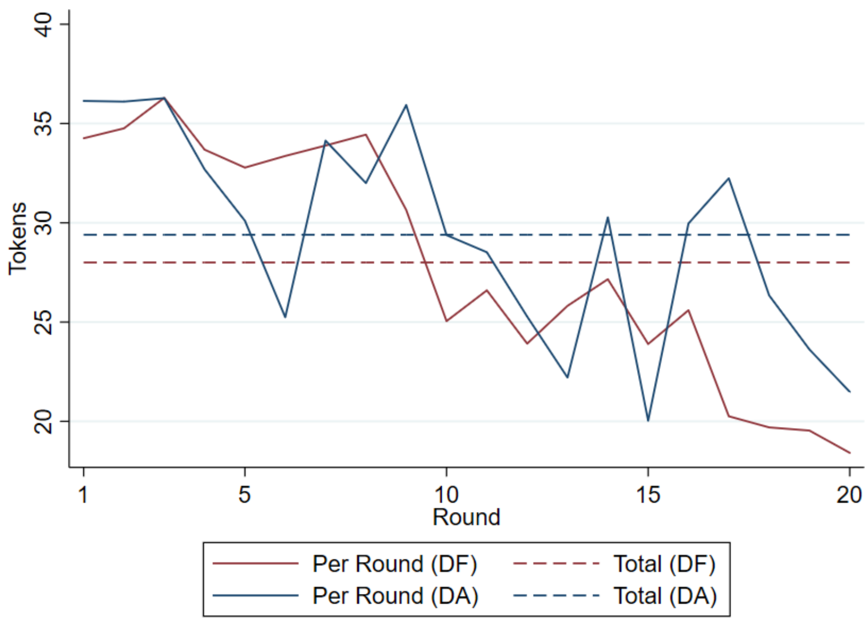
Finally, respondents were asked about their patience and attitude towards risk. Patience can take on values going from 1 to 10, with lower values corresponding to being a very impatient person and higher values a very patient one. Similarly, Risk ranges from 1, which corresponds to an unwillingness to take risks, to 10, that stands for being fully prepared to take risks. As it can be seen, the experimental sample consists of moderately patient and risk-tolerant individuals.

Table 2.1 also reports the results of t-tests to compare the samples from the two experimental versions over the various demographic characteristics. Overall, there are no relevant differences in terms of individual characteristics between the two versions. The t-tests fail to reject the null of equality of the means in almost every case. The only exceptions are represented by education and employment status. It appears that participants in the DA version are slightly more educated and more likely to be employed. However, for education the difference is fairly small compared to the dimension of the variable, and it is only significant at the 10%

level. While for employment, although the difference is more consistent and significant, the share of employed participants remains considerably below 50% in both experimental versions. In addition, the sample size of the DA version is much smaller than that of the DF one, which could, at least partially, explain these differences. All this considered, it is seems reasonable to assume that the two samples are comparable in terms of individuals' characteristics.

This comparability also extends to the contribution patterns in the PGs game. Figure 2.2 displays average contributions per round and in total for both experimental versions. In both cases, participants contribute between 35-40% of their endowment in the first rounds. The contribution levels decrease as the game progresses, but they do not reach the Nash equilibrium of complete free riding, which is consistent with previous findings in the literature on linear public goods (Bernasconi et al., 2009; Blackwell and McKee, 2003; Fellner and Lünser, 2014; Isaac and Walker, 1998; Ledyard, 1995).

Figure 2.2: Average contributions by round and experimental version



Notes: This graph displays average contribution levels of participants in the two experimental versions. The solid lines represent mean contributions in each round. Whereas the dashed horizontal lines are the average contributions across the entire game.

2.4 Empirical Strategy

To investigate whether having donated leads to the emergence of spillover effects, the following model is estimated:

$$Contribution_{its} = \alpha + \beta Donation_i + \delta Type_{ts} + \boldsymbol{X_i'}\gamma + \tau_t + \sigma_s + \varepsilon_{its}. \qquad (2.1)$$

The dependent variable is the contribution of player i in round t of experimental session s. The regressor of interest is $Donation_i$, a dummy variable representing whether the player has decided to donate to the WWF. The presence of a positive(negative) behavioural spillover would be represented by the coefficient β being significantly greater(smaller) than zero. The variable $Type_{ts}$ takes on value 1 if the PG in round t of session s is an environmental PG, and value 2 if it is a generic one. The vector $\boldsymbol{X_i}$ controls for time-unvarying individual characteristics which could affect the decision to donate and the level of contribution. It includes age, education, gender, whether the participant is in a relationship, whether she/he lives alone, whether she/he has a background in economics, whether she/he has experience with economic experiments, income, overall environmental attitude (Total ECCA), patience and attitude towards risk. The round and experimental session fixed effects, τ_t and σ_s, absorb confounding factors like the propensity to reduce contributions as the game progresses and the pool of participants or the order in which different types of PGs are presented in a given experimental session. Since the amount participants can contribute to the public account is restricted to be between 0 and 100 tokens, a Tobit estimation for double-censored data is used (Tobin, 1958). To account for serial correlation standard errors are clustered at the participant level.

As already mentioned, since performing the initial PEB is a free and conscious choice made by the individuals, the model outlined above potentially suffers from endogeneity. This paper employs a coarsened exact matching approach methodology to cope with endogeneity issues. A recent similar study adopts an instrumental variable approach, using an intention-to-treat variable as an instrument for the decision to engage in the initial PEB (Alacevich et al., 2021). The rationale for using coarsened exact matching rather than IV approach is the following.

In this setting, the experimental version — i.e. whether participants are asked if they would like to donate before or after the PGs game — would represent the intention-to-treat instrument (Z), whereas the actual treatment would be the decision to make the donation (D). Since participants who are not asked to donate before the PGs game cannot make the donation, the treatment variable has to be 0 for this entire group (namely, $Pr(D = 0|Z = 0) = 1$), which creates a series of complications in terms of model specifications. In particular, it prevents the use of a Tobit model in the second stage; and it requires the use of a linear probability model to estimate the first stage, which leads to unreasonable outcomes (negative predicted values). In addition, it is not possible to exclude that there are no defiers, which constitutes one of the required assumptions for the identification of a causal effect for a relevant subgroup of the population (Angrist et al., 1996). In this case, a defier would be someone who did not donate when asked at the beginning but would have if not asked (which is unlikely but not impossible), or someone who donated when not asked but would not have donated if asked (which cannot be observed given the experimental design).

Therefore, to account for the concerns regarding the endogeneity of the initial PEB, a matching estimation is adopted instead, using coarsened exact matching (CEM, Iacus et al., 2012) to compare participants who donated to others with a similar environmental attitude but who were not asked to make the donation, thus deriving the average treatment effect on the treated.

2.5 Results

2.5.1 Priming effect

Firstly, it is important to establish whether simply the fact of having had the opportunity to make a donation could affect individuals' behaviour in the PGs game. To investigate this priming effect a t-tests on the equality of average contribution across the experimental versions is conducted:

$$H0 : \text{Av. Contribution}_{DA} = \text{Av. Contribution}_{DF}$$

$$H1 : \text{Av. Contribution}_{DA} \neq \text{Av. Contribution}_{DF}.$$

Table 2.2 reports the average contributions in the two experimental versions and the results of Welch's t-tests for their equality[19]. The first row considers all 20 rounds, while the second row focuses on round 1 since a priming effect could manifest itself more strongly in the very first round of the game. In both cases it is not possible to reject the null hypothesis of equality of means, thus suggesting that simply having seen the donation at the beginning of the PGs game does not significantly affect contribution behaviours.

Table 2.2: Priming effect

	Obs.	Av.Contr. DA	Av.Contr. DF	Diff. DA–DF	Std.Err.	t	p-value
All Rounds	2220	29.40	28.00	1.40	1.169	1.195	0.232
Round 1	111	36.14	34.26	1.88	5.250	0.358	0.722

Notes: This table reports the average contributions of participants in the two experimental versions and the difference between them. Average contributions and the differences between them are expressed in tokens. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[19]The Levene's test for the equality of variance (Levene, 1960), reported in Appendix A.3.1, rejects the null hypotheses of equal variances in the two experimental versions.

2.5.2 Behavioural spillover

Even though it has been shown that the two experimental versions are comparable in terms of sample characteristics and overall contribution levels, the remainder of the analysis will focus solely on the donation-first version. The only exception is represented by the matching estimation presented in Section 2.5.4.

Table 2.3 reports the main results based on the empirical strategy outlined in Section 2.4.

Table 2.3: Behavioural spillover

	Contribution (1)	Contribution (2)	Contribution (3)	Contribution (4)
Donation	16.35***	11.27***	13.57***	12.62***
	(4.597)	(4.335)	(4.615)	(4.475)
Donation × Type			-6.60*	
			(4.006)	
Log likelihood	-6526.81	-6454.74	-6452.21	-6460.58
Pseudo R^2	0.028	0.039	0.039	0.038
Observations	1640	1640	1640	1640
Controls				
Round FE	✓	✓	✓	✓
Session FE	✓	✓	✓	✓
PG Type		✓	✓	✓
Ind. Charact.		✓	✓	✓
Total ECCA bins				✓

Notes: This table presents the results of Tobit regressions of the amount contributed by player i in round t on the set of controls and FEs listed at the bottom. All regressions focus on the donation-first sample. Standard errors, clustered at the participant level, are reported in parenthesis. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Column (1) is a simple regression of the contribution made by player i in round t on the donation dummy, with round and session FEs but without additional controls. In Column (2) the type of the PG and the characteristics of the player are included in the model, thus estimating Equation 2.1. The positive coefficient of the donation dummy confirms the presence of a positive spillover. Specifically, participants who decided to donate €2 to the WWF prior to the PGs game, then contribute 11.27 tokens more to the public accounts than participants who did not donate, *ceteris paribus*, which corresponds to roughly 40% of the average contribution level. This effect is considerably bigger than those found in the majority of previous studies (Maki et al., 2019). As mentioned in Section 2.1, those papers relied on self-reported performance of PEBs, and, in some cases, those behaviours displayed ceiling effects[20]. Not only are those approaches ill fit to analyse behavioural spillover, but they can also pose limits to the magnitude of the effect that gets detected. Conversely, the design adopted in the current study, by leaving individuals free to decide whether to engage in the initial PEB and then monitoring the actual performance of subsequent behaviours, may allow for more substantial effects to emerge.

To investigate whether the positive spillover is the same across various domains, the model is extended including an interaction between the donation dummy and the PG type dummy:

$$\begin{aligned} Contribution_{its} = \alpha + &\beta Donation_i + \delta Type_{ts} \\ &+ \theta \left(Donation_i \times Type_{ts} \right) + \boldsymbol{X_i'} \gamma + \tau_t + \sigma_s + \varepsilon_{its}. \end{aligned} \tag{2.2}$$

In this case, the coefficient β represents the effect of having donated on the contribution to environmental PGs, while the effect on generic ones is given by the sum $\beta + \theta$. A positive(negative) spillover would be represented by β and $\beta + \theta$ being positive(negative) and significant. In addition, a significant θ coefficient would suggest that a difference exists in how the effect of the initial PEB spills over to the environmental and other domains.

Column (3) reports the results of the model contained in Equation 2.2. The coefficient of Donation represents the effect of having donated on the contribution to

[20]The PEB was already performed very frequently at baseline, thus leaving little room for improvement.

environmental PGs. The sign of this coefficient shows that there is a positive effect in the same domain as the one of the initial behaviour. In addition, its magnitude is greater than in Column (2), suggesting that the same-domain effect is stronger. The coefficient of the interaction, on the other hand, is negative and significant (albeit only at the 10% level), highlighting that a difference does exist in terms of spillover effect between domains. In particular, the impact of having donated on the contribution to generic PGs is 6.60 tokens less than to environmental ones. The corresponding marginal effect of the donation on the contribution to generic PGs is 6.97, but this is not statistically different from zero[21], meaning that the initial PEB does not generate spillover effects in other domains.

This differs from previous studies that evidenced the presence of cross-domain spillovers (Mazar and Zhong, 2010; Sachdeva et al., 2009). In those cases, however, the effect was a negative one, and the analysis was rooted in the more general context of pro-social behaviour and the characterization of such an effect as a cross-domain one was not done by the authors but it has been applied *ex-post* in the current discussion. What is more important, those studies rely on imaginary or abstract initial tasks[22] which do not involve any of the elements that are at the core of the current research — namely, the purposefulness and concreteness of the initial PEB and the trade-off that it entails. Hence, the combination of differences in purpose of the research, experimental design and direction of the effect could explain this discrepancy.

To better control for the impact of an individual's environmental attitude, in Column (4) the variable Total ECCA included in the $\boldsymbol{X_i}$ vector of individual characteristics is broken up into a full set of dummy variables corresponding to bins

[21]The 90% confidence interval for the marginal effect is [-1.18 ; 15.11].

[22]Sachdeva et al. (2009) asked participants to write a story about themselves using positive or negative traits and found that those in the positive(negative) trait version were less(more) likely to behave pro-environmentally in a subsequent imaginary task. Mazar and Zhong (2010) assigned players to stores which differed in the number of environmental items available and saw that those who purchased from the green store were more likely to engage in unethical behaviour in an unrelated task.

of the distribution[23]. This approach takes the difference in contribution between people who did and did not donate in each bin, and then weighs these differences by the variation in "treatment" — i.e. having made the donation — within bins. The results are equivalent to those in Column (2), but slightly higher in magnitude, with participants who donated now contributing 12.62 tokens more than those who did not.

Overall, these outcomes point to the presence of a strong positive behavioural spillover following an initial PEB. However, the knock-on effect seems to be limited to the same domain as the original behaviour and does not spread across different domains.

As a robustness check, all models presented in Table 2.3 are re-estimated using standard ordinary least squares (OLS) models. The results, reported in Appendix A.3.2, are qualitatively identical, albeit slightly lower in magnitude.

2.5.3 Persistence of the spillover effect

A key contribution of the current analysis is the ability to investigate the persistence of the behavioural spillover. As mentioned in the Introduction, the multi-round nature of the PGs game allows to assess whether the effect influences only the very first rounds following the initial PEB or if it lasts longer throughout the game. This is done by extending Equation 2.1, including an interaction between the donation dummy and round dummies:

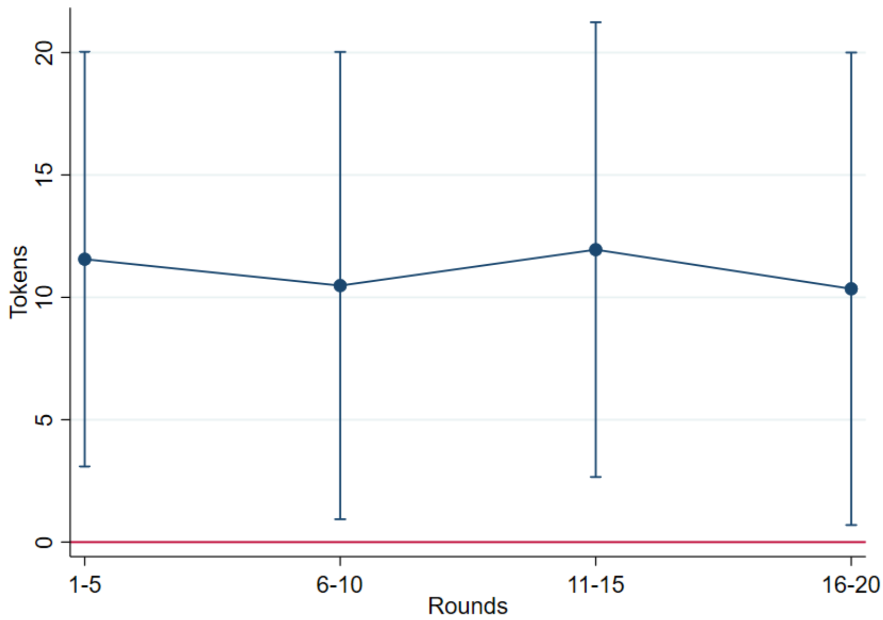
$$\begin{aligned} Contribution_{its} = \alpha + \beta Donation_i + \boldsymbol{\eta} \left(Donation_i \times \boldsymbol{Round'_t} \right) \\ + \delta Type_{ts} + \boldsymbol{X'_i} \gamma + \tau_t + \sigma_s + \varepsilon_{its}. \end{aligned} \tag{2.3}$$

To enhance the power, the twenty rounds have been divided into blocks of five, hence $\boldsymbol{Round'_t}$ is a set of round blocks dummy variables; τ_t also represents a round

[23] The bins used are: Total ECCA $\leq$ 45; 46 $\leq$ Total ECCA $\leq$ 50; 51 $\leq$ Total ECCA $\leq$ 55; 56 $\leq$ Total ECCA $\leq$ 60; 61 $\leq$ Total ECCA $\leq$ 65.

block FE. The marginal effects of having donated on the contribution in each round block are presented in Figure 2.3, together with their 95% confidence intervals. In Appendix A.3.3 are reported also the graphs with the marginal effects in each individual round. The results are qualitatively very similar, albeit with more variation in terms of the size and significance of the marginal effects due to the limited number of observations in each round.

Figure 2.3: Persistence of the behavioural spillover effect



Notes: This graph displays the marginal effects, with the 95% confidence intervals, of having made the donation on contribution levels in each round block of the PGs game. The estimation refers to the donation-first sample.

Figure 2.3 highlights that the behavioural spillover generated by having made the donation is rather persistent throughout the various stages of the PGs game. The magnitude of the marginal effects fluctuates in the 10 to 12 range, meaning that participants who donated tend to contribute between 10 and 12 tokens more than those who did not donate. This magnitude is comparable to the coefficients reported in Table 2.3. Crucially, there is no evidence of mean reversion. The effect of having donated remains at the same level also in the final rounds of the game. This suggests that the behavioural spillover generated by the initial PEB does not fade out after a few subsequent actions but it tends to have long lasting effects.

Considering individual rounds offers a very similar picture. Figure A.1a in Appendix A.3.3 shows that the marginal effects go from a minimum of 4.55 (in round 6) to a maximum of 16.26 (in round 3), with most of them fluctuating between 10 and 14. Ten out if the 20 marginal effects are significant at the 95% level, with an additional 4 being significant at the 90% level (Figure A.1b)[24]. Once again, there seems to be no evidence of mean reversion, thus reinforcing the claim of a persistent spillover effect.

This is somewhat different from the findings of Alacevich et al. (2021) who, investigating the effect of a waste separation program on households' waste production, highlighted that the positive spillover tended to disappear between 5-to-8 months after the intervention, with the amount of waste reverting back to pre-program levels. A difference of Alacevich and co-authors' study with respect to the current analysis is the level of effort characterising the initial PEB. The new waste separation scheme introduced by the policy entailed a decrease in waste collection costs, whereas making the donation does not provide any monetary benefit but, conversely, a reduction of a participant's final payoff. Namely, deciding to make the donation requires a greater level of effort which may lead to more long-lasting consequences. In addition, the reversion of waste amounts to pre-program levels could be partly due to an income effect. While the new scheme is effective in reducing waste production immediately after its implementation, as the time goes by and households become more accustomed to the new waste costs, they could be inclined to revert to producing a more conspicuous amount of waste since they can still benefit from an overall expenditures reduction, in a sort of rebound effect (Binswanger, 2001; Khazzoom, 1980). Once again, this does not apply to the PEBs considered in current study. An alternative explanation could be that the time period considered by Alacevich et al. (2021) is actually considerably longer than

[24]The marginal effects significant at the 90% level are those corresponding to rounds one, fourteen, seventeen and nineteen.

the 20 rounds of the PGs game. Thus, if more rounds were to be conducted, it would be possible that a disappearance of the spillover effect would take place also in the context of the current analysis.

2.5.4 Coping with endogeneity: matching estimation

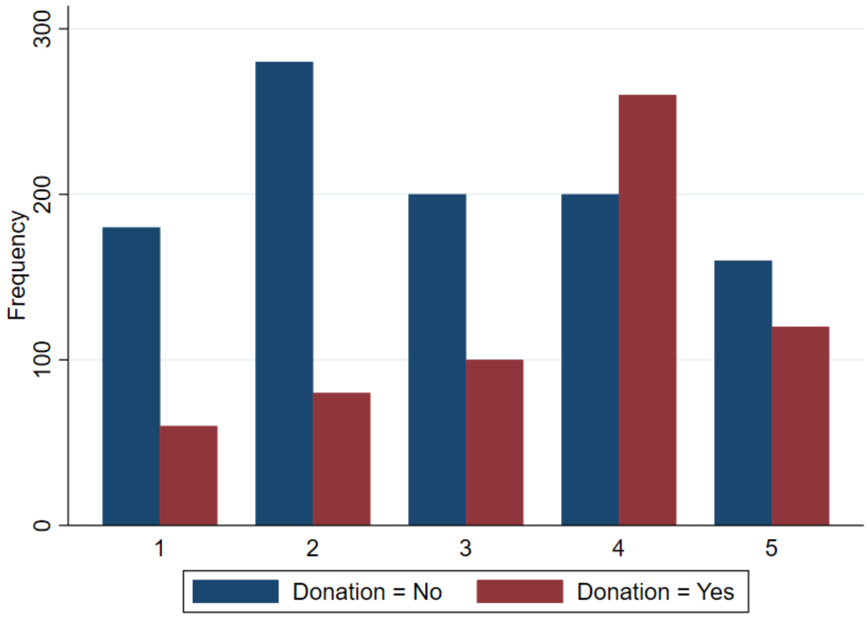
In Section 2.2 it was stated that a potential concern for the validity of the analysis is that the decision of donating to the WWF might depend on an individual's environmental self-identity, meaning that only environmentally-friendly individuals will decide to donate. Since environmentally-friendly people are also more likely to contribute to the provision of environmental PGs, this could confound the results.

The strategy adopted in Column (4) of Table 2.3 tries to limit this issue by comparing participants who fall in the same bin of the distribution of the Total ECCA variable. However, if the variation of individuals who engaged in the initial PEB is higher in some parts of the distribution this approach still does not accomplish the task. As it can be seen from Figure 2.4, there is in fact an uneven variation throughout the Total ECCA distribution of participants who decided to make the donation, with a higher frequency in the top two distribution bins.

This being the case, a matching approach is to be preferred to the bin strategy adopted in Table 2.3. The donation-after version of the experiment allows the creation of a fictitious counterfactual scenario which can be used to investigate what a player who engaged in the initial PEB would have done *had she/he not seen* the initial PEB.

This is achieved through Coarsened Exact Matching (CEM, Iacus et al., 2012), a Monotonoic Imbalance Bounding (MIB, Iacus et al., 2011) matching method. This method relies on temporarily coarsening a designated variable or set or variables — $X \to X^*$, with the coarsening cut-points to be defined by the investigator *ex-*

Figure 2.4: Frequency of participants within each bin of the Total ECCA distribution by Donation



Notes: This graph displays the frequency of participants who did and did not make the donation, in the donation-first version, across the bins of the distribution of the Total ECCA variable. The bins are the same used to estimate the model reported in Column (4) of Table **??**: 1 corresponds to Total ECCA ≤ 45; 2 corresponds to $46 \leq$ Total ECCA ≤ 50; 3 corresponds to $51 \leq$ Total ECCA ≤ 55; 4 corresponds to $56 \leq$ Total ECCA ≤ 60; 5 corresponds to $61 \leq$ Total ECCA ≤ 65.

ante based on the characteristics of the data —, apply exact matching on the coarsened X^*, then sort the observations into strata corresponding the the unique values of X^* and prune those strata with no "treated" or no "control" units in them. After removing the observations belonging to the aforementioned strata, the uncoarsened data is used to estimate the average treatment effect on the treated (ATT), based on weights assigned to controls in each stratum to balance with the treated.

Here, individuals in the donation-first version who decided to donate €2 to the WWF are matched with individuals from the donation-after version with a similar level of Total ECCA. The coarsening cut-points are the same cut-points used to construct the bins adopted in the model reported in Column (4) of Table 2.3. This results in a total of 6 strata, with 5 having valid matched observations[25], leaving

[25]In the pruned stratum there are 2 participants from the donation-after version but none from the donation-first version. The 2 control observations in this stratum are thus not included in the

58 individuals, 31 from the donation-fist version and 27 from the donation-after one.

Equation 2.1 is then estimated over the matched sample, and the results are displayed Table 2.4.

Table 2.4: Behavioural Spillover - Matching

	Clustered S.E. (1)	Conventional S.E. (2)	Robust S.E. (3)	Jacknife S.E. (4)	Bootstrapped S.E. (5)
Donation	11.65	11.65***	11.65***	11.65***	8.81**
	(7.871)	(3.723)	(3.818)	(3.981)	(3.859)
Log likelihood	-4885.94	-4885.94	-4885.94	-4885.94	-4892.97
Pseudo R^2	0.040	0.040	0.040	0.040	0.036
Observations	1160	1160	1160	1160	1160
Controls					
Round FE	✓	✓	✓	✓	✓
Session FE	✓	✓	✓	✓	✓
PG Type	✓	✓	✓	✓	✓
Ind. Charact.	✓	✓	✓	✓	✓

Notes: This table presents the results of Tobit regressions of the amount contributed by player i in round t on the set of controls and FEs listed at the bottom. All regressions focus on the CEM matched sample and make use of the appropriate weights. Standard errors are reported in parenthesis. Column (1) uses standard errors clustered at the participant level as in Table 2.3. Columns (2)-(5) adopt alternative methods to compute standard errors, as represented by the respective column headers. In Column (4), jacknife standard errors are obtained for 1160 replications. In Column (5), bootstrapped standard errors are obtained for 1000 replications. Since bootstrapping does not allow external weights to be used in the estimation, as is the case following CEM, the outcome in Column (5) is produced by a simple Tobit regression on the matched sample. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Column (1) adopts standard errors clustered at the participant level as done in previous regressions. As it can be seen, the Donation coefficient retains a positive sign and presents a very similar magnitude to that in Column (2) of Table 2.3, suggesting that having donated increases contribution levels by 11.65 tokens. However, the effect does not appear to be statistically significant, thus pointing to the absence of a behavioural spillover.

estimation.

Yet, a consideration is needed. This estimation focuses an a smaller sample size than the one used in the previous analysis, which comprises of a total of just 58 individuals. Therefore, it could be that, for such a small sample, clustering is not the most adequate method to compute standard errors and it results in these being over-inflated[26]. For this reason, Columns (2)-(5) of Table 2.4 re-estimate the model on the matched sample adopting alternative methods to compute standard errors. Specifically, Columns (2) uses conventional standard errors; Column (3) adopts the correction for heteroskedasticity robustness; Column (4) considers the jacknife estimation with 1160 replications; and Column (5) derives them with bootstrapping using 1000 replications. In each of these instances the size of the standard errors is much smaller than the clustered ones, and the Donation coefficient reverts to being statistically significant. It is worth noting that the coefficient in Column (5) is different from the other four. This is because with the bootstrap method it is not possible to use the CEM weights. Therefore, in this specific instance, the results are produced by a simple Tobit regression on the matched sample. Since the purpose of this exercise is to assess the goodness of the standard errors rather than the effect itself, the fact that the coefficient is derived through a different method is not a concern.

In light of these results, there should be enough confidence to affirm that, even in the matched sample, engaging in an initial PEB generates a positive behavioural spillover. That is, if an individual decides to make the donation, she/he then contributes more to the public accounts than had she/he not made the donation.

2.6 Discussion and Conclusion

To cope with the pressing challenges posed by climate change, governments and international organizations call for individual action to engage in more pro-

[26]Typically, a number of clusters greater than 40 is considered enough for unbiased inference, but there is no universally accepted threshold (Angrist and Pischke, 2009; Wooldridge, 2003).

environmental behaviours and shift one's lifestyle towards a greater environmental awareness and friendliness.

To promote the uptake of PEBs, a key concept is that of behavioural spillover. However, the literature on environmental behaviour and spillover has found mixed results (Carrico et al., 2018; Thøgersen, 1999; Thøgersen and Ölander, 2003), and it is not clear whether a positive (Lacasse, 2017; Lanzini and Thøgersen, 2014; Thøgersen and Noblet, 2012; Thomas et al., 2016) or negative (Geng et al., 2016; Tiefenbeck et al., 2013; Truelove et al., 2016) effect is likely to emerge following an initial environmentally-friendly action, if any effect at all.

Many of the previous studies suffer from various conceptual or empirical limitations which could affect the validity of the analysis and potentially explain the inconsistent results. In fact, they often resort to using stated performance measures, or artificially incentivising a certain behaviour. These approaches are prone to several biases — like incorrect recollection, false reporting or misrepresentation —, and are not best suited to investigate spillover effects. In the everyday life, deciding to act pro-environmentally involves a trade-off, and there is an inherent causal relation between this decision and the knock-on effects on future behaviour. Therefore, investigating the purposeful engagement in an effortful PEB and its consequences is crucial to accurately answer the question at hand.

Two aspects that have so far been overlooked are the persistence of the spillover effect and its ability to span across different domains. Most papers consider the impacts on a single subsequent behaviour, but to achieve the lifestyle changes required to cope with today's environmental challenges it is important to understand if a positive spillover can be sustained for or a negative spillover is likely to discourage the performance of multiple subsequent behaviours. In addition, since our societies are threatened by more than just climate change and environmental problems, assessing if the effects of a PEB can spill over to other domains is not trivial.

This paper aims to fill in those gaps in the literature by means of a lab experiment in which participants can freely decide whether they want to engage in a PEB — the donation of €2 to the World Wildlife Fund from their final payoff — before playing a multi-round public goods game, which contains environmental as well as generic PGs.

The results show that people who decided to donate contribute significantly more for the provision of PGs than those who did not donate, highlighting the presence of a positive behavioural spillover. The effect appears to be rather persistent throughout the entire PGs game, but to be restricted primarily to the environmental domain. The size of this effect is noteworthy. While other studies had found mainly relatively small effects, the current analysis goes in the opposite direction, suggesting a positive impact in the order of 40% of the baseline value. It therefore seems possible that the experimental design is effective in overcoming some of the limitations affecting previous research.

A potential concern is that only environmentally-friendly individuals are likely to make the donation, and since they are also more likely to contribute to environmental PGs the analysis might not actually be capturing a spillover effect. To cope with that, the experiment included a control version in which participants were not asked to donate before the PGs game. This affords the opportunity to match individuals according to their environmental friendliness, thus creating a counterfactual scenarios to see how a person who donated would have behaved in the PGs game had she/he not made the donation. The outcomes suggest that donating increases the contribution levels also in the matched sample.

Overall, these findings offer a strong support to the presence of a positive behavioural spillover following a purposeful and effortful PEB. They also suggest that its effects can influence numerous subsequent environmentally-friendly actions, which is reassuring in light of the fundamental lifestyle changes people will have to make in order to tackle the pressing environmental problems. All of this is particularly relevant since many interventions to cope with said problems present public goods' characteristics or are implemented by pro-environmental organizations like the one considered in this paper.

The biggest limitation of this study is represented by the small sample size. As mentioned in Section 2.2, due to the emergence of unforeseen external circumstances which prevented further sessions to be conducted, the total sample consists of 111 individuals, of which 82 participated in the donation-first version and 29 donation-after one. These numbers are smaller than the typical sample sizes in behavioural spillover experiments (Carrico et al., 2018; Geng et al., 2016; Lacasse, 2017; Lanzini and Thøgersen, 2014; Tiefenbeck et al., 2013; Truelove et al., 2016). This can pose various problems.

Firstly, a smaller sample means less representativeness. Although the experiment was mainly intended to members of the University community, which in and of itself is not representative of the general population, in principle participation was not prohibited to outside individuals, and it was surely not limited to students as it is often the case with laboratory experiments. Therefore, such a reduced sample implies missing out on a considerable variation in terms of age composition, ethnicity, educational background, or living conditions and habits, which would be important for a study that aims to provide informative and generalisable findings.

Second, a limited sample size generates problems in terms of the statistical power and reliability of the analysis, which can undermine the validity of the results. Based on the meta-analysis by (Maki et al., 2019), the analysis does appear to be under-powered, although the magnitude of the effect evidenced in the current paper is considerably greater than that found in the majority of previous studies, which discovered primarily modest effects. However, the (potentially) over-inflated standard errors in the matched sample — when the usual clustered standard errors were roughly twice as big as those derived with alternative approaches — do remain a cause of concern.

Therefore, future research could try to replicate the current study collecting more observations, which would help to clarify some of the doubts regarding the robustness of the analysis while, at the same time, improving the representativeness of the results.

A second potential limitation is represented by the use of the PGs game to measure subsequent PEBs. Although the contribution to the public accounts does constitute a free and intentional choice by the participants and it does involves that already mentioned trade-off that inherently characterises acting pro-environmentally, the way in which the game and public goods themselves have been designed may be too abstract when compared to everyday-life circumstances. Hence, it would be possible to expand the scope of this research by considering alternative environmentally-friendly behaviours, potentially adopting a field study, which is not prone to several biases affecting lab experiments (Galizzi1 and Whitmarsh, 2019), to further improve the capability to assess real-world implications.

Thirdly, although the empirical analysis controls for several individual characteristics and attitudes that may influence the willingness to perform PEBs, it does not take into consideration measures of trust, altruism and reciprocity, which have been found to significantly affect environmentally relevant behaviour (Ziegler, 2021). Therefore, future studies should try to incorporate said measures to enhance the reliability and representativeness of the results.

Another potential limitation is that it is not possible to disentangle habits developed outside the game from spillovers within the game. However, there is no *a priori* reason to believe this would be different between the control and treatment groups.

Finally, it is important to note that this paper does not investigate the mechanisms that are driving the effects which have been uncovered. Nevertheless, it is still possible to reason on which ones they could be. First of all, given that a positive spillover has emerged, it is possible to rule out moral licensing (Merritt et al., 2010), which is related to negative spillover.

If self-perception theories (Bem, 1972) were to apply, it would be the case that the decision to donate, rather than being necessarily made by individuals who *ex-ante* are more environmentally-friendly — which has been extensively controlled for with the matching procedure —, leads people to perceive themselves as *being* environmentally-friendly. Namely, it would establish their environmental self-identity. And it is then through the lens of this perception of themselves that they would be inclined to contribute more for the provision of environmental PGs.

Another potential explanation is that, after having decided to donate, people would perceive as inconsistent not engaging in subsequent PEBs. To avoid this sense of cognitive dissonance (Festinger, 1957), they would once again tend to contribute more to environmental PGs. A key aspect in this sense is that the initial and subsequent behaviours are viewed as similar (Thøgersen, 2004). And in fact, in support of this hypothesis, the results show that the positive spillover is confined to the environmental domain.

Conversely, mechanisms like action-based learning (Nigg et al., 1999) are less likely to be at play here, since it seems unlikely that individuals would have derived meaningful information regarding the mechanisms of the PGs game — of which contribution is a key element — from the act of having donated. Any possible learning that might have taken place seems to relate more to one's self-identity or usual behaviour, thus pointing once again to self-perception and cognitive dissonance theories respectively.

Therefore, future research could continue the work of previous studies that have investigated the potential channels through which the behavioural spillover emerges (Baca-Motes et al., 2013; Carrico et al., 2018; Cornelissen et al., 2008; Geng et al., 2016; Lacasse, 2017; Thøgersen, 1999; Thomas et al., 2016; Truelove et al., 2016; van der Werff et al., 2014), extending it to experimentally test the purposeful engagement in an effortful pro-environmental action, to provide a better understanding of the drivers of spillover effects. This is particularly relevant from a policy perspective. In fact, after having established that when people freely decide to perform a "costly" PEB they tend to do more environmentally-friendly actions in the future, it is important to evaluate the most effective interventions to induce people, especially those who still tend to be agnostic about environmental problems, to take that first step towards a more environmentally-attentive lifestyle.

3 Putting a new 'spin' on energy information: Measuring the impact of reframing energy efficiency information on tumble dryer choices in a multi-country experiment

3.1 Introduction

World energy consumption has been increasing over the past three decades (International Energy Agency, 2019), and, with a growing population condensed primarily in developing countries (The World Bank Group, 2020), this trend is likely to continue in the future. The residential sector contributes to more than 20% of global energy consumption (International Energy Agency, 2019), with shares close to 21% in the United States (U.S. Energy Information Administration, 2020) and above 27% in the European Union (European Environmental Agency, 2020). Governments and public administrations have seen energy efficiency as a powerful tool to combat these issues. While more energy-efficient products have a higher upfront cost, their lower consumption has the potential to make them better investments over their lifetime. However, the literature has documented the existence of an *energy efficiency gap* (Jaffe and Stavins, 1994), whereby agents' inability to recognize such trade-offs leads to an underinvestment in more energy-efficient technologies.

Although the existence and relevance of the 'energy-efficiency gap' has been questioned (Allcott and Greenstone, 2012), energy efficiency remains a key policy focus for many Governments. In an effort to improve agents' awareness and understanding of energy efficiency, various information tools and programs have been deployed[1]. Among the most well-known and widely adopted are energy efficiency labels (Collaborative Labeling and Appliance Standards Program, 2005). Examples include the U.S. "EnergyGuide", the EU "Energy Label", Australia's star "Energy Rating" and the "EnergyStar" logo. The motivation for energy labels rests on the assumption that making energy information more readily available to consumers facilitates the comparison among different products as well as between purchasing price and operating costs, ultimately leading to better energy investment decisions.

Energy efficiency labels generally provide consumption estimates in physical units (kWh/annum) based on average energy use and prices. However, it has been shown that people are likely to make mistakes when translating physical consumption into expenditures and savings (Allcott, 2011a, 2013; Brounen et al., 2013; Davis and Metcalf, 2016; Heinzel, 2012; Sammer and Wüstenhagen, 2006). Also, energy prices may vary substantially within regions. Over the past decade, several studies have been conducted to assess the efficacy of energy labels and whether reframing energy information improves effectiveness (Andor et al., 2020; Carroll et al., 2021; Davis and Metcalf, 2016; Heinzel, 2012; Heinzle and Wüstenhagen, 2012; Jain et al., 1994; Newell and Siikamäki, 2014; Shen and Saijo, 2009). Results from these studies on the effectiveness of reframing energy consumption have been mixed. This might be due to the specific contexts in which these studies were

[1]See Dranove and Jin (2010) for a review of the literature on disclosure of quality information and certifications.

conducted. As Allcott and Greenstone (2012) note, there have not been large-scale evaluations on the impact of energy efficiency labelling on consumer choices.

This paper tries to fill this gap by answering the following questions: "Does providing long-term average energy consumption information in monetary terms increase uptake of more efficient technologies?"; "Does providing personalised long-term energy consumption information in monetary terms increase uptake of more efficient technologies?"; and "Does the effect differ across countries?". To do so, we run an online randomised discrete choice experiment (RDCE) in four countries using the same methodology, to investigate whether different ways of framing energy efficiency/consumption affect consumers' willingness-to-pay (WTP) for energy efficiency and whether this effect is the same for all countries. Specifically, we ask respondents from Canada, the Republic of Ireland[2], the United Kingdom and the United States to express their preferences for tumble dryers which vary over a number of attributes.

Information on energy efficiency/consumption is reported in three forms. As our benchmark, we use the letter scale of the EU Energy Label (for Ireland and the United Kingdom) and the EnergyStar logo (for Canada and the United States), with products being assigned to an energy class (from A+++ to C), or being given the EnergyStar, based on their physical energy consumption (kWh/annum).

In a first manipulation, we convert this physical value into its monetary counterpart (the 10-years energy costs), based on average usage and national electricity prices. In a second manipulation, we derive individual-specific energy consumption according to self-reported use patterns. Also in this case, we express it in monetary terms for a 10-years time span. To the best of our knowledge, this is the first paper to test the provision of long-term monetary energy consumption information using the same experiment in a multi-country setting.

[2]In the remainder of the paper 'Ireland' will be used in place of 'Republic or Ireland'.

Since one of the core motivations behind efficiency labels is that inducing consumers to purchase more energy-efficient products will make them better off, irrespective of the external impact on the environment (Allcott and Knittel, 2019), providing energy information in a clear and accessible way is of fundamental importance. For this reason, we reframe energy consumption in the form of the long-term cost of electricity, which should represent a more meaningful representation of energy information for individuals than physical energy consumption. The choice to focus on tumble dryers stems from the fact that it is one of the highest energy-consuming household appliances. Its consumption depends solely on actual usage and derives from just one "fuel", namely electricity. To make a comparison, appliances like refrigerators or most TV sets consume energy even when people are not actively using them, while washing machines or dishwashers require water inputs in addition to electricity to function. Also, tumble dryers present the broadest range of ratings on the market with models carrying a 'C-rating' still available for purchase at the time of the experiment (2018). This is not the case for other appliances where the lowest available rating is often A+. On top of that, none of the countries in the study have monetary labels for tumble dryers. The European Energy Label is a color-coded letter scale based on physical consumption, and while the United States and Canada provide annual energy cost labels for several appliances this is not the case for tumble dryers. Therefore, the treatments that we introduce present new information in all contexts.

The outcomes of our mixed logit models suggest that, in general, displaying energy consumption in monetary terms does little to improve the uptake of more efficient technologies, irrespective of whether consumption is based on average or individual-specific use. However, there are two exceptions. In the Canadian sample we detect a negative and significant effect of our treatments, with the WTP for an improvement in energy efficiency actually decreasing by between Can$118 and Can$126 for tumble dryers in the three higher-efficiency classes — with no statistical differences between the two treatments. On the other hand, in the United

Kingdom, personalised energy information has a positive effect. Respondents receiving personalised information are willing to pay £81 more than those seeing the letter scale of the EU Energy Label and £64 more than those receiving generic cost information to purchase a product from the three higher-efficiency classes. The results remain substantially unaltered if we adopt different models or we split the sample based on various individual characteristics.

Previous studies have investigated the effect of providing monetary energy information in a variety of contexts: from TV sets and refrigerators in Germany (Andor et al., 2020; Heinzel, 2012), to cars in the United States (Allcott and Knittel, 2019), to refrigerators in India (Jain et al., 1994), to the residential market in Ireland (Carroll et al., 2021, 2020). For those focusing on tumble dryers, Kallbekken et al. (2013) show that lifetime electricity costs reduce the average energy consumption of purchased products at retail stores in Norway only if coupled with staff training, the UK Department of Energy and Climate Change (2014) does not observe any effect in the United Kingdom, and Carroll et al. (2016b) find no significant improvement of providing 5-years energy expenditures in Ireland. It is not clear whether these mixed results are attributable to the different core products, methodologies or countries. By adopting a common framework which considers the same product and the same treatments in four countries, we are able to overcome this ambiguity. In particular, our findings point at the specific national context in which the intervention is implemented as a key determinant of its effectiveness. This suggests that, when designing new tools to provide energy efficiency information, there is no one-size-fits-all solution, and policy makers should carefully evaluate which approach is best suited for their country or region.

Our paper builds on two main strands of literature. First, it is related to the literature on energy efficiency information (Allcott, 2011b, 2013; Allcott and Rogers, 2014; Allcott and Sweeney, 2016; Allcott and Taubinsky, 2015; Ayres et al., 2009; Brounen et al., 2013); and, more specifically, to that focusing on energy labels and their effectiveness (Andor et al., 2020; Carroll et al., 2016a; Heinzle and Wüstenhagen, 2012; Newell and Siikamäki, 2014; Sammer and Wüstenhagen, 2006; Shen and Saijo, 2009).

Second, our work draws from the literature on the framing of information and its impact on intertemporal choices (Kahneman and Tversky, 1984; Lowenstein, 1988; Lowenstein and Prelec, 1992; Lowenstein and Thaler, 1989; Tversky and Kahneman, 1981). Over the years, research on information framing has been applied to several contexts, including health (Block and Keller, 1995; Meyers-Levy and Maheswaran, 2004; Rothman and Salovey, 1997; Rothman et al., 1993), tax compliance (Hasseldine and Hite, 2003; Holler et al., 2009), and environmental behaviour (Loroz, 2007; Ropert Homar and Cvelbar, 2021; de Velde et al., 2010). In the context of energy efficiency, studies have investigated the effect of providing physical versus monetary energy information (Anderson and Claxton, 2014; Andor et al., 2020; Jain et al., 1994; McNeill and Wilkie, 1979), short-term versus long-term cost forecasts (Carroll et al., 2021; Heinzel, 2012), generic versus state-specific energy prices (Davis and Metcalf, 2016), and personalised information (Allcott and Knittel, 2019). Our paper contributes to the current debate by helping to shed light on the reasons behind the mixed effects evidenced by previous studies.

The remainder of the paper is organized as follows. Section 3.2 introduces the discrete choice theory and our experimental design. Section 3.3 describes the data and investigates the differences between the four countries in our sample. Section 3.4 presents the results of the analysis; and Section 3.5 concludes.

3.2 Methodology

3.2.1 DCE overview

Discrete choice experiments (DCE) have gained popularity as a tool to elicit agents' preferences for goods and services, since they help overcome some of the limitations presented by revealed preferences (RP) data. DCEs are a stated preferences (SP) method, usually involving surveys in which respondents are presented with repeated choice situations (called choice sets) comprising the comparison between two or more alternatives that vary over several attributes.

This type of experiment facilitates the measurement of non-use values, as well as the utility attached to individual attributes, which can be difficult to retrieve from revealed preferences data that often suffers from collinearity between attributes (Adamowicz et al., 1994; Carroll et al., 2021). In addition, it gives the experimenter a greater degree of control and flexibility than RP methods, coupled with the possibility to accommodate for the randomization between various treatments. The main drawback, as for any SP method, is represented by the hypothetical nature of the task. In most cases, the decisions people make do not have any real-world consequence (e.g. they do not actually purchase the product they selected among the array of alternatives), which introduces the possibility of hypothetical bias.

DCEs can be used to evaluate willingness-to-pay, to assess non-monetary valuation, to provide insights on consumers' preferences, and to test the effectiveness of new policies. They were initially developed in the marketing literature (Louviere and Woodworth, 1983). Over the years, they have been applied to a number of other fields, including health (see Ryan et al., 2008, for a review of the literature), transport economics (Greene and Hensher, 2003; Hensher and Louviere, 1983), or environmental economics (Adamowicz et al., 1994; Aravena et al., 2014; Hanley et al., 1998). In the energy economics literature, DCEs have been used to study preferences for power generation (Rivers and Jaccard, 2005) and fuel mix (Komarek and Kaplowitz, 2011); to investigate WTP for energy efficiency improvements (Banfi et al., 2008; Carroll et al., 2016a) and financial instruments to encourage their adoption (Revelt and Train, 1998); and to evaluate the effectiveness of energy efficiency information and labelling (Davis and Metcalf, 2016; Heinzel, 2012; Heinzle and Wüstenhagen, 2012; Newell and Siikamäki, 2014; Sammer and Wüstenhagen, 2006; Shen and Saijo, 2009).

3.2.2 Empirical strategy

DCEs are based on Lancaster's characteristics theory of demand (Lancaster, 1966), according to which agents derive utility not from the good or service *per se* but from its characteristics (Lancsar and Louviere, 2008). Their empirical analysis follows random utility theory (McFadden, 1974), which posits that the utility consumer i derives from choosing good j can be decomposed into an explainable component (V_{ij}) and a random component (ε_{ij}):

$$U_{ij} = V_{ij} + \varepsilon_{ij}. \tag{3.1}$$

The explainable or systematic component can then be expressed as a function of the good's attributes (or at least some of them, X_{ij}) and the consumer's individual characteristics (Z_i):

$$V_{ij} = X'_{ij}\beta + Z'_i\gamma, \tag{3.2}$$

where β and γ are vectors of marginal utilities coefficients to be estimated.

While utility is not directly observed (it remains a latent quantity), we can assume that consumers choose the alternative that gives them the greatest utility out of all the available options. Therefore, the probability that agent i chooses alternative k

is:

$$
\begin{aligned}
P(Y_i = k) &= P(U_{ik} > U_{ij}) \\
&= P(V_{ik} + \varepsilon_{ik} > V_{ij} + \varepsilon_{ij}) \\
&= P(V_{ik} - V_{ij} > \varepsilon_{ij} - \varepsilon_{ik}), \forall j \neq k.
\end{aligned}
\tag{3.3}
$$

For this to be estimable, a joint probability distribution for ε_{ij} needs to be specified. Typically, the error component is assumed to be independently and identically distributed as an extreme value type 1 random variable, thus resulting in a conditional logit form for the choice probabilities:

$$
P(Y_i = k) = \frac{e^{\mu V_{ik}}}{\Sigma_{j=1}^{J} e^{\mu V_{ij}}} = \frac{e^{\mu X'_{ik}\beta + Z'_i\gamma}}{\Sigma_{j=1}^{J} e^{\mu X'_{ij}\beta + Z'_i\gamma}}, \tag{3.4}
$$

where μ is a scale parameter inversely proportional to the variance of the error distribution which cannot be identified and is conventionally set to 1[3] (Lancsar and Louviere, 2008).

The standard conditional logit, however, presents some limitations. The assumption of the error term being iid implies that independence of irrelevant alternatives (IIA) is a key feature of the model. In addition, the preference parameters (βs) are assumed to be the same for all agents. Over the years, different models have been adopted to overcome these limitations. We decide to use a mixed logit model for our analysis in light of its flexibility. As McFadden and Train (2000) demonstrate, any random utility model can be approximated with a mixed logit model.

The mixed multinomial logit model (or mixed logit for simplicity) relaxes IIA[4] and allows for heterogeneity of attribute coefficients across individuals (while keeping them constant for the same individual). In addition, it is also efficient

[3]This implies that we do not estimate the parameters β and γ, but their ratio to the variance of the error distribution σ_ε, e.g. β/σ_ε (Adamowicz et al., 1994; Lancsar and Louviere, 2008).

[4]Other models that relax IIA include nested logit models, multinomial probit models, latent class models, or heteroscedastic error variance models (Lancsar and Louviere, 2008).

with repeated choices and therefore can accommodate the panel structure of the data thanks to its flexible substitution patterns which allow for within subject correlation (Lancsar and Louviere, 2008; Revelt and Train, 1998). The individual parameters are obtained by including an individual-specific stochastic component (δ_i):

$$\beta_i = \overline{\beta} + \delta_i, \tag{3.5}$$

where $\overline{\beta}$ is the population mean (Lancsar and Louviere, 2008). Since, differently from the standard conditional logit model, the mixed logit does not have a closed form solution, it is estimated through maximum simulated likelihood.

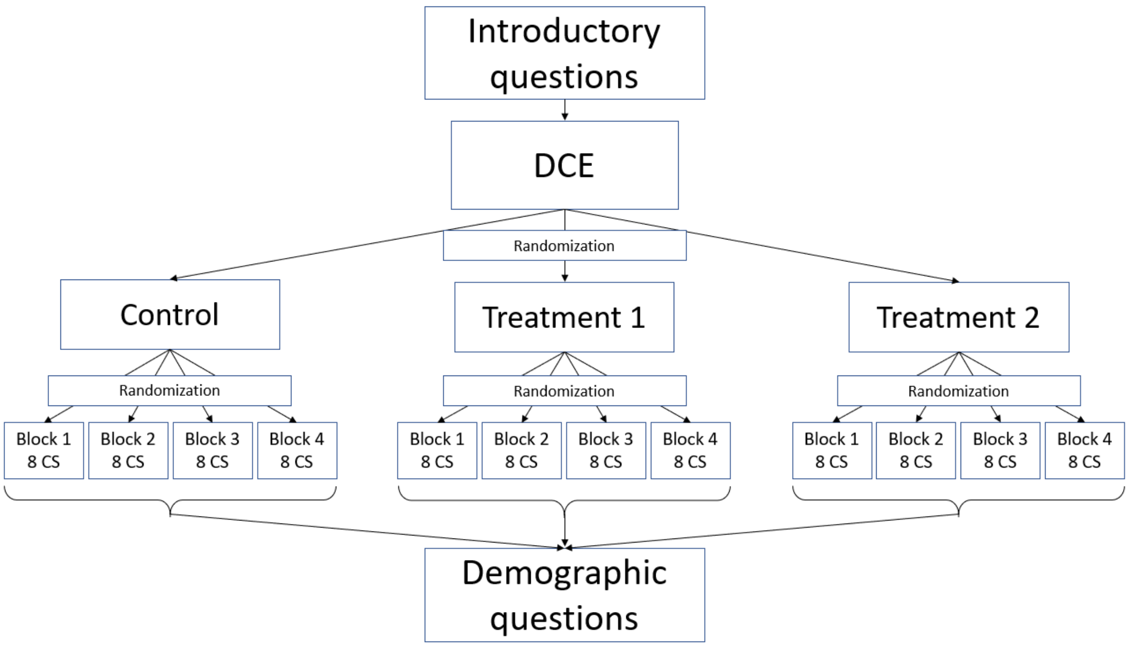
3.2.3 Experiment design

The DCE experimental design was carried out in JMP using the software's Bayesian procedures, which allow for assumptions regarding the direction and variance of utility for each attribute. In particular, with JMP, we assume a utility range of one (split evenly across attribute levels) and a variance of 0.25. There were no dominant alternatives. Such a design enables us to assume, for example, that price is negatively correlated with utility whereas the number of stars in consumer rating is positively correlated.

The final design contained 32 pairs of choices — called choice sets (CS) — which were split across four blocks. Each choice set consisted of two tumble dryers and an opt-out alternative. Including an "opt-out" or "neither" alternative is desirable in contexts where respondents are presented with hypothetical pairs, since its absence would force them to choose between potentially unappealing options, a choice that might not be made in a real world scenario (Lancsar and Louviere, 2008).

Respondents were randomly assigned to one of three groups which differed in the way in which energy information is displayed — namely control with the customary energy information, treatment 1 with generic energy expenditures, and treatment 2 with personalised energy expenditures. In addition, they were also randomly assigned into one of the four blocks, leading to eight choices per respondent. Figure 3.1 reports the structure of the DCE[5], highlighting the points of randomization.

Figure 3.1: Structure of the discrete choice experiment



The tumble dryers presented in each choice set vary over five attributes, which were chosen on the basis of previous research on household electric appliances (Carroll et al., 2016b; Heinzel, 2012; Heinzle and Wüstenhagen, 2012; Sammer and Wüstenhagen, 2006; Shen and Saijo, 2009), through focus groups and in consultations with salespersons at retail stores[6,7]. The selected attributes are:

(i) *Price*. Price is based on the range of models available on the market on electrical retailer websites[8] in each country at the time of experimental design

[5]In Figure 3.1 CS stands for choice set.

[6]The DCE was run in parallel with a field experiment.

[7]One of the questions in the survey that accompanied the DCE asked participants to rate the importance of several characteristics in the hypothetical purchase of a new tumble dryer. Responses confirm that the selected attributes are also those considered more important by the individuals in our sample (results available from the authors upon request). The question can be found in Appendix B.2.

[8]Like Argos, Best Buy, Currys and The Home Depot.

(2018).

(ii) *Brand*. Brand is characterized as either "established" or "new". An established brand is one with more than 5 years of activity that has developed a solid relationship with its customers. A new brand is one which has been operating for less than 5 years and has still not developed a solid relationship with the customers. Such a categorization was chosen to facilitate comparisons, as the survey was conducted in four different countries with different leading brands and attitudes.

(iii) *Capacity*. Capacity is measured in kilograms (kg) for the Irish and British versions and cubic feet (cu ft) for the Canadian and American ones.

(iv) *Customer rating*. Customer rating takes the form of a typical star rating[9].

(v) *Energy efficiency*. Energy efficiency is based on physical energy consumption (kWh/annum), also consistent with typical products available on electrical retailer websites.

At the beginning of the DCE, all attributes were presented and described to respondents with the aid of images. A summary of the attributes and their levels in each country is reported in Table 3.1. The way these attributes and levels were introduced to respondents is displayed in Figures B.1-B.8 in Appendix B.1.

Participants were randomly assigned to one of three groups that differed in the way in which the energy efficiency attribute is presented. In the control group, energy efficiency is presented in the form of the typical energy label customary in the respective country: that is, the letter scale of the EU Energy Label for Ireland and the United Kingdom, and the EnergyStar logo for Canada and the United States. Tumble dryers were assigned a letter from C to A+++ or the EnergyStar

[9]On electrical retailer websites there are almost no products with less than 3-star ratings. Therefore we use the range 3-5 stars in the experiment.

Table 3.1: Attributes and levels by country and treatment groups

Attributes	Country	Levels					
Price	IRE(€),UK(£)	200	400	600	800	1000	1200
	CAN(CAN$)	400	600	800	1000	1200	1400
	USA($)	300	500	700	900	1100	1300
Brand	All	Established			New		
Capacity	IRE,UK(kg)	7	8	9	10		
	CAN,USA(cu ft)	6	7	8	9		
Customer rating	All(# stars)	3	4	5			
Energy efficiency (control, based on kWh/annum)	IRE,UK	C	B	A	A+	A++	A+++
	CAN,USA	No	No	No	Yes	Yes	Yes
Energy cost: 10-years cost based on average usage (treatment 1)	IRE(€)	1100	950	800	650	500	350
	UK(£)	950	825	700	575	450	325
	CAN(CAN$)	930	810	690	570	450	330
	USA($)	850	730	640	490	370	250
Energy cost: 10-years cost based on personalised usage (treatment 2)	IRE(€)	Based on respondent's self-reported use and national average electricity prices					
	UK(£)						
	CAN(CAN$)						
	USA($)						

Notes: Energy efficiency and energy costs are based on the underlying level of physical energy consumption (kWh/annum). E.g. for the EU Energy Label: C = 636 kWh/a, B = 551 kWh/a, A = 466 kWh/a, A+ = 381 kWh/a, A++ = 296 kWh/a and A+++ = 211 kWh/a. An equivalent relationship applies to the EnergyStar logo. For energy costs the relations are expressed by Equations 3.6 and 3.7. In the case of generic energy costs the same average usage was assumed in all countries, the variation comes from differences in electricity prices. Conversely, in the case of personalised energy costs variation comes from both usage and electricity prices.

logo based on their physical energy consumption (kWh/a) as shown in Table 3.1.

Treatment 1 frames energy efficiency as the 10-years energy costs according to the formula:

$$Energy\ cost = kWh/a \times national\ electricity\ price \times 10\ years, \tag{3.6}$$

where the physical energy consumption is considered for an average of 160 cycles per year[10]. In treatment 2, we still present energy efficiency as the the 10-years energy costs, however this is now based on individual-specific self-reported usage[11]:

$$\begin{aligned} Energy\ cost =& \frac{kWh/a}{160} \times individual\text{-}specific\ weekly\ use \\ & \times 52\ weeks \times national\ electricity\ price \times 10\ years. \end{aligned} \tag{3.7}$$

Figures B.6-B.8 in Appendix B.1 provide examples of the descriptions of the energy efficiency attribute given to participants in each of the three groups, and Figure B.9 of the choice sets.

Estimation strategy

As mentioned in Section 3.2.2, the mixed multinomial logit model allows to distinguish between parameters that are constant for all respondents (*non-random parameters*), and parameters that vary by respondent (*random parameter*). Therefore, the X_{ij} vector consists of both attributes with a constant impact on utility (N_{ij}), and attributes which impact varies by individual (R_{ij}).

We keep price and consumer rating as constant for all individuals, since it is reasonable to assume that everyone prefers products with a lower price and a

[10] Average usage is based on the assumptions underlying the EU Energy Label.
[11] Electricity prices are €0.17 in Ireland, £0.15 in the United Kingdom; CAN$0.1465 in Canada and $0.1312 in the United States. They all include VAT.

higher star rating. On the other hand, capacity, brand and energy efficiency are allowed to vary by respondent, since it is possible that different individuals have different preferences over these attributes. We relax this categorization in the robustness checks reported in Appendix B.3.

In all specifications we define energy efficiency as a dichotomous variable "high efficiency versus low efficiency" (EE_{ij}), based on the underlying level of physical energy consumption used in the experimental design[12]. More specifically, the variable takes the value 1 for the three least efficient classes (which correspond to the higher consumption levels), and value 2 for the three most efficient classes (corresponding to lower consumption levels)[13]. The effect of reframing energy efficiency in monetary terms is captured by an interaction between the energy efficiency variable and treatment dummies.

We estimate the following model in each country:

$$U_{ij} = \alpha_j + N'_{ij}\beta_N + R'_{ij}\beta_{Ri} + \beta_{EET1i}(EE_{ij} \times T1_i) + \beta_{EET2i}(EE_{ij} \times T2_i) + Z'_i\gamma + \varepsilon_{ij}, \tag{3.8}$$

where α_j is an opt-out alternative-specific constant; Nij is the vector of non-random parameters and β_N a vector of their coefficients; R_{ij} is the vector of random parameters (including energy efficiency) and β_{Ri} a vector of their individual-specific coefficients; $T1_i$ and $T2_i$ are dummy variables for treatment 1 (the generic 10-years cost of electricity) and treatment 2 (the personalised 10-years cost of electricity), respectively; and Z_i is the vector of individual characteristics[14].

As aforementioned, the treatment effects are captured by an interaction between energy efficiency and the treatment dummies. This ensures that the coefficient of

[12] See the notes of Table 3.1.

[13] We code the energy efficiency variable, like other binary attributes, as (1-2) rather than (0-1) because all attributes take value zero for the opt-out alternative.

[14] In order to include individual characteristics in the Stata routine `mixlogit` they have to be interacted with alternative-specific constants.

energy efficiency alone gives an indication of the baseline value of this attribute on individuals' utility, and the interaction terms represent the incremental effect generated by our treatments. The coefficients of the interaction terms (β_{EET1i} and β_{EET2i}) are also assumed to be individual-specific.

The models are estimated through maximum simulated likelihood using 1000 Halton draws. Standard errors are clustered at the individual level.

3.3 Data description

The DCE was embedded in a survey distributed in November 2018 by the market research company ResearchNow in all four countries. Our target was individuals who own and utilize a tumble dryer in their everyday life. Therefore, at the beginning of the survey, we screen out participants who do not have a tumble dryer in their home, or who never use it. The survey included demographic quotas based on National Census information to ensure a representative sample in each country.

The initial sample consisted of a total of 2,676 individual observations. However, we exclude respondents who did not provide any demographic information, who did not complete all 8 choice sets in the DCE[15], or who gave an extreme answer to the question: "Approximately how many times a week do you use your tumble dryer?"[16]. This leaves us with 634 valid respondents in the Canadian sample (214 in the control group, 205 in treatment 1 and 215 in treatment 2); 581 in Ireland (198 in the control group, 189 in treatment 1 and 194 in treatment 2); 655 in the United Kingdom (220 in the control group, 218 in treatment 1 and 217 treatment 2); and 657 in the United States (208 in the control group, 228 in treatment 1 and 221 in

[15]There was one participant in the UK sample and one participant in the US sample for whom we have answers to just 4 choice sets.

[16]Specifically, we exclude from the analysis participants who report to use the tumble dryer, on average, more than 21 times per week (7 respondents). This exclusion leaves unaltered the final outcomes of the analysis.

treatment 2).

As a first step, we want to test whether there are significant differences between the four countries in our sample, or if it is possible to pool them together in our analysis. For this reason, we conduct likelihood-ratio Chow tests to verify if it is possible to pool Ireland and the United Kingdom in a European group, Canada and the United States in an American group, as well as all countries together.

Table 3.2: Likelihood-ratio test for pooled and country groups data

	Log likelihood	Degrees of freedom	LR test statistic	p-value
Ireland	-3,420.718			
UK	-4,156.447			
Europe	-7,609.117	29	63.902	0.0002
Canada	-4,008.181			
USA	-4,174.241			
America	-8,214.446	29	64.048	0.0002
Canada	-4,008.181			
Ireland	-3,420.718			
UK	-4,156.447			
USA	-4,174.241			
Pooled	-15885.673	87	252.169	0.0000

Notes: The log-likelihoods are derived from mixed logit models with the same specifications as those used in Table 3.4.

Table 3.2 reports the results of the tests. As we can see, for all the combinations considered, it is possible to reject the null hypothesis that pooling the countries together is the same as treating them individually. Therefore, in the remainder of the analysis we will run separate models for each country. A possible alternative would be to run the model on the pooled sample and include interactions with a country variable. Such an approach has not been chosen because it would require adding two triple interactions between energy efficiency, country and treatment dummies[17] to Equation 3.8, which would make the interpretation of the treatment

[17] In addition, the model should also include all the two-way interactions as well as the individual variables.

effects considerably more cumbersome.

As mentioned in Section 3.2.3, participants to the experiment were randomly assigned to the control group or one of the two treatments. We want to control if, in the various countries, there are differences between the three groups in terms of their demographics and other relevant individual characteristics. The Levene's tests for homogeneity of variances report no significant differences in most of the cases[18]. This is also largely confirmed by the pairwise *t*-tests reported in Table 3.3, which do not show major differences in the averages between the control and treatment groups. The most notable differences are represented by a greater proportion of participants who hold a degree in the personalised energy cost treatment in Canada, and in the control group in Ireland. It is worth noting that most of the differences are relatively small compared to the dimension of the corresponding variable. Overall, these results suggest that the three groups (control group, treatment 1 and treatment 2) in each country do not present fundamental differences and are, therefore, comparable.

Table 3.3: Descriptive statistics and pairwise comparisons by country and treatment groups

	Control	Treatment 1	Treatment 2	Difference C - T1	Difference C - T2	Difference T1 - T2
A. IRELAND						
Age	3.500	3.545	3.531	-0.045	-0.031	0.014
	(1.688)	(1.733)	(1.689)			
Female	0.535	0.460	0.490	0.075	0.046	-0.029
	(0.500)	(0.500)	(0.501)			
Marital status	1.924	1.947	1.928	-0.023	-0.004	0.019
	(0.799)	(0.867)	(0.908)			
Degree	0.803	0.672	0.624	0.131***	0.179***	0.048
	(0.399)	(0.471)	(0.486)			
Working	0.742	0.667	0.675	0.076	0.067	-0.009
	(0.438)	(0.473)	(0.469)			
Env. Concern	3.914	4.011	3.948	-0.096	-0.034	0.062
	(1.174)	(1.135)	(1.203)			

[18]The results of the Levene's tests are not reported in the paper but they are available from the authors upon request.

Table 3.3 — continued

	Control	Treatment 1	Treatment 2	Difference C - T1	Difference C - T2	Difference T1 - T2
Income	3.081	2.899	2.943	0.181*	0.138	-0.044
	(0.880)	(0.992)	(1.014)			
Impatience	6.439	6.503	5.959	-0.063	0.481**	0.544**
	(2.107)	(2.170)	(2.280)			
Risk	5.646	5.603	5.304	0.043	0.342*	0.299
	(1.932)	(2.123)	(2.117)			
Tumble dryer use	3.576	3.657	3.572	-0.081	0.004	0.085
	(2.414)	(2.976)	(2.687)			
B. UNITED KINGDOM						
Age	3.732	3.766	3.677	-0.034	0.054	0.089
	(1.748)	(1.797)	(1.792)			
Female	0.545	0.564	0.502	-0.019	0.043	0.062
	(0.499)	(0.497)	(0.501)			
Marital status	1.968	1.995	1.963	-0.027	0.005	0.032
	(0.665)	(0.823)	(0.907)			
Degree	0.627	0.606	0.604	0.022	0.024	0.002
	(0.485)	(0.490)	(0.490)			
Working	0.636	0.610	0.636	0.026	0.000	-0.026
	(0.482)	(0.489)	(0.482)			
Env. Concern	3.623	3.789	3.544	-0.166	0.079	0.245**
	(1.212)	(1.234)	(1.239)			
Income	3.173	3.275	3.309	-0.103	-0.136	-0.034
	(1.001)	(1.015)	(1.010)			
Impatience	6.232	6.408	6.300	-0.176	-0.068	0.109
	(2.132)	(2.134)	(2.092)			
Risk	5.173	5.266	5.290	-0.093	-0.118	-0.024
	(2.108)	(2.327)	(2.170)			
Tumble dryer use	4.109	3.720	3.323	0.389	0.787***	0.398
	(3.247)	(2.954)	(2.668)			
C. CANADA						
Age	3.715	3.761	3.726	-0.046	-0.011	0.035
	(1.757)	(1.767)	(1.781)			
Female	0.505	0.478	0.540	0.027	-0.035	-0.061
	(0.501)	(0.501)	(0.500)			
Marital status	1.986	1.976	2.009	0.010	-0.023	-0.034
	(0.947)	(0.942)	(0.922)			
Degree	0.673	0.668	0.772	0.005	-0.099**	-0.104**
	(0.470)	(0.472)	(0.420)			
Working	0.607	0.600	0.665	0.007	-0.058	-0.065
	(0.489)	(0.491)	(0.473)			
Env. Concern	3.883	3.893	3.842	-0.010	0.041	0.051

Table 3.3 — continued

	Control	Treatment 1	Treatment 2	Difference C - T1	Difference C - T2	Difference T1 - T2
	(1.230)	(1.162)	(1.145)			
Income	3.341	3.195	3.270	0.146	0.071	-0.075
	(1.007)	(1.058)	(1.010)			
Impatience	6.528	6.498	6.340	0.030	0.189	0.158
	(2.096)	(2.069)	(2.021)			
Risk	5.528	5.415	5.474	0.113	0.054	-0.060
	(2.157)	(2.200)	(2.055)			
Tumble dryer use	3.220	3.215	3.381	0.005	-0.162	-0.167
	(2.570)	(2.106)	(2.622)			
D. UNITED STATES						
Age	3.553	3.632	3.498	-0.079	0.055	0.134
	(1.713)	(1.717)	(1.752)			
Female	0.538	0.491	0.575	0.047	-0.036	-0.083*
	(0.500)	(0.501)	(0.496)			
Marital status	2.005	2.070	2.023	-0.065	-0.018	0.048
	(0.909)	(0.941)	(0.881)			
Degree	0.702	0.675	0.633	0.026	0.068	0.042
	(0.459)	(0.469)	(0.483)			
Working	0.606	0.605	0.624	0.001	-0.019	-0.019
	(0.490)	(0.490)	(0.485)			
Env. Concern	3.909	3.794	3.828	0.115	0.081	-0.034
	(1.257)	(1.286)	(1.320)			
Income	3.188	3.281	3.222	-0.093	-0.034	0.059
	(1.120)	(1.150)	(1.120)			
Impatience	6.370	6.500	6.448	-0.130	-0.078	0.052
	(2.172)	(2.385)	(2.128)			
Risk	5.620	5.785	5.457	-0.165	0.163	0.328
	(2.287)	(2.347)	(2.160)			
Tumble dryer use	4.069	4.132	4.195	-0.063	-0.126	-0.063
	(3.300)	(3.300)	(3.121)			

Notes: Columns 1-3 report means and standard deviations (in parenthesis) of the various demographics for each treatment group in each country. Columns 4-6 report pairwise mean differences and the statistical significance of the *t*-tests. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The list of demographic questions included in the experiment is reported in Appendix B.2.

The age of respondents in our samples is in line with national averages: mean age ranges between 3 (from 35 to 44 years of age) and 4 (from 45 to 54 years of age), and average age is 37.4 in Ireland (Central Statistics Office, 2016a), 40.3 in the United Kingdom (Office for National Statistics, 2019b), 41.1 in Canada (Statistics Canada, 2020a), and 38.5 (median) in the United States (U.S. Census Bureau, 2019a). Also the gender ratio mirrors national averages, being close to 50% with a slight prevalence of females, in general. And so does the percentage of individuals in our samples that is working, albeit with some differences. Employed respondents range from 66-74% in the Irish sample, with a participation rate in the labour force of 61.4% in the country (Central Statistics Office, 2016b). They are around 60% in both the Canadian and US samples, with percentages of population in the labour force of 64.9% (Statistics Canada, 2016) and 63% (U.S. Census Bureau, 2019c), respectively. For the UK, the employment rate between 16 and 64 years of age was 75.2% in 2020 (Office for National Statistics, 2021), which is greater than the percentage of respondents who report to be employed in the British sample, between 61% and 63%, although these measures are not immediately comparable.

However, we also detect some discrepancies. First of all, the percentage of participants with tertiary education or higher in each sample is considerably greater than the respective country average — 42% in Ireland (Central Statistics Office, 2016c) and the UK (Office for National Statistics, 2017), 54% in Canada (Statistics Canada, 2016), and 32.1% in the United States (U.S. Census Bureau, 2019b). In addition, it also seems that individuals in our sample are more likely to be in a relationship than the corresponding national population. In fact, the percentage of respondents stating to be married or in a domestic relationship in the Irish sample is 63.7%, against a national average of married couples of 37.6% (Central Statistics Office, 2017); it is 72.2% in the UK, against the percentage of people being in a couple of 60% at the national level (with 50.4% being married or in a civic relationship; Office for National Statistics (2019a)); 61.6% in Canada, where the percentage of people married or living together is 47.6%[19]; and 64% in the USA, where 52% of individuals older than 15 are married[20].

It should be noted that while this sample is broadly representative of the main national population in each country, we do not have information on the population of 'typical tumble dryer owners'.

3.4 Results

Table 3.4 presents the results of mixed logit regressions for the four countries separately. These are considered over the whole sample for each country, with the inclusion of interaction variables to account for treatment groups one and two as shown in Equation 3.8. In Appendix B.3 we report separate models for the control group, the generic cost information treatment and the personalised cost information treatment: results were qualitatively identical. All regressions control for income, gender, living area, whether the individual holds a degree, environmental concern, impatience, risk attitude and tumble dryer usage[21].

Although the magnitude of the coefficients does not have an immediate interpretation, their sign gives us an indication of the effect on the utility function. As it can be seen, attributes have the expected effect on utility, with, for example, price being negative — signifying that respondents would prefer cheaper products —, and star rating and capacity being positive – meaning that people would rather purchase a tumble dryer with better reviews and that can accommodate more clothes[22].

Brand takes value 1 for an established brand and 2 for a new one, hence the nega-

[19]Own calculations based on data from Statistics Canada (2020b).

[20]Own calculations based on data from the U.S. Census Bureau (2021).

[21]These coefficients are not displayed in Table 3.4 for ease of presentation. However, later in the paper we investigate if the effect of our manipulations differs by individual characteristics.

[22]In the experimental instructions, participants were told that each model would fit the space they have available, so the size of the tumble dryer, which is connected to its capacity, does not represent an issue when selecting the preferred option.

Table 3.4: Mixed logit models

	(1) Ireland	(2) UK	(3) Canada	(4) USA
Non-Random Parameters in Utility Function				
Constant (neither option)	0.4281	0.7840	1.7037	2.3724**
	(1.0975)	(0.9339)	(1.1259)	(0.9462)
Price	-0.0031***	-0.0031***	-0.0026***	-0.0027***
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
Stars	0.5559***	0.5493***	0.8139***	0.7344***
	(0.0483)	(0.0469)	(0.0459)	(0.0463)
Random Parameters in Utility Function				
Capacity	0.2338***	0.1046***	0.1959***	0.2348***
	(0.0291)	(0.0270)	(0.0294)	(0.0276)
Brand	-0.2587***	-0.2174***	-0.4637***	-0.1405**
	(0.0680)	(0.0691)	(0.0640)	(0.0650)
Energy efficiency	1.0191***	0.5424***	1.3251***	0.7712***
	(0.1028)	(0.0987)	(0.1090)	(0.0925)
EE × T1	-0.0235	0.0516	-0.3036**	-0.0483
	(0.1428)	(0.1321)	(0.1476)	(0.1270)
EE × T2	-0.0477	0.2506*	-0.3240**	-0.1085
	(0.1468)	(0.1300)	(0.1325)	(0.1248)
Model statistics				
Observations	13944	15720	15216	15768
Clusters	581	655	634	657

Notes: This table reports the results of mixed logit regressions of respondents' choices in each country separately. Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. All regressions control for income, gender, living area, whether the individual holds an degree, environmental concern, impatience, risk attitude and tumble dryer usage. Standard errors are clustered at the participant level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

tive sign of the coefficients represents the fact that respondents prefer products of established brands. Energy efficiency presents positive and significant coefficients for all countries: more efficient models have a positive impact on utility. However, the interaction terms are insignificant in most of the cases, which means that presenting energy efficiency information in monetary terms (treatment 1) does not have any relevant effects on people's choices, nor does personalising this information (treatment 2) produce any appreciable difference. There are however two exceptions. In the Canadian sample we find negative and statistically significant coefficients for the two interactions terms. This suggests that displaying energy efficiency information in monetary terms, rather than the simple EnergyStar logo, reduces utility. Conversely, for the UK, we detect a positive and significant effect (at the 10% level) of personalised energy costs information.

It is important to notice that the negative effect of out treatments in the Canadian sample can still be consistent with monetary information helping respondents making better investment decisions. The attribute values generated in JMP yield a composition of the generic energy information treatment in which the more efficient option has the highest total lifetime cost — given by the sum of the purchasing price and the 10-years cost of electricity — in 14 out to 32 choice sets (43.75% if the times)[23]. Respondents in the Canadian sample where shown the block with the greater frequency of choice sets presenting this characteristic more often than other countries. In addition, also in the personalised energy information treatment, Canada has the highest percentage of choice sets with the more efficient option having the highest lifetime cost. However, said percentage is still below 50% and not considerably different from that of the other countries[24]. Some extra considerations are needed. First, the experimental design excludes

[23]Block 1 presents an even distribution. Block 2 has 3 choice sets with the more efficient option being more expensive, and 5 choice sets with the less efficient option being more expensive. Block 3 has 5 choice sets with the more efficient option being more expensive, and 3 choice sets with the less efficient option being more expensive. Block 4 has 2 choice sets with the more efficient option being more expensive, and 6 choice sets with the less efficient option being more expensive.

[24]It is 42.6% in Canada, 41.9% in the United Kingdom, 39.9% in the United States and 39.5% in Ireland.

Table 3.5: Mixed logit models willingness to pay

	Ireland	UK	Canada	USA
Stars	182.19	178.45	317.21	274.12
	[150.95 ; 213.43]	[149.01 ; 207.88]	[277.84 ; 356.57]	[238.53 ; 309.72]
Capacity	76.64	33.99	76.35	87.66
	[57.38 ; 95.89]	[16.51 ; 51.46]	[53.59 ; 99.10]	[66.80 ; 108.52]
Brand	-84.79	-70.61	-180.71	-52.46
	[-127.35 ; -42.23]	[-114.09 ; -27.12]	[-228.70 ; -132.72]	[-99.42 ; -5.50]
EE	334.00	176.18	516.44	287.89
	[264.60 ; 403.39]	[112.91 ; 239.46]	[428.89 ; 603.99]	[219.38 ; 356.40]
EE $\times$ T1	-7.69	16.77	-118.31	-18.03
	[-99.46 ; 84.08]	[-67.27 ; 100.82]	[-230.99 ; -5.63]	[-110.91 ; 74.84]
EE $\times$ T2	-15.65	81.40	-126.27	-40.50
	[-109.94 ; 78.64]	[-1.31 ; 164.11]	[-227.95 ; -24.59]	[-131.82 ; 50.81]

Notes: This table reports the willingness to pay of respondents in each country for the tumble dryer's attributes. Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. The 95% confidence intervals are reported in brackets.

the possibility of strictly dominated options. Hence, for example, it makes perfect sense that the more efficient tumble drier presents a higher total lifetime cost if it has a substantially greater capacity, since that would imply a higher purchasing price as well as operating costs. Second, the choice sets did not report the total lifetime cost, only price and energy cost information separately. It is therefore not possible to know whether respondents carried out such a calculation when taking their decisions. In light of this, we cannot confidently say if the negative effect of monetary energy information in the Canadian sample is the result of people correctly choosing the option with the lower lifetime cost or if it is attributable to other factors. The analysis presented later in this section and in Appendix B.4 tries to shed light on some of these potential factors.

Table 3.5 presents respondents' willingness-to-pay[25] for the various attributes. As it can be seen, energy efficiency is the attribute with the highest WTP in all countries. When energy information is presented as the EU Energy Label's letter

[25]Willingness-to-pay for attribute a is obtained as the ratio between the attribute's coefficient and the price coefficient, $WTP_a = -\frac{\beta_a}{\beta_{price}}$.

scale, Irish participants are willing to pay €334.98 more for a tumble dryer from the three most efficient classes with respect to one from the three least efficient ones, and British participants £182.51 more. Similarly, respondents are willing to pay $283.51 more in the United States and Can$516.44 more in Canada for a product with the EnergyStar certification.

Consistent with the results in Table 3.4, our manipulations of the way in which energy efficiency information is displayed have limited impact on WTP[26]. Once again, we highlight a negative effect of both treatments in the Canadian sample, where the willingness-to-pay for energy efficiency decreases by roughly Can$118 when generic energy costs are provided, and by Can$126 with personalised energy costs. Whereas, in the United Kingdom, personalised energy information based on self-reported use patterns increases consumers' willingness-to-pay for energy efficiency by more than £81 with respect to the baseline level under the current letter scale framing[27], making consumers willing to pay almost £258 more for a tumble dryer in the three most efficient classes.

In a series of robustness checks we have relaxed the definition of random and non-random parameters in two ways. First, we allow all attributes except for price, as well as the opt-out alternative-specific constant, to be individual-specific, hence estimating an error component model. Second, we adopt the opposite approach and restrict all coefficients to be constant for all respondents, which yields the classic conditional logit model. The results, reported in Appendix B.3, are substantially in line with the mixed logit estimations presented in Table 3.4 and Table 3.5.

One aspect that is has not been specifically accounted for in the current analysis are differences in weather conditions between the four countries. Weather conditions are likely to play a role in determining tumble drier usage in general and their effect

[26] In almost every case, the 95% confidence intervals (in brackets) include 0.
[27] The 90% confidence interval is [11.99 ; 150.82].

may differ from country to country depending on the variation in temperatures, precipitations and solar radiation from season to season. However, while not explicitly captured, weather is implicitly reflected in the electricity prices in each country, which reflect electricity supply and demand (both of which are impacted by weather and seasons in each country, among many other things). Therefore, to the extent that country-specific electricity prices are a major component of Equations 3.6 and 3.7, which are used to compute generic and personalised energy costs respectively, the fact that we do not explicitly control for differences in weather conditions between countries does not represent a major concern for our results.

The overall absence of a positive effect of providing personalised energy information, although contrary to our prior beliefs, is not unprecedented. Considering the automobile sector, Allcott and Knittel (2019) evidence a limited impact of personalised fuel costs on individuals' purchasing decisions, which tended to disappear a few months after the intervention. A possible explanation in the context of our analysis is that tumble dryer usage in the sample could be fairly limited. If this was the case, the EU Energy Label and the EnergyStar logo, by somewhat shrouding the actual monetary value of energy costs, might induce people to overvalue energy efficiency.

It is therefore important to investigate whether the effect of framing energy information in alternative ways differs for various subgroups based on personal attitudes and demographics. One hypothesis, which has already been introduced, is that a limited average usage of the tumble dryer might make energy costs less relevant than the more general information contained in the current labels. I.e. if a household has a low usage of their tumble dryer, then monetary labelling may be less effective than the more general kWh energy label. Second, in light of the evidence, highlighted by previous studies, suggesting that people are typically not very good at translating physical consumption into energy expenditures, one could expect that the provision of more explicit information might benefit mainly those with lower levels of education. A third hypothesis is that people concerned about the environment will tend to chose the most efficient product irrespective of the way in which energy information is framed, while those less concerned will pay more attention to the monetary aspects of energy consumption. Finally, income-constrained individuals can benefit more from energy information reported in monetary terms if energy bills are a considerable proportion of their expenditures.

With this in mind, we run our models splitting the samples on the basis of the levels of self-reported weekly tumble dryer usage, educational attainments, environmental concern and income. For tumble dryer usage, we define as low usage values smaller than or equal to the median of the respective country, mid-high usage between the median and the 90th percentile, while very-high usage corresponds to the top 10th percentile in each country[28]. For education, we distinguish between respondents with and without a degree. For environmental concern we split the sample into participants who say to be concerned or extremely concerned about the environment, and those who are slightly concerned, not concerned or do not know. Lastly, we separate between people stating to live comfortably or very comfortably on current income, and those who do not live comfortably or are coping on current income.

We defer results of these estimations and the corresponding WTP to Tables B.10-B.17 in Appendix B.4. Here we present a discussion of their implications.

For all countries, and in particular for Canada, there are considerably more people in the low usage category. In fact, it is for this subgroup that personalised energy

[28]In all four countries the median is equal to 3 weekly cycles. The 90th percentile is 6 in Canada and 7 in Ireland, the United Kingdom and the United States. Therefore, low usage corresponds to 3 or fewer weekly cycles; mid-high usage is between 4 and 7 (both included) weekly cycles in Ireland, the UK and the US and between 4 and 6 in Canada; very-high usage is given by more than 7 cycles in Ireland, the UK and the US and more than 6 in Canada.

costs lead to a significant decrease in consumers' utility in the Canadian sample. On the other hand, for respondents in the top 10th percentile of the respective distribution, personalised information presents positive coefficients in all countries, with the effect being significant for the UK. Both these instances seem to confirm that the results in Tables 3.4 and 3.5 could be due, at least in part, to a limited average usage of the tumble dryer in our sample, which would make the current labels more salient than actual energy expenditures. In addition, as mentioned before, it is possible that for the low-usage subgroup the less efficient option is more convenient in terms of total lifetime costs.

Conversely, the hypothesis that providing more accurate and personalised energy information should benefit mostly those with lower levels of education is not substantiated. Although the coefficients of our two treatments become positive (but insignificant) for respondents without a degree in the Canadian sample, this effect does not apply to the other countries. If anything, we observe outcomes that are somewhat contrary to this belief. The positive effect of the personalised energy costs treatment in the United Kingdom comes from the subgroup of respondents who hold a degree. While generic energy costs generate a negative effect for Irish participants without a degree.

In each country, participants who state they are concerned about the environment have a higher WTP for energy efficiency than those who say they are not. In the subgroup of less concerned respondents, providing energy costs has a general positive impact on consumers' utility, which represents a statistically significant improvement with respect to the EU Energy Label's letter scale in the Irish and British samples. Presumably, these households care more about the consequences that energy use has on their wallet than on the environment, so reframing energy information in monetary terms improves their WTP for energy efficiency since it allows them to realise how much less they would spend with a more efficient model. On the other hand, monetary information has a generally negative effect for individuals concerned about the environment, which is statistically significant in the Canadian sample. People who care about the environment would buy the more efficient model to reduce their environmental impact, and this impression might be conveyed more strongly by a graphical representation like the EnergyStar logo or the colour-coded letter scale scheme used in current labels than by energy expenditures, especially if usage is low. Hence, providing monetary information seems to crowd out those who would buy a more energy efficient tumble dryer for environmental motivations.

Finally, we do not detect a clear impact of individuals' income on the effectiveness of our treatments. Personalised energy costs information does increase utility for income-constrained people in the United Kingdom, but this effect does not translate to the other countries — apart from a positive but insignificant coefficient in the Irish sample. In addition, the negative effect of our treatments in the Canadian sample evidenced in Table 3.4 interests both more and less wealthy individuals.

3.5 Conclusion

It has been asserted that the current kWh information reported on energy labels might not be sufficient to help consumers make well-informed energy efficiency investments. The literature has documented that individuals often struggle to interpret energy information when provided in physical units. Reframing energy information in monetary terms could allow them to make better and more informed purchasing decisions. Prior studies have investigated the effect of providing monetary energy information in several contexts. Outcomes have been mixed, and it is not clear whether this is to be attributed to the use of different core products, the employment of different methodologies, or the fact that they were conducted in different countries. This paper represents the first attempt to clarify that ambiguity by examining the impact of lifetime energy expenditures employing the same experiment in a multi-country setting.

With an online randomized discrete choice experiment we study individuals' preferences for tumble dryers in Canada, the Republic of Ireland, the United Kingdom and the United States. Energy information is presented in three forms. In the control group it follows the typical framing of the energy labels in each country. The first treatment converts the physical consumption at the basis of energy labels into the 10-years energy costs, assuming a uniform usage of 160 cycles per year. The second treatment adopts self-reported use patterns to derive individual-specific 10-years energy costs.

Our findings show that monetary information has different impacts in different countries. In Ireland and the United States we fail to detect any significant effect of providing lifetime energy expenditures. In Canada, both generic and personalised monetary information reduce the willingness-to-pay for energy efficiency with respect to the classic EnergyStart logo. Whereas in the United Kingdom the individual-specific 10-years energy costs has a positive impact on people's preferences for energy efficiency. Disentangling the effect based on demographic and socio-economic characteristics highlights that the negative effect comes primarily from individuals who make less frequent use of the tumble dryer, and that monetary information seems to crowd out respondents who would buy a more efficient model for environmental motivations.

While framing information in monetary terms appears as a promising and easy to implement option to favour the uptake of more efficient appliances, the results of this paper suggest that a generic measure would have little impact. This is not to say that there is no scope for improvements on how to convey energy efficiency information, but that an effective intervention should be tailored to the characteristics of the context where it is to be implemented. Countries and individuals differ for a plethora of reasons, and policymakers should carefully consider these peculiarities to design effective policy solutions. Moreover, often times there are also considerable differences between regions of the same country — for example in terms of energy prices (Davis and Metcalf, 2016) — which is important to consider and evaluate. Unfortunately, our paper does not have a large enough sample size to accommodate a regional within-country approach. Therefore, future research with larger samples could conduct such an analysis.

Our work paves the way for new research to examine additional products. Examples already exist as stand-alone analyses for refrigerators (Andor et al., 2020; Jain et al., 1994; Kallbekken et al., 2013), TV sets (Heinzel, 2012), washing machines (Department of Energy and Climate Change, 2014), cars (Allcott and Knittel, 2019) and the housing market (Carroll et al., 2021). In addition, labelling is only one means of promoting investments in energy efficiency — others include, but are not limited to, direct regulation, tax reductions, financial incentives, etc. Therefore, future efforts should be devoted to develop structured, multi-country studies to understand what is the most effective intervention in each specific context.

Finally, although stated preferences methods represent an invaluable tool to investigate consumers' behaviour in a variety of contexts and to assess the effectiveness of new policies thanks to their flexibility and ease of implementation, some studies have found differences in effects between online and field trials (Allcott and Taubinsky, 2015). So, future research should consider the value coupling survey data with field experiments and revealed preferences data.

4 Air Pollution Affects Decision-Making: Evidence from the Ballot Box

4.1 Introduction

Poor air quality affects many domains of life. Besides well-documented adverse impacts on health and the environment, it has important psychological, economic and social consequences (see Lu, 2020, for a recent review). A high level of air pollution can impair cognitive functioning, thereby limiting people's ability to process information. It can also trigger negative emotions such as anxiety and anger, thus leading to impulsive and seemingly irrational decisions (Chen, 2019; Graff Zivin and Neidell, 2018; Trushna et al., 2020). Recent findings suggest that these psychological impacts have knock-on effects on decision-making. For example, studies have shown that air pollution affects professional investors' willingness to take risks (Heyes et al., 2016; Levy and Yagil, 2011), and leads to more aggressive behavior and higher criminal activity (Bondy et al., 2020; Burkhardt et al., 2019; Herrnstadt et al., 2020). However, despite providing compelling evidence, this literature faces the challenge of external validity. Whereas most findings are based on specific groups and decision-making contexts, little is known about the effect of air pollution on decision-making among the population at large.

In this paper, we study the effect of air pollution on a decision made by millions of people at a time, namely voting in parliamentary elections. An election can be seen as a large-scale real-world laboratory, whereby people decide on the same issue on the same day but in different locations, which means that they are exposed to different levels of air pollution when making their decision. We use county-level data on 64 federal and state elections in Germany between 2000 and 2018, which contains information on election results, turnout, as well as daily measures of air pollution and weather conditions. Our measure of local air pollution is the daily average concentration of PM10, one of the most frequently-used indicators for suspended particles in the air. Voting is a high-stakes decision: which parties and leaders are in government determines politics for several years, and has important consequences for people's lives. This holds for federal as well as state elections, because many important political decisions in Germany are made at the state level.

To capture the effect of pollution on voting through several plausible pathways, we choose the vote share of the incumbent parties as the main outcome. Voting for incumbent parties can be seen as an expression of support for the status quo. Relative to voting for opposition parties, it also represents the less risky option. Therefore, if air pollution affects either people's support for the status quo or their willingness to take risks, it may affect voting for the incumbent parties. *A priori*, this effect could work in either direction: if pollution reduces the willingness to take risks, we would expect greater support for the incumbent parties, while if pollution worsens people's mood and increases anxiety, these emotions may reduce the support for the status quo, resulting in a lower vote shares for the incumbent parties. Which of these effects dominates is an empirical question.

We identify a causal effect by exploiting idiosyncratic variation in air pollution within the same county across election dates, by exploiting deviations from the typical level of pollution in a given county. The identification assumption underlying this strategy is that within a given county, the level of air pollution on election day is independent of the local political situation or any other factors that determine individual voting behavior. While politics can influence air pollution in the long run through environmental policies, it is nearly impossible to affect pollution on election day. We isolate within-county variation by including county and election date fixed effects, which absorb average differences in air pollution and election results across counties as well as trends that are common to all counties. In addition, we control for weather conditions that could simultaneously affect voting decisions and pollution levels. Moreover, to alleviate concerns about potentially endogenous exposure to poor air quality, we use an instrumental variable strategy that leverages daily variation in wind directions (Deryugina et al., 2019).

Our main finding is a negative effect of air pollution on the vote share of the incumbent parties and a corresponding increase in the vote share of established opposition parties. An increase in the ambient concentration of PM10 by ten micrograms per cubic meter ($\mu g/m^3$) — around two within-county standard deviations — reduces the vote share of the incumbent parties by 2 percentage points, increases the vote share of established opposition parties by 2.8 percentage points, and reduces the vote share of other parties by 0.8 percentage points. We find no significant effect on voter turnout.

These effects are strong given observed changes in the incumbent vote share in federal elections. A useful benchmark is the change in the incumbent vote share when Angela Merkel was elected as chancellor in 2005. Relative to the previous election in 2002, the incumbent center-left government lost 4.8 percentage points of its vote share. Compare that to the effect of a higher concentration in air pollution on election day by $5\mu g/m^3$ – one within-county standard deviation, and thus not uncommon – which reduces the vote share of the incumbent by one percentage point. This means that a higher level of air pollution on election day leads to a reduction in the incumbent vote share equivalent to 21% of the observed drop in the incumbent's share in 2005.

To corroborate our identification assumptions and exclude the notion that our results are spurious, we perform balancing tests, placebo tests and pursue an instrumental variable strategy. Balancing tests show that that pollution neither predicts changes in population nor local GDP nor local employment rates. To further rule out that our results are contaminated by omitted variables or time trends, we perform placebo tests based on pollution levels on days before and after the actual election. We find that the effects of air pollution are sizable and statistically significant on days before the election but are very small and mostly insignificant on days immediately following the election day. Significant effects several days before the election are plausible because people may make their voting decisions before the election. These results suggest that our identification strategy uncovers a real effect: they show that pollution only matters on days when it *can* affect decision-making but not on days when it cannot. Likewise, when we run permutation tests and assign each observation the pollution level from different election dates, we consistently obtain estimates close to zero and far away from our estimates based on the correct election dates. This suggests that our estimates pick up a real effect. In a further step, we perform an instrumental variable analysis that exploits daily variation in wind directions (Deryugina et al., 2019). The idea behind this strategy is that wind directions have an effect on daily levels of air pollution, but they are independent of local political or economic factors that could determine voting. The instrumental variable estimates confirm our main results, which suggests that our results are not contaminated by omitted variable bias.

We generalize our results by documenting similar effects in two large-scale representative surveys. The first is a monthly opinion poll — *Politbarometer* — carried out on behalf of the German public television since 1977. On days with higher pollution in a respondent's region, we find that respondents report a weaker intention to vote for the incumbent federal government, and a stronger intention to vote for the opposition. At the same time, the results indicate a weaker approval of the current government's policies, while approval of the opposition is unaffected. A second piece of evidence comes from the German Socio-Economic Panel (SOEP). The panel structure of the SOEP allows us to exploit fluctuations in air pollution across interviews by the same respondent. Again, on interview days with higher air pollution, respondents show weaker identification with the current federal government and stronger identification with the opposition.

In theory, our findings can be explained by two behavioral channels. One is that voters rationally punish the incumbent for a high local level of air pollution. Another is a behavioral bias: voters subconsciously vote for the opposition because air pollution happens to be higher on election day. Given the variation in air pollution that we exploit as well as the nature of particulate matter, we view our results as evidence of a behavioral bias. The difference in pollution levels *between* places may be salient for voters; they likely notice that a large industrial city is more polluted than a rural area. However, we exploit variation *within* the same place, namely we exploit that on the election day the level of air pollution happens to be higher or lower than it would normally be in the same place. Unlike variation in rainfall or temperature, such fluctuations in air pollution are hardly noticeable. Therefore, we consider it unlikely that people deliberately choose to change their voting behavior because they are exposed to high air pollution. A more plausible explanation is that air pollution has an unconscious effect; for example, by affecting a person's emotions or health, which in turn affects how they process information and make decisions. Using the survey data, we find evidence that emotions are an important underlying channel. On days with elevated levels of air pollution, respondents are more likely to be worried, feel angry and sad, and they are less likely to feel happy. By contrast, we find no evidence that air pollution affects people's perceptions of the current state of the economy or their own economic situation.

This paper contributes to three strands of literature. First, it adds new evidence to the literature in behavioral economics on the role of incidental factors in high-stakes decisions. Numerous studies have shown that factors that are unrelated to a given decision influence important decisions, often through subconscious changes in behavior (for reviews, see DellaVigna, 2009; Lerner et al., 2015). An example is judges' decisions in court cases: research has shown that sentencing decisions are influenced by temperatures (Heyes and Saberian, 2019), wins of the local football team (Eren and Mocan, 2018), or whether a decision is made before or after a judge's lunch break (Danziger et al., 2011). Similar influences of incidental factors have been documented in other contexts, such as stock trading (Edmans et al., 2007; Hirshleifer and Shumway, 2003; Kamstra et al., 2003) or students' enrollment decisions (Simonsohn, 2010). By affecting emotions and in turn decision-making, air pollution can be seen as an incidental factor. However, despite many studies documenting an effect of ambient air pollution on health and well-being (e.g. Manisalidis et al., 2020; Zhang et al., 2017), there is limited evidence of its effect on high-stakes decisions. Perhaps the most compelling evidence is provided by studies on specific groups such as stock traders (Heyes et al., 2016; Huang et al., 2020; Levy and Yagil, 2011; Meyer and Pagel, 2017), chess players (Klingen and van Jos N. Ommeren, 2020; Künn et al., 2021), baseball umpires (Archsmith et al., 2018), or criminals (Bondy et al., 2020; Burkhardt et al., 2019; Herrnstadt et al., 2020). These studies document that high air pollution leads to systematic biases in decision-making. However, there is scant evidence of air pollution affecting decision-making in the population at large. Exceptions are Chang et al. (2018), who show that air pollution affects people's health plan choices in a way that is inconsistent with rational choice theory, and Qin et al. (2019), who show that homes in Beijing sell for significantly more on days with high air pollution. Our paper adds important evidence to this literature by documenting significant behavioral effects among millions of people going to the ballot on the election day. We use elections as a real-world laboratory to show that air pollution changes decision-making, and provide evidence that affective emotions are an important mechanism through which this effect operates.

Second, this paper contributes to the growing literature on the economic and social consequences of air pollution. While earlier work has mainly focused on the effect of air pollution on health and the environment, recent evidence shows that its impact unfolds in many domains of life. Even short-run fluctuations in air quality have measurable consequences[1]. Several studies show that poor air quality reduces the productivity of workers in manual and cognitive tasks (Chang et al., 2016, 2019; Graff Zivin and Neidell, 2012; Lichter et al., 2017) and increases the frequency of worker absences (Aragón et al., 2017; Hanna and Oliva, 2015). Moreover, poor air quality has substantial consequences for education by increasing absences (Balakrishnan and Tsaneva, 2021; Chen et al., 2018; Currie et al., 2009) and reducing academic performance (Ebenstein et al., 2016; Heissel et al., 2021; Stafford, 2015). The novelty of our study is documenting an effect of air pollution on *political* outcomes. Our analysis yields a strong effect of air pollution, suggesting that high levels of air pollution may tip the scale in favor of opposition parties. Therefore, pollution may have a substantial impact on people's lives by affecting which parties are in government[2].

Third, this paper contributes to the literature on the determinants of voting. While large parts of voters' choices are determined by political factors such as party programs or the popularity of candidates, there is growing evidence that incidental factors outside the political or economic sphere — often irrelevant for the voting decision itself — can affect voting decisions. A commonly-studied incidental factor is rainfall, which may affect the cost of voting as well as voters' emotions

[1] The studies summarized here consider the impact of *short-run* fluctuations in air pollution. There is a separate literature on the impact of *long-run* exposure to air pollution – often during pregnancy or early childhood – on later-life outcomes. See Graff Zivin and Neidell (2013) for a review.

[2] To our knowledge, one of the few papers on air pollution and politics is Heyes et al. (2019), who use text analysis to show that on days with high air pollution members of parliament give speeches of lower quality.

(Arnold and Freier, 2016; Gomez et al., 2007; Hansford and Gomez, 2010; Meier et al., 2019). Studies have also shown that voters respond to events such as natural disasters (Eriksson, 2016; Healy and Malhotra, 2010), shark attacks (Achen and Bartels, 2004), or wins of the local football team (Healy et al., 2010). Although some of these results are contested (Fowler and Hall, 2018; Fowler and Montagnes, 2015), the overall evidence points to an important impact of incidental factors in voting, because either past events trigger negative emotions or voters deliberately punish the government for these events. Our paper contributes to this literature by identifying air pollution as an important determinant of voting outcomes. On days with high air pollution, voters are systematically less likely to vote for the incumbent parties compared to days with lower pollution. However, the nature of the effect is different from the effect of natural disasters or rainfall, which are salient. Air pollution is not salient, and the fact that we find an effect highlights that incidental factors can affect voting behavior even if voters do not observe them.

4.2 Conceptual Framework: Air Pollution and Voting Behavior

The literature in medicine, epidemiology and psychology highlights several plausible channels through which air pollution can affect decision-making. Broadly speaking, these can be split in two categories, namely conscious and subconscious reactions to air pollution. If people notice that air pollution is high, they may make a conscious decision to change their behavior. For example, they may punish the current government for not doing enough to reduce air pollution, and thus vote for an opposition party. However, even if people do not notice air pollution, it can prompt them to subconsciously change their behavior. In this case, ambient air pollution is an incidental influence in the decision process; it is a transient factor that is unrelated to the decision itself yet indirectly affects the decision (Loewenstein et al., 2003). Examples of incidental influences include environmental factors such as the weather, or emotional cues such as whether one's favorite football team has won a match, or whether a decision occurs on a person's birthday. Although in most contexts these factors are unrelated with the decision, there is ample evidence of people deciding differently on sunny days or days after their football team has won. With respect to air pollution, the literature has highlighted physiological as well as psychological pathways through which it affects decisions (see Chen, 2019, for a comprehensive review).

Given our empirical strategy and the nature of particulate matter, our estimation results are more likely to represent a subconscious reaction to air pollution rather than a deliberate choice. At levels commonly observed in Germany, people cannot see or smell air pollution, but rather only feel it indirectly through symptoms such as cough or irritation of the airways. Only at high levels of air pollution — such as levels observed in parts of China — is air pollution actually observable to humans (Barwick et al., 2019). Moreover, we exploit variation in air pollution within the same county across election dates. This means that any effect that we may find is due to air pollution being higher or lower than its normal level in the same place. Because such fluctuations of PM10 are hardly noticeable for voters, it is unlikely that voters deliberately punish the government simply because air pollution happens to be higher on the election day than it normally is. Daily fluctuations in air pollution can affect people's decision-making through three main channels, namely physiological effects, emotions, and cognitive functioning.

4.2.1 Physiological effects

Air pollution has both immediate and chronic effects on human health, which in turn may affect a person's decision-making. Air pollution may affect several different systems and organs, ranging from minor irritations of the upper respiratory tract to chronic respiratory and heart disease, lung cancer, acute respiratory infections and asthmatic attacks (Kampa and Castanas, 2008). The general consensus from medical studies indicates that the mechanisms of air pollution-induced health effects involve an inflammatory response and oxidative stress in the lungs, the vascular system, the heart tissues and the central nervous system (Lodovici and Bigagli, 2011). These effects are stronger among older people and tend to be stronger for people in worse general health (Bell et al., 2013). In the short run, these health effects can lead to fatigue and lower well-being, which can affect decision-making.

4.2.2 The role of emotions.

Recent studies have explored pathways through which air pollution influences the human brain and affects mental health. Pre-clinical and clinical studies have shown that air pollution induces oxidative stress and increases the occurrences of headaches and depression (Lim et al., 2012; Salvi and Salim, 2019, among others). This can have knock-on effects on people's mental health. It is well documented that exposure to high levels of air pollution has a negative effect on people's mood, reduces people's happiness (Levinson, 2012; Li et al., 2014; Zhang et al., 2017; Zheng et al., 2019) and well-being (Luechinger, 2009), and increases anxiety (Trushna et al., 2020).

In turn, the link between mental health, emotions, and decision-making has been documented in a large body of literature in psychology. A review by Lerner et al. (2015) and a meta-analysis by Angie et al. (2011) cite many examples of incidental factors that lead to systematic biases in decision-making. These effects are mostly non-conscious: an incidental factor like air pollution affects a person's emotions, which changes their judgment, and in turn affects their decision-making.

4.2.3 Cognitive functioning

An additional channel through which air pollution can affect decision-making is cognitive functioning. Exposure to air pollution can cause inflammation and oxidative stress, which may affect the development and operation of brain cells, and in turn affect how people process information and make judgments. Although the literature has not yet reached a consensus on the exact biological mechanisms, there is ample evidence that long-run exposure to high levels of air pollution impairs cognitive functioning (e.g. Weuve et al., 2012; Zhang et al., 2018). In particular, it slows down the cognitive development among young people and accelerates the cognitive decline among older people (Clifford et al., 2016). Studies also show that short-run fluctuations in air pollution can affect cognitive performance. Examples include Künn et al. (2021), who find that chess players make more erroneous moves on days with high air pollution, and Archsmith et al. (2018), who find that baseball umpires make significantly more incorrect calls. Other studies document negative effects in cognitive tests. Powdthavee and Oswald (2020) use a representative survey in England and show that people exposed to higher levels of NO_2 on the day of an interview perform significantly worse on a memory test. Bedi et al. (2021) document similar effects among students undertaking cognitive tests in Brazil: a higher concentration of particulate matter significantly reduces performance on a fluid reasoning test.

4.2.4 Main outcome: incumbent vote share

We choose voting for the incumbent government parties as the main outcome, as it reflects voters' support for the current political status quo and may be indicative of voters' risk preferences. If air pollution increases negative emotions, this may affect voters' willingness to change the status quo, which has implications for the support for the incumbent government. In general, the status quo operates as a reference point from which change is considered and people assign more weight to losses than to equally-sized gains (Kahneman and Tversky, 1979). The higher the loss aversion, the more sizable the status quo bias, increasing the relative support for the status quo (Alesina and Passarelli, 2019; Attanasi et al., 2017). Given that increased unhappiness fosters impatience and induces a desire to change (e.g. Lerner et al., 2013, 2004), this might reduce loss aversion and the status quo bias. In the context of voting decisions, this implies withdrawing support from the incumbent government. Consistent with this reasoning, there is evidence that happier people are more likely to vote for incumbents (Liberini et al., 2017; Ward, 2015), whereas unhappy people are more likely to vote for the opposition (Nowakowski, 2021; Ward et al., 2020).

It has also been shown that anxiety has a direct impact on citizens' political behavior, determining the strategies that they use to construct their political judgments (Marcus et al., 2000, 2007; Valentino et al., 2008). Voters who are anxious are found to reduce their reliance on political habits and heuristics (e.g., party identification) and devote more attention to contemporary information. Accordingly, one could expect anxious voters to acquire more information about candidates' policy stands, rely less on partisanship and more on policy preferences and therefore reduce their support for the status quo.

Negative emotions may also have an indirect effect on voting behavior through their impact on risk attitudes (Bruyneel et al., 2009; Grable and Roszkowski, 2008; Hockey et al., 2000; Kliger and Levy, 2003; Lepori, 2015; Otto and Eichstaedt, 2018, among others). As Shepsle (1972) posits, "*the act of voting, like that of gambling or purchasing insurance, is one involving 'risky' alternatives*". Following this view, a substantial body of literature in political science has analyzed the link between risk aversion and candidate choice, *incumbent advantage* (Kam and Simas, 2012; Morgenstern and Zechmeister, 2003, among others), and policy choices, *status quo bias* (Eckles and Schaffner, 2011; Ehrlich and Maestas, 2010, among others). In particular, more recently Eckles et al. (2014) have shown empirically that citizens who are more risk averse are more likely to support incumbent candidates in US congress elections, while citizens who are more risk tolerant are more likely to vote for challengers. Similarly, Liñeira and Henderson (2019); Morisi (2018); Sanders and Jenkins (2016) show that more risk-averse individuals are more likely to vote for the "status quo" policy in the recent UK "*Leave*" and Scotland "*Independence*" referenda, respectively.

In sum, the literature shows that air pollution has impacts on the human body at various levels. By affecting the brain and increasing oxidative stress, it can negatively affect people's mood and emotions. These feelings in turn affect individuals' decision-making. Overall, we should expect that exposure to air pollution has *some* effect on voting outcomes. However, whether it increases or reduces support for the incumbent government is an empirical question. If air pollution increases risk aversion, we would expect a positive effect. On the other hand, if it mainly affects voters' mood, we would expect a negative effect.

4.3 Main Data and Descriptive Statistics

To study the effect of air pollution on voting outcomes, we focus on parliamentary elections in Germany. The country has regular elections at the federal and state level, and elections at both levels have important consequences for all political domains. For our analysis, we combine county-level data on federal and state elections with data on pollution, weather conditions and socio-economic characteristics. The sample period runs from 2000 — the first year in which pollution measures are available — to 2018, the most recent year in which GDP data is available.

4.3.1 Election data

We use county-level voting data from the German Statistical Office, which covers five federal and 64 state elections from 2000 to 2018. Both election types are administered by municipalities in a uniform procedure. The national parliament (*Bundestag*) is elected for a four-year term, with elections typically taking place on a Sunday in September or October. The state parliaments (*Landtage*) are elected for five years[3]. State elections are typically held on a Sunday in spring or fall. In the sample period, there have been five early elections, one at the federal (2005) and four at the state level (Hamburg in 2004, Nordrhein-Westfalen, Saarland and Schleswig-Holstein in 2012)[4].

In all elections, voters have two votes: the first vote is cast for a direct candidate in a local electoral district, and the second for a state-wide party list. The seats in parliament are distributed to directly elected candidates as well as candidates on the party lists. With some minor exceptions, the proportionality of parties in parliament is governed by the second vote (*Zweitstimme*). Voters do not need to give both votes to the same party. It is allowed — and not uncommon — that people give their first vote to a candidate from one party and the second vote to another party. In our analysis, we focus on the second votes because they are representative of people's party preferences, whereas the first votes are often strategically given to candidates from large parties who have a higher chance of winning. For each election, we observe the date and type, the number of eligible, valid and invalid votes as well as the number of votes for each party.

The vast majority of votes are cast at the polling stations on election day. However, it is possible to vote by mail, and this option has become increasingly popular in recent years. For example, in federal elections the share of mail voters increased from 13.4% in 1994 to 28.6% in 2017 (Bundeswahlleiter, 2017). Our data do not contain separate county-level information on the voting behavior of ballot voters

[3]The exception is Bremen, where the term is four years.

[4]There was also an early election for the *Landtag* in Hessen in 2009. The regular election took place in 2008, but the negotiation for the formation of a government failed and new elections were held in 2009. Since no government came out of the 2008 elections — i.e. no incumbent and opposition — we do not consider the 2009 elections in our analysis. See Appendix C.1.2 for the list of all elections taking place in the 2000-2019 period and their distribution across calendar months.

vs. mail voters. In Appendix C.2, we discuss the implications of mail voting for our estimation.

To test whether air pollution leads to changes in voting decisions, we consider three vote shares as outcomes:

1. *Incumbent parties*: Parties that are part of the governing coalition on the day before the election. The exact classification depends on whether the election is at the federal or state level (the complete breakdown is reported in Appendix C.1.1). At the federal level, we consider incumbent parties that form the federal government, and analogously at the state level we consider parties that form the state goverment.

2. *Established opposition*: We define all parties that are not part of the government on the day before the election as opposition. Furthermore, we split the opposition parties into two categories. One category is the *established opposition*, which includes parties that have been regularly represented in the German Bundestag over the sample period: CDU/CSU, SPD, Greens, FDP, Linke. The exact classification depends on the specific federal or state election, and is reported in Appendix C.1.1.

3. *Other parties*: These are smaller opposition parties, many of which are not frequently represented in the federal or state parliaments. To enter the federal or a state parliament as a faction, a party has to get at least 5% of all valid votes. Parties in this category are those that whose vote shares are typically below 5%. This category also includes the far-right party AfD, which only entered the Bundestag in 2017 but was not regularly represented in parliaments over the sample period.

Besides vote shares, we also consider turnout, which may explain the observed changes in voting patterns. We estimate the effect of pollution on turnout and subsequently use turnout as a control variable when estimating the effect of pollution on vote shares.

Based on the county-level election data, we construct a panel dataset for all counties and elections. In thirteen out of sixteen states, the definition of counties remained stable over the sample period. Three states — namely, Mecklenburg-Vorpommern, Saxony, and Saxony-Anhalt — had territorial reforms between 2007 and 2011, during which some counties were merged or dissolved, meaning that county-level data from before and after the reform are not comparable. To obtain consistent panel data for these three states, we apply the post-reform county definition and construct the vote shares for pre-reform years based on municipality-level voting data[5]. We explain the construction of the pre-reform data in greater detail in Appendix C.3.2, where we also perform robustness checks omitting these three states from the analysis.

4.3.2 Pollution and weather data

The pollution data is provided by the German Federal Environment Agency (*Umweltbundesamt*) and it comprises geo-coded daily average measures of ground-level concentration of several pollutants from 1,170 measuring stations. Our measure for air pollution is the daily average concentration of particles smaller than ten micrometers in ambient air (PM10), one of the most frequently-used indicators for suspended particles in the air (World Health Organization, 2005)[6]. Particulate matter is a broad definition used to characterize a mixture of solid and liquid particles that significantly vary in their size. PM10 includes particles with

[5]The same is applied for demographic and socio-economic characteristics when the original datasets do not already include observations for the post-reform county definitions.

[6]Our analysis is based on measurements of PM10 concentrations in ambient air. We do not use the concentration of PM2.5, which only captures the concentration of very fine particles with a diameter not exceeding 2.5 μm and therefore most harmful for the human body by being able to penetrate very deep in the lungs and brain. Unfortunately, the measurement of PM2.5 in Germany only started in 2008 and with a much lower geographic coverage than PM10, which substantially reduces our sample size. However, the concentration of PM10 also captures fine particles and is consequently strongly correlated with PM2.5.

an aerodynamic diameter smaller than ten micrometers (μm). The World Health Organization (WHO) and the European Environmental Agency (EEA) recommend a 24-hour average concentration of no more than 50 $\mu g/m^3$ (European Environment Agency, 2016). Across Germany, PM10 has been consistently monitored since 2000 and measurements are conducted through gravimetry, which is the standard method in the EU.

In our preferred specification, we control for ozone levels. Unlike PM10, ozone is not directly emitted into the atmosphere but emerges from certain combinations of temperature and solar radiation[7]. Ozone concentrations tend to be particularly high during summer months, whereas particulate matter is lowest in summer. Given that most elections happen in spring or autumn, the level of ozone is negatively correlated with PM10, and thus may confound the estimation of the effect of PM10.

In order to control for weather, we obtained geo-coded weather data from the German Meteriological Service (*Deutscher Wetterdienst*). These comprise various measures of temperature (°C), relative humidity (%), wind (m/s), precipitation (mm/m^2), solar radiation (h), air pressure (hpa) and dew point (°C). As documented by a large body of literature in the natural and social sciences, meteorological conditions affect concentration levels of pollution as well as voting behavior, which is why we include these variables as controls (see, for example Eisinga et al., 2012a,b; Sforza, 2014). All pollution and weather variables are measured as 24-hours averages, apart from precipitation, which is the total amount over 24 hours.

We link the election, pollution and weather data based on the county centroid and the election date. For each county, we calculate the pollution and weather measures

[7]The WHO suggests a maximum daily eight-hour mean concentration of 100 $\mu g/m^3$, while the EEA's target is set at 120 $\mu g/m^3$, not to be exceeded on more than 25 days per year (European Environment Agency, 2016). Measurement is carried out by UV absorption.

at the centroid as the inverse distance-weighted average across all stations within a certain radius. The choice of the radius comes with a trade-off between the accuracy of the measurement within a county and the number of counties that can be included. A smaller radius yields more accurate measures at each centroid but some centroids would not be sufficiently close to any measuring station and therefore cannot be included in the dataset. In our main analysis, we choose a radius of 30km.

4.3.3 Demographic and economic data

We also collect data to control for demographic and economic characteristics that could simultaneously affect the concentration of PM10 as well as voter preferences. The data are provided by the German Statistics Office (*Statistisches Bundesamt*) and include county-level observations of population by gender and age group, gross domestic product and gross value-added by economic sector as well as employment by sector for the 2000–2018 period. In our preferred specification, we control for total population, GDP per capita, and the employment rate as the ratio of the total number of employed persons over the population aged between 15 and 65 years. Note that this ratio may exceed 100% for counties characterized by a high share of inbound commuters.

4.3.4 Estimation sample and descriptive statistics

We restrict our estimation sample to all county-election observations for which PM10, voting, weather, demographic and economic data are available. For our preferred data linkage based on a radius of 30km, this leaves us with 2770 observations (356 counties and a total number of 60 elections)[8].

Table 4.1 reports the descriptive statistics of the main variables in our analysis.

[8] For some counties, we do not have observations for one or more of the pollution or weather variables for the entire sample period. In addition, most state elections in the county-states of Berlin and Hamburg are singletons, and they are thus dropped from the fixed effect estimation.

The within-standard deviation is the standard deviation of the residuals after conditioning on election and county fixed effects. The table reveals a strong degree of variation in the voting and pollution data. The variation in the number of eligible voters and valid votes reflects the fact that the population per county strongly differs between rural and urban areas. Large cities such as Berlin, Hamburg, Munich or Cologne coincide with counties, whereas other counties comprise the surroundings of large cities or rural areas. The average turnout is 69%, with little variation. The statistics for incumbent parties show that their average share in any election is higher than that of the established opposition, although the vote share considerably varies from 17% to 79%. The average vote share of the established opposition is 42%, and it varies to a similar degree as the vote share of incumbent parties.

Table 4.1: Descriptive Statistics

	Mean	SD(total)	SD(within)	min	max	N
Voting data						
Eligible voters	159,376	159,294	5,247	26,396	2,505,718	2,770
Valid votes	109,548	116,304	16,955	13,132	1,872,133	2,770
Turnout	0.69	0.09	0.02	0.38	0.87	2,770
Share incumbent parties	0.48	0.10	0.07	0.17	0.79	2,770
Share established opposition parties	0.42	0.12	0.07	0.13	0.82	2,770
Share other parties	0.10	0.07	0.02	0.01	0.44	2,770
Pollution data						
PM10 ($10\mu g/m^3$)	1.90	0.85	0.47	0.26	6.79	2,770
Ozone ($10\mu g/m^3$)	4.20	1.54	0.81	1.36	16.21	2,770
Weather data						
Temperature (°C)	11.22	4.01	0.83	-7.60	21.12	2,770
Relative humidity (%)	80.02	9.12	4.45	47.40	99.58	2,770
Wind speed (m/s)	2.72	1.63	0.84	0.10	11.87	2,770
Precipitation (mm)	1.34	3.18	2.14	0.00	34.80	2,770
Demographic and economic data						
Population	214,510	228,459	10,663	34,084	3,613,495	2,770
GDP per capita	31,128	14,902	3,417	12,481	172,437	2,770
Employment rate	0.76	0.22	0.03	0.37	1.97	2,770

Notes: This table displays the descriptive statistics for the estimation sample. SD(within) represents the standard deviation of the residuals after removing election and municipality fixed effects. Pollution and weather measurements are computed based on a radius of 30 km. The employment rate is based on yearly average number of employed persons in a given county divided by its total population .

4.3.5 Variation in air pollution levels and incumbent vote shares

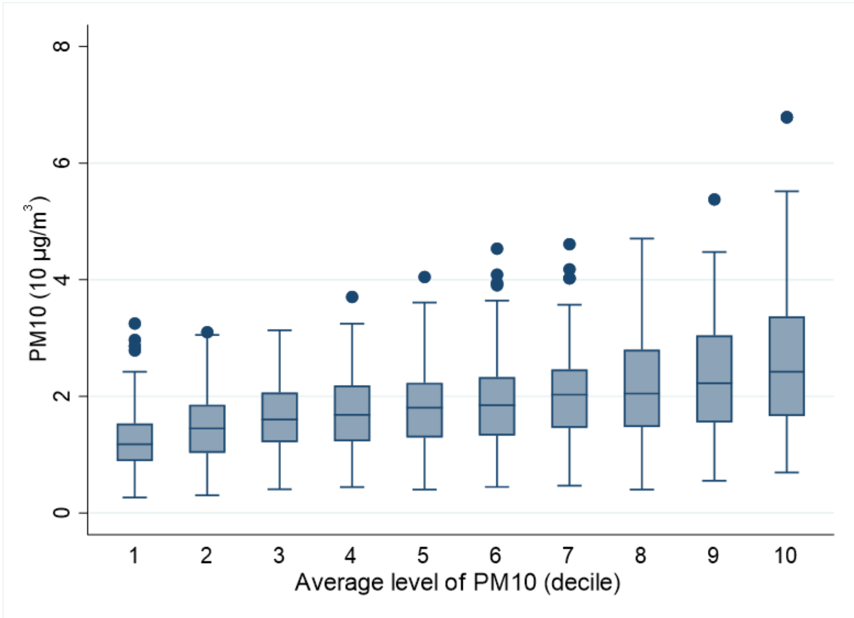
The descriptive statistics on the ambient concentration of PM10 in Table 4.1 show that the mean on election days (always on a Sunday) is $19\mu g/m^3$ and also indicate a strong variation in pollution. The within-standard deviation — which is close to our identifying variation — accounts for more than 50% of the total variation in PM10 and around 25% of its mean. Day to day variation in ambient concentrations of particulate matter in a specific location results from a combination of emissions from various sources and local atmospheric conditions[9]. This means that short-term variation in local pollution does not simply reflect variation in the amount of air pollutants emitted to ambient air (e.g., from industrial activity or traffic volumes) but also crucially depends on fluctuations in local weather conditions. For a given level of emissions, exposure to pollution is significantly lower on rainy or windy days as precipitation and wind reduce ambient concentrations of particles. Also, during temperature inversion episodes warmer air at higher altitude traps air pollutants emitted at the ground (Jans et al., 2018). In addition, depending on the direction from where the wind blows, particles emitted at other locations may be transported over long distances and increase air pollution independently of local economic activities (Deryugina et al., 2019).

Figure 4.1 shows the variation in PM10 across election dates for different levels of average air pollution in a county. The graph illustrates a significant degree of variation in all deciles — regardless of the average level of pollution, there is a large amount of variation in pollution levels *within* each decile. Despite the substantial variation in daily average concentrations of PM10, the levels we observe in Germany over the period under investigation do not exceed $70\mu g/m^3$. These levels are low in global comparison and substantially lower than in highly polluted industrial cities in India or China. In particular, the levels observed in

[9]While particulate pollution may have natural sources (e.g. wildfires, sandstorms or volcano eruptions), there are various man-made emission sources such as automobile exhaust, electricity generation or any other industrial activity involving combustion processes.

Germany are substantially lower than levels that inhibit visibility, such that it is essentially impossible for voters to visually observe pollution levels they are exposed to[10]. This means that, different from what has been shown for weather conditions such as rainfall, sunshine, temperature or wind speed (Gomez et al., 2007), pollution levels are not at all salient and we can thus confidently exclude the notion that there may be a selection of specific types of voters going to the polls on more polluted days compared to low pollution days.

Figure 4.1: Variation of PM10 by Decile of the Average Level of PM10



Notes: This graph displays boxplot charts visualizing the variation of particulate matter (PM10) on the day of election across different counties grouped by their average level of PM10.

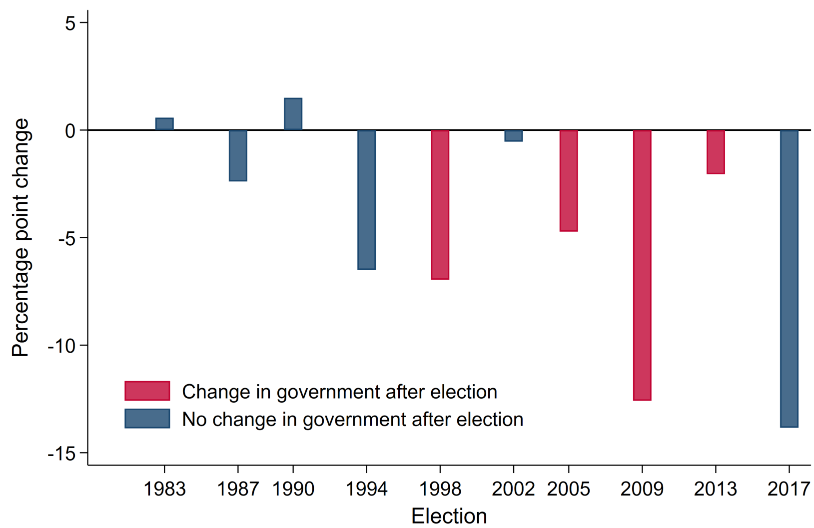
In Appendix C.3.1, we present further illustrations of the variation in PM10 within and across counties. The results show that the average level of PM10 fluctuates considerably from year to year, although the process appears to be mean-reverting, i.e. there is no noticeable time trend in the level of PM10. We also document the extent of within-county variation in PM10, which represents our identifying variation. Counties differ in their within-variation, but in most counties the within-

[10]Research in atmospheric sciences has documented a non-linear relationship between visibility and particulate pollution which may differ by the level of relative humidity. Sun et al. (2020) report that the negative relationship between visibility and PM2.5 concentrations becomes particularly strong beyond a threshold of $76\mu g/m^3$ and $49\mu g/m^3$ for relative humidity levels from 60% to 80% and from 80% to 90% respectively. Over the period from 2010 to 2020 the mean PM2.5 concentration was in the of 10–$20\mu g/m^3$, see `www.umweltbundesamt.de/publikationen/luftqualitaet-2020` (page 11, last accessed: 19 Oct, 2021).

standard deviation across election dates lies between 5 and $10\mu g/m^3$. Overall the descriptive analysis shows that our estimation strategy relies on a large amount of identifying variation that is spread across the country.

Figure 4.2 illustrates the typical fluctuations in voting for the incumbent parties in federal elections since the 1980s. Each bar indicates the change in the vote share of the incumbent government in a given election relative to the previous election. In most elections, the incumbent government performed considerably worse than in the previous election. In some years, these losses led to a change in government. An example is the change from Chancellor Gerhard Schröder's center-left (SPD/Greens) to Angela Merkel's Grand Coalition (CDU/CSU/SPD) government in 2005, following a drop in the vote share of the incumbent parties by close to five percentage points.

Figure 4.2: Changes in Incumbent Vote Shares in Federal Elections



Notes: This graph displays the change in the vote share of the incumbent parties in federal elections. The change is measured for the incumbent parties on the day of the election relative to the vote share of the same parties in the previous election. Red bars indicate a change in the government coalition. In 1998 and 2005, the change in the government coalition coincided with a change of the chancellor. Source: own calculations based on data from the German Statistical Office.

4.4 Empirical Strategy

Our goal is to study the effect of poor air quality on the election day on voting outcomes. In an ideal experiment, we would randomly assign air pollution levels on the election day to local areas. In this case, we could interpret the difference in voting results between areas with a high and low concentration of PM10 as a causal effect. In the absence of such an experiment, we exploit quasi-experimental variation in pollution levels within counties over time. The underlying thought experiment is that on a given election day — for random reasons — the level of air pollution in a county is higher or lower compared to its normal level. This strategy allows us to identify a causal effect under the maintained assumption that the variation in pollution across election dates within a county can be considered as good as random. In this section, we explain our identification strategy under this assumption and point to some potential challenges. We will address these challenges through placebo and robustness checks as well as an instrumental variable strategy after having presented the main results.

4.4.1 Regression model

To exploit variation in the concentration of PM10 within counties across election dates, we estimate a two-way fixed effect regression,

$$y_{it} = \alpha + \beta PM10_{it} + \boldsymbol{X'_{it}}\gamma + \delta_i + \tau_t + \varepsilon_{it}, \tag{4.1}$$

whereby the outcome variable y_{it} denotes an election outcome in county i at election date t. The regressor of interest is $PM10_{it}$, the air concentration of PM10 (in tens of $\mu g/m^3$) measured on the day of the election. The vector X_{it} controls for two types of time-varying confounders, namely weather (temperature, relative humidity, wind speed, precipitation and ozone levels) and demographic variables (total population, GDP per capita and the employment rate). The county and election date fixed effects, δ_i and τ_t, absorb all confounding factors that are constant within a county over time as well as those that are common to all counties during the same election.

The error term ε_{it} summarizes all determinants of election outcomes that are not absorbed by the controls and fixed effects. To account for serial correlation within counties, we cluster the standard errors at the county level.

4.4.2 Identification

Our parameter of interest, β, measures the contemporaneous effect of a change in air pollution on election outcomes within the same county. A causal interpretation of β requires that pollution be uncorrelated with any determinants of election outcomes conditional on controls and fixed effects, i.e.

$$E(\varepsilon_{it}|PM10_{it}, \boldsymbol{X_{it}}, \delta_i, \tau_t) = 0. \tag{4.2}$$

Given the two-way fixed effects, β is causally identified if the fluctuation in pollution levels *within a given county* is uncorrelated with time-varying determinants of voting in the same county. Our controls account for several common challenges to identification. Weather conditions may affect who votes as well as for what party — for example, by affecting turnout or people's mood on election day — while being potentially correlated with air pollution. To address this challenge, we control for a set of potential confounders, namely temperature, relative humidity, wind speed and precipitation. When the outcome is a vote share, we also control for voter turnout. This helps with the interpretation of the effect: conditional on turnout, β represents the effect on pollution on vote shares rather than the number of votes.

We address several identification challenges in robustness checks. One challenge could be local economic shocks or public policies that may simultaneously affect pollution and voting. For example, the closure of a nearby factory or changes in local environmental regulations may reduce pollution levels while leading to a response among voters. We address this challenge in three ways. First, we perform balancing tests whereby we regress economic outcomes on pollution and condition on fixed effects and controls. Insignificant coefficients suggest that the fluctuations in pollution used for identification are unrelated with fluctuations in economic variables. Second, we show regressions with placebo election dates before and after the actual date. The idea is that profound local shocks or policy changes should affect pollution levels both before and after the actual election. However, if we do not observe significant effects after the election date, this indicates that our results are not confounded by local shocks. As a second robustness check, we follow Deryugina et al. (2019) and instrument for air pollution with changes in local wind directions, which are plausibly orthogonal to local economic shocks.

A further challenge is voting by mail, whose share among all eligible votes in the sample period stands between 13% and 28% (Bundeswahlleiter, 2017). Because we neither observe the time at which mail voters send their ballot papers nor the place in which they cast their vote, we likely assign the incorrect level of air pollution to mail voters. We assign the concentration of particulate matter on the election day despite the fact that they have cast their vote up to one month before the election, and potentially in a different place. In Appendix C.2, we show that the absence of detailed data on mail voting is akin to a measurement error problem, which — under reasonable assumptions — leads to attenuation bias. A back-of-the-envelope calculation shows that the estimates are attenuated by around 15–20%.

4.5 Results

4.5.1 Main results

Table 4.2 displays our estimates for the effect of air pollution on voting. Each coefficient is the result of a separate regression of the outcome listed on top on the concentration of PM10 as well as the controls and fixed effects listed at the bottom. PM10 is measured in $10\mu g/m^3$, whereas the outcomes are shares that are bounded between zero and one. A coefficient of 0.01 means that an increase in PM10 by $10\mu g/m^3$ increases the respective outcome by one percentage point. An increase in PM10 by $10\mu g/m^3$ in turn is equivalent to an increase by two within-county standard deviations in the concentration of PM10[11].

In Panel A, we condition on county and election date fixed effects, whereas in Panel B we additionally control for weather variables, demographics, and election type fixed effects, which absorb systematic differences between county-, state- and federal-level elections. Since pollution may also affect turnout, we control for turnout in Columns (1)-(3) of Panel B. Overall, the results in Panels A and B are very similar, suggesting that the fluctuation in air pollution across election days is uncorrelated with more salient fluctuations in the weather, or changes in demographics.

In Columns (1)-(3), we find strong and statistically significant effects of pollution on voting outcomes. The results from both panels indicate that an increase in pollution shifts votes away from the incumbent parties and towards established opposition parties. In Panel B, for an increase in PM10 by $10\mu g/m^3$, the vote share of the incumbent parties drops by roughly two percentage points, which is more than 4% of the mean. By contrast, the vote share of the established opposition increases by almost three percentage points, which is equivalent to 6.9% of the

[11] As shown in Table 4.1, the within-standard deviation of PM10 measured in $10\mu g/m^3$ is $sd = 0.47$.

mean. We also find a negative effect on the vote share of smaller parties. An increase in the PM10 concentration by $10\mu g/m^3$ reduces the vote share of smaller parties by 0.8 percentage points, which is equivalent to 8% of the mean. In Column (4), we find no effect of pollution on voter turnout. The coefficients from both panels are small in magnitude and insignificant, which suggests that air pollution does not affect people's decision on whether to vote or not. This result is important for the interpretation of the effects on voting outcomes, which reflect changes in voting behavior rather than changes in turnout.

In Appendices C.3.2 and C.3.3 we re-estimate the models presented in Table 4.2 removing those states that experienced territorial reforms during the estimation period and including states fixed effects, respectively. The results remain quantitatively and qualitatively identical.

Non-linear dose-response relationship

In Table 4.3, we explore the dose-response relationship between air pollution and voting outcomes. For this purpose, we replace the regressor $PM10_{it}$ in Equation (4.1) with indicators for four levels of PM10 (0–$10\mu g/m^3$, 10–$20\mu g/m^3$, 20–$30\mu g/m^3$, $>30\mu g/m^3$). The coefficients of these indicators are to be interpreted as the difference in voting results between a given level of pollution and the base level of less than $10\mu g/m^3$. The results indicate that the effect of pollution on most outcomes is approximately linear. For example, consider the effect on the vote share of the incumbent parties in Column (2). In Panel B, the difference between pollution at the base level (0–$10\mu g/m^3$) and the next higher level (10–$20\mu g/m^3$) is around -0.02, between the second and third level (10–$20\mu g/m^3$ vs 20–$30\mu g/m^3$) it is around -0.017, and between the third and fourth level (20–$30\mu g/m^3$ vs $>30\mu g/m^3$) it is around -0.018. In addition, we fail to detect any significant impact of air pollution on turnout, even at the highest concentration level.

How strong are these effects?

While the magnitude of our estimates does not imply landslide shifts in election results, it shows that pollution plays a role in affecting voting behavior. To understand the magnitude of the effect, consider first an increase in the concentration of PM10 by one within-standard deviation, which is equivalent to an increase in the concentration of PM10 in the same county by around $5\mu g/m^3$ relative to its normal level. Our estimates suggests that an increase in the level of PM10 by $5\mu g/m^3$ reduces the vote share of the incumbent government by one percentage point. Now compare this one-percent decrease in voting for the incumbent to the overall drop in votes for the incumbent in a federal election. For example, in 2005, when Angela Merkel came to power, the incumbent government's vote share had dropped by 4.8 percentage points (see Figure 4.2). An increase in pollution levels by one within-county standard deviation – i.e. an increase that frequently happens – leads to a drop in the incumbent vote share that was 20% of the overall decrease in 2005. By no means do we claim that air pollution brought Merkel into power, but this exercise shows that the effects of air pollution are socially significant.

4.5.2 Robustness checks and discussion

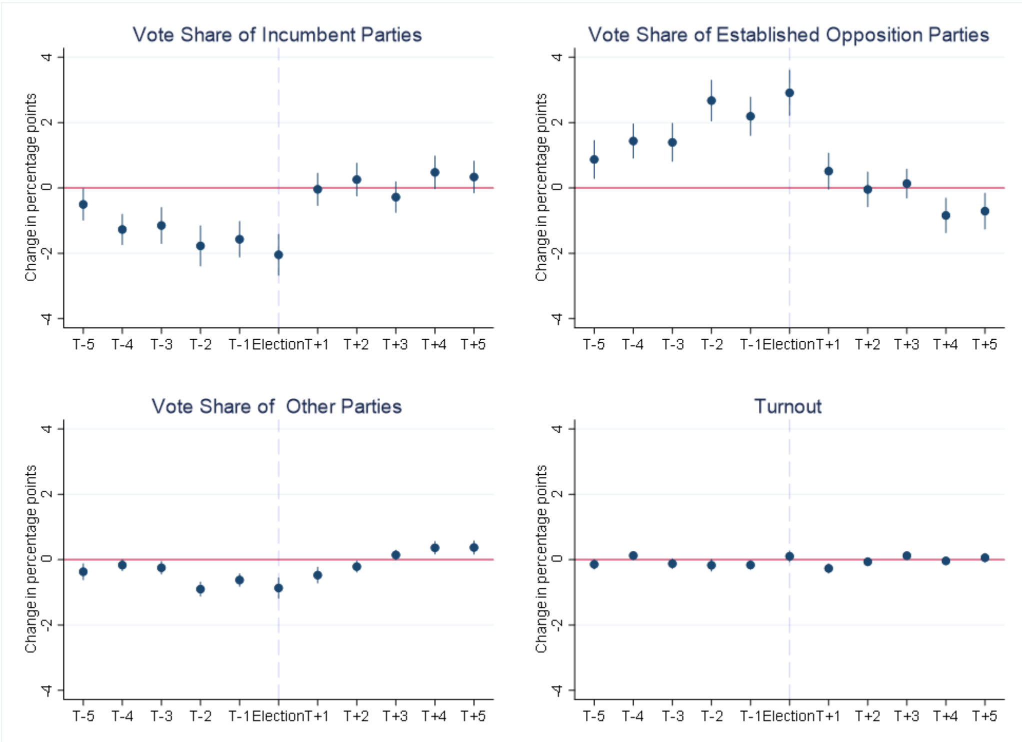
Estimation with placebo election dates

To assess the plausibility of our identification strategy, we run regressions with placebo election dates. We re-estimate our main specification from Panel B in Table 4.2 but construct the regressor and controls based on the measurements on different days. For example, instead of using the concentration of PM10 on the election day, we use PM10 on a day before or after the election. Ideally, we should not find significant effects of PM10 *after* the election, as pollution in the future cannot affect voting outcomes today. By contrast, significant effects before the election day are possible, as voters may make their voting decision several days in advance and/or cast their vote by mail.

The results in Figure 4.3 corroborate our identification strategy. Each displayed coefficient is the result of a separate regression of the outcome listed at the top on the concentration of PM10 on a given day as well as controls and fixed effects. In the run-up to the election, we see results that are significantly different from zero and have the same sign as the estimated effect based on pollution on the election day. We find significant negative effects on the vote share of incumbent parties and significant positive effects of pollution on the vote share of established opposition parties, whereas we find no effect on turnout and small effects on the vote share of other parties. Reassuringly, we only find very small and mostly insignificant effects of air pollution on days immediately *after* the election. The pattern of the estimates before the election – which become larger the closer to the election date – is consistent with the literature on political campaigns, which shows that events closer to an election have a stronger effect than events further in the past (Gerber et al., 2011).

We view the results in Figure 4.3 as evidence in favor of the identification assumption. The results suggest that our results represent real effects rather than a noise pattern that emerges by chance. If the main result was the result of noise — a false positive — a pattern like the one in Figure 4.3 would be unlikely to emerge. Instead, we would expect to see similar estimates before and after the election, or estimates that significantly fluctuate. The placebo tests also suggest that our results are not contaminated by omitted variable bias. An omitted variable – for example, the closure of a local factory – would affect pollution and voting in the same way regardless of whether pollution is measured before or after the election. The same holds for diverging regional trends. If the results were driven by diverging trends in pollution and voting, we would expect to see similar estimates before and after the election. The fact that we see insignificant results immediately after the election suggests that the estimates are not confounded by omitted variables or time trends.

Figure 4.3: Effect of Pollution on Voting with Placebo Election Dates



Notes: This graph displays the point estimates and 95% confidence intervals for OLS regressions of the outcomes listed at the top on the air concentration of PM10 (in $10\mu g/m^3$) on a given day. The lead terms $(T-1, T-2, \ldots)$ refer to days before the election, the lag terms $(T+1, T+2, \ldots)$ to days after. For example, in $T-2$, the regressor is the air concentration of PM10 two days before the election. In all regressions, we condition on the election date, county and election type fixed effects and control for ozone, weather, demographic and economic variables measured on the same day as those used in the main analysis in Table 4.2. Standard errors clustered at the county level.

Balancing and permutation tests

In Appendices C.3.4 and C.3.5, we perform balancing and permutation tests that corroborate our identification strategy. One concern is that fluctuations in pollution are systematically related to fluctuations in economic variables. To address this concern, we regress three economic variables — population, GDP per capita, and the employment rate — on the level of PM10 on election day, conditioning on two-way fixed effects and weather controls. We do not find any significant relationship between the level of PM10 and the economic variables. We view this result as one piece of evidence in favour of our identification assumption.

Another concern is that our estimates are the result of fitting noise rather than extracting a signal. In Appendix C.3.5, we address this concern through permutation tests. Within each county, we randomly re-shuffle the level of pollution across election dates and otherwise run the same regressions as in Table 4.2. In none of 500 permutations do we find a placebo estimate that is more extreme — that is, larger in absolute value — than our estimates based on the true level of pollution. With all three main outcomes, our estimates are far away from the distribution of placebo estimates. We view this finding as evidence that our results pick up signal rather than noise.

IV estimation exploiting exogenous variation in wind directions

In addition to the placebo tests, we also perform an instrumental variable estimation that leverages plausibly exogenous variation in wind directions on the day of the election. An instrument that is both valid and sufficiently strong can help us to overcome two potential challenges, namely omitted variable bias and measurement error. Omitted variable bias could be present when local shocks such as weather shocks or economic shocks simultaneously affect air pollution and voting. Although our set of controls comprises many potential confounders, we cannot be certain that it comprises all of them.

A second — and perhaps more important — challenge is measurement error in the regressor PM10. We match county-level voting results to local levels of pollution via the county centroid. The matching introduces two types of measurement error. One stems from the geographic interpolation of PM10, as the location of the county centroid and the measuring stations rarely coincide. Instead, we interpolate the measure of PM10 at the centroid as a weighted average of measures taken within a 30km radius. A second type of measurement error is present because pollution exposure is likely not uniform within a county. The pollution level where most voters live may be different from the pollution level at the centroid.

To overcome both challenges, we follow the work of Deryugina et al. (2019) and employ an instrumental variable strategy that leverages exogenous variation in wind directions. Pollution particles can be transported by winds, so there is a strong correlation between the concentration of air pollution at a certain location and the direction in which the wind blows. The idea behind this instrument is that the wind direction affects the level of air pollution in a given location due to its physical and economic geography. For example, a county to the west of an industrial center has higher air pollution levels on days with east wind than on days with west wind. As Deryugina et al. (2019) show, wind direction is a strong predictor of pollution levels, and the variation thus generated comes primarily from non-local transport (i.e. from outside-county sources) which should have a uniform effect on the entire county, thus alleviating the measurement error. At the same time, the wind direction on a particular day can be considered orthogonal to economic, social or environmental conditions and shocks, and should therefore have no impact on voting decisions other than its effect on pollution levels, which makes it a suitable candidate for a valid instrument.

The first-stage regression predicts the level of PM10 in a given county on a given day based on the wind direction in the county's state on the same day,

$$PM10_{it} = \sum_{s \in S} \sum_{b=0}^{2} \theta_b \times \mathbb{1}[i \in s] \times WINDDIR_{it}^{90b} + \boldsymbol{X}'_{it}\boldsymbol{\lambda} + \kappa_i + \kappa_t + \eta_{it}. \quad (4.3)$$

We choose variation at the state level as opposed to county level to reduce the computational burden of the IV estimation and increase statistical power. Equation (4.3) comprises 36 instruments: for each of twelve states s, we have three indicators for wind directions b[12]. Each instrument is the interaction of a state indicator — one if county i lies within state s — and indicators for wind direction which equal one if the wind direction in county i on day t lies within one of the intervals [0,90) degrees, [90,180) degrees or [180, 270) degrees. Wind directions falling into the interval [270,360) degrees are the reference category. Data on daily variation

[12]For this purpose, we merged the three city-states of Berlin, Hamburg and Bremen as well as the small state of Saarland with their larger neighboring states — Berlin with Brandenburg, Hamburg and Bremen with Lower Saxony, and Saarland with Rhineland-Palatinate — yielding twelve states in total.

in wind directions comes from the German Meteriological Service (*Deutscher Wetterdienst*).

The IV estimates are shown in Table 4.4. The first stage is sufficiently strong, with a Kleibergen-Paap F-statistic between 11.53 and 12.75. The point estimates are considerably larger in absolute value compared to the main results in Table 4.2; in Columns (1) and (2) of Panel A, they are more than three times the size of their OLS counterparts. In Panel B, when we control for weather, ozone levels and demographics, the coefficients become slightly smaller but the overall pattern remains. Unlike with the OLS results, the effect on turnout is now statistically significant, although the effect size is small relative to the mean turnout of 69%.

We view the difference between the OLS and IV estimates – as well as the difference between IV estimates with and without controls – primarily as evidence of bias from measurement error. The results in Panel A are consistent with attenuation bias, similar to what Deryugina et al. (2019) find in their study on air pollution and mortality in the US. Although we cannot formally prove this, the reduction in the size of the coefficients in Panel B may also be the result of measurement error. If the measurement error of the weather variables is correlated with the measurement error of PM10, the overall amount of measurement error in PM10 is reduced once we include the weather variables as controls. In turn, a smaller measurement error in PM10 means that the difference between OLS and IV is smaller.

Table 4.2: Main Results — Air Pollution and Voting

Outcome:	Vote Share of Incumbent Parties (1)	Vote Share of Established Opposition Parties (2)	Vote Share of Other Parties (3)	Turnout (4)
A. Without controls				
PM10 ($10\mu g/m^3$)	-0.0198***	0.0278***	-0.0080***	0.0012
	(0.003)	(0.004)	(0.002)	(0.001)
Mean dep. var.	0.48	0.42	0.10	0.69
R^2	0.576	0.685	0.893	0.961
N	2770	2770	2770	2770
Controls				
County FE	✓	✓	✓	✓
El. Date FE	✓	✓	✓	✓
B. With controls				
PM10 ($10\mu g/m^3$)	-0.0205***	0.0291***	-0.0087***	0.0010
	(0.003)	(0.004)	(0.002)	(0.001)
Mean dep. var.	0.48	0.42	0.10	0.69
R^2	0.604	0.705	0.902	0.961
N	2770	2770	2770	2770
Controls				
County FE	✓	✓	✓	✓
El. Date FE	✓	✓	✓	✓
Weather	✓	✓	✓	✓
Ozone	✓	✓	✓	✓
Demographics	✓	✓	✓	✓
El. Type FE	✓	✓	✓	✓
Turnout	✓	✓	✓	

Notes: This table displays the results of OLS regressions of the outcomes listed at the top on the air concentration of PM10 (in $10\mu g/m^3$) and the controls listed at the bottom of each panel. Standard errors clustered at the county level are displayed in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4.3: Dose-Response Relationship: Air Pollution and Voting Results

Outcome:	Vote Share of Incumbent Parties (1)	Vote Share of Established Opposition Parties (2)	Vote Share of Other Parties (3)	Turnout (4)
A. Without controls				
PM10 10-20 $\mu g/m^3$	-0.0154***	0.0209***	-0.0055***	0.0012
	(0.005)	(0.006)	(0.002)	(0.002)
PM10 20-30 $\mu g/m^3$	-0.0343***	0.0478***	-0.0135***	0.0022
	(0.007)	(0.007)	(0.003)	(0.002)
PM10 $>$ 30 $\mu g/m^3$	-0.0509***	0.0721***	-0.0212***	0.0016
	(0.009)	(0.009)	(0.004)	(0.002)
Mean dep. var.	0.48	0.42	0.10	0.69
R^2	0.575	0.684	0.893	0.961
N	2770	2770	2770	2770
Controls				
County FE	✓	✓	✓	✓
El. Date FE	✓	✓	✓	✓
B. With controls				
PM10 10-20 $\mu g/m^3$	-0.0198***	0.0266***	-0.0068***	0.0001
	(0.006)	(0.006)	(0.002)	(0.002)
PM10 20-30 $\mu g/m^3$	-0.0365***	0.0503***	-0.0138***	0.0012
	(0.007)	(0.007)	(0.003)	(0.002)
PM10 $>$ 30 $\mu g/m^3$	-0.0546***	0.0768***	-0.0222***	0.0011
	(0.009)	(0.009)	(0.003)	(0.002)
Mean dep. var.	0.48	0.42	0.10	0.69
R^2	0.603	0.704	0.901	0.961
N	2770	2770	2770	2770
Controls				
County FE	✓	✓	✓	✓
El. Date FE	✓	✓	✓	✓
Weather	✓	✓	✓	✓
Ozone	✓	✓	✓	✓
Demographics	✓	✓	✓	✓
El. Type FE	✓	✓	✓	✓
Turnout	✓	✓	✓	

Notes: This table displays the results of OLS regressions of the outcomes listed at the top on three binary indicators for different levels of PM10 (10-20$\mu g/m^3$, 20-30$\mu g/m^3$, >30$\mu g/m^3$). A PM10 concentration smaller than or equal to 10$\mu g/m^3$ is used as the reference category. Standard errors clustered at the county level are displayed in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4.4: Air Pollution and Voting: Instrumental Variable Estimates

Outcome:	Vote Share of Incumbent Parties (1)	Vote Share of Established Opposition Parties (2)	Vote Share of Other Parties (3)	Turnout (4)
A. Without controls				
PM10 ($10\mu g/m^3$)	-0.0777***	0.0862***	-0.0085	0.0262***
	(0.011)	(0.012)	(0.005)	(0.005)
Mean dep. var.	0.48	0.42	0.10	0.69
R^2	-0.146	-0.115	0.024	-0.377
N	2770	2770	2770	2770
Cragg-Donald F	13.01	13.01	13.01	13.01
Kleibergen-Paap F	12.75	12.75	12.75	12.75
Kleibergen-Paap LM	76.59	76.59	76.59	76.59
Controls				
County FE	✓	✓	✓	✓
El. Date FE	✓	✓	✓	✓
B. With Controls				
PM10 ($10\mu g/m^3$)	-0.0575***	0.0775***	-0.0201***	0.0161***
	(0.011)	(0.012)	(0.006)	(0.004)
Mean dep. var.	0.48	0.42	0.10	0.69
R^2	0.021	-0.002	0.058	-0.111
N	2770	2770	2770	2770
Cragg-Donald F	13.17	13.17	13.17	13.32
Kleibergen-Paap F	11.53	11.53	11.53	11.85
Kleibergen-Paap LM	74.45	74.45	74.45	74.33
Controls				
County FE	✓	✓	✓	✓
El. Date FE	✓	✓	✓	✓
Weather	✓	✓	✓	✓
Ozone	✓	✓	✓	✓
Demographics	✓	✓	✓	✓
El. Type FE	✓	✓	✓	✓
Turnout	✓	✓	✓	

Notes: This table displays the results of second-stage regressions where PM10 (in $10\mu g/m^3$) has been instrumented with wind directions following Deryugina et al. (2019). Standard errors clustered at the county level are displayed in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

4.6 Mechanisms and Additional Results

In the previous section, we have documented that ambient air pollution reduces the electoral support for the incumbent coalition parties in favor of the established opposition. In this section, we use data from surveys and opinion polls to generalize our findings and shed light on the underlying mechanisms. The main mechanisms highlighted in the literature are impaired cognitive functioning and mood effects, which may operate through greater anxiety and lower levels of happiness. With the data at hand, we can shed light on the importance of mood effects by looking at measures of affective well-being. In addition, we use data on perceptions of the economy as well as political interest to investigate whether the effect of pollution on voting represents a conscious or unconscious choice. For example, if higher pollution leads to a stronger interest in politics or changes perceptions about the economy, this may be seen as evidence for a conscious choice: pollution changes how people see the world, which also changes how they vote. The absence of such effects would suggest that the effect is more likely to operate through unconscious choices. Pollution may affect people's emotions, which in turn changes how they make decisions, although the observed change in decisions is not a conscious choice.

4.6.1 Data

Monthly opinion poll data

We use data from the *Politbarometer*, a monthly opinion poll that has been run and presented by a national TV station (*Zweites Deutsches Fernsehen*, ZDF) since the 1970s. The poll focuses on opinions and attitudes of the electorate in Germany. In addition to surveying opinions on current political topics and individual politicians, the questionnaire comprises a number of questions that have been surveyed over a long period of time, including the assessment of the current (federal) government and opposition. We use the *Politbarometer* microdata from Forschungsgruppe Wahlen (2020) over the period from 2003 to 2019, which are repeated cross-sections of 1,500–2,000 respondents per month. The data only indicate the week of the interview and the respondent's state of residence, and is thus, less precise than the election data in terms of location and time. We merge average levels of pollution by state and week to the opinion poll and use a binary indicator for state-by-week PM10 concentrations exceeding $20\mu g/m^3$ (roughly the median) as a proxy for elevated pollution exposure.

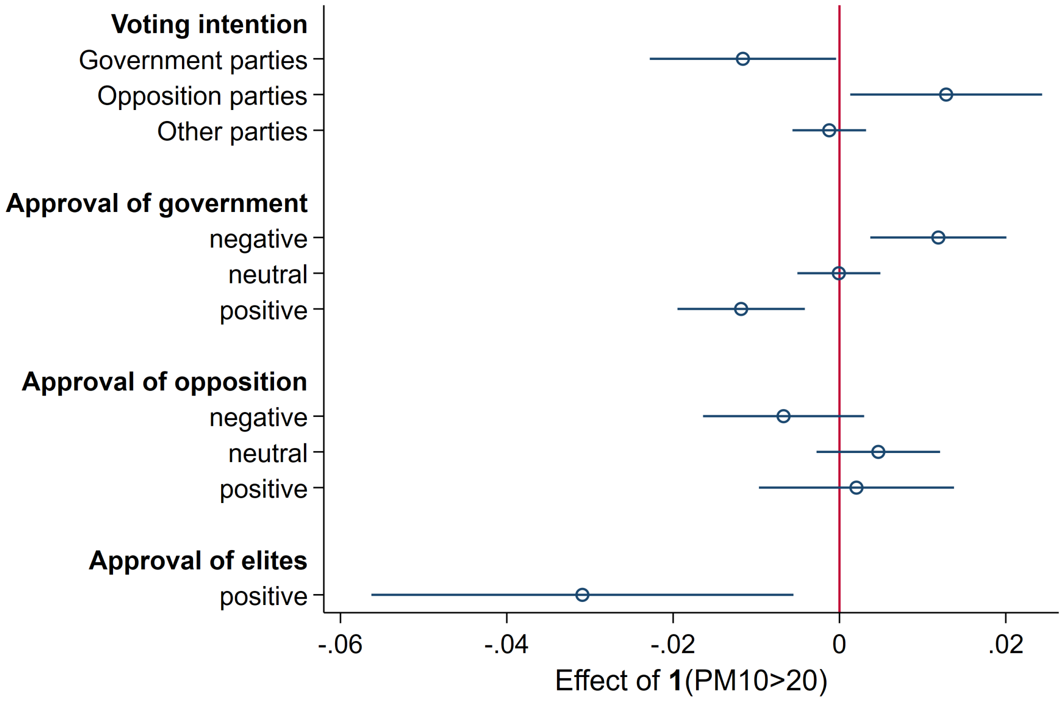
Panel survey data

We also use data from the Socio-Economic Panel (SOEP), a long-running representative panel survey since 1984, covering individuals living in private households in Germany (SOEP, 2019). The questionnaire has included questions on individuals' political attitudes since the mid-1980s as well as their affective well-being since 2007. To be consistent with the analysis of the *Politbarometer* data, we merge average levels of pollution by county of residence over the seven days preceding the interview date and use a binary indicator for county-level PM10 concentrations exceeding $20\mu g/m^3$ as a proxy for elevated pollution exposure.

4.6.2 Results from opinion polls

Figure 4.4 shows the results for voting intention for the parties forming the federal coalition government at the time of the interview as well as voting intentions for the opposition and other parties. Respondents are asked about their voting intention: if the federal election were to take place the following Sunday (*Sonntagsfrage*), would they vote, and — if so — for which party. We group parties into incumbent, opposition and other parties using the same classification as in Section 4.3.1. Each coefficient is the result of a separate regression of the outcomes listed on the left on an indicator for elevated levels of PM10 and controls for individual characteristics of respondents (gender, age, education, urban area, marital status, employment status, occupational status), weather controls (temperature, wind speed, precipitation) as well as year-by-week and state-by-month fixed effects.

Figure 4.4: Effect of Air Pollution on Voting Intentions and Government Approval



Notes: This graph displays the point estimates and 95% confidence intervals for OLS regressions of survey questions from a weekly opinion poll (*Politbarometer*) on a binary indicator for the PM10 concentration being above $20\mu g/m^3$ in a respondent's state in the week of the interview. The construction of the binary indicators is described in the text of Section 4.6.2. The regressions control for individual characteristics of respondents (gender, age, education, urban area, marital status, employment status, occupational status), weather controls (temperature, wind speed, precipitation) as well as year-by-week and state-by-month fixed effects. Standard errors clustered at the state-by-year level.

The results in Figure 4.4 confirm our main findings in Section 4.5, namely that exposure to higher levels of air pollution reduces support for the current government and increases support for the opposition. A higher concentration of PM10 is associated with lower voting intentions for the incumbent parties and higher voting intentions for the opposition. People exposed to elevated levels of PM10 report a 1.2 percentage point higher likelihood of voting for the established opposition and an equivalently lower likelihood of voting for the current government. This result amounts to 2.3% of the mean vote share of the government (52%) and 2.9% of the mean vote share of the opposition (42%). We do not observe any changes for other parties.

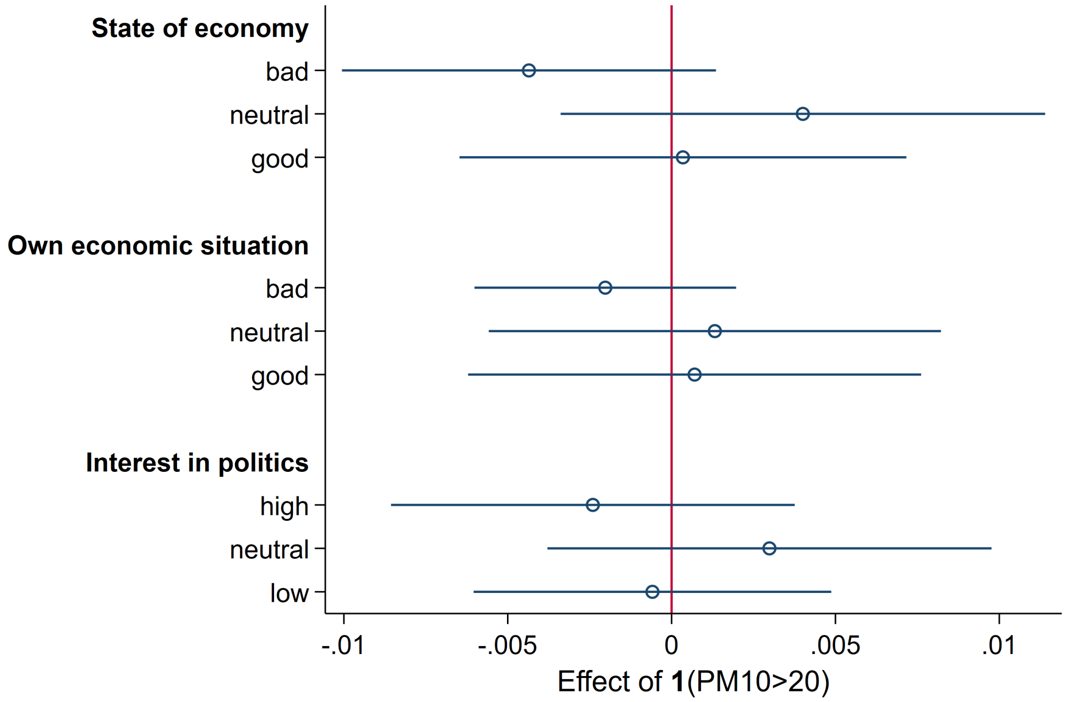
Figure 4.4 shows that the results for voting intentions are concurrent with changes in voters' approval of the government. Respondents can state their approval on an eleven-point scale from -5 (very dissatisfied) to 5 (very satisfied). We construct binary indicators for a positive approval (approval > 0), a negative approval (approval < 0), and a neutral approval (approval $= 0$). Individuals exposed to higher levels of PM10 are 1.2 percentage point more likely to express negative approval for the government and 1.2 percentage point less likely to express positive approval, relative to a mean of 36% (negative approval rating) and 51% (positive approval rating) respectively. By contrast, we find no difference in respondents' approval of the opposition. At the same time, we observe a large negative and statistically significant effect on the approval of elites in general[13]. The effect is sizable with a reduction of elite approval by about 3.1 percentage points relative to a mean of 27.7%. These results suggest that the observed effects on voting — fewer votes for the current government, more for the opposition — are driven by voters' dissatisfaction with the government and the elites rather than a change in people's views about the opposition.

In Figure 4.5, we investigate the role of people's perceptions of the economy and their interest in politics in general. The survey asks about respondents' assessment of the state of the German economy and their own economic situation on a five-point scale from 1 (very good) to 5 (very bad). The outcomes *State of economy bad*/*Own economic situation bad* are binary indicators for responses *bad/very bad*, while *State of economy good*/*Own economic situation good* indicate responses *good/very good*. We find no evidence that high levels of air pollution affect people's perception of the state of the economy or their own economic situation. There is also no evidence that it affects people's interest in politics. Respondents are further asked

[13]The survey asks whether respondents believe that currently in Germany, by and large, the right people are in leading positions or not. The outcome *Approval of elites* is a binary indicator for responding "yes".

about their extent of interest in politics on a five-point scale from 1 (very strong) to 5 (none). *High political interest* is a binary indicator for responses *strong/very strong*, whereas *Low political interest* indicates responses *none/little*. The stated interest in politics is no different between people exposed to a high versus low concentration of PM10.

Figure 4.5: Effect of Air Pollution on Interest in Politics and Perceptions of the Economy



Notes: This graph displays the point estimates and 95% confidence intervals for OLS regressions of survey questions from a weekly opinion poll (*Politbarometer*) on a binary indicator for the PM10 concentration being above $20\mu g/m^3$ in a respondent's state in the week of the interview. The construction of the binary indicators is described in the text of Section 4.6.2. The regressions control for individual characteristics of respondents (gender, age, education, urban area, marital status, employment status, occupational status), weather controls (temperature, wind speed, precipitation) as well as year-by-week and state-by-month fixed effects. Standard errors clustered at the state-by-year level.

Overall, these findings suggest that pollution affects people's voting intentions and behavior through their dissatisfaction with the current government rather than their dissatisfaction with the opposition or their own economic situation. Moreover, the fact that we do not see an effect on political interest supports the notion that the overall effect of air pollution on voting operates through subconscious channels.

4.6.3 Results from the Socio-Economic Panel

We complement the results from the opinion poll with data from the SOEP. The main advantage of the SOEP is its panel structure. Respondents are repeatedly asked the same questions, including questions about affective well-being and political preferences. This allows us to run regressions with individual fixed effects and compare the answers of the *same* person who was exposed to different levels of air pollution on different interview dates.

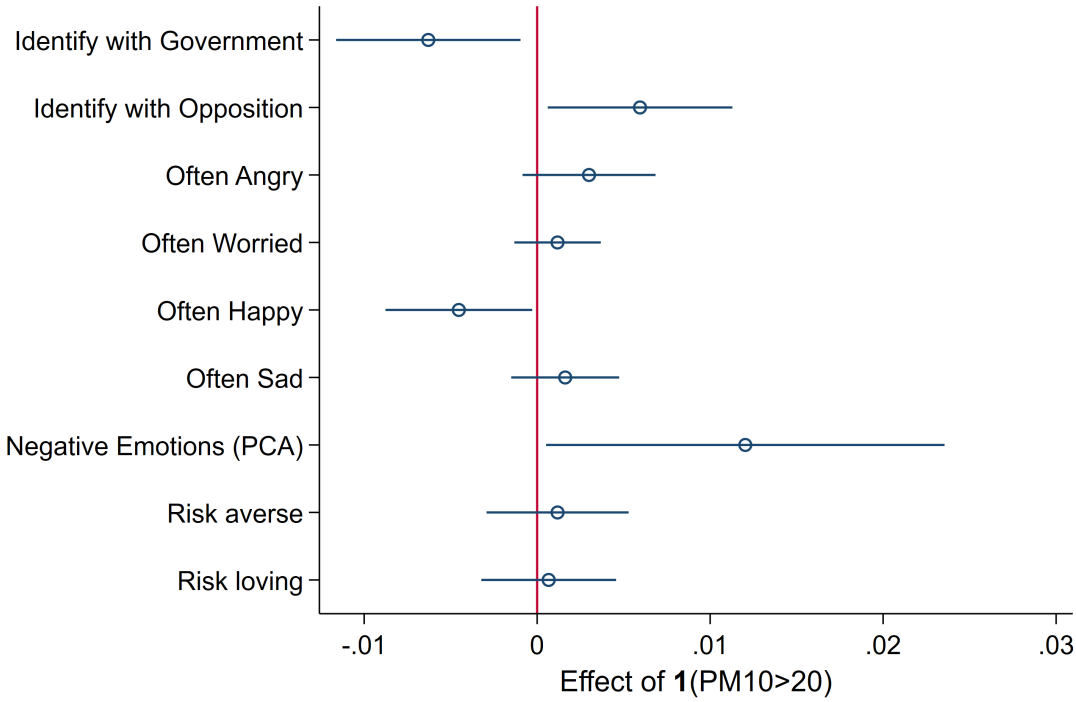
We use the SOEP to generalize our main findings, as well as to illuminate the role of emotions and risk preferences in explaining the overall effect. Our main outcomes are party identification, affective well-being and risk attitudes. To measure support for the incumbent government and the opposition, we use survey questions on party identification. Survey respondents are asked whether they lean towards a specific party in the long run and – if so – to which party they lean. Based on the responses, we construct binary indicators for *Party identification government/opposition* depending on which coalition was in power at the time of the interview date. To measure affective well-being, we use survey questions asking respondents how often they felt angry, worried, happy, or sad in the last four weeks. Based on the responses — ranging from 1 (*very rarely*) to 5 (*very often*) — we construct binary indicators that equal one if a respondent answers *often* or *very often*. As these four dimensions of affective well-being are strongly correlated, we further combine them into one dimension which we call "Negative emotions" using principal component analysis (PCA). Specifically, we run a PCA on the four binary indicators of affective well-being and use the first principal component as a summary outcome of affective well-being.

As discussed in Section 4.2, risk attitudes may also play an important role in voters' decision-making. In light of recent evidence that variation in people's emotions over time predict changes in risk attitudes (Meier, 2021), we additionally use information on self-reported risk attitudes from the SOEP. Respondents are asked whether they are generally willing to take risks on a scale from 0 (not at all willing to take risks) to 10 (very willing to take risks). This question about risk taking in general has been shown to be very predictive of risky behavior (Dohmen et al., 2011). Based on this survey question, we create two binary variables for risk attitudes. The indicator "Risk averse" takes on a value of one if individuals report values between 0 and 4 and zero otherwise, while the binary indicator "Risk loving" is one for responses between 6 and 10 and zero otherwise.

Figure 4.6 displays the results from individual fixed effects regressions for the 2000-2019 period. The coefficients are based on separate regressions of the binary indicators listed on the left on a binary indicator that equals one if the the average concentration of PM10 in a respondent's county of residence was above $20\mu g/m^3$ on the seven days preceding the interview. The regressions control for individual characteristics (age, marital status, number of children, education, net household income), weather controls (temperature, wind speed, precipitation) as well as year, quarter and state fixed effects. The results confirm the pattern found in voting data in Section 4.5. On days with higher air pollution, the same person shows less support for the government and more support for the opposition compared to days with low levels of air pollution.

The data on affective well-being allow us to test whether emotions are an important mechanism explaining the effect of air pollution on voting. As laid out in Section 4.2, anger, anxiety and (un-)happiness may affect electoral decision-making by changing people's conscious or unconscious perception of the status quo. Negative emotions have been shown to reduce the bias for the status quo, lead to greater willingness to change and reduce reliance on heuristics in decision-making. The results in Figure 4.6 are consistent with this mechanism. We find that higher levels of PM10 increase the likelihood of the negative emotions such as anger,

Figure 4.6: Effect of Air Pollution on Political Preferences and Affective Well-being



Notes: This graph displays the point estimates and 95% confidence intervals for OLS regressions of binary indicators of party identification, affective well-being, the first principal component of the four indicators of affective well-being as well as binary indicators for risk attitudes on a binary indicator for average PM10 concentrations being above $20\mu g/m^3$ over the seven days preceding the interview date. The results are based on survey responses from the SOEP over the period from 2000 to 2019. The regressions control for individual fixed effects as well as further characteristics of respondents (age, marital status, number of children, education, net household income), weather controls (temperature, wind speed, precipitation) as well as year, quarter and state fixed effects. The variable *Negative Emotions (PCA)* is the first principal component of four dimensions of affective well-being, namely anger, worry, happiness, and sadness. A larger value of this index indicates more negative emotions.

worry and sadness and at the same time reduce happiness. Effect sizes imply changes around 0.3 to 0.4 percentage points relative to means of 22% for being angry often or very often, 7% for being worried, 59% for feeling happy and 12% for feeling sad. Although these effect sizes are small relative to their respective means, they indicate that higher air pollution can change people's emotions. To gain statistical power, we construct an index of affective well-being by taking the first principal component of the four dimensions of well-being, namely anger, worry, happiness, and sadness. The results show that air pollution has a significant effect on emotions: the positive coefficient means that a high level of air pollution is associated with more negative emotions. Finally, we do not find any evidence that exposure to poor air quality shifts individuals' self-reported risk attitudes. Both binary indicators for being rather risk averse and risk loving respectively are small and statistically insignificant. This means that the channel via which negative emotions affect voting outcomes is unlikely to be risk attitudes. Overall, these findings are consistent with a psychological mechanism linking exposure to poor air quality with voting decisions. People who are exposed to elevated levels of particulate air pollution feel worried and unhappy more often, which may translate to a reduction in the status quo bias and therefore a reduction in the political support for the incumbent government in favor of the opposition.

4.7 Conclusion

In this paper, we have used parliamentary elections as a real-world laboratory to show that ambient air pollution affects decision-making. Using county-level voting outcomes from federal and state elections in Germany over a nineteen-year period as well as data on ambient concentrations of particulate matter and weather conditions on the election day, we find that higher concentrations of particulate matter reduce the electoral support of the incumbent government coalition's parties and increases the vote share of the opposition. We find similar results based on a weekly opinion poll and a large panel survey.

Our empirical setup as well as additional evidence from a survey suggest that our findings represent a behavioral bias. Our identification strategy exploits deviations in air pollution on election day from the usual level of pollution in a given county. Unlike changes in rainfall or temperature, such fluctuations in air pollution are not noticeable for voters. It is thus unlikely that our results reflect a deliberate choice, such as voters punishing the government for poor air quality. Our results are more likely to represent a subconscious change in behavior. Pollution can affect emotions and cognitive functioning, which in turn can lead to unintended changes in voting behavior. Based on survey data, we find that a plausible psychological mechanism is the impact of air pollution on people's emotions such as anxiety and unhappiness, which may reduce the support for the political status quo.

Our findings show that air pollution can have important effects on society. Parliamentary elections determine government formation as well as policy setting, which has a substantial impact on individual voters and society at large. Our results that air pollution affects the decision-making of the population at large in a high-stakes real-world setting, where people are faced with a decision concerning whether to retain or abandon the political status quo.

This finding opens up several avenues for future research. First, it would be welcome to enhance our understanding of the mechanisms through which air pollution affects decision-making. Although we are able to shed some light on potential mechanisms, this remains only suggestive evidence as we are limited by the granularity of our data in terms of the analysis we can conduct. Moreover, our survey data also do not allow us to illuminate the neurological and psychological responses to air pollution that affect decision-making. Existing research suggests that people appear to be less willing to take risks when exposed to elevated levels of air pollution in specific settings (investment decisions), while they show more impulsive and aggressive behavior in other situations (violent criminal behavior). It would be important to understand why air pollution triggers different responses in different contexts. Therefore, future research should further investigate the link between air pollution and emotions, focusing on the role of the decision-making environment and the individual returns associated with these decisions. In particular, the design of a dedicated study would enable researchers to provide a more robust and convincing evidence on the relationship between pollution and emotions. And it would also allow the testing of people's knowledge and awareness of air pollution, a topic that remains underinvestigated.

A second avenue for future research is to understand the effect of long-run changes in air pollution on voting. Our research identifies the effect of short-run fluctuations in air pollution, which points to subconscious changes in voting behavior. However, it would be equally important to know whether long-run changes in air pollution — namely changes that people actually notice — lead to deliberate changes in people's voting behavior.

5 Conclusion

Below are summarised the main findings uncovered by this thesis. These are first conceived independently and then they are reconciled together to form a unified and coherent picture. Subsequently, through the lens of this unified view, the discussion highlights the policy implications and, in light of the limitations mentioned in the previous Chapters, delineates some promising avenues for future research.

5.1 Key findings

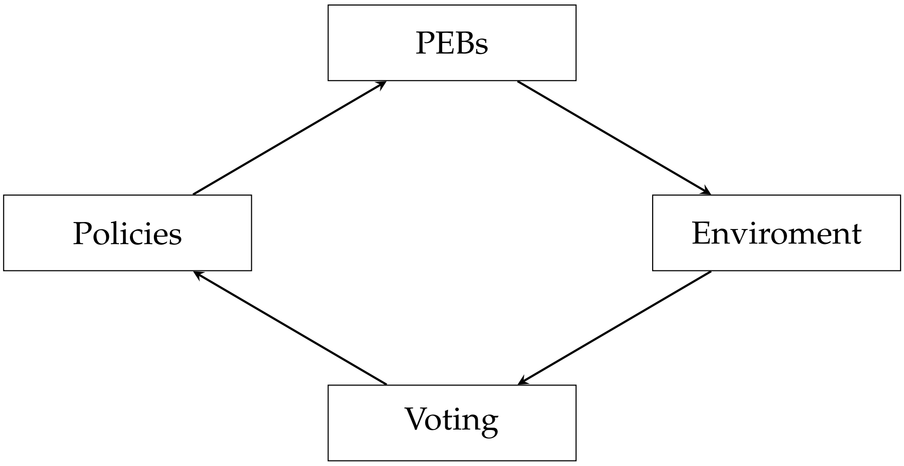
Chapter 2 shows that voluntarily performing an effortful pro-environmental behaviour generates a positive behavioural spillover. Individuals who decided to donate part of their final payoff to a pro-environmental organization have higher contribution levels in a subsequent multi-round public goods game. The study also highlights that the spillover effect is persistent (it remains at the same level throughout the entire game and does not disappear after a few rounds), and that it interests primarily the same domain as the initial behaviour (contributions increase for environmental public goods but not for generic ones). The combination of these results is reassuring in light of the lifestyle changes that people will have to make to tackle climate change. The investigated measures are also particularly relevant since many of the mitigation or adaptation interventions have public goods characteristics or are provided by pro-environmental organizations.

Chapter 3 highlights that reframing the energy consumption information displayed on appliances labels in monetary terms does not have the same impact everywhere. In Canada, both generic and personalised cost information reduce the WTP for energy efficiency of tumble driers. In the United Kingdom, personalised information increases the WTP for energy efficiency, while generic information has no effect. In Ireland and the United States, neither generic nor personalised energy cost information have a significant effect on the WTP for energy efficiency. The effect also differs based on individual characteristics. The negative effect of monetary information comes from the subgroup of households with lower usage of the appliance, while for those in the top percentile it is always positive and significant in the United Kingdom. In addition, although participants who are concerned about the environment have higher WTP for energy efficiency in all countries, reporting energy information in monetary terms has at best no effect for this subgroup. Conversely, monetary information improves WTP for energy efficiency for the subgroup of household not concerned about the environment in Ireland and the United Kingdom. There is no evidence of differential effects due to education or income levels.

Chapter 4 demonstrates that environmental conditions affect voting decisions through their effects on emotions, rather than by altering people's perception of the economy or their risk preferences. An increase by $10\mu g/m^3$ in the concentration of PM10 on election day reduces the vote share for incumbent parties by 2 percentage points, at the expenses of established opposition parties. On the other hand, no effect on voter turnout is detected. Hence, it appears that pollution does not change the propensity to vote but the decision of who to vote for. These results are corroborated by two nationally representative surveys. In addition, a higher concentration of PM10 favours the emergence of negative emotions, which lead to a reduced support for the incumbent, while it has no effect on the perception of the socio-economic situation nor on risk aversion or tolerance.

Although the three studies have been conducted independently and can rightfully exist as stand-alone analyses, their results are interconnected and can be read as a unified picture of mutual influences as reported in Figure 5.1.

Figure 5.1: Interconnection of the studies and phenomena presented in this thesis.



Pro-environmental behaviours, of which donations to pro-environmental organizations, contributions for the provision of environmental public goods and investments in energy efficiency are clear examples, have repercussions on environmental conditions. Among other things, they can have more or less direct impacts on the amount of pollution that gets either emitted or avoided. The state of the environment has been proven to affect individual decision-making. In particular, ambient air pollution influences voting behaviour, creating knock-on effects on the policy setting. Elected governments design and put in place policy interventions which then determine the performance of pro-environmental behaviours. In this way the circle will close and start again.

Like other policy topics, PEBs can affect governments' approval and their chances of re-election. However, unlike most other topics, their effect is more subtle since it can act through subconscious channels (people's emotions) and even when the reasons are not immediately apparent (neither the average pollution levels nor their variations considered in Chapter 4 are clearly visible and recognizable without the specific measurement techniques). Therefore, policymakers ought to carefully consider the ways in which different interventions can impact individuals' choices and the consequences that these will have on the environment.

5.2 Policy implications and avenues for future research

Figure 5.1 is a self-fuelling circle: PEBs can impact policy settings through affective influences of the environment on voting decisions, and, in turn, policy settings have an effect on the performance of PEBs. This scenario highlights that designing effective policies aimed at promoting the uptake of PEBs should be a key point in governments' agendas, since worse environmental conditions might translate in lower re-election chances.

Energy efficiency investments are a type of pro-environmental behaviour, and based on the findings of Chapter 2 if people decide to purchase energy efficient products they are likely to perform more environmentally-friendly actions in the future. This should reduce ambient air pollution which, as documented in Chapter 4, plays in favour of the government. However, Chapter 3 shows that framing energy consumption information in monetary terms improves the willingness-to-pay for energy efficiency in certain contexts but not in others. Therefore, further research is needed to assess the potential of alternative ways to promote the adoption of more energy efficient products in various contexts. In this sense, social comparison might prove fruitful since it has been shown to have positive effects on energy use either as a stand-alone intervention or in combination with other treatments (Alberts et al., 2016; Allcott, 2011b; Allcott and Rogers, 2014; Ayres et al., 2013; Costa and Kahn, 2013; Ferraro and Price, 2013; Mizobuchi and Takeuchi, 2013; Peschiera et al., 2010; Schultz et al., 2015; Seyranian et al., 2015; Tiefenbeck et al., 2013). Future studies could try to incorporate a social comparison component in the energy labels — e.g. reporting the consumption of the most commonly bought model of a certain product class — which should, at worst, discourage the purchase of less efficient products.

Although Chapter 2 has documented the emergence of positive behavioural spillovers, it has not delved into the the mechanisms which are responsible for this phenomenon. Previous studies have considered self-perception, cognitive dissonance, goal activation, action-based learning and moral licensing as possible drivers (Thøgersen and Noblet, 2012), with the first two appearing more plausible within the specific analysis presented in this thesis. Nevertheless, a more clear knowledge of which one(s) of these mechanisms is at play in various contexts is needed to ensure that policy interventions, for example those aimed at promoting energy efficiency investments, employ measures which favour, or at least do not hinder, their functioning. If spillover effects are due to either self-perception or cognitive dissonance, it once again appears that measures based on social comparison could be a valuable option. In the first case, if one purchases a product which is more efficient than that used by the majority of other people, this person might perceive herself/himself as more environmentally-friendly than the majority and thus behave accordingly in the future. In the second case, an individual who purchases a product which is more efficient than the one used by the majority of other people should continue to buy more efficient products to avoid the discomfort generated by a sense of dissonance with past behaviour.

Many of the considerations above are tied together by the premise that governments will want to improve environmental conditions to limit the risk of electoral defeat. However, the analysis presented in Chapter 4 only considers a specific country (Germany) and a specific pollutant (PM10). Albeit informative, this offers but a partial representation of the more complex phenomenon of affective influences of environmental factors on voting decisions. Further investigation is required that examines different countries, potentially with different electoral systems. Other pollutants should also be considered. Due to the spatial and temporal distribution of the German monitoring network, PM2.5 could not be used. Yet, according to the World Health Organization (2005), it might represent a more accurate indicator of anthropogenic suspended particles. Hence, in light of the mutual relationship between human behaviour and environment which has been documented in this thesis, it is important that future studies try to incorporate it into their analysis.

Bibliography

Abrahamse, W., Steg, L., Vlek, C. and Rothengatter, T. (2007). The effect of tailored information, goal setting, and tailored feedback on household energy use, energy-related behaviors, and behavioral antecedents. *Journal of Environmental Psychology* 27: 265–276, doi:10.1016/j.jenvp.2007.08.002.

Achen, C. H. and Bartels, L. M. (2004). Blind Retrospection: Electoral Responses to Drought, Flu, and Shark Attacks. *Princeton University* doi:10.2139/ssrn.3609237.

Adamowicz, W., Louvier, J. and Williams, M. (1994). Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management* 26: 271–292, doi:10.1006/jeem.1994.1017.

Agarwal, S., Qin, Y., Shi, L., Wei, G. and Zhu, H. (2021). Impact of temperature on morbidity: New evidence from china. *Journal of Environmental Economics and Management* 109: 102495, doi:10.1016/j.jeem.2021.102495.

Alacevich, C., Bonev, P. and Söderberg, M. (2021). Pro-environmental interventions and behavioral spillovers: Evidence from organic waste sorting in sweden. *Journal of Environmental Economics and Management* 108: 102470, doi:10.1016/j.jeem.2021.102470.

Alberts, G., Gurguc, Z., Koutroumpis, P., Martin, R., Muuls, M. and Napp, T. (2016). Competition and norms: a self-defeating combination? *Energy Policy* 96: 504–523, doi:10.1016/j.enpol.2016.06.001.

Alesina, A. and Passarelli, F. (2019). Loss Aversion in Politics. *American Journal of Political Science* 63: 936–947, doi:10.1111/ajps.12440.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine. *Econometrica* 21: 503–546.

Allcott, H. (2011a). Consumers' perceptions and misperceptions of energy costs. *American Economic Review* 101: 98–104, doi:10.1257/aer.101.3.98.

Allcott, H. (2011b). Social norms and energy conservation. *Journal of Public Economics* 95: 1082–1095, doi:10.1016/j.jpubeco.2011.03.003.

Allcott, H. (2013). The welfare effects of misperceived product costs: Data and calibrations from the automobile market. *American Economic Journal: Economic Policy* 5: 30–66.

Allcott, H. and Knittel, C. (2019). Are consumers poorly informed about fuel economy? evidence from two experiments. *American Economic Journal: Economic Policy* 11: 1–37, doi:10.1257/pol.20170019.

Allcott, H. and Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review* 104: 3003–3037, doi:10.1257/aer.104.10.3003.

Allcott, H. and Sweeney, R. (2016). The role of sales agents in information disclosure: Evidence from a field experiment. *Management Science* 63: 1–19, doi: 10.1287/mnsc.2015.2327.

Allcott, H. and Taubinsky, D. (2015). Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market. *American Economic Review* 105: 2501–2538, doi:10.1257/aer.20131564.

Allcott, N. and Greenstone, M. (2012). Is there an energy efficiency gap? *The Journal of Economic Perspectives* 26: 3–28, doi:10.1257/jep.26.1.3.

Anderson, C. D. and Claxton, J. D. (2014). Barriers to consumer choice of energy efficient products. *Journal of Consumer Research* 9: 163–170, doi:10.1086/208909.

Andor, M. A. and Fels, K. M. (2018). Behavioral Economics and Energy Conservation – A Systematic Review of Non-price Interventions and Their Causal Effects. *Ecological Economics* 148: 178–210, doi:10.1016/j.ecolecon.2018.01.018.

Andor, M. A., Gerster, A. and Sommer, S. (2020). Consumer inattention, heuristic thinking and the role of energy labels. *The Energy Journal* 40: 83–112, doi:10.5547/01956574.41.1.mand.

Angie, A. D., Connelly, S., Waples, E. P. and Kligyte, V. (2011). The influence of discrete emotions on judgement and decision-making: A meta-analytic review. *Cognition and Emotion* 25: 1393–1422, doi:10.1080/02699931.2010.550751.

Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91(434): 444–455, doi:10.2307/2291629.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Aragón, F. M., Miranda, J. J. and Oliva, P. (2017). Particulate matter and labor supply: The role of caregiving and non-linearities. *Journal of Environmental Economics and Management* 86: 295–309, doi:10.1016/j.jeem.2017.02.008, special issue on environmental economics in developing countries.

Aravena, C., Martinsson, P. and Scarpa, R. (2014). Does money talk? - The effect of a monetary attribute on the marginal values in a choice experiment. *Energy Economics* 44: 483–491, doi:10.1016/j.eneco.2014.02.017.

Archsmith, J., Heyes, A. and Saberian, S. (2018). Air Quality and Error Quantity: Pollution and Performance in a High-Skilled, Quality-Focused Occupation. *Journal of the Association of Environmental and Resource Economists* 5: 827–863, doi:10.1086/698728.

Arnold, F. and Freier, R. (2016). Only conservatives are voting in the rain: Evidence from German local and state elections. *Electoral Studies* 41: 216–221, doi:https://doi.org/10.1016/j.electstud.2015.11.005.

Attanasi, G., Corazzini, L. and Passarelli, F. (2017). Voting as a lottery. *Journal of Public Economics* 146: 129–137, doi:10.1016/j.jpubeco.2016.11.

Ayres, I., Raseman, S. and Shih, A. (2009). Evidence from Two Large Field Experiments That Peer Comparison Feedback Can Reduce Residential Energy Usage. Working Paper 15386, National Bureau of Economic Research.

Ayres, I., Raseman, S. and Shih, A. (2013). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *Journal of Law, Economics & Organizations* 29: 992–1022, doi:10.1093/jleo/ews020.

Baca-Motes, K., Brown, A., Gneezy, A., Keenan, E. A. and Nelson, L. D. (2013). Commitment and behavior change: Evidence from the field. *Journal of Consumer Research* 39: 1070–1084, doi:10.1086/667226.

Balakrishnan, U. and Tsaneva, M. (2021). Air pollution and academic performance: Evidence from India. *World Development* 146: 105553, doi:10.1016/j.worlddev.2021.105553.

Banfi, S., Farsi, M., Filippini, M. and Jakob, M. (2008). Willingness to pay for energysaving measures in residential buildings. *Energy Economics* 30: 503–516, doi:10.1016/j.eneco.2006.06.001.

Barwick, P. J., Li, S., Lin, L. and Zou, E. (2019). From Fog to Smog: the Value of Pollution Information. Working Paper 26541, National Bureau of Economic Research, doi:10.3386/w26541.

Bedi, A. S., Nakaguma, M. Y., Restrepo, B. J. and Rieger, M. (2021). Particle Pollution and Cognition: Evidence from Sensitive Cognitive Tests in Brazil. *Journal of the Association of Environmental and Resource Economists* 8: 443–474, doi:10.1086/711592.

Bell, M. L., Zanobetti, A. and Dominici, F. (2013). Evidence on vulnerability and susceptibility to health risks associated with short-term exposure to particulate matter: A systematic review and meta-analysis. *American Journal of Epidemiology* 178: 865–876, doi:10.1093/aje/kwt090.

Bem, D. J. (1972). Self-perception theory. In Berkowitz, L. (ed.), *Advances in Experimental Social Psychology*. New York: Academic Press, 1–62.

Bernasconi, M., Corazzini, L., Kube, S. and Maréchal, M. A. (2009). Two are better than one! individuals' contributions to "unpacked" public goods. *Economic Letters* 104: 31–33, doi:10.1016/j.econlet.2009.03.015.

Bernedo, M., Ferraro, P. J. and Price, M. (2014). The persistent impacts of norm-based messaging and their implications for water conservation. *Journal of Consumer Policy* 37: 437–452, doi:10.1007/s10603-014-9266-0.

Binswanger, M. (2001). Technological progress and sustainable development: what about the rebound effect? *Ecological Economics* 36: 119–132, doi:10.1016/S0921-8009(00)00214-7.

Blackwell, C. and McKee, M. (2003). Only for my own neighborhood? preferences and voluntary provision of local and global public goods. *Journal of Economic Behavior & Organization* 52: 115–131, doi:10.1016/S0167-2681(02)00178-6.

Blanken, I., Ven, N. van de and Zeelenberg, M. (2015). A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin* 41(4): 540–558, doi:10.1177/0146167215572134.

Block, L. G. and Keller, P. A. (1995). When to accentuate the negative: The effects of perceived efficacy and message framing on intentions to perform a health related behavior. *Journal of Marketing Research* 32: 192–203, doi:10.2307/3152047.

Bondy, M., Roth, S. and Sager, L. (2020). Crime Is in the Air: The Contemporaneous Relationship between Air Pollution and Crime. *Journal of the Association of Environmental and Resource Economists* 7: 555–585, doi:10.1086/707127.

Brounen, D., Kok, N. and Quigley, L. M. (2013). Energy literacy, awareness, and conservation behavior of residential households. *Energy Economics* 38: 42–50, doi:10.1016/j.eneco.2013.02.008.

Brown, M. and Forsythe, A. (1974). Robust test for the equality of variances. *Journal of the American Statistical Association* 69: 364–367, doi:10.2307/2285659.

Bruyneel, S. D., Dewitte, S., Franses, P. H. and Dekimpe, M. G. (2009). I felt low and my purse feels light: depleting mood regulation attempts affect risk decision making. *Journal of Behavioral Decision Making* 22: 153–170, doi:10.1002/bdm.619.

Bundeswahlleiter (2017). Anteil der Briefwählerinnen und Briefwähler bei den Bundestagswahlen 1994 bis 2017.

Burkhardt, J., Bayham, J., Wilson, A., Carter, E., Berman, J. D., O'Dell, K., Ford, B., Fischer, E. V. and Pierce, J. R. (2019). The effect of pollution on crime: Evidence from data on particulate matter and ozone. *Journal of Environmental Economics and Management* 98: 102267, doi:10.1016/j.jeem.2019.102267.

Camerer, C., Issacharoff, S., Loewenstein, G., O'Donoghue, T. and Rabin, M. (2003). Regulation for conservatives: Behavioral economics and the case for "asymmetric paternalism". *University of Pennsylvania Law Review* 151: 1211–1254, doi:10.2307/3312889.

Carrico, A. R., Raimi, K. T., Truelove, H. B., and Eby, B. (2018). Putting your money where your mouth is: An experimental test of pro-environmental spillover from reducing meat consumption to monetary donations. *Environment and Behavior* 50(7): 723–748, doi:10.1177/0013916517713067.

Carroll, J., Aravena, C., Boeri, M. and Denny, E. (2021). The energy cost information gap and the effects of short and long-term monetary labels on household decisions. *The Energy Journal, forthcoming* .

Carroll, J., Aravena, C. and Denny, E. (2016a). Low energy efficiency in rental properties: Asymmetric information or low willingness-to-pay? *Energy Policy* 96, doi:10.1016/j.enpol.2016.06.019.

Carroll, J., Denny, E. and Lyons, R. C. (2020). Better energy cost information changes household property investment decisions: Evidence from a nation-wide experiment. Trinity Economics Papers tep1520, Trinity College Dublin, Department of Economics.

Carroll, J., Denny, E. and Lyons, S. (2016b). The effects of energy cost labelling on appliance purchasing decisions: Trial results from ireland. *Journal of Consumer Policy* 39: 23–40, doi:10.1007/s10603-015-9306-4.

Central Statistics Office (2016a). Census 2016. Available at: `https://www.cso.ie/en/csolatestnews/presspages/2017/census2016profile3-anageprofileofireland/`. Accessed on February 8, 2021.

Central Statistics Office (2016b). Census 2016. Available at: `https://www.cso.ie/en/releasesandpublications/ep/p-cp11eoi/cp11eoi/`. Accessed on February 8, 2021.

Central Statistics Office (2016c). Census 2016. Available at: `https://www.cso.ie/en/releasesandpublications/ep/pcp10esil/p10esil/`. Accessed on February 8, 2021.

Central Statistics Office (2017). Households and families. Available at: `https://www.cso.ie/en/media/csoie/releasespublications/documents/population/2017/Chapter_4_Households_and_families.pdf`. Accessed on February 8, 2021.

Chang, T. Y., Graff Zivin, J., Gross, T. and Neidell, M. (2016). Particulate Pollution and the Productivity of Pear Packers. *American Economic Journal: Economic Policy* 8: 141–169.

Chang, T. Y., Graff Zivin, J., Gross, T. and Neidell, M. (2019). The Effect of Pollution on Worker Productivity: Evidence from Call Center Workers in China. *American Economic Journal: Applied Economics* 11: 151–172.

Chang, T. Y., Huang, W. and Wang, Y. (2018). Something in the air: Pollution and the demand for health insurance. *The Review of Economic Studies* 85: 1609–1634, doi:10.1093/restud/rdy016.

Chen, D., Schonger, M. and Wickens, C. (2016). otree - an open-source platform for laboratory, online and field experiments. *Journal of Behavioral and Experimental Finance* 9: 88–97, doi:10.1016/j.jbef.2015.12.001.

Chen, S., Guo, C. and Huang, X. (2018). Air Pollution, Student Health, and School Absences: Evidence from China. *Journal of Environmental Economics and Management* 92: 465–497, doi:https://doi.org/10.1016/j.jeem.2018.10.002.

Chen, X. (2019). Smog, Cognition and Real-World Decision Making. *International Journal of Health Policy and Management* 8: 76–80.

Clark, C. F., Kotchen, M. J. and Moore, M. R. (2003). Internal and external in‡uences on pro-environmental behavior: Participation in a green electricity program. *Journal of Environmental Psychology* 23: 237–246, doi:10.1016/S0272-4944(02)00105-6.

Clements, J., McCright, A., Dietz, T. and Marquart-Pyatt, S. (2015). A behavioural measure of environmental decision-making for social surveys. *Environmental Sociology* 1: 27–37, doi:10.1080/23251042.2015.1020466.

Clifford, A., Lang, L., Chen, R., Anstey, K. J. and Seaton, A. (2016). Exposure to air pollution and cognitive functioning across the life course–a systematic literature review. *Environmental Research* 147: 383–398, doi:10.1016/j.envres.2016.01.018.

Collaborative Labeling and Appliance Standards Program (2005). Global S&L Database, `https://clasp.ngo/publications/s-l-guidebook-english-version`. Accessed on September 2, 2020.

Corazzini, L., Cotton, C. and Valbonesi, P. (2015). Donor coordination in project funding: Evidence from a threshold public goods experiment. *Journal of Public Economics* 128: 16–29, doi:10.1016/j.jpubeco.2015.05.005.

Cornelissen, G., Pandelaere, M., Warlop, L. and Dewitte, S. (2008). Positive cueing: Promoting sustainable consumer behavior by cueing common environmental behaviors as environmental. *International Journal of Research in Marketing* 25: 46–55, doi:10.1016/j.ijresmar.2007.06.002.

Costa, D. L. and Kahn, M. E. (2013). Energy Conservation "Nudges" and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment. *The Review of Economics and Statistics* 11: 680–702, doi: doi.org/10.1111/jeea.12011.

Currie, J., Hanushek, E. A., Kahn, E. M., Neidell, M. and Rivkin, S. G. (2009). Does Pollution Increase School Absences? *The Review of Economics and Statistics* 91: 682–694.

Danziger, S., Levav, J. and Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences* 108: 6889–6892, doi: 10.1073/pnas.1018033108.

Davis, L. W. and Metcalf, G. E. (2016). Does better information lead to better choices? evidence from energy-efficiency labels. *Journal of the Association of Environmental and Resource Economists* 3: 589–625, doi:10.1086/686252.

DEFRA (2008). A framework for pro-environmental behaviours. London: Department for Environment, Food and Rural Affairs. January 2008.

Della Giusta, M., Jewell, S. and McCloy, R. (2012). Good enough? pro environmental behaviors, climate change and licensing. Economics & Management Discussion Papers em-dp2012-03, Henley Business School, Reading University.

DellaVigna, S. (2009). Psychology and Economics: Evidence from the Field. *Journal of Economic Literature* 47: 315–72, doi:10.1257/jel.47.2.315.

Department of Energy and Climate Change (2014). Evaluation of the decc/john lewis energy labelling trial. Available at: `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/350282/John_Lewis_trial_report_010914FINAL.pdf`. Accessed on March 12, 2021.

Deryugina, T., Heutel, G., Miller, N. H., Molitor, D. and Reif, J. (2019). The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction. *American Economic Review* 109: 4178–4219, doi:10.1257/aer.20180279.

Dhar, R. and Simonson, I. (1999). Making complementary choices in consumption episodes: highlighting versus balancing. *Journal of Marketing Research* 36(1): 29–44, doi:10.1177/002224379903600103.

Dietz, T., Gardner, G. T., Gilligan, J., Stern, P. C. and Vandenbergh, M. P. (2009). Proceedings of the national academy of sciences of the united states of america. *Journal of Experimental Social Psychology* 106(44): 18452–18456, doi:10.1073/pnas.0908738106.

Dohmen, T., Armin, F., David, H., Uwe, S., Juergen, S. and Gert, G. W. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of theEuropean Economic Association* 9: 522–550.

Dolan, P. and Galizzi, M. M. (2015). Like ripples on a pond: Behavioral spillovers and their implications for research and policy. *Journal of Economic Psychology* 47: 1–16, doi:10.1016/j.joep.2014.12.003.

Dranove, D. and Jin, G. Z. (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature* 48: 953–963, doi:10.1257/jel.48.4.935.

Ebenstein, A., Lavy, V. and Roth, S. (2016). The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution. *American Economic Journal: Applied Economics* 8: 36–65, doi:10.1257/app.20150213.

Eckles, D. and Schaffner, B. (2011). Risk Tolerance and Support for Potential Military Interventions. *The Public Opinion Quarterly* 75, doi:10.2307/41288400.

Eckles, D. L., Kam, C. D., Maestas, C. D. and Schaffner, B. F. (2014). Risk Attitudes and the Incumbency Advantage. *Political Behavior* 36: 731–749, doi:10.1007/s11109-013-9258-9.

Edmans, A., García, D. and Norli, Ø. (2007). Sports Sentiment and Stock Returns. *The Journal of Finance* 62: 1967–1998, doi:10.1111/j.1540-6261.2007.01262.x.

Ehrlich, S. and Maestas, C. (2010). Risk Orientation, Risk Exposure, and Policy Opinions: The Case of Free Trade. *Political Psychology* 31: 657–684, doi:10.1111/j.1467-9221.2010.00774.x.

Eisinga, R., Te Grotenhuis, M. and Pelzer, B. (2012a). Weather conditions and political party vote share in Dutch national parliament elections, 1971–2010. *International Journal of Biometeorology* 56: 1161–1165, doi:10.1007/s00484-011-0504-8.

Eisinga, R., Te Grotenhuis, M. and Pelzer, B. (2012b). Weather conditions and voter turnout in Dutch national parliament elections, 1971–2010. *International Journal of Biometeorology* 56: 783–786, doi:10.1007/s00484-011-0477-7.

Ellsberg, D. (1961). Risk, ambiguity and the savage axioms. *Quarterly Journal of Economics* 75: 643–669.

Elster, J. (1989). Social norms and economic-theory. *Journal of Economic Perspectives* 3: 99–117.

Eren, O. and Mocan, N. (2018). Emotional Judges and Unlucky Juveniles. *American Economic Journal: Applied Economics* 10: 171–205, doi:10.1257/app.20160390.

Eriksson, L. M. (2016). Winds of change: voter blame and Storm Gudrun in the 2006 Swedish parliamentay elections. *Elecoral Studies* 41: 129–142, doi:10.1016/j.electstud.2015.12.003.

European Environment Agency (2016). Air quality standards under the Air Quality Directive, and WHO air quality guidelines.

European Environmental Agency (2020). Final energy consumption by fuel type and sector. Available at: https://www.eea.europa.eu/data-and-maps/daviz/final-energy-consumption-of-fuel-1#tab-chart_1_filters=%7B%22rowFilters%22%3A%7B%7D%3B%22columnFilters%22%3A%7B%22pre_config_nrg_bal_label%22%3A%5B%22commercial%20and%20public%20services%22%3B%22households%22%3B%22industry%22%3B%22other%22%3B%22transport%22%5D%3B%22pre_config_siec_label%22%3A%5B%22Total%22%5D%7D%7D.

European Parliament (2018). Directive 2018/2002/eu amending directive 2012/27/eu on energy efficiency. Official Journal of the European Union, 328.

Farrow, K., Grolleau, G. and Ibanez, L. (2017). Social norms and pro-environmental behavior: A review of the evidence. *Psychological Science* 140: 1–13, doi:10.1016/j.ecolecon.2017.04.017.

Feinberg, M. and Willer, R. (2011). Apocalypse soon? dire messages reduce belief in global warming by contradicting just-world beliefs. *Psychological Science* 21: 34–38, doi:10.1177/0956797610391911.

Fellner, G. and Lünser, G. (2014). Cooperation in local and global groups. *Journal of Economic Behavior & Organization* 108: 364–373, doi:10.1016/j.jebo.2014.02.007.

Ferraro, P. J., Miranda, J. J. and Price, M. K. (2011). The persistence of treatment effects with norm-based policy instruments: Evidence from a randomized environmental policy experiment. *American Economic Review* 101: 318–322, doi:10.1257/aer.101.3.318.

Ferraro, P. J. and Price, M. K. (2013). Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *The Review of Economics and Statistics* 95: 64–73, doi:10.1162/REST_a_00344.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Evanstone, IL: Row, Peterson and Company.

Forschungsgruppe Wahlen (2020). Politbarometer 1977–2019 (Partielle Kumulation). GESIS Datenarchiv, Köln. ZA2391 Datenfile Version 12.0.0, https://doi.org/10.4232/1.13631, doi:10.4232/1.13631.

Fowler, A. and Hall, A. B. (2018). Do Shark Attacks Influence Presidential Elections? Reassessing a Prominent Finding on Voter Competence. *The Journal of Politics* 80: 1423–1437, doi:10.1086/699244.

Fowler, A. and Montagnes, B. P. (2015). College football, elections, and false-positive results in observational research. *Proceedings of the National Academy of Sciences* 112: 13800–13804, doi:10.1073/pnas.1502615112.

Galizzi1, M. M. and Whitmarsh, L. (2019). How to measure behavioral spillovers: A methodological review and checklist. *Frontiers in Psychology* 10: 342, doi:10.3389/fpsyg.2019.00342.

Geng, L., Cheng, X., Tang, Z., Zhou, K. and Ye, L. (2016). Can previous pro-environmental behaviours influence subsequent environmental behaviours? the licensing effect of pro-environmental behaviours. *Journal of Pacific Rim Psychology* 10(9): 1–9, doi:10.1017/prp.2016.6.

Gerber, A. S., Gimpel, J. G., Green, D. P. and Shaw, D. R. (2011). How Large and Long-lasting Are the Persuasive Effects of Televised Campaign Ads? Results from a Randomized Field Experiment. *American Political Science Review* 105: 135–150, doi:10.1017/S000305541000047X.

Gomez, B. T., Hansford, T. G. and Krause, G. A. (2007). The Republicans Should Pray for Rain: Weather, Turnout, and Voting in U.S. Presidential Elections. *Journal of Politics* 69: 649–663, doi:10.1111/j.1468-2508.2007.00565.x.

Gonzales, M. H., Aronson, E. and Costanzo, M. A. (1988). Using social cognition and persuasion to promote energy conservation: a quasi-experiment. *Journal of Applied Social Psychology* 18: 1049–1066, doi:10.1111/j.1559-1816.1988.tb01192.x.

Grable, J. E. and Roszkowski, M. J. (2008). The influence of mood on the willingness to take financial risks. *Journal of Risk Research* 11: 905–923.

Graff Zivin, J. and Neidell, M. (2012). The Impact of Pollution on Worker Productivity. *American Economic Review* 102: 3652–73, doi:10.1257/aer.102.7.3652.

Graff Zivin, J. and Neidell, M. (2013). Environment, Health, and Human Capital. *Journal of Economic Literature* 51: 689–730, doi:10.1257/jel.51.3.689.

Graff Zivin, J. and Neidell, M. (2018). Air pollution's hidden impacts. *Science* 359: 39–40, doi:10.1126/science.aap7711.

Grazzini, L., Rodrigo, P., Aiello, G. and Viglia, G. (2018). Loss or gain? The role of message framing in hotel guests' recycling behaviour. *Journal of Sustainable Tourism* 26: 1944–1966, doi:10.1080/09669582.2018.1526294.

Greene, W. H. and Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Resource Part B: Methodology* 37: 681–698, doi:10.1016/S0191-2615(02)00046-2.

Hale, S. (2010). The new politics of climate change: why we are failing and how we will succeed. *Environmental Politics* 19(2): 255–275, doi:10.1080/09644010903576900.

Hanley, N., Wright, R. and Adamowicz, V. (1998). Using choice experiments to value the environment. *Environmental Resource Economics* 11: 413–428, doi:10.1023/A:1008287310583.

Hanna, R. and Oliva, P. (2015). The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City. *Journal of Public Economics* 122: 68–79, doi:10.1016/j.jpubeco.2014.10.004.

Hansford, T. G. and Gomez, B. T. (2010). Estimating the electoral effects of voter turnout. *The American Political Science Review* 104: 268–288.

Harding, M. and Hsiaw, A. (2014). Goal setting and energy conservation. *Journal of Economic Behavior & Organization* 107: 209–227.

Hasseldine, J. and Hite, P. A. (2003). Framing, gender and tax compliance. *Journal of Economic Psychology* 24: 517–533, doi:10.1016/S0167-4870(02)00209-X.

Healy, A. and Malhotra, N. (2010). Random Events, Economic Losses, and Retrospective Voting: Implications for Democratic Competence. *Quarterly Journal of Political Science* 5: 193–208, doi:10.1561/100.00009057.

Healy, A. J., Malhotra, N. and Mo, C. H. (2010). Irrelevant events affect voters' evaluations of government performance. *Proceedings of the National Academy of Sciences* 107: 12804–12809, doi:10.1073/pnas.1007420107.

Heinzel, S. (2012). Disclosure of energy operating cost information: A silver bullet for overcoming the energy-efficiency gap? *Journal of Consumer Policy* 35, doi: 10.1007/s10603-012-9189-6.

Heinzle, S. and Wüstenhagen, R. (2012). Dynamic Adjustment of Eco-labeling Schemes and Consumer Choice - the Revision of the EU Energy Label as a Missed Opportunity? *Business Strategy and the Environment* 21: 60–70, doi: 10.1002/bse.722.

Heissel, J., Persico, C. and Simon, D. (2021). The Impact of Accumulated and Acute Exposure to Traffic Pollution on Child Academic Outcomes. *Journal of Human Resources* 56: 406–445.

Hensher, D. A. and Louviere, J. (1983). On the design and analysis of simulated choice or allocation choice in travel choice modelling. *Transportation Research* 890: 1–7.

Herrnstadt, E., Heyes, A., Muehlegger, E. and Saberian, S. (2020). Air Pollution and Criminal Activity: Microgeographic Evidence from Chicago. *American Economic Journal: Applied Economics* Forthcoming.

Heyes, A., Neidell, M. and Saberian, S. (2016). The Effect of Air Pollution on Investor Behavior: Evidence from the S&P 500. Working Paper 22753, National Bureau of Economic Research, doi:10.3386/w22753.

Heyes, A., Rivers, N. and Schaufele, B. (2019). Pollution and Politician Productivity: The Effect of PM on MPs. *Land Economics* 95: 157–173.

Heyes, A. and Saberian, S. (2019). Temperature and decisions: Evidence from 207,000 court cases. *American Economic Journal: Applied Economics* 11: 238–65, doi:10.1257/app.20170223.

Hirshleifer, D. and Shumway, T. (2003). Good day sunshine: Stock returns and the weather. *The Journal of Finance* 58: 1009–1032.

Hockey, G. R. J., John Maule, A., Clough, P. J. and Bdzola, L. (2000). Effects of negative mood states on risk in everyday decision making. *Cognition & Emotion* 14: 823–855.

Holler, M., Hoelzl, E., Kirchler, E., Leder, S. and Mannetti, L. (2009). Framing of information of the use on public finances, regulatory fit of recipients and tax compliance. *Journal of Economic Psychology* 29: 597–611, doi:10.1016/j.joep.2008.01.001.

Huang, J., Xu, N. and Yu, H. (2020). Pollution and performance: Do investors make worse trades on hazy days? *Management Science* 66: 4455–4476, doi:10.1287/mnsc.2019.3402.

Iacus, S. M., King, G. and Porro, G. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association* 106: 345–361, doi:10.1198/jasa.2011.tm09599.

Iacus, S. M., King, G. and Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis* 20(1): 1–24, doi:10.1093/pan/mpr013.

International Energy Agency (2019). Iea world energy balances 2019. Available at: `https://www.iea.org/subscribe-to-data-services/world-energy-balancesand-statistics`. Accessed on September 9, 2020.

IPCC (2014). Climate change 2014: Synthesis report. contribution of working groups i, ii and iii to the fifth assessment report of the intergovernmental panel on climate change. [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

IPCC (2018). Global Warming of 1.5°C. Summary for Policymakers. Incheon, Republic of Korea. `http://report.ipcc.ch/sr15/pdf/sr15_spm_final.pdf`.

Isaac, R. M. and Walker, J. M. (1998). Nash as an organizing principle in the voluntary provision of public goods: Experimental evidence. *Experimental Economics* 1: 191–206, doi:10.1023/A:1009996324622.

Jaeger, C. M. and Schultz, P. W. (2017). Coupling social norms and commitments: Testing the underdetected nature of social influence. *Journal of Environmental Psychology* 51: 199–208, doi:10.1016/j.jenvp.2017.03.015.

Jaffe, A. B. and Stavins, R. N. (1994). The energy-efficiency gap what does it mean? *Energy Policy* 22: 804–810, doi:10.1016/0301-4215(94)90138-4.

Jain, M., Rao, A. B. and Patwardhan, A. (1994). Energy cost information and consumer decisions: Results from a choice experiment on refrigerator purchases in india. *The Energy Journal* 42: 253–272, doi:10.5547/01956574.42.2.mjai.

Jans, J., Johansson, P. and Nilsson, J. P. (2018). Economic status, air quality, and child health: Evidence from inversion episodes. *Journal of Health Economics* 61: 220–232, doi:10.1016/j.jhealeco.2018.08.002.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47: 263–291.

Kahneman, D. and Tversky, A. (1984). Choices, values and frames. *American Psychologist* 39: 341–350, doi:10.1037/0003-066X.39.4.341.

Kallbekken, S., Sælen, H. and Hermansen, E. A. T. (2013). Bridging the energy efficiency gap: A field experiment on lifetime energy costs and household appliances. *Journal of Consumer Policy* 36: 1–16, doi:10.1007/s10603-012-9211-z.

Kam, C. and Simas, E. (2012). Risk attitudes, candidate characteristics, and vote choice. *The Public Opinion Quarterly* 76: 747–760, doi:10.2307/41684598.

Kampa, M. and Castanas, E. (2008). Human health effects of air pollution. *Environmental Pollution* 151: 362 – 367.

Kamstra, M. J., Kramer, L. A. and Levi, M. D. (2003). Winter blues: A sad stock market cycle. *American Economic Review* 93: 324–343, doi:10.1257/000282803321455322.

Khazzoom, J. D. (1980). Economic implications of mandated efficiency in standards for household appliances. *The Energy Journal* 1(4): 21–40.

Kliger, D. and Levy, O. (2003). Mood-induced variation in risk preferences. *Journal of Economic Behavior & Organization* 52: 573–584, doi:10.1016/S0167-2681(03)00069-6.

Klingen, J. and Jos N. Ommeren van (2020). Risk Attitude and Air Pollution: Evidence From Chess. Tinbergen Institute Discussion Paper 2020-027/VIII, doi:10.2139/ssrn.3609237.

Komarek, T. M. and Kaplowitz, F. L. M. D. (2011). Valuing energy policy attributes for environmental management. Choice experiment evidence from a research institution. *Energy Policy* 39: 5105–5115, doi:10.1016/j.enpol.2011.05.054.

Koop, S. H. A., Dorssen, A. J. V. and Brouwer, S. (2019). Enhancing domestic water conservation behaviour: A review of empirical studies on influencing tactics. *Journal of Environmental Management* 247: 867–876, doi:10.1016/j.jenvman.2019.06.126.

Künn, S., Palacios, J. and Pestel, N. (2021). Indoor air quality and strategic decision making. Working Paper (earlier version available as IZA Discussion Paper No. 12632).

Lacasse, K. (2017). Can't hurt, might help: Examining the spillover effects from purposefully adopting a new pro-environmental behavior. *Environment and Behavior* 51(3): 259–287, doi:10.1177/0013916517748164.

Lancaster, K. (1966). A new approach to consumer theory. *Journal of Political Economy* 74: 132–157, doi:10.1086/259131.

Lancsar, E. and Louviere, J. (2008). Conducting Discrete Choice Experiments to Inform Healthcare Decision Making. A User's Guide. *Pharmacoeconomics* 26: 661–677, doi:10.2165/00019053-200826080-00004.

Landon, A. C., Woodward, R. T., Kyle, G. T. and Kaiser, R. A. (2018). Evaluating the efficacy of an information-based residential outdoor water conservation program. *Journal of Cleaner Production* 195: 56–65, doi:10.1016/j.jclepro.2018.05.196.

Lanzini, P. and Thøgersen, J. (2014). Behavioural spillover in the environmental domain: An intervention study. *Journal of Environmental Psychology* 40: 381–390, doi:10.1016/j.jenvp.2014.09.006.

Lavy, V., Ebenstein, A. and Roth, S. (2014). The Impact of Short Term Exposure to Ambient Air Pollution on Cognitive Performance and Human Capital Formation. Working Paper 20648, National Bureau of Economic Research, doi:10.3386/w20648.

Ledyard, J. (1995). Public goods: A survey of experimental research. In Kagel, J. and Roth, A. (eds), *The Handbook of Experimental Economics*. Princeton, NJ: Princeton University Press, 111–194.

Leiserowitz, A., Maibach, E., Roser-Renouf, C., Feinberg, G. and Howe, P. (2013). Climate change in the american mind: Americans' global warming beliefs and attitudes in april, 2013. Yale University and George Mason University. New Haven, CT: Yale Project on Climate Change Communication.

Lepori, G. M. (2015). Positive mood and investment decisions: Evidence from comedy movie attendance in the U.S. *Research in International Business and Finance* 34: 142–163, doi:10.1016/j.ribaf.2015.02.001.

Lerner, J. S., Li, Y., Valdesolo, P. and Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology* 66: 799–823, doi:10.1146/annurev-psych-010213-115043.

Lerner, J. S., Li, Y. and Weber, E. U. (2013). The financial costs of sadness. *Psychological Science* 24: 72–79, doi:10.1177/0956797612450302, pMID: 23150274.

Lerner, J. S., Small, D. A. and Loewenstein, G. (2004). Heart strings and purse strings: Carryover effects of emotions on economic decisions. *Psychological Science* 15: 337–341, doi:10.1111/j.0956-7976.2004.00679.x, pMID: 15102144.

Levene, H. (1960). Robust tests for equality of variances. In Olkin, I., Ghurye, S., Hoeffding, W., Madow, W. and Mann, H. (eds), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Menlo Park, CA: Stanford University Press, 278–292.

Levin, I. P., Schneider, S. L. and Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes* 76: 149–188, doi:10.1006/obhd.1998.2804.

Levinson, A. (2012). Valuing public goods using happiness data: The case of air quality. *Journal of Public Economics* 96: 869–880, doi:10.1016/j.jpubeco.2012.06.007.

Levy, T. and Yagil, J. (2011). Air pollution and stock returns in the US. *Journal of Economic Psychology* 32: 374–383.

Li, Z., Folmer, H. and Xue, J. (2014). To what extent does air pollution affect happiness? the case of the jinchuan mining area, china. *Ecological Economics* 99: 88–99, doi:https://doi.org/10.1016/j.ecolecon.2013.12.014.

Liberini, F., Redoano, M. and Proto, E. (2017). Happy voters. *Journal of Public Economics* 146: 41–57.

Lichter, A., Pestel, N. and Sommer, E. (2017). Productivity effects of air pollution: Evidence from professional soccer. *Labour Economics* 48: 54–66, doi:10.1016/j.labeco.2017.06.002.

Lim, Y.-H., Kim, H., Kim, J. H., Bae, S., Park, H. Y. and Hong, Y.-C. (2012). Air pollution and symptoms of depression in elderly adults. *Environmental Health Perspectives* 120: 1023–1028, doi:10.1289/ehp.1104100.

Liñeira, R. and Henderson, A. (2019). Risk attitudes and independence vote choice. *Political Behavior* doi:10.1007/s11109-019-09560-x.

Lodovici, M. and Bigagli, E. (2011). Oxidative stress and air pollution exposure. *Journal of toxicology* 2011: 487074–487074, doi:10.1155/2011/487074.

Loewenstein, G., O'Donoghue, T. and Rabin, M. (2003). Projection bias in predicting future utility. *The Quarterly Journal of Economics* 118: 1209–1248.

Lord, K. R. (1994). Motivating recycling behavior: a quasiexperimental investigation of message and source strategies. *Psychology & Marketing* 11: 341–358, doi:10.1002/mar.4220110404.

Loroz, P. S. (2007). The interaction of message frames and reference points in prosocial persuasive appeals. *Psychology & Marketing* 24: 1001–1023, doi:10.1002/mar.20193.

Louviere, J. and Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregated data. *Journal of Marketing Research* 20: 350–367, doi:10.2307/3151440.

Lowenstein, G. (1988). Frames of mind in intertemporal choice. *Management Science* 34: 200–214, doi:10.1287/mnsc.34.2.200.

Lowenstein, G. and Lerner, J. S. (2003). The role of affect in decision making. In Davidson, R. J., Scherer, K. R. and Goldsmith, H. H. (eds), *Handbook of Affective Scineces*. Oxford, UK: Oxford University Press, 619–642.

Lowenstein, G. and Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics* 107: 573–597, doi:10.2307/2118482.

Lowenstein, G. and Thaler, R. H. (1989). Anomalies: Intertemporal choice. *The Journal of Economomic Perspectives* 3: 181–193, doi:10.1257/jep.3.4.181.

Lu, J. G. (2020). Air pollution: A systematic review of its psychological, economic, and social effects. *Current Opinion in Psychology* 32: 52–65.

Luechinger, S. (2009). Valuing Air Quality Using the Life Satisfaction Approach. *The Economic Journal* 119: 482–515, doi:10.1111/j.1468-0297.2008.02241.x.

MacKay, D. J. (2009). *Sustainable energy - without the hot air*. Cambridge, England: UIT Cambridge Ltd.

Maki, A., Carrico, A. R., Raimi, K. T., Truelove, H. B., Araujo, B. and Yeung, K. L. (2019). Meta-analysis of pro-environmental behaviour spillover. *Nature Sustainability* 2: 307–315, doi:10.1038/s41893-019-0263-9.

Manisalidis, I., Stavropoulou, E., Stavropoulos, A. and Bezirtzoglou, E. (2020). Environmental and health impacts of air pollution: A review. *Frontiers in Public Health* 8: 14, doi:10.3389/fpubh.2020.00014.

Mankiw, N. G. and Taylor, M. P. (2014). *Economics. Third Edition.*. Andover, United Kingdom: Cengage Learning EMEA.

Marcus, G., Neuman, W. R. and MacKuen, M. (2000). *Affective Intelligence and Political Judgment*. Chicago: The University of Chicago Press.

Marcus, G., Neuman, W. R. and MacKuen, M. (eds) (2007). *The Affect Effect: Dynamics of Emotion in Political Thinking and Behavior*. Chicago: The University of Chicago Press.

Margetts, E. A. and Kashima, Y. (2017). Spillover between pro-environmental behaviours: The role of resources and perceived similarity. *Journal of Environmental Psychology* 49: 30–42, doi:10.1016/j.jenvp.2016.07.005.

Mazar, N. and Zhong, C. (2010). Do green products make us better people? *Psychological Science* 21(4): 494–498, doi:10.1177/0956797610363538.

McCalley, L. T. and Midden, C. J. H. (2002). Energy conservation through product-integrated feedback: the roles of goal-setting and social orientation. *Journal of Economic Psychology* 23: 589–603, doi:10.1016/S0167-4870(02)00119-8.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P. (ed.), *Frontiers of econometrics*. New York: Academic Press, 105–142.

McFadden, D. and Train, K. (2000). Mixed mnl models for discrete response. *Journal of Applied Econometrics* 15: 447–470, doi:10.1002/1099-1255(200009/10)15:5%3C447::AID-JAE570%3E3.0.CO;2-1.

McKinsey (2010). Energy efficiency: A compelling global resource. McKinsey Sustainability & Resource Productivity. `https://www.mckinsey.com/~/media/mckinsey/dotcom/client_service/Sustainability/PDFs/A_Compelling_Global_Resource.ashx`.

McNeill, D. L. and Wilkie, W. L. (1979). Public policy and consumer information:

Impact of the new energy labels. *Journal of Consumer Research* 6: 1–11, doi: 10.1086/208743.

Meier, A. N. (2021). Emotions and Risk Attitudes. *American Economic Journal: Applied Economics, forthcoming* .

Meier, A. N., Schmid, L. and Stutzer, A. (2019). Rain, emotions and voting for the status quo. *European Economic Review* 119: 434–451, doi:10.1016/j.euroecorev.2019.07.014.

Merritt, A., Effron, D. and Monin, B. (2010). Moral self-licensing: When being good frees us to be bad. *Social and Personality Psychology Compass* 4: 344–357, doi:10.1111/j.1751-9004.2010.00263.x.

Meyer, S. and Pagel, M. (2017). Fresh Air Eases Work–The Effect of Air Quality on Individual Investor Activity. Working Paper 24048, National Bureau of Economic Research, doi:10.3386/w24048.

Meyers-Levy, J. and Maheswaran, D. (2004). Exploring message framing outcomes when systematic, heuristic, or both types of processing occur. *Journal of Consumer Psychology* 14: 159–167, doi:10.1207/s15327663jcp1401&218.

Minton, A. P. and Rose, R. L. (1997). The effects of environmental concern on environmentally friendly consumer behavior: An exploratory study. *Journal of Business Research* 40: 37–48, doi:10.1016/S0148-2963(96)00209-3.

Mizobuchi, K. and Takeuchi, K. (2013). The influences of financial and non-financial factors on energy-saving behaviour: a field experiment in japan. *Energy Policy* 63: 775–787, doi:10.1016/j.enpol.2013.08.064.

Morgenstern, S. and Zechmeister, E. (2003). Better the Devil You Know than the Saint You Don't? Risk Propensity and Vote Choice in Mexico. *Journal of Politics* 63: 93–119, doi:10.1111/0022-3816.00060.

Morisi, D. (2018). Choosing the Risky Option: Information and Risk Propensity in Referendum Campaigns. *Public Opinion Quarterly* 82: 447–469, doi:10.1093/poq/nfy033.

Nabi, R. L., Gustafson, A. and Jensen, R. (2018). Framing climate change: exploring the role of emotion in generating advocacy behavior. *Science Communication* 40: 442–468, doi:10.1177/1075547018776019.

Newell, R. G. and Siikamäki, J. (2014). Nudging energy efficiency behavior: The role ¨ of information labels. *Journal of the Association of Environmental and Resource Economists* 1, doi:10.1086/679281.

Nigbur, D., Lyons, E. and Uzzell, D. (2010). Attitudes, norms, identity and environmental behaviour: using an expanded theory of planned behaviour to predict participation in a kerbside recycling programme. *British Journal of Social Psychology* 49: 259–284, doi:10.1348/014466609X449395.

Nigg, C., Burbank, P., Padula, C., Dufresne, R., Rossi, J., Velicer, W., Laforge, R. and Prochaska, J. (1999). Stages of change across ten health risk behaviors for older adults. *Gerontologist* 39(4): 473–482, doi:10.1093/geront/39.4.473.

Nisan, M. and Horenczyk, G. (1990). Moral balance: The effect of prior behaviour on decision in moral conflict. *British Journal of Social Psychology* 29(1): 29–42, doi:10.1111/j.2044-8309.1990.tb00884.x.

Nowakowski, A. (2021). Do unhappy citizens vote for populism? *European Journal of Political Economy* 68: 101985, doi:10.1016/j.ejpoleco.2020.101985.

Office for National Statistics (2017). Graduates in the uk labour market: 2017. Available at: `https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/graduatesintheuklabourmarket/2017#steady-increase-in-the-number-of-graduates-in-the-uk-over-the-past-decade`. Accessed on February 8, 2021.

Office for National Statistics (2019a). Population estimates by marital status and living arrangements, england and wales: 2019. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/populationestimatesbymaritalstatusandlivingarrangements/latest. Accessed on February 8, 2021.

Office for National Statistics (2019b). Population estimates for the uk, england and wales, scotland and northern ireland: mid-2019. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2019estimates. Accessed on February 8, 2021.

Office for National Statistics (2021). Labour market overview, uk: January 2021. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/uklabourmarket/latest#employment-unemployment-and-economic-inactivity. Accessed on February 8, 2021.

Otto, A. R. and Eichstaedt, J. C. (2018). Real-world unexpected outcomes predict city-level mood states and risk-taking behavior. *PloS one* 13: e0206923.

Peschiera, G., Taylor, J. E. and Siegel, J. A. (2010). Response–relapse patterns of building occupant electricity consumption following exposure to personal, contextualized and occupant peer network utilization data. *Energy and Buildings* 42: 1329–1336, doi:10.1016/j.enbuild.2010.03.001.

Poortinga, W. and Whitaker, L. (2018). Promoting the use of reusable coffee cups through environmental messaging, the provision of alternatives and financial incentives. *Sustainability* 10: 873, doi:10.3390/su10030873.

Powdthavee, N. and Oswald, A. J. (2020). Is there a link between air pollution and impaired memory? Evidence on 34,000 English citizens. *Ecological Economics* 169: 106485, doi:10.1016/j.ecolecon.2019.106485.

Qin, Y., Wu, J. and Yan, J. (2019). Negotiating housing deal on a polluted day: Consequences and possible explanations. *Journal of Environmental Economics and Management* 94: 161–187, doi:10.1016/j.jeem.2019.02.00.

Revelt, D. and Train, K. (1998). Mixed logit with repeated choices: Households' choices of appliance efficiency level. *The Review of Economics and Statistics* 80: 647–657, doi:10.1162/003465398557735.

Rivers, N. and Jaccard, M. (2005). Combining top-down and bottom-up approaches to energy-economy modelling using discrete choice methods. *The Energy Journal* 26: 83–106.

Ropert Homar, A. and Cvelbar, L. K. (2021). The effects of framing on environmental decisions: A systematic literature review. *Ecological Economics* 183: 106950, doi:10.1016/j.ecolecon.2021.106950.

Rothman, A. J. and Salovey, P. (1997). Shaping perceptions to motivate healthy behavior: The role of message framing. *Psychological Bulletin* 121: 3–19, doi: 10.1037/0033-2909.121.1.3.

Rothman, A. J., Salovey, P., Antone, C., Keough, K. and Martin, C. D. (1993). The influence of message framing on intentions to perform health behaviors. *Journal of Experimental Social Psychology* 29: 408–433, doi:10.1006/jesp.1993.1019.

Ryan, M., Gerard, K. and Amaya-Amaya, M. (2008). Using discrete choice experiments to value health and health care. Dordrecth: Springer.

Sachdeva, S., Iliev, R. and Medin, D. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science* 20(4): 523–528, doi: 10.1111%2Fj.1467-9280.2009.02326.x.

Salvi, A. and Salim, S. (2019). Neurobehavioral consequences of traffic-related air pollution. *Frontiers in neuroscience* 13: 1232–1232, doi:10.3389/fnins.2019.01232.

Sammer, K. and Wüstenhagen, R. (2006). The influence of eco-labelling on consumer behaviour - Results of a discrete choice analysis for washing machines. *Business Strategy and the Environment* 15: 185–199, doi:10.1002/bse.522.

Samuelson, P. (1938). A note on the pure theory of consumer's behavior. *Econometrica* 5: 61–71.

Sanders, J. G. and Jenkins, R. (2016). Weekly fluctuations in risk tolerance and voting behaviour. *PLOS ONE* 11: 1–12, doi:10.1371/journal.pone.0159017.

Savage, L. J. (1954). *The foundations of statistics*. New York, NY: Wiley.

Schultz, P., Estrada, M., Schmitt, J., Sokoloski, R. and Silva-Send, N. (2015). Using in-home displays to provide smart meter feedback about household electricity consumption: a randomized control trial comparing kilowatts, cost, and social norms. *Energy* 90: 351–358, doi:10.1016/j.energy.2015.06.130.

Seyranian, V., Sinatra, G. M. and Polikoff, M. S. (2015). Comparing communication strategies for reducing residential water consumption. *Journal of Environmental Psychology* 41: 81–90, doi:10.1016/j.jenvp.2014.11.009.

Sforza, A. (2014). The Weather Effect: estimating the effect of voter turnout on electoral outcomes in Italy. *Banco de Portugal Working Paper* 5.

Shen, J. and Saijo, T. (2009). Does an energy efficiency label alter consumers' purchasing decisions? a latent class approach based on a stated choice experiment in shanghai. *Journal of Environmental Management* 90: 3561–3573, doi:10.1016/j.jenvman.2009.06.010.

Shepsle, K. A. (1972). The Strategy of Ambiguity: Uncertainty and Electoral Competition. *American Political Science Review* 66: 555–568.

Simonsohn, U. (2010). Weather to go to college. *The Economic Journal* 120: 270–280, doi:10.1111/j.1468-0297.2009.02296.x.

Sitarz, D. (1993). Agenda 21: The earth summit strategy to save our planet. United States.

SOEP (2019). Socio-Economic Panel (SOEP), data for years 1984-2018, version 35. doi:10.5684/soep.v35.

Sparks, P. and Shepherd, R. (1992). Self-identity and the theory of planned behavior: Assesing the role of identification with"green consumerism". *Social Psychology Quarterly* 55(4): 388–399, doi:10.2307/2786955.

Stafford, T. M. (2015). Indoor air quality and academic performance. *Journal of Environmental Economics and Management* 70: 34–50, doi:https://doi.org/10.1016/j.jeem.2014.11.002.

Statistics Canada (2016). The daily. education in canada: Key results from the 2016 census. Available at: `https://www150.statcan.gc.ca/n1/dailyquotidien/171129/dq171129a-eng.htm?indid=14431-3&indgeo=0`. Accessed on February 8, 2021.

Statistics Canada (2020a). The daily. Available at: `https://www150.statcan.gc.ca/n1/daily-quotidien/200929/dq200929b-eng.htm`. Accessed on February 8, 2021.

Statistics Canada (2020b). Estimates of population as of july 1st, by marital status or legal marital status, age and sex. Available at: `https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710006001`. Accessed on February 8, 2021.

Stavins, R. N. (2008). Addressing climate change with a comprehensive us cap-and-trade system. *Oxford Review of Economic Policy* 24(4): 298–321, doi:10.1093/oxrep/grn017.

Steg, L. and Vlek, C. (2009). Encouraging pro-environmental behaviour: An integrative review and research agenda. *Journal of Environmental Psychology* 29(3): 309–317, doi:10.1016/j.jenvp.2008.10.004.

Stern, N. (2008). The economics of climate change. *American Economic Review* 98: 1–37, doi:10.1257/aer.98.2.1.

Stewart, N., Chater, N. and Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology* 53: 1–26.

Sun, X., Zhao, T., Liu, D., Gong, S., Xu, J. and Ma, X. (2020). Quantifying the Influences of PM2.5 and Relative Humidity on Change of Atmospheric Visibility over Recent Winters in an Urban Area of East China. *Atmosphere* 11, doi:10.3390/atmos11050461.

von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Thaler, R. H. and Sunstein, C. R. (2009). *Nudge: Improving Decisions about Health, Wealth and Happiness*. Penguin.

The World Bank Group (2020). Population growth (annual `https://data.worldbank.org/indicator/SP.POP.GROW`. Accessed on September 9, 2020.

Thøgersen, J. (1999). Spillover processes in the development of a sustainable consumption pattern. *Journal of Economic Psychology* 20: 53–81, doi:10.1016/S0167-4870(98)00043-9.

Thøgersen, J. (2004). A cognitive dissonance interpretation of consistencies and inconsistencies in environmentally responsible behavior. *Journal of Environmental Psychology* 24: 93–103, doi:10.1016/S0272-4944(03)00039-2.

Thøgersen, J. and Crompton, T. (2009). Simple and painless? the limitations of spillover in environmental campaigning. *Journal of Consumer Policy* 32: 141–163, doi:10.1007/s10603-009-9101-1.

Thøgersen, J. and Noblet, C. (2012). Does green consumerism increase the acceptance of wind power? *Energy Policy* 51: 854–862, doi:10.1016/j.enpol.2012.09.044.

Thøgersen, J. and Ölander, F. (2003). Spillover of environment-friendly consumer behaviour. *Journal of Environmental Psychology* 23: 225–236, doi:10.1016/S0272-4944(03)00018-5.

Thomas, G. O., Poortinga, W. and Sautkina, E. (2016). The welsh single-use carrier bag charge and behavioural spillover. *Journal of Environmental Psychology* 47: 126–135, doi:10.1016/j.jenvp.2016.05.008.

Tiefenbeck, V., Staake, T., Roth, K. and Sachs, O. (2013). For better or for worse? empirical evidence of moral licensing in a behavioral energy conservation campaign. *Energy Policy* 57: 160–171, doi:10.1016/j.enpol.2013.01.021.

Tijs, M. S., Karremans, J. C., Veling, H., de Lange, M. A., van Meegeren, P. and Lion, R. (2017). Saving water to save the environment: contrasting the effectiveness of environmental and monetary appeals in a residential water saving intervention. *Social Influence* 12: 69–79, doi:10.1080/15534510.2017.1333967.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrics* 26(1): 24–36, doi:10.2307/1907382.

Truelove, H., Carrico, A., Weber, E., Raimi, K. and Vandenbergh, M. (2014). Positive and negative spillover of pro-environmental behavior: an integrative review and theoretical framework. *Global Environmental Change* 29: 127–138, doi:10.1016/j.gloenvcha.2014.09.004.

Truelove, H., Yeung, K. L., Carrico, A., Gillis, A. J. and Raimi, K. (2016). From plastic bottle recycling to policy support: An experimental test of pro-environmental spillover. *Journal of Environmental Psychology* 46: 55–66, doi:10.1016/j.jenvp.2016.03.004.

Trushna, T., Dhiman, V., Raj, D. and Tiwari, R. R. (2020). Effects of ambient air pollution on psychological stress and anxiety disorder: a systematic review and meta-analysis of epidemiological evidence. *Reviews on Environmental Health* doi:doi:10.1515/reveh-2020-0125.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science* 185: 1124–1131.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science* 211: 453–458.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5: 297–323.

U.S. Census Bureau (2019a). Census bureau releases 2020 demographic analysis estimates. Available at: `https://www.census.gov/newsroom/pressreleases/2020/2020-demographic-analysis-estimates.html`. Accessed on Februray 8, 2021.

U.S. Census Bureau (2019b). Population estimates 2019. bachelor's degree or higher, percent of persons age 25 years+, 2015-2019. Available at: `https://www.census.gov/quickfacts/fact/table/US/PST045219`. Accessed on December 28, 2020.

U.S. Census Bureau (2019c). Population estimates 2019. in civilian labor force, total, percent of population age 16 years+, 2015-2019. Available at: `https://www.census.gov/quickfacts/fact/table/US/PST045219`. Accessed on December 28, 2020.

U.S. Energy Information Administration (2020). Annual energy outlook 2020. Washington, DC.

Valentino, N. A., Hutchings, V. L., Banks, A. J. and Davis, A. K. (2008). Is a Worried Citizen a Good Citizen? Emotions, Political Information Seeking, and Learning via the Internet. *Political Psychology* 29: 247–273.

Velde, L. V. de, Verbeke, W., Popp, M. and Huylenbroeck, G. V. (2010). The importance of message framing for providing information about sustainability and environmental aspects of energy. *Energy Policy* 38: 5541–5549, doi: 10.1016/j.enpol.2010.04.053.

Viscusi, W. K., Huber, J. and Bell, J. (2014). Private recycling values, social norms, and legal rules. *Revue d'économie politique* 12: 159–178.

Waichman, I., Requate, T., Karde, M. and Milinski, M. (2021). Challenging conventional wisdom: Experimental evidence on heterogeneity and coordination in avoiding a collective catastrophic event. *Journal of Environmental Economics and Management* 109: 102502, doi:10.1016/j.jeem.2021.102502.

Ward, G. (2015). Is Happiness a Predictor of Election Results? CEP Discussion Papers, Centre for Economic Performance, LSE.

Ward, G., Neve, J. D., Ungar, L. and Eichstaedt, J. (2020). (Un)happiness and voting in U.S. presidential elections. *Journal of Personality and Social Psychology* 120: 370–383.

Werff, E. van der, Steg, L. and Keizer, K. (2014). Follow the signal: When past pro-environmental actions signal who you are. *Journal of Environmental Psychology* 40: 273–282, doi:10.1016/j.jenvp.2014.07.004.

Weuve, J., Puett, R. C., Schwartz, J., Yanosky, J. D., Laden, F. and Grodstein, F. (2012). Exposure to particulate air pollution and cognitive decline in older women. *Archives of Internal Medicine* 172: 219–227, doi:10.1001/archinternmed.2011.683.

White, K., Macdonnell, R. and Dahl, D. W. (2011). It's the mind-set that matters: the role of construal level and message framing in influencing consumer efficacy and conservation behaviors. *Journal of Marketing Research* 48: 472–485, doi:10.1509/jmkr.48.3.472.

Willy, D. K. and Holm-Müller, K. (2013). Social influence and collective action effects on farm level soil conservation effort in rural kenya. *Ecological Economics* 90: 94–103, doi:10.1016/j.ecolecon.2013.03.008.

Winett, R. A., Hatcher, J. W., Fort, T. R., Leckliter, I. N., Love, S. Q., Riley, A. W. and Fishback, J. F. (1982). The effects of videotape modeling and daily feedback on residential electricity conservation home temperature and humidity, perceived comfort and clothing worn: winter and summer. *Journal of Applied Behavior Analysis* 15: 381–402, doi:10.1901/jaba.1982.15-381.

Winett, R. A., Neale, R. A. and Grier, H. C. (1979). Effects of self-monitoring and feedback on residential electricity consumption. *Journal of Applied Behavior Analysis* 12: 173–184, doi:10.1901/jaba.1979.12-173.

Wing, I. S., Cian, E. D. and Mistry, M. N. (2021). Global vulnerability of crop yields to climate change. *Journal of Environmental Economics and Management* 109: 102462, doi:10.1016/j.jeem.2021.102462.

WMO (2020). State of global climate 2020. WMO-No. 1264.

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review* 93(2): 133–138, doi:10.1257/000282803321946930.

World Health Organization (2005). Air Quality Guidelines. Global Update 2005.

Zhang, X., Chen, X. and Zhang, X. (2018). The impact of exposure to air pollution on cognitive performance. *Proceedings of the National Academy of Sciences* 115: 9193–9197, doi:10.1073/pnas.1809474115.

Zhang, X., Zhang, X. and Chen, X. (2017). Happiness in the air: How does a dirty sky affect mental health and subjective well-being? *Journal of Environmental Economics and Management* 85: 81–94, doi:10.1016/j.jeem.2017.04.00.

Zheng, S., Wang, J., Sun, C., Zhang, X. and Kahn, M. E. (2019). Air pollution lowers Chinese urbanites' expressed happiness on social media. *Nature Human Behaviour* 3: 237–243, doi:10.1038/s41562-018-0521-2.

Ziegler, A. (2021). New ecological paradigm meets behavioral economics: On the relationship between environmental values and economic preferences. *Journal of*

Environmental Economics and Management 109: 102516, doi:10.1016/j.jeem.2021.102516.

A Appendix Chapter 2

A.1 Experiment

A.1.1 Recruiting and pre-game procedures

Participants were recruited through fliers posted in all main buildings of Trinity College Dublin's campus, as well as advertisement in some large undergraduate classes in the Departments of Economics and Social Sciences. The fliers reported the following text:

> With a computerised multi-rounds experiment, we are studying how people behave when they have to coordinate with other individuals and no communication is allowed;

together with some basic information on the duration and times of the sessions, the maximum possible payment and the email of the experimenter. The same pieces of information were given to students during the presentation in the undergraduate classes. Hence, perspective participants were ignorant regarding the actual purpose of the experiment.

People who were interested in participating were asked to fill in a Doodle form to express their preference on the time slots. They were told they could select how many slots they wanted, and they would then be assigned to one of them depending on the experimenter's needs and the availability of other perspective participants. After assigning people to a certain slot, the experimenter would contact them individually to verify they were still available for that date and time, and, in case they were not, to reschedule to a more suitable slot. In addition, periodic reminders were sent to participants as the agreed-upon date approached.

In the main corpus of this article it was said that participants were randomly assigned to one of the two experimental versions. This was done through randomization of the order of sessions done by the experimenter prior to the launch of the experiment. Specifically, a list with an arbitrarily big number of sessions was generated, going from session 1 to session N. The order of the session in the list was then randomly reshuffled. In parallel, and equally numerous list of session versions was generated, containing $N/2$ donation sessions and $N/2$ control sessions. Also in this case, the list was randomly reshuffled. Finally, the elements of the two reshuffles lists were matched in cardinal order, namely, the first element with the first element, the second with the second, and so on. Hence, although players could decide in which slot to participate, they had no control over the experimental version that was assigned to that slot. Therefore, their participation in the donation or control version was as good as randomly determined.

The experimental sessions were conducted in the Psychology Lab at Trinity College Dublin from February 2019 to February 2020. Upon arrival, participants were assigned to a computer station and were advised to take their decision independently and to avoid any sort of communication with other players. An experimenter was surveilling the room to prevent illicit behaviours, to answer participants' questions and to fix any technical issue that might emerge.

The experiment was designed to be conducted with multiples of 5, since the groups in the public goods game consist of 5 members. For this reason, sessions consisted of 10 or 15 people.

At the beginning of each session, participants were presented with a brief introduction detailing the structure of the experiment. They were also told that their final payoff would be derived base on 2 round randomly selected by the experimental software at the end of the PGs game and that it would be paid to the in the form of a One4All voucher.

A.1.2 Donation-First version

Below are reported the brief introduction and the instructions of the PGs game participants were presented in the donation version of the experiment. The instruction were displayed on the computer screens and read aloud by the experimented. Participants could ask questions in case anything was unclear. They would see the instructions a second time, after the 2-rounds practice session and before the 20-rounds PGs game.

Introduction

Thank you for agreeing to take part in this experiment conducted by Trinity researchers. This project provides you an opportunity to earn some money: an €5 show-up fee plus what you earn during the experiment, which will be paid to you in the form of a One4All voucher.

So, you should be careful to follow directions and make good decisions. Therefore, it is important for you (and for our research!) that you take your time to understand the instructions.

This research has NO commercial purpose. Your answers, your details and opinions are 100% CONFIDENTIAL and will never be given to a third party. Your participation in this survey is entirely voluntary and you may withdraw at any time and for any reason without penalty. In the extremely unlikely event that illicit activity is reported during this study, the lead researcher will be obliged to report it to the appropriate authorities. Please do not name third parties in any open text field of the experiment. Any such replies will be anonymised.

It is important that you follow the instructions for the duration of the experiment. All your decisions have to be taken individually, so please do not communicate with the other participants for the duration of the experiment. If you have any questions, please raise your hand and we will answer them. Instructions will be provided on your computer and will be read aloud by the supervisor. You will make your decisions using the computer workstation, which will also provide you with feedback about the outcomes of those decisions.

Throughout the experiment we will use tokens rather than Euros. At the end of the experiment your token earnings will be converted to Euros at an exchange rate of 1 token = 3.25 cents (approximately 31 tokens = 1 Euro).

Today's experiment consists of a coordination game comprised of multiple Rounds. You will receive the instructions for the coordination game on your computer prior to its beginning. Before the game starts, you will be asked whether you want to complete a certain action. This will also be explained to you in a moment. After you have completed the coordination game, you will be asked to provide some general information about yourself. Again, please remember that your answers are 100% confidential and will not be shared with any third party.

At the end of the experiment the computer will randomly choose 2 out of the multiple Rounds and you will be paid (tokens*0.0325) Euros based on your payoffs in those two Rounds. Therefore, you should seek to maximize your tokens in each round.

Coordination game instructions

At the beginning of each Round of the coordination game, you will be matched with four other people, randomly selected from the participants in the room. You and the people you are matched with will form a Group. Your earnings will depend on your decisions and on the decisions of the other members of your Group. Because the composition of Groups is randomly determined at the beginning of each Round, the identity of the people you are matched with will change from Round to Round. Otherwise, each set of Rounds is identical (that is, all Groups see the same Rounds in the same order).

At the beginning of each Round you will receive 100 tokens. We will refer to that amount as your 'Endowment'. Your tokens do not carry over between Rounds; that is, you will always start off with an endowment of 100 tokens in every Round.

Your task is to decide how many tokens from your Endowment you would like to allocate to your Private Account and how many to a Public Account. Your decision is made anonymously and independently: no other participant can associate you with your decision and you must take this decision on your own. Likewise, the other four Group members can allocate tokens to their Private Accounts and to the Public Account.

Specifically, on the decision screen on the computer for each Round you are asked: "How many of your 100 tokens do you want to contribute to the Public Account?" In the input box, type in the number of tokens you want to contribute: any number between 0 and 100, both inclusive. You can change your mind any time prior to clicking the "Next" button. When you are satisfied with your choice, click "Next" and a new page will appear showing your result for that Round. Any tokens you do not place in the Public Account are placed in your Private Account. The number of tokens in your Private Account belong solely to you whereas the tokens in the Public Account will be shared by all Group members. The tokens you put in the Public Account are no longer available for the Private Account.

After all Group members have clicked the "Next" button, the computer software will calculate the total Group contribution to the Public Account in that Round. Let us call this number M. All 5 members of your Group, even those who did not put any tokens in the Public Account, will earn additional tokens based on the total Group contribution. Specifically, each member will earn $(0.4 * M)$ tokens from the Public Account. In addition, each member will also get the tokens allocated to the Private Account, and the payoff for the Round is the sum of the two.

So, your total Payoff for each Round is the sum of two items:

- The number of tokens that remain in your Private Account: $100 - c$, where c is the number of tokens that you allocated to the Public Account;

- The earnings from the Public Account: You and every other member of your Group earns an additional $(0.4 * M)$ tokens; that is, for each token that is allocated to the Group Account (by you or by any of the other members of your Group) every Group member receives 0.4 tokens.

In the coordination game there are two types of Public Accounts: Generic and Environmental. The two typologies are identical in terms of earnings: irrespective of your personal preferences, both generate an earning of $(0.4 * M)$ tokens for each Group member. They only differ in terms of the services they can provide.

Think about the Public Account's earnings $(0.4 * M)$ as a service the Public Account can generate: the money (tokens) contributed to the Public Account are used to create a service which provides benefits (earnings) to the whole society (Group). So, even if the benefits (earnings) are the same for the two types, the services that generate them are different.

- **Environmental Public Account.** Goods and services provided may include, but are not limited to, national parks, oceans and shores' clean-up, waste recycling, afforestation, greenhouse gas emissions' reductions, renewable electricity power plants, and so on.
- **Generic Public Account.** Goods and services provided may include, but are not limited to, street lighting, public radio and television broadcasts, national defense, research on diseases' treatments, the Red Cross, and so on.

The coordination game consists of 20 Rounds. And each Round consists of the decision of how many tokens to allocate to a Public Account (either one of the two types). So, in every Round, you will be presented with either an Environmental Public Account or a Generic Public Account, and you have to decide how many of your 100 tokens to put in that Public Account, while the rest remains in your Private Account. The order in which the two types of Public Accounts are presented is randomly determined by the software at the beginning of the experimental session. However, it is the same for all Groups.

Also, remember that the composition of the Groups is randomly determined at the beginning of each Round.

After Rounds 7 and 14 you will be presented with an unrelated task, which will give you a bit of a break from the coordination game. These tasks will not count towards your final Payoff.

At the end of each Round your computer screen will report back to you:

- The Endowment;
- The typology of Public Account;
- The number of tokens you contributed to the Public Account, c;
- The total Group contribution to the Public Account (including you), M;
- Your earnings from the Public Account, $0.4 * M$;
- The total Payoff (in tokens) for the Round, as well as the fractions you keep for yourself and you donate.

At the end of the coordination game, 2 Rounds will be randomly selected for payment, which is given by the sum of the Payoffs from each of them. Finally, assuming earlier in this experiment you have decided to donate €2 to the WWF, the part of the final Payoff you keep for yourself is: $FinalPayoff -$ €2.

A.1.3 Donation-After version

Below are reported the brief introduction and the instructions of the PGs game participants were presented in the control version of the experiment. The instruction were displayed on the computer screens and read aloud by the experimented. Participants could ask questions in case anything was unclear. They would see the instructions a second time, after the 2-rounds practice session and before the 20-rounds PGs game.

Introduction

Thank you for agreeing to take part in this experiment conducted by Trinity researchers. This project provides you an opportunity to earn some money: an €5 show-up fee plus what you earn during the experiment, which will be paid to you in the form of a One4All voucher.

So, you should be careful to follow directions and make good decisions. Therefore, it is important for you (and for our research!) that you take your time to understand the instructions.

This research has NO commercial purpose. Your answers, your details and opinions are 100% CONFIDENTIAL and will never be given to a third party. Your participation in this survey is entirely voluntary and you may withdraw at any time and for any reason without penalty. In the extremely unlikely event that illicit activity is reported during this study, the lead researcher will be obliged to report it to the appropriate authorities. Please do not name third parties in any open text field of the experiment. Any such replies will be anonymised.

It is important that you follow the instructions for the duration of the experiment. All your decisions have to be taken individually, so please do not communicate with the other participants for the duration of the experiment. If you have any questions, please raise your hand and we will answer them. Instructions will be provided on your computer and will be read aloud by the supervisor. You will make your decisions using the computer workstation, which will also provide you with feedback about the outcomes of those decisions.

Throughout the experiment we will use tokens rather than Euros. At the end of the experiment your token earnings will be converted to Euros at an exchange rate of 1 token = 3.25 cents (approximately 31 tokens = 1 Euro).

Today's experiment consists of a coordination game comprised of multiple Rounds. You will receive the instructions for the coordination game on your computer prior to its beginning. After you have completed the coordination game, you will be asked to provide some general information about yourself. Again, please remember that your answers are 100% confidential and will not be shared with any third party.

At the end of the experiment the computer will randomly choose 2 out of the multiple Rounds and you will be paid (tokens*0.0325) Euros based on your payoffs in those two Rounds. Therefore, you should seek to maximize your tokens in each round.

Coordination game instructions

At the beginning of each Round of the coordination game, you will be matched with four other people, randomly selected from the participants in the room. You and the people you are matched with will form a Group. Your earnings will depend on your decisions and on the decisions of the other members of your Group. Because the composition of Groups is randomly determined at the beginning of each Round, the identity of the people you are matched with will change from Round to Round. Otherwise, each set of Rounds is identical (that is, all Groups see the same Rounds in the same order).

At the beginning of each Round you will receive 100 tokens. We will refer to that amount as your ’Endowment’. Your tokens do not carry over between Rounds; that is, you will always start off with an endowment of 100 tokens in every Round.

Your task is to decide how many tokens from your Endowment you would like to allocate to your Private Account and how many to a Public Account. Your decision is made anonymously and independently: no other participant can associate you with your decision and you must take this decision on your own. Likewise, the other four Group members can allocate tokens to their Private Accounts and to the Public Account.

Specifically, on the decision screen on the computer for each Round you are asked: "How many of your 100 tokens do you want to contribute to the Public Account?" In the input box, type in the number of tokens you want to contribute: any number between 0 and 100, both inclusive. You can change your mind any time prior to clicking the "Next" button. When you are satisfied with your choice, click "Next" and a new page will appear showing your result for that Round. Any tokens you do not place in the Public Account are placed in your Private Account. The number of tokens in your Private Account belong solely to you whereas the tokens in the Public Account will be shared by all Group members. The tokens you put in the Public Account are no longer available for the Private Account.

After all Group members have clicked the "Next" button, the computer software will calculate the total Group contribution to the Public Account in that Round. Let us call this number M. All 5 members of your Group, even those who did not put any tokens in the Public Account, will earn additional tokens based on the total Group contribution. Specifically, each member will earn $(0.4 * M)$ tokens from the Public Account. In addition, each member will also get the tokens allocated to the Private Account, and the payoff for the Round is the sum of the two.

So, your total Payoff for each Round is the sum of two items:

- The number of tokens that remain in your Private Account: $100 - c$, where c is the number of tokens that you allocated to the Public Account;
- The earnings from the Public Account: You and every other member of your Group earns an additional $(0.4 * M)$ tokens; that is, for each token that is allocated to the Group Account (by you or by any of the other members of your Group) every Group member receives 0.4 tokens.

In the coordination game there are two types of Public Accounts: Generic and Environmental. The two typologies are identical in terms of earnings: irrespective of your personal preferences, both generate an earning of $(0.4 * M)$ tokens for each Group member. They only differ in terms of the services they can provide.

Think about the Public Account's earnings $(0.4 * M)$ as a service the Public Account can generate: the money (tokens) contributed to the Public Account are used to create a service which provides benefits (earnings) to the whole society (Group). So, even if the benefits (earnings) are the same for the two types, the services that generate them are different.

- **Environmental Public Account.** Goods and services provided may include, but are not limited to, national parks, oceans and shores' clean-up, waste recycling, afforestation, greenhouse gas emissions' reductions, renewable electricity power plants, and so on.

- **Generic Public Account.** Goods and services provided may include, but are not limited to, street lighting, public radio and television broadcasts, national defense, research on diseases' treatments, the Red Cross, and so on.

The coordination game consists of 20 Rounds. And each Round consists of the decision of how many tokens to allocate to a Public Account (either one of the two types). So, in every Round, you will be presented with either an Environmental Public Account or a Generic Public Account, and you have to decide how many of your 100 tokens to put in that Public Account, while the rest remains in your Private Account. The order in which the two types of Public Accounts are presented is randomly determined by the software at the beginning of the experimental session. However, it is the same for all Groups.

Also, remember that the composition of the Groups is randomly determined at the beginning of each Round.

After Rounds 7 and 14 you will be presented with an unrelated task, which will give you a bit of a break from the coordination game. These tasks will not count towards your final Payoff.

At the end of each Round your computer screen will report back to you:

- The Endowment;
- The typology of Public Account;
- The number of tokens you contributed to the Public Account, c;
- The total Group contribution to the Public Account (including you), M;
- Your earnings from the Public Account, $0.4 * M$;

- The total Payoff (in tokens) for the Round.

At the end of the coordination game, 2 Rounds will be randomly selected for payment, which is given by the sum of the Payoffs from each of them.

A.2 Demographics

Below are reported the questions included in the survey at the end of the experiment (and that were used in the analysis).

- What age are you? Answer options: [Under 18; 18-24; 25-34; 35-44; 45-54; 55-64; 65-74; 75-84; 85 or older; Prefer not to say].
- What is the highest level of school you have completed or the highest degree you have received? Answer options: [No formal education; Primary school; Secondary school; Some college but no degree; Lower degree (certificate or diploma); Higher degree; Master's degree; Doctoral degree; Other (please specify); Prefer not to say].
- What is your gender? Answer options: [Male; Female; Prefer not to say].
- What is your employment status? Answer options: [Self-employed; Employed; House persons and carers; Unemployed; Retired; Student; Unable to work (e.g. disability); Other (please specify); Prefer not to say]
- What is your marital status? Answer options: [Married; Widowed; Divorced; Separated; Single; Never married; Other (please specify); Prefer not to state].
- Which of the following best describes your living situation? Answer options: [Living with my spouse/partner (with or without children); Living with my parents or other relatives; Living alone; Sharing a property with non-family members; Single parent; Prefer not to say].
- Which discipline most closely aligns with the content of your degree? (If you have more than one degree in different disciplines, please specify all disciplines in the "Other" option). Answer options: [Economics/Business; Arts; Humanities; Social

Sciences; Natural Sciences; Computer Science; Engineering/ Mathematics; Health Sciences; Other (please specify)]

- Had you ever participated to an economic experiment before? If yes, please specify the kind of experiment (e.g. public goods game, dictator game, etc.).

- How would you describe your current income situation? (If you are married or in a domestic partnership consider your combined income). Answer options: [Finding it very difficult to live on current income; Finding it difficult to live on current income; Coping on current income; Living comfortably on current income; Living very comfortably on current income; Prefer not to state].

- Here are a number of pro-environmental statements that may or may not apply to you. Please, state how much you agree/disagree with the following statements by selecting the appropriate answer from the drop down menus below Answer options: [Strongly disagree; Somewhat disagree; Neither agree nor disagree; Somewhat agree; Strongly agree].
  - I think of myself as someone who is very concerned with environmental issues. (Adapted from Sparks and Shepherd (1992).)
  - I feel I have to limit my energy consumption to protect the environment.
  - I feel I have to limit my environmental impact.
  - I feel I have to do whatever I can to help improve the environment. (Adapted from Minton and Rose (1997).)
  - Natural resources must be preserved even if people must do without some products. (Adapted from Minton and Rose (1997).)
  - Doing pro-environmental deeds requires too much time/effort. (Adapted from Della Giusta et al. (2012). Reverse scale.)
  - A single individual's actions are irrelevant in affecting the environment. (Reverse scale.)

- Here are a number of climate change statements. Please, state how much you agree/disagree with the following statements by selecting the appropriate answer from the drop down menus below. Answer options: [Strongly disagree; Somewhat disagree; Neither agree nor disagree; Somewhat agree; Strongly agree].

  – Climate change is happening (Adapted from Leiserowitz et al. (2013).)
  – Much more fuss is being made about climate change than is really justified. (Adapted from Minton and Rose (1997). Reverse scale.)
  – Climate change will be a major threat already in the next 30 years. (Adapted from Della Giusta et al. (2012).)
  – Climate change will only become a serious threat in more than 30 years from now. (Adapted from Della Giusta et al. (2012). Reverse scale.)
  – Climate change is mainly due to human activity. (Adapted from Leiserowitz et al. (2013).)
  – Science will be able to solve climate change and environmental problems without individuals having to make big changes in their lives. (Adapted from Leiserowitz et al. (2013). Reverse scale.)

- Are you generally an impatient person, or someone who always shows great patience? Please rate yourself on a scale of 1 to 10, where 1 represents 'very impatient' and 10 is 'very patient'.
- Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please rate yourself on a scale of 1 to 10, where 1 represents 'unwilling to take risks' and 10 represents 'fully prepared to take risks'.

A.3 Additional Checks

A.3.1 Descriptives

In this section are reported the results of Levene's tests for the equality of variance in average contribution levels between various groups. Table A.1 compares participants from the two experimental version; while Table A.2 compares participants who did and did not donate. The last three columns of both tables contains different specifications of the test statistic. W0 is the nonnormality robust Levene's test statistic. W50 and W10 are alternative formulations of the Levene's test statistic proposed by Brown and Forsythe (1974), which replace the mean with the median and the 10% trimmed mean, respectively.

From Table A.1 it is possible to see that the Levene's test rejects the null hypothesis of equality of variance in average contribution levels between the two experimental versions. According to the alternative formulations of the test statistic, this rejection can only be made at the 10% level of significance.

Table A.1: Equality of variance of contribution levels between experimental versions

Av.Contr. DA	Av.Contr. DF	Std.Dev. DA	Std.Dev. DF	W0	W50	W10
29.40	28.00	23.42	26.24	7.49***	2.99*	2.92*

Notes: This table reports the mean and standard deviation of contributions of participants in the two experimental versions, and the test statistics from the test of equality of variance. Average contributions and standard deviations are expressed in tokens. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

From Table A.2 it is possible to see the variance of average contribution levels differs between participants who did and did not donate. Irrespective of whether one considers only the donation-first sample (Panel A) or the full sample (Palen B) the Levene's test rejects the null hypothesis of equality of variance at any conventional level of significance and for all formulations of the test statistic.

Table A.2: Equality of variance of Contribution between participants who did and did not donate

Av.Contr. Don	Av.Contr. No Don	Std.Dev. Don	Std.Dev. No Don	W0	W50	W10
A. Donation-first sample						
36.90	22.59	26.89	24.31	20.50***	20.54***	22.73***
B. Full sample						
36.90	25.06	26.89	24.21	18.79***	19.24***	19.51***

Notes: This table reports the mean and standard deviation of contributions of participants who did and did not donate, and the test statistics from the test of equality of variance. In Panel B, the "No Don" group is formed by individuals in the donation treatment who did not make the donation plus all players in the control group. Average contributions and standard deviations are expressed in tokens. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.3.2 Behavioural spillover

Table A.3 replicates the regressions reported in Table 2.3 using ordinary least squares (OLS) instead of Tobit to estimate the effect of having donated to the WWF on contribution levels. The results are comparable to those obtained with Tobit models, albeit slightly lower in magnitude. They confirm the presence of a positive behavioural spillover generated by the initial PEB, and also that this positive spillover interests primarily environmental PGs.

Table A.3: Behavioural spillover - Regressions (OLS)

	Contribution (1)	Contribution (2)	Contribution (3)	Contribution (4)
Donation	13.03***	8.36**	10.55***	9.68***
	(3.746)	(3.605)	(3.923)	(3.679)
Donation × Type			-6.25*	
			(3.383)	
R^2	0.207	0.277	0.280	0.273
Observations	1640	1640	1640	1640
Controls				
Round FE	✓	✓	✓	✓
Session FE	✓	✓	✓	✓
PG Type		✓	✓	✓
Ind. Charact.		✓	✓	✓
Total ECCA bins				✓

Notes: This table reports the results of OLS regressions of the amount contributed by player i in round t on the set of controls and FEs listed at the bottom. All regressions focus on the donation-first sample. Standard errors, clustered at the participant level, are reported in parenthesis. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

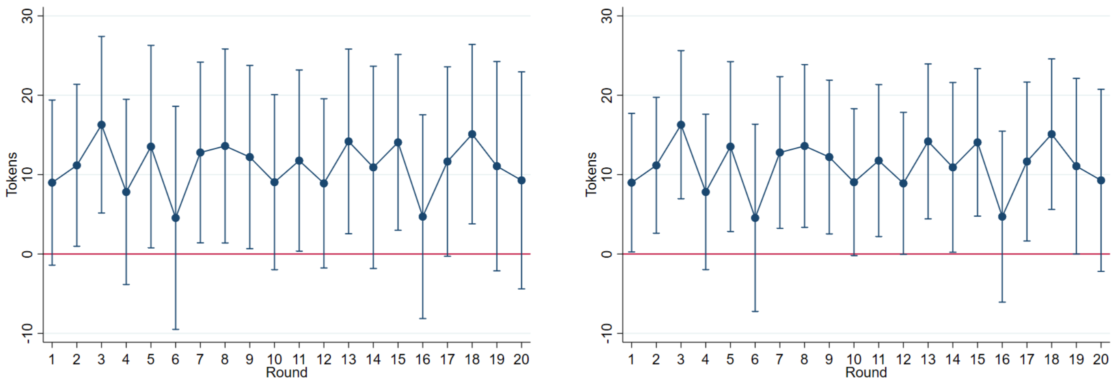
A.3.3 Persistence

In this section are reported two marginal effects graphs that complement the analysis of the persistence of the behavioural spillover presented in Section 2.5.3. The vector of round dummies, $\boldsymbol{Round'_t}$, of Equation 2.3 now considers all twenty rounds individually rather than dividing them into blocks of five. Panel (a) of Figure A.1 displays the 95% confidence intervals, while Panel (b) the 90% ones.

The graphs confirms the general trend represented in Figure 2.3 that the behavioural spillover generated by having made the donation is rather persistent. The magnitude of the marginal effects goes from a minimum of 4.55 (in round 6) to a maximum of 16.26 (in round 3), but most of them fluctuate in the 10 to 14 range, meaning that participants who donated tend to contribute between 10 and 14 tokens more than those who did not donate. This magnitude is comparable to the coefficients reported in Table 2.3.

Ten out if the 20 marginal effects are significant at the 95% level, with an additional 4 being significant at the 90% level — the marginal effects significant at the 90% level are those corresponding to rounds one, fourteen, seventeen and nineteen. In general, those that are not statistically different from zero are those with a lower magnitude. With all this considered, there seems to be enough evidence to establish that the behavioural spillover generated by the initial PEB does not fade out after a few subsequent actions but it tends to have long lasting effects.

Figure A.1: Persistence of the behavioural spillover effect



(a) 95% CI

(b) 90% CI

Notes: This graph displays the marginal effects of having made the donation on contribution levels in each round of the PGs game. Panel (a) reports the 95% confidence intervals, while Panel (b) the 90% confidence intervals. The estimation refers to the donation-first sample.

B Appendix Chapter 3

B.1 Experiment

B.1.1 Attributes and levels

Figures B.1-B.8 display the images with the description of each attribute and their levels that participants were shown at the beginning of the discrete choice experiment. All images are taken from the Irish version of the experiment. Figure B.6 reports the energy efficiency attribute for the control group in all countries.

B.1.2 Choice sets

Table B.1 reports all the 32 choice sets that were used in the experiment and their division into the 4 blocks. Energy efficiency is displayed according to the letter scale of the EU Energy Label. This was appropriately reframed in each specific treatment group and country version following the scheme shown in Table 3.1. As mentioned in Section 3.2.3, each choice set included an opt-out option and there was no dominant alternative. Figure B.9 presents examples of choice sets for all three groups (control, treatment 1 and treatment 2) taken from the Irish version of the experiment.

Table B.1: Full list of choice sets

Block	Choice Set	Alternative	Price	Capacity	Brand	Stars	Efficiency
1	1	1	400	New	7	4	B
1	1	2	1000	New	8	3	A+++

Table B.1 — continued

Block	Choice Set	Alternative	Price	Capacity	Brand	Stars	Efficiency
1	2	1	600	New	7	3	A
1	2	2	800	New	8	4	C
1	3	1	600	New	9	4	A++
1	3	2	400	New	7	3	A+
1	4	1	1200	New	8	4	A+++
1	4	2	600	Established	8	3	A++
1	5	1	800	Established	10	4	B
1	5	2	1200	Established	10	4	A+
1	6	1	400	Established	8	5	A
1	6	2	800	Established	9	3	C
1	7	1	200	New	8	4	A+++
1	7	2	1200	Established	8	5	A+
1	8	1	600	New	7	3	B
1	8	2	800	New	10	4	A+
2	9	1	1000	New	10	5	A++
2	9	2	1200	Established	9	4	A+++
2	10	1	400	New	9	3	A++
2	10	2	1000	New	8	4	B
2	11	1	1200	New	8	5	A++
2	11	2	1000	New	9	5	C
2	12	1	1000	Established	9	4	A
2	12	2	600	New	7	3	A+
2	13	1	200	New	10	4	A++
2	13	2	400	Established	9	3	A+++
2	14	1	600	New	8	4	A++
2	14	2	800	New	10	3	A
2	15	1	200	Established	10	3	A++
2	15	2	600	New	9	5	A+++
2	16	1	400	Established	9	3	A+

Table B.1 — continued

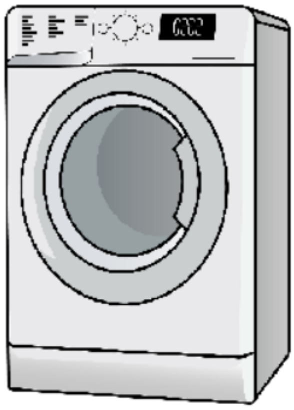
Block	Choice Set	Alternative	Price	Capacity	Brand	Stars	Efficiency
2	16	2	800	Established	7	5	A
3	17	1	1000	Established	7	5	A++
3	17	2	600	Established	8	5	C
3	18	1	800	Established	9	5	A+
3	18	2	1000	Established	8	3	A++
3	19	1	400	New	9	4	A+
3	19	2	600	New	10	5	A+++
3	20	1	200	New	7	5	C
3	20	2	1000	Established	8	4	B
3	21	1	800	New	7	5	A++
3	21	2	400	New	10	5	B
3	22	1	800	Established	7	5	C
3	22	2	1200	New	9	5	B
3	23	1	1200	Established	8	3	A
3	23	2	400	Established	7	4	A+++
3	24	1	1000	Established	9	5	A+
3	24	2	400	Established	8	5	C
4	25	1	400	New	10	5	A++
4	25	2	200	Established	9	5	A
4	26	1	200	Established	7	3	B
4	26	2	400	New	8	4	A
4	27	1	1000	Established	9	5	C
4	27	2	800	Established	7	4	A+
4	28	1	600	Established	10	4	C
4	28	2	200	New	9	5	A++
4	29	1	1200	Established	9	5	A
4	29	2	1000	Established	8	5	A+
4	30	1	800	Established	9	3	B
4	30	2	1000	Established	7	4	A

Table B.1 — continued

Block	Choice Set	Alternative	Price	Capacity	Brand	Stars	Efficiency
4	31	1	1200	Established	7	4	C
4	31	2	800	Established	8	3	A
4	32	1	1200	Established	8	5	B
4	32	2	600	New	9	4	A

Figure B.1: Discrete choice experiment intro

Choose your favourite tumble dryer



Imagine you are looking to buy a new tumble dryer.

In a moment you will be shown a number of tumble dryers. They all fit the space you have, look exactly the same and have the same display features and functions.

Simply look at the characteristics of each, **think about what you can afford** and choose your favourite.

Figure B.2: Price attribute

PRICE (purchase price in €)
- will be one of the following ...

€ 200	€ 400	€ 600	€ 800	€ 1,000	€ 1,200

Figure B.3: Brand attribute

Note: An "Established" brand is one that has been on the market for over five years and has developed a reputation with its customers. A "New" brand is on the market less than five years and hasn't yet built up a reputation with customers.

Figure B.4: Capacity attribute

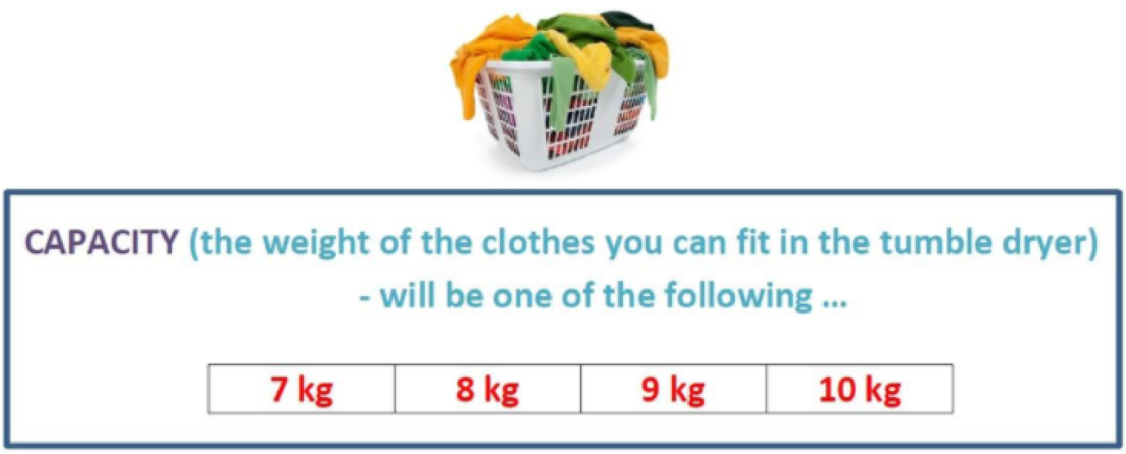
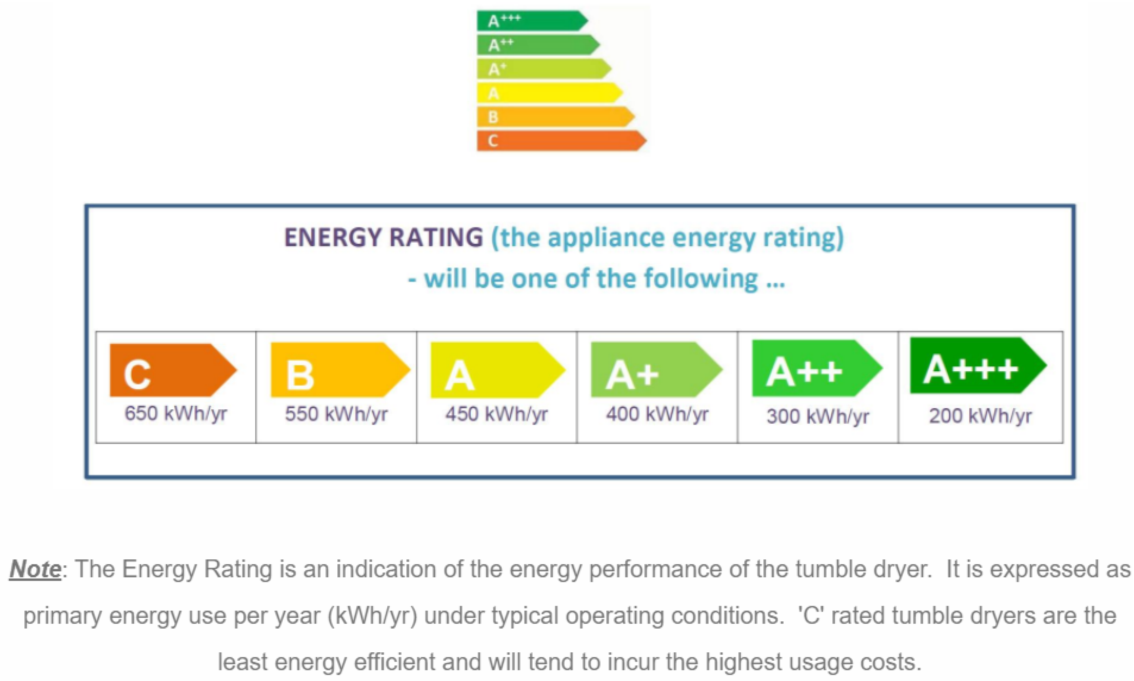


Figure B.5: Customer rating attribute

Figure B.6: Control group energy efficiency attribute

(a) Ireland ond United Kingdom



(b) Canada and United States

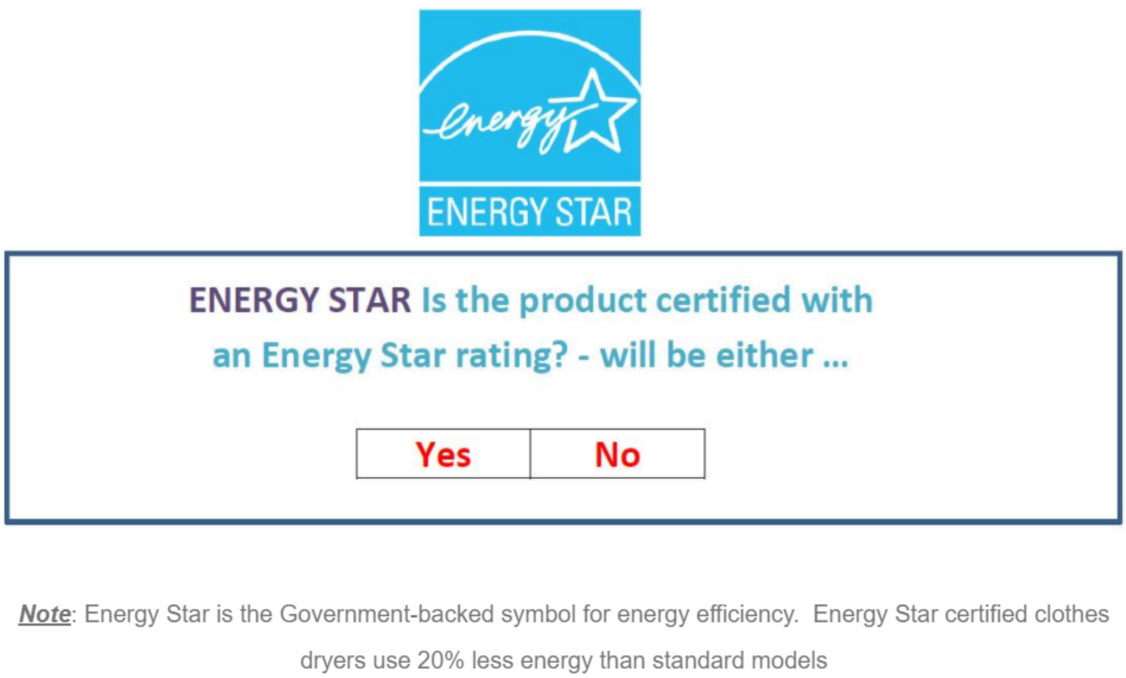


Figure B.7: Treatment 1 energy efficiency attribute

10 YEAR ENERGY COST (the electricity cost associated with using the appliance for 10 years): - will be one of the following …

€350	€500	€650	€800	€950	€1,100

Note: Costs are estimated based on the energy efficiency of the tumble dryer under typical operating conditions and constant electricity prices.

The most efficient tumble dryers consume the least amount of electricity and therefore involve the lowest energy cost.

Figure B.8: Treatment 2 energy efficiency attribute

YOUR 10 YEAR ENERGY COST

(the electricity cost associated with your household using the appliance 2 times per week for 10 years):

- will be one of the following …

€ 234	**€ 328**	**€ 422**	**€ 514**	**€ 608**	**€ 702**

Note: Costs are estimated based on the energy efficiency of the tumble dryer and your answers to previous questions, in particular, **the number of times you stated you use your tumble dryer per week.** These are 10 year costs and assume constant electricity prices. The most efficient tumble dryers consume the least amount of electricity and therefore involve the lowest energy cost.

Figure B.9: Example of choice sets

(a) Control

Which of these two tumble dryers would you choose?

	Tumble dryer 1	Tumble dryer 2
Brand	New	New
Capacity	9 kg	10 kg
Energy rating	A+	A+++
Customer rating	★★★★☆	★★★★★
Price	€ 400	€ 600

	Tumble dryer 1	Tumble dryer 2	Neither
I would choose:	○	○	○

(b) Treatment 1

Which of these two tumble dryers would you choose?

	Tumble dryer 1	Tumble dryer 2
Brand	New	New
Capacity	9 kg	8 kg
10 year Energy Cost	€ 500	€ 950
Customer rating	★★★☆☆	★★★★☆
Price	€ 400	€ 1,000

	Tumble dryer 1	Tumble dryer 2	Neither
I would choose:	○	○	○

(c) Treatment 2

Which of these two tumble dryers would you choose?

	Tumble dryer 1	Tumble dryer 2
Brand	New	Established
Capacity	10 kg	9 kg
Your 10 year Energy Cost	€ 1148	€ 819
Customer rating	★★★★★	★★★★☆
Price	€ 1,000	€ 1,200

	Tumble dryer 1	Tumble dryer 2	Neither
I would choose:	○	○	○

B.2 Demographics

Below are reported the questions included in the survey that accompanied the DCE (and that were used in the analysis).

- What age are you? Answer options: [Under 18; 18-24; 25-34; 35-44; 45-54; 55-64; 65-74; 75-84; Prefer not to say].
- Do you have a tumble dryer in your home? Answer options: [Yes; No].
- Approximately how many times a week do you use your tumble dryer? Enter number.
- What is your gender? Answer options: [Male; Female; Prefer not to say].
- How many people (including yourself) live in your household? Answer options: [1; 2; 3; 4; 5; 6; 7; 8; More than 8; Prefer not to say].
- Which of the following best describes your relationship status? Answer options: [Single (never married); Married or in a domestic partnership; Widowed; Divorced; Separated; Prefer not to say].
- What is your highest level of education? Answer options: [No formal education; Primary school/Elementary school; Secondary school/High school; Lower degree (certificate or diploma)/Associate Degree or Certificate; Higher degree/Bachelor's degree; Masters or higher; Other (please specify); Prefer not to say].
- What is your employment status? Answer options: [Self-employed; Employed; House persons and carers; Unemployed; Retired; Student; Unable to work (e.g. disability); Prefer not to say]
- Which of the following best describes your living situation? Answer options: [Living with my spouse/partner (with or without children); Living with my parents or other relatives; Living alone; Sharing a property with non-family members; Single parent; Prefer not to say].
- How many ADULTS live in your household? Enter number.

- Which of the following best describes the area you live in? Answer options: [Urban; Suburban; Rural; Prefer not to say].
- Suppose you are purchasing a new tumble dryer for your home. Please rate the importance of each of the following characteristics in making your decision on which model to buy. Answer options: [Not at all important; Slightly important; Moderately important; Very important; Don't know].
  - Price
  - Brand
  - Energy efficiency/energy consumption
  - Water consumption
  - Load capacity
  - Dimensions (height, weight, depth)
  - Features (timer, digital display)
  - Aesthetics (colour, design)
- In relation to the energy efficiency of electrical appliances, please state whether you disagree or agree with the following statements. answer options: [Strongly disagree; Slightly disagree; Slightly agree; Strongly agree; Don't know].
  - Buying more energy efficient appliances would reduce my household's environmental impact
  - All new appliances have similar energy efficiency levels
  - More energy efficient appliances are less reliable
  - I am willing to take a chance on new technologies to reduce my energy consumption
  - I am aware of energy prices; that is, the price of fuels such as gas, oil and electricity
  - I understand how much money I would save if I bought a more energy efficient appliance
  - I would like to buy more energy efficient appliances but I cannot afford them

- Please rate how concerned you are about the environment (for example, pollution, global warming and climate change). Answer options: [Not concerned; Slightly concerned; Concerned; Extremely concerned; Don't know].

- How would you describe your current income situation? (If you are married or in a domestic partnership consider your combined income). Answer options: [Finding it very difficult to live on current income; Finding it difficult to live on current income; Coping on current income; Living comfortably on current income; Living very comfortably on current income; Prefer not to state].

- Are you generally an impatient person, or someone who always shows great patience? Please rate yourself on a scale of 1 to 10, where 1 represents 'very impatient' and 10 is 'very patient'.

- Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please rate yourself on a scale of 1 to 10, where 1 represents 'unwilling to take risks' and 10 represents 'fully prepared to take risks'.

B.3 Robustness

In this section are reported the results of our robustness checks to complement the mixed multinomial logit analysis presented in Section **??** of the paper. The approach we follow is threefold. First, we estimate separate models for the three experimental groups in each country. Second, we allow all attributes' coefficients to vary between individuals, apart from the price coefficient, which corresponds to estimating an error component model. Thirdly, we implement the opposite exercise, restricting all coefficients to be constants for all individuals, thus reverting back to the conditional logit model.

B.3.1 Split samples models

The mixed logit model estimated for the control group, the generic information treatment and the personalised information treatment separately present similar patterns to the polled models reported in Table 3.4 in all four countries.

Table B.2: Split samples models - Ireland

	(1) Control	(2) Treatment 1	(3) Treatment 2
Non-Random Parameters in Utility Function			
Constant (neither option)	0.2182	1.0185	0.6450
	(1.8587)	(1.7074)	(1.8173)
Price	-0.0034***	-0.0029***	-0.0030***
	(0.0003)	(0.0003)	(0.0003)
Stars	0.5618***	0.4889***	0.6213***
	(0.0803)	(0.0909)	(0.0816)
Random Parameters in Utility Function			
Capacity	0.2311***	0.3060***	0.1781***
	(0.0491)	(0.0600)	(0.0434)
Brand	-0.1332	-0.2875**	-0.3375***
	(0.1107)	(0.1316)	(0.1101)
Energy efficiency	1.1161***	0.9490***	0.9124***
	(0.1241)	(0.1319)	(0.1187)
Model statistics			
Observations	4,752	4,536	4,656
Clusters	198	189	194

Notes: This table reports the results of mixed logit regressions of respondents' choices in the Irish sample. Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. All regressions control for income, gender, living area, whether the individual holds an degree, environmental concern, impatience, risk attitude and tumble dryer usage. Standard errors are clustered at the participant level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.3: Split samples models - UK

	(1) Control	(2) Treatment 1	(3) Treatment 2
Non-Random Parameters in Utility Function			
Constant (neither option)	0.2548	0.2512	1.0233
	(1.7875)	(1.6404)	(1.5598)
Price	-0.0033***	-0.0033***	-0.0027***
	(0.0003)	(0.0003)	(0.0003)
Stars	0.4932***	0.5691***	0.5999***
	(0.0776)	(0.0851)	(0.0845)
Random Parameters in Utility Function			
Capacity	0.1461***	0.0786*	0.0944**
	(0.0453)	(0.0473)	(0.0479)
Brand	-0.0521	-0.2743**	-0.3180***
	(0.1144)	(0.1330)	(0.1143)
Energy efficiency	0.5331***	0.5345***	0.8739***
	(0.1185)	(0.1236)	(0.1139)
Model statistics			
Observations	5,280	5,232	5,208
Clusters	220	218	217

Notes: This table reports the results of mixed logit regressions of respondents' choices in the UK sample. Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. All regressions control for income, gender, living area, whether the individual holds an degree, environmental concern, impatience, risk attitude and tumble dryer usage. Standard errors are clustered at the participant level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.4: Split samples models - Canada

	(1) Control	(2) Treatment 1	(3) Treatment 2
Non-Random Parameters in Utility Function			
Constant (neither option)	6.3177***	2.4083	-2.3835
	(1.8550)	(1.8314)	(1.8814)
Price	-0.0028***	-0.0031***	-0.0020***
	(0.0002)	(0.0002)	(0.0002)
Stars	1.0001***	0.8187***	0.6732***
	(0.0882)	(0.0820)	(0.0700)
Random Parameters in Utility Function			
Capacity	0.3461***	0.1643***	0.1143**
	(0.0532)	(0.0516)	(0.0482)
Brand	-0.7446***	-0.4648***	-0.2503**
	(0.1210)	(0.1144)	(0.1019)
Energy efficiency	1.7476***	0.9594***	0.7426***
	(0.1436)	(0.1394)	(0.1076)
Model statistics			
Observations	5,136	4,920	5,160
Clusters	214	205	215

Notes: This table reports the results of mixed logit regressions of respondents' choices in the Canadian sample. Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. All regressions control for income, gender, living area, whether the individual holds an degree, environmental concern, impatience, risk attitude and tumble dryer usage. Standard errors are clustered at the participant level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.5: Split samples models - USA

	(1) Control	(2) Treatment 1	(3) Treatment 2
Non-Random Parameters in Utility Function			
Constant (neither option)	2.1495	2.6941**	2.1098
	(1.4845)	(1.2846)	(2.0091)
Price	-0.0034***	-0.0024***	-0.0024***
	(0.0003)	(0.0002)	(0.0002)
Stars	0.8890***	0.7339***	0.6353***
	(0.0913)	(0.0742)	(0.0782)
Random Parameters in Utility Function			
Capacity	0.3489***	0.1902***	0.1987***
	(0.0485)	(0.0469)	(0.0487)
Brand	-0.4426***	-0.1392	0.1071
	(0.1108)	(0.1081)	(0.1183)
Energy efficiency	1.1864***	0.6593***	0.4005***
	(0.1328)	(0.1124)	(0.1100)
Model statistics			
Observations	4,992	5,472	5,304
Clusters	208	228	221

Notes: This table reports the results of mixed logit regressions of respondents' choices in the USA sample. Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. All regressions control for income, gender, living area, whether the individual holds an degree, environmental concern, impatience, risk attitude and tumble dryer usage. Standard errors are clustered at the participant level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.3.2 Error component models

The error component models do not present any meaningful difference with respect to the mixed logit models reported in the main corpus of the paper, both in terms of coefficients and WTP. All attributes retain the same effect on individuals' utility. Energy efficiency remains the most valuable attribute in all samples, followed by customer rating. Presenting energy information in monetary terms does not significantly increase WTP in Ireland and the United States, while it reduces it in Canada; personalised energy costs lead to a marginally significant improvement in the United Kingdom.

Table B.6: Error component models

	(1) Ireland	(2) UK	(3) Canada	(4) USA
Non-Random Parameters in Utility Function				
Price	-0.0031***	-0.0031***	-0.0026***	-0.0027***
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
Random Parameters in Utility Function				
Constant (neither option)	0.2831	0.9386	0.8568	2.2544**
	(1.2345)	(0.9836)	(1.3095)	(1.0058)
Stars	0.5621***	0.5366***	0.8144***	0.7300***
	(0.0514)	(0.0470)	(0.0476)	(0.0493)
Capacity	0.2598***	0.1403***	0.2253***	0.2662***
	(0.0290)	(0.0264)	(0.0292)	(0.0280)
Brand	-0.2609***	-0.2243***	-0.4683***	-0.1419**
	(0.0700)	(0.0693)	(0.0651)	(0.0655)
Energy efficiency	1.0910***	0.5647***	1.4621***	0.8528***
	(0.1112)	(0.1029)	(0.1155)	(0.1034)
EE $\times$ T1	-0.0796	0.0018	-0.4453***	-0.1429
	(0.1500)	(0.1398)	(0.1486)	(0.1416)
EE $\times$ T2	-0.1264	0.2693*	-0.4962***	-0.2185
	(0.1556)	(0.1384)	(0.1451)	(0.1381)
Model statistics				
Observations	13,944	15,720	15,216	15,768
Clusters	581	655	634	657

Notes: This table reports the results of error component models of respondents' choices in each country separately. Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. All regressions control for income, gender, living area, whether the individual holds an degree, environmental concern, impatience, risk attitude and tumble dryer usage. Standard errors are clustered at the participant level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.7: Error components models willingness to pay

	Ireland	UK	Canada	USA
Stars	178.60	170.44	309.26	266.01
	[148.25 ; 208.96]	[142.18 ; 198.71]	[271.47 ; 347.05]	[231.57 ; 300.46]
Capacity	82.57	44.58	85.55	96.99
	[63.81 ; 101.32]	[27.65 ; 61.50]	[63.30 ; 107.79]	[76.19 ; 117.79]
Brand	-82.89	-71.25	-177.83	-51.69
	[-125.13 ; -40.65]	[-113.66 ; -28.84]	[-225.00 ; -130.67]	[-97.71 ; -5.67]
EE	346.66	179.36	555.18	310.74
	[274.64 ; 418.67]	[114.91 ; 243.82]	[464.35 ; 646.01]	[236.60 ; 384.87]
EE × T1	-25.28	0.57	-169.08	-52.08
	[-118.88 ; 68.32]	[-86.46 ; 87.60]	[-280.18 ; -57.98]	[-153.07 ; 48.91]
EE × T2	-40.16	85.55	-188.42	-79.63
	[-136.94 ; 56.62]	[-0.55 ; 171.65]	[-297.15 ; -79.69]	[-178.28 ; 19.01]

Notes: This table reports the willingness to pay of respondents in each country for the tumble dryer's attributes. Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. The 95% confidence intervals are reported in brackets.

B.3.3 Conditional logit models

The conditional logit models tell a somewhat different story. Firstly, the coefficient for the neither option is now positive and significant in all countries, suggesting that respondents are more inclined to select neither the two alternatives. But the main differences are represented by the coefficients of the two treatments, which become positive in all countries, with personalised energy costs having a slightly positive effect in the United States (significant at the 10% level).

Table B.8: Conditional logit models

	(1) Ireland	(2) UK	(3) Canada	(4) USA
Constant (neither option)	1.6090**	1.5402***	2.1789***	2.8473***
	(0.6389)	(0.5413)	(0.5915)	(0.5608)
Price	-0.0025***	-0.0024***	-0.0020***	-0.0023***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Stars	0.4272***	0.4009***	0.6013***	0.6309***
	(0.0409)	(0.0390)	(0.0381)	(0.0377)
Capacity	0.1736***	0.0843***	0.1339***	0.1842***
	(0.0230)	(0.0206)	(0.0232)	(0.0229)
Brand	-0.1664***	-0.1165**	-0.2847***	-0.1145**
	(0.0572)	(0.0571)	(0.0531)	(0.0547)
Energy efficiency	0.7183***	0.4471***	0.7563***	0.4534***
	(0.0821)	(0.0758)	(0.0812)	(0.0769)
EE × T1	0.0336	0.0210	0.0852	0.1320
	(0.1179)	(0.0992)	(0.1043)	(0.0965)
EE × T2	0.0165	0.0273	0.0664	0.1745*
	(0.1197)	(0.1007)	(0.1035)	(0.0968)
Model statistics				
Observations	13944	15720	15216	15768
Clusters	581	655	634	657

Notes: This table reports the results of conditional logit regressions of respondents' choices in each country separately. Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. All regressions control for income, gender, living area, whether the individual holds an degree, environmental concern, impatience, risk attitude and tumble dryer usage. Standard errors are clustered at the participant level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.9: Conditional logit willingness to pay

	Ireland	UK	Canada	USA
Stars	170.04	165.84	297.12	277.80
	[138.56 ; 201.52]	[135.03 ; 196.65]	[255.76 ; 338.49]	[242.29 ; 313.31]
Capacity	69.08	34.88	66.15	81.11
	[50.75 ; 87.41]	[18.04 ; 51.72]	[43.49 ; 88.81]	[60.83 ; 101.39]
Brand	-66.25	-48.18	-140.67	-50.42
	[-109.47 ; -23.03]	[-93.76 ; -2.59]	[-191.01 ; -90.34]	[-96.97 ; -3.88]
EE	285.91	184.94	373.70	199.65
	[217.18 ; 354.64]	[122.61 ; 247.26]	[289.53 ; 457.87]	[132.40 ; 266.90]
EE $\times$ T1	13.39	8.70	42.11	58.13
	[-78.56 ; 105.33]	[-71.74 ; 89.13]	[-59.15 ; 143.37]	[-25.40 ; 141.65]
EE $\times$ T2	6.55	11.29	32.83	76.84
	[-86.82 ; 99.92]	[-70.34 ; 92.92]	[-67.30 ; 132.95]	[-6.94 ; 160.62]

Notes: This table reports the willingness to pay of respondents in each country for the tumble dryer's attributes. Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. The 95% confidence intervals are reported in brackets.

B.4 Individual characteristics

Another important thing to investigate is whether the effect of our manipulations varies for different types of people. To do so, we run our models splitting the samples on the basis of the levels of various individual characteristics.

B.4.1 Tumble dryer usage

A first consideration we can make is that a limited average usage of the tumble dryer would translate into small energy expenditures, thus making energy cost information, especially personalised one, less salient than the current EU Energy Label and the EnergyStar logo. To investigate this, we have divided respondents based on their self-reported weekly usage, considering as low usage values smaller than or equal to the median of the respective country, mid-high usage between the median and the 90th percentile, while very-high usage corresponds to the top 10th percentile in each country. Specifically, in all four countries, the median is equal to 3 weekly cycles. The 90th percentile is 6 in Canada and 7 in Ireland, the United Kingdom and the United States. Therefore, low usage corresponds to 3 or fewer weekly cycles; mid-high usage is between 4 and 7 (both included) weekly cycles in Ireland, the UK and the US and between 4 and 6 in Canada; and very-high usage is given by more than 7 cycles in Ireland, the UK and the US and more than 6 in Canada.

From Table B.10 we see that there are considerably more people in the low usage category[1], which may sustain our prior claim. From Panels B and C it appears that personalised energy information no longer has a negative impact on Canadian respondents. In fact, the coefficient of the interaction between energy efficiency and treatment 2 is insignificant for the subgroup of mid-high usage respondents, and it becomes positive for those with high usage. In addition, in this last subgroup, personalised energy costs information presents positive coefficients in all four countries. However, we still fail to detect a significant effect of our treatments apart from the United Kingdom. Also, the number of clusters is

[1]The variable reporting tumble dryer usage right-skewed, with a median of 3 weekly cycles and a mean between 3 and 4 depending on the country.

borderline for unbiased inference.

Table B.10: Mixed logit models by tumble dryer usage

	Ireland	UK	Canada	USA
A. LOW USAGE				
Non-Random Parameters in Utility Function				
Constant (neither option)	1.5920	0.8043	1.8160	0.9276
	(1.2597)	(1.3551)	(1.3097)	(1.4794)
Price	-0.0033***	-0.0033***	-0.0027***	-0.0030***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Stars	0.5796***	0.6075***	0.8279***	0.6922***
	(0.0639)	(0.0637)	(0.0566)	(0.0639)
Random Parameters in Utility Function				
Capacity	0.2018***	0.0735**	0.1846***	0.1618***
	(0.0372)	(0.0356)	(0.0358)	(0.0382)
Brand	-0.3016***	-0.3122***	-0.4446***	-0.1719*
	(0.0891)	(0.0895)	(0.0796)	(0.0922)
Energy efficiency	1.0637***	0.6575***	1.3082***	0.6678***
	(0.1323)	(0.1386)	(0.1341)	(0.1514)
EE $\times$ T1	-0.0605	-0.0083	-0.0845	-0.0411
	(0.1784)	(0.1819)	(0.1796)	(0.1845)
EE $\times$ T2	-0.2177	0.1955	-0.3870**	-0.2370
	(0.1707)	(0.1754)	(0.1637)	(0.2057)
Model statistics				
Observations	8,448	9,600	10,440	8,328
Clusters	352	400	435	347
B. MID-HIGH USAGE				
Constant (neither option)	-1.9033	1.3134	3.7114	7.7983***
	(2.8505)	(1.9506)	(2.7317)	(1.5629)
Price	-0.0027***	-0.0028***	-0.0026***	-0.0025***
	(0.0003)	(0.0003)	(0.0003)	(0.0002)
Stars	0.5576***	0.4940***	0.8583***	0.8486***
	(0.0832)	(0.0808)	(0.0978)	(0.0771)
Random Parameters in Utility Function				
Capacity	0.2499***	0.1426***	0.2671***	0.3012***
	(0.0497)	(0.0475)	(0.0614)	(0.0430)
Brand	-0.1498	-0.0598	-0.5985***	-0.1650
	(0.1198)	(0.1262)	(0.1289)	(0.1029)
Energy efficiency	0.9393***	0.3638**	1.5738***	1.0815***
	(0.1853)	(0.1596)	(0.2271)	(0.1333)
EE $\times$ T1	0.0259	0.2747	-0.9670***	-0.2160
	(0.2769)	(0.2251)	(0.2856)	(0.2015)
EE $\times$ T2	0.2253	0.1012	-0.4814	-0.2350
	(0.2982)	(0.2356)	(0.2933)	(0.1700)
Model statistics				
Observations	4,656	4,848	3,480	6,000
Clusters	194	202	145	250

Table B.10 — continued

	Ireland	UK	Canada	USA
C. HIGH USAGE				
Constant (neither option)	-2.2415	0.9739	2.2848	-0.2994
	(4.5573)	(3.1072)	(2.9170)	(2.5441)
Price	-0.0032***	-0.0027***	-0.0014***	-0.0018***
	(0.0005)	(0.0005)	(0.0003)	(0.0004)
Stars	0.3814**	0.4361***	0.6426***	0.6034***
	(0.1555)	(0.1321)	(0.1353)	(0.1447)
Random Parameters in Utility Function				
Capacity	0.5290***	0.1992**	0.1283	0.3849***
	(0.1129)	(0.0855)	(0.0976)	(0.0861)
Brand	-0.4541*	-0.1782	-0.2599	0.0529
	(0.2451)	(0.2190)	(0.1930)	(0.2058)
Energy efficiency	1.0598***	0.5789*	1.0051***	0.3765
	(0.3254)	(0.3185)	(0.3199)	(0.3066)
EE $\times$ T1	0.3266	-0.0827	-0.0532	0.3446
	(0.4079)	(0.3856)	(0.3737)	(0.4286)
EE $\times$ T2	0.7602	1.6218***	0.2201	0.6026
	(0.8590)	(0.6023)	(0.3145)	(0.4394)
Model statistics				
Observations	840	1,272	1,296	1,440
Clusters	35	53	54	60

Notes: Panel A reports the results of mixed logit models for respondents who report a low tumble dryer usage (less than the median), Panel B for mid-high usage (between the median and the 90th percentile), and Panel C for very-high usage (90th percentile). Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. All regressions control for income, gender, living area, whether the individual holds an degree, environmental concern, impatience, risk attitude and tumble dryer usage. Standard errors are clustered at the participant level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.4.2 Education

A second consideration we can make is whether the effect differs for more or less educated people. In particular, one might argue that the provision of more explicit information on

Table B.11: Willingness to pay by tumble dryer usage

	Ireland	UK	Canada	USA
A. LOW USAGE				
Stars	176.07	182.91	302.68	229.75
	[138.68 ; 213.46]	[146.58 ; 219.24]	[257.63 ; 347.73]	[186.96 ; 272.53]
Capacity	61.30	22.12	67.49	53.69
	[38.06 ; 84.54]	[0.95 ; 43.28]	[41.42 ; 93.56]	[28.84 ; 78.55]
Brand	-91.63	-94.01	-162.53	-57.04
	[-142.63 ; -40.64]	[-145.54 ; -42.49]	[-218.89 ; -106.17]	[-116.17 ; 2.08]
EE	323.15	197.96	478.28	221.63
	[238.61 ; 407.69]	[115.54 ; 280.38]	[378.03 ; 578.54]	[123.35 ; 319.91]
EE $\times$ T1	-18.38	-2.51	-30.91	-13.66
	[-124.81 ; 88.05]	[-109.86 ; 104.83]	[-159.36 ; 97.54]	[-133.58 ; 106.27]
EE $\times$ T2	-66.13	58.88	-141.48	-78.66
	[-168.30 ; 36.04]	[-44.68 ; 162.44]	[-259.36 ; -23.60]	[-212.33 ; 55.02]
B. MID-HIGH USAGE				
Stars	206.42	175.56	326.99	336.64
	[144.41 ; 268.42]	[119.78 ; 231.35]	[250.20 ; 403.77]	[272.02 ; 401.26]
Capacity	92.53	50.68	101.75	119.48
	[57.78 ; 127.27]	[16.75 ; 84.61]	[55.33 ; 148.18]	[83.97 ; 154.99]
Brand	-55.46	-21.24	-228.02	-65.46
	[-141.64 ; 30.73]	[-109.02 ; 66.54]	[-320.37 ; -135.66]	[-144.08 ; 13.16]
EE	347.72	129.29	599.56	429.06
	[208.14 ; 487.29]	[18.44 ; 240.14]	[410.51 ; 788.61]	[318.01 ; 540.11]
EE $\times$ T1	9.58	97.62	-368.41	-85.69
	[-191.16 ; 210.31]	[-59.50 ; 254.74]	[-599.33 ; -137.49]	[-244.29 ; 72.91]
EE $\times$ T2	83.42	35.95	-183.40	-93.24
	[-132.66 ; 299.49]	[-127.87 ; 199.78]	[-403.79 ; 36.99]	[-227.17 ; 40.68]
C. VERY-HIGH USAGE				
Stars	117.60	159.90	453.49	328.45
	[22.95 ; 212.25]	[54.27 ; 265.52]	[184.66 ; 722.32]	[160.12 ; 496.78]
Capacity	163.10	73.03	90.55	209.54
	[81.95 ; 244.25]	[3.72 ; 142.33]	[-42.38 ; 223.48]	[86.98 ; 332.10]
Brand	-140.03	-65.33	-183.41	28.82
	[-280.94 ; 0.87]	[-221.36 ; 90.69]	[-451.33 ; 84.51]	[-193.83 ; 251.48]
EE	326.79	212.23	709.26	204.96
	[110.53 ; 543.04]	[-23.49 ; 447.95]	[235.48 ; 1183.05]	[-120.63 ; 530.56]
EE $\times$ T1	100.70	-30.31	-37.52	187.61
	[-144.67 ; 346.07]	[-307.73 ; 247.11]	[-552.70 ; 477.66]	[-275.58 ; 650.80]
EE $\times$ T2	234.40	594.61	155.32	328.01
	[-299.25 ; 768.05]	[145.85 ; 1043.36]	[-288.65 ; 599.28]	[-183.88 ; 839.91]

Notes: Panel A reports the WTP for respondents who report a low tumble dryer usage (less than the median), Panel B for mid-high usage (between the median and the 90th percentile), and Panel C for very-high usage (90th percentile). Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. 95% confidence intervals in brackets.

energy costs might benefit mainly those with lower levels of education. To check this, we run separate models distinguishing between participants who hold a degree and those who do not. Remember that, from Table 3.3 in the main corpus of the paper, this was the variable that showed the most differences between the three groups, although not in the same direction for all countries.

The results in Table B.12 do not give any particular indication that monetary information — either generic of personalised — has a bigger effect for less educated people. Although the coefficients of our two treatments become positive (but insignificant) for respondents without a degree in the Canadian sample, this effect does not apply to the other countries. If anything, we observe outcomes contrary to this belief, with the coefficient for the generic energy costs treatment being negative and significant for Irish participants without a degree.

Table B.12: Mixed logit models by education

	Ireland	UK	Canada	USA
A. WITH A DEGREE				
Non-Random Parameters in Utility Function				
Constant (neither option)	0.5585	0.3637	1.5939	1.8199*
	(1.2586)	(1.1128)	(1.2611)	(1.0719)
Price	-0.0031***	-0.0028***	-0.0024***	-0.0027***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Stars	0.5556***	0.5385***	0.8160***	0.7488***
	(0.0580)	(0.0595)	(0.0540)	(0.0569)
Random Parameters in Utility Function				
Capacity	0.2323***	0.1226***	0.2175***	0.2602***
	(0.0345)	(0.0331)	(0.0353)	(0.0343)
Brand	-0.2247***	-0.1798**	-0.4571***	-0.1956**
	(0.0832)	(0.0810)	(0.0763)	(0.0814)
Energy efficiency	0.9458***	0.5543***	1.4095***	0.7284***
	(0.1156)	(0.1274)	(0.1361)	(0.1084)
EE × T1	0.1397	0.0709	-0.4901***	-0.0562
	(0.1739)	(0.1724)	(0.1834)	(0.1527)
EE × T2	-0.0856	0.3541**	-0.4932***	-0.0577
	(0.1686)	(0.1734)	(0.1641)	(0.1508)
Model statistics				
Observations	9768	9624	10728	10560
Clusters	407	401	447	440
B. WITHOUT A DEGREE				
Constant (neither option)	1.0626	0.8344	0.4247	3.5218**
	(2.0298)	(1.5751)	(1.5433)	(1.3973)

Table B.12 — continued				
	Ireland	UK	Canada	USA
Price	-0.0031***	-0.0037***	-0.0030***	-0.0027***
	(0.0003)	(0.0003)	(0.0003)	(0.0002)
Stars	0.5644***	0.5800***	0.8202***	0.7201***
	(0.0878)	(0.0744)	(0.0888)	(0.0807)
Random Parameters in Utility Function				
Capacity	0.2348***	0.0787*	0.1499***	0.1873***
	(0.0539)	(0.0473)	(0.0536)	(0.0469)
Brand	-0.3327***	-0.2985**	-0.4741***	-0.0237
	(0.1177)	(0.1266)	(0.1169)	(0.1091)
Energy efficiency	1.3164***	0.5423***	1.2016***	0.8657***
	(0.2128)	(0.1543)	(0.1887)	(0.1745)
EE × T1	-0.4954*	0.0358	0.0907	-0.0014
	(0.2563)	(0.2060)	(0.2264)	(0.2260)
EE × T2	-0.1098	0.1077	0.0043	-0.2457
	(0.2633)	(0.2005)	(0.2280)	(0.2268)
Model statistics				
Observations	4176	6096	4488	5208
Clusters	174	254	217	217

Notes: Panel A reports the results of mixed logit models for respondents who hold a bachelor's degree (or a corresponding title for Canada and the United States) or higher. Panel B presents the results for those how do not have one. Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. All regressions control for income, gender, living area, environmental concern, impatience, risk attitude and tumble dryer usage. Standard errors are clustered at the participant level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.4.3 Environmental concern

Another possible consideration is that people concerned about the environment will always tend to chose the most efficient product, irrespectively of how energy information is framed. This is confirmed in Table B.15, where the WTP for energy efficiency is consistently higher in the subgroup of respondents who are concerned about the environment.

Table B.13: Willingness to pay by education

	Ireland	UK	Canada	USA
A. WITH A DEGREE				
Stars	181.44	195.75	335.37	278.63
	[143.88 ; 219.00]	[155.25 ; 236.25]	[286.42 ; 384.32]	[235.25 ; 322.01]
Capacity	75.86	44.56	89.37	96.82
	[53.95 ; 97.76]	[20.35 ; 68.76]	[60.62 ; 118.13]	[71.00 ; 122.65]
Brand	-73.38	-65.36	-187.84	-72.79
	[-125.16 ; -21.59]	[-121.92 ; -8.80]	[-247.78 ; -127.91]	[-131.33 ; -14.24]
EE	308.90	201.49	579.27	271.02
	[231.88 ; 385.92]	[109.66 ; 293.32]	[463.31 ; 695.24]	[191.19 ; 350.85]
EE $\times$ T1	45.64	25.77	-201.42	-20.93
	[-65.39 ; 156.67]	[-96.97 ; 148.52]	[-349.58 ; -53.25]	[-132.28 ; 90.42]
EE $\times$ T2	-27.96	128.72	-202.71	-21.48
	[-136.00 ; 80.07]	[5.37 ; 252.08]	[-336.75 ; -68.67]	[-131.50 ; 88.54]
B. WITHOUT A DEGREE				
Stars	184.19	156.08	277.83	266.57
	[129.89 ; 238.49]	[114.02 ; 198.13]	[213.01 ; 342.65]	[204.49 ; 328.66]
Capacity	76.63	21.17	50.78	69.35
	[38.51 ; 114.74]	[-4.02 ; 46.36]	[14.40 ; 87.17]	[34.51 ; 104.20]
Brand	-108.59	-80.34	-160.60	-8.77
	[-181.53 ; -35.65]	[-147.45 ; -13.23]	[-238.98 ; -82.21]	[-87.66 ; 70.12]
EE	429.58	145.94	407.03	320.47
	[284.16 ; 575.00]	[63.54 ; 228.34]	[275.42 ; 538.63]	[189.63 ; 451.32]
EE $\times$ T1	-161.68	9.62	30.72	-0.52
	[-328.86 ; 5.49]	[-98.87 ; 118.12]	[-119.78 ; 181.22]	[-164.49 ; 163.45]
EE $\times$ T2	-35.85	28.98	1.44	-90.97
	[-203.97 ; 132.28]	[-76.74 ; 134.70]	[-149.94 ; 152.82]	[-255.86 ; 73.91]

Notes. Panel A reports the WTP for respondents who hold a degree. Panel B presents the results for those without a degree. Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. 95% confidence intervals in brackets.

As we can see from Panel B of Table B.14, presenting energy information in monetary terms seems to have a positive effect for those who are less concerned about the environment, especially in Ireland and the UK; while no appreciable impact can be detected for people more concerned about environmental problems. A possible explanation could be that the first group is more interested in how much money they are going to spend for energy consumption rather than its environmental impact, and the treatments are more effective in making this information easily available and understandable.

Table B.14: Mixed logit models by environmental concern

	Ireland	UK	Canada	USA
A. CONCERNED ABOUT THE ENVIRONMENT				
Non-Random Parameters in Utility Function				
Constant (neither option)	1.9209	1.9707	1.2735	2.6740
	(1.8422)	(2.1897)	(1.8231)	(1.6723)
Price	-0.0029***	-0.0031***	-0.0025***	-0.0026***
	(0.0002)	(0.0002)	(0.0001)	(0.0002)
Stars	0.5221***	0.6042***	0.8373***	0.7843***
	(0.0533)	(0.0571)	(0.0527)	(0.0544)
Random Parameters in Utility Function				
Capacity	0.2410***	0.1133***	0.1650***	0.2424***
	(0.0327)	(0.0330)	(0.0336)	(0.0325)
Brand	-0.2633***	-0.2078**	-0.4121***	-0.1337*
	(0.0752)	(0.0822)	(0.0744)	(0.0756)
Energy efficiency	1.1475***	0.7177***	1.3710***	0.8563***
	(0.1172)	(0.1286)	(0.1248)	(0.1122)
EE $\times$ T1	-0.1831	-0.0161	-0.4124***	-0.1146
	(0.1654)	(0.1655)	(0.1599)	(0.1533)
EE $\times$ T2	-0.2406	0.1637	-0.3718**	-0.2275
	(0.1736)	(0.1753)	(0.1597)	(0.1487)
Model statistics				
Observations	10,992	10,872	11,952	11,928
Clusters	458	453	498	497
B. NOT CONCERNED ABOT THE ENVIRONMENT				
Non-Random Parameters in Utility Function				
Constant (neither option)	-5.7161**	1.2830	2.6564	0.0196
	(2.7743)	(1.6268)	(2.3223)	(1.8580)
Price	-0.0038***	-0.0031***	-0.0027***	-0.0028***
	(0.0004)	(0.0003)	(0.0003)	(0.0003)
Stars	0.7259***	0.4424***	0.7503***	0.5897***
	(0.1228)	(0.0819)	(0.0956)	(0.0885)
Random Parameters in Utility Function				
Capacity	0.2008***	0.0959**	0.3183***	0.2159***
	(0.0669)	(0.0481)	(0.0606)	(0.0530)
Brand	-0.2560*	-0.2603**	-0.6579***	-0.1661

Table B.14 — continued

	Ireland	UK	Canada	USA
	(0.1550)	(0.1295)	(0.1278)	(0.1337)
Energy efficiency	0.5544***	0.2048	1.2165***	0.4375***
	(0.2130)	(0.1395)	(0.2210)	(0.1695)
EE × T1	0.5490*	0.1470	-0.0477	0.2897
	(0.3118)	(0.2000)	(0.2796)	(0.2266)
EE × T2	0.7490**	0.4169**	-0.1586	0.2802
	(0.3174)	(0.1790)	(0.2478)	(0.2500)
Model statistics				
Observations	2,928	4,848	3,264	3,816
Clusters	122	202	159	159

Notes: Panel A reports the results of mixed logit models for respondents who state to be concerned or extremely concerned about the environment. Panel B presents the results for those who are slightly concerned, not concerned or they don't know. Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. All regressions control for income, gender, living area, whether the individual holds an degree, environmental concern, impatience, risk attitude and tumble dryer usage. Standard errors are clustered at the participant level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.4.4 Income

A final important factor to be taken into account is income, since making energy costs more explicit could benefit mostly income-constrained people. However, dividing respondents on the basis of their income — whether they live comfortably on their current income or not — does not provide a clear indication in this sense. The coefficients suggest that this seems to be the case only in the UK, where the personalised energy costs information increases utility and the WTP for energy efficiency. We also detect a positive impact of personalised energy costs in the Irish sample, but this is not significant. On the other hand, we find a negative effect for both subgroups in Canada, and no significant effect for either of them in the United States. This might indicate that if energy costs are relatively

Table B.15: Willingness to pay by environmental concern

	Ireland	UK	Canada	USA
A. CONCERNED ABOUT THE ENVIRONMENT				
Stars	181.34	196.37	329.39	295.96
	[144.75 ; 217.93]	[160.12 ; 232.61]	[283.94 ; 374.84]	[253.10 ; 338.81]
Capacity	82.45	36.82	64.92	92.57
	[59.93 ; 104.98]	[15.04 ; 58.60]	[38.98 ; 90.85]	[67.54 ; 117.59]
Brand	-91.96	-67.55	-162.13	-50.00
	[-141.32 ; -42.61]	[-119.52 ; -15.58]	[-218.93 ; -105.34]	[-105.73 ; 5.73]
EE	392.37	233.27	539.36	325.34
	[308.69 ; 476.05]	[150.19 ; 316.35]	[436.84 ; 641.87]	[241.18 ; 409.51]
EE $\times$ T1	-53.34	-5.24	-162.22	-44.69
	[-161.07 ; 54.40]	[-110.68 ; 100.21]	[-286.16 ; -38.27]	[-163.26 ; 73.87]
EE $\times$ T2	-79.80	53.19	-146.27	-85.02
	[-190.99 ; 31.38]	[-58.64 ; 165.02]	[-269.91 ; -22.63]	[-195.33 ; 25.30]
B. NOT CONCERNED ABOUT THE ENVIRONMENT				
Stars	190.18	141.50	276.69	208.73
	[132.49 ; 247.87]	[90.88 ; 192.12]	[199.64 ; 353.73]	[147.49 ; 269.98]
Capacity	52.61	30.66	117.37	76.44
	[16.58 ; 88.63]	[0.83 ; 60.50]	[69.27 ; 165.46]	[38.88 ; 113.99]
Brand	-67.06	-83.26	-242.62	-58.79
	[-144.71 ; 10.59]	[-163.01 ; -3.51]	[-329.68 ; -155.57]	[-151.14 ; 33.57]
EE	145.25	65.50	448.60	154.86
	[39.36 ; 251.14]	[-20.16 ; 151.16]	[285.00 ; 612.20]	[39.62 ; 270.10]
EE $\times$ T1	143.84	47.01	-17.59	102.57
	[-18.83 ; 306.51]	[-78.72 ; 172.73]	[-219.30 ; 184.13]	[-57.93 ; 263.06]
EE $\times$ T2	196.24	133.34	-58.49	99.17
	[26.74 ; 365.75]	[19.97 ; 246.72]	[-238.40 ; 121.42]	[-76.93 ; 275.27]

Notes: Panel A reports the WTP for respondents who state to be concerned or extremely concerned about the environment. Panel B presents the results for those who are slightly concerned, not concerned or they don't know. Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. 95% confidence intervals in brackets.

small, making them more explicit has a limited impact on the importance attached to energy efficiency, especially for people who are not in financial hardship; while more explicit information could benefits mainly less wealthy households if energy bills are a considerable proportion of their expenditures.

Table B.16: Mixed logit models by income

	Ireland	UK	Canada	USA
A. LIVING COMFORTABLY ON CURRENT INCOME				
Non-Random Parameters in Utility Function				
Constant (neither option)	8.0536*	5.7640**	3.7666	1.5771
	(4.8066)	(2.4475)	(2.6573)	(2.0732)
Price	-0.0026***	-0.0028***	-0.0025***	-0.0023***
	(0.0003)	(0.0002)	(0.0002)	(0.0002)
Stars	0.6529***	0.6075***	0.9551***	0.7904***
	(0.0906)	(0.0725)	(0.0678)	(0.0694)
Random Parameters in Utility Function				
Capacity	0.2424***	0.1443***	0.1907***	0.2207***
	(0.0532)	(0.0434)	(0.0436)	(0.0418)
Brand	-0.2912**	-0.3030***	-0.5775***	-0.1718*
	(0.1241)	(0.1082)	(0.0885)	(0.0955)
Energy efficiency	1.2609***	0.7366***	1.3391***	0.7343***
	(0.1928)	(0.1634)	(0.1567)	(0.1362)
EE $\times$ T1	-0.0685	-0.0544	-0.1615	-0.0933
	(0.3092)	(0.1981)	(0.2287)	(0.1787)
EE $\times$ T2	-0.4533*	-0.0942	-0.3462*	-0.2030
	(0.2593)	(0.2303)	(0.1939)	(0.1764)
Model statistics				
Observations	4,008	6,408	6,840	7,008
Clusters	167	267	285	292
B. NOT LIVING COMFORTABLY ON CURRENT INCOME				
Non-Random Parameters in Utility Function				
Constant (neither option)	0.5798	0.7468	0.7873	2.5198**
	(1.5435)	(1.3352)	(1.5236)	(1.2553)
Price	-0.0033***	-0.0032***	-0.0027***	-0.0031***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Stars	0.5129***	0.5139***	0.7072***	0.6974***
	(0.0591)	(0.0617)	(0.0623)	(0.0635)
Random Parameters in Utility Function				
Capacity	0.2284***	0.0819**	0.2067***	0.2463***
	(0.0355)	(0.0343)	(0.0397)	(0.0371)
Brand	-0.2378***	-0.1566*	-0.3681***	-0.1077
	(0.0846)	(0.0895)	(0.0914)	(0.0904)
Energy efficiency	0.9117***	0.4428***	1.3314***	0.8021***
	(0.1170)	(0.1261)	(0.1474)	(0.1287)
EE $\times$ T1	0.0063	0.1049	-0.4211**	0.0036
	(0.1610)	(0.1731)	(0.1772)	(0.1828)

Table B.16 — continued

	Ireland	UK	Canada	USA
EE × T2	0.1338	0.4733***	-0.3350*	-0.0288
	(0.1721)	(0.1684)	(0.1803)	(0.1775)
Model statistics				
Observations	9,912	9,312	8,376	8,736
Clusters	413	388	364	364

Notes: Panel A reports the results of mixed logit models for respondents who state they live comfortably or very comfortably on current income. Panel B presents the results for those who are copying on current income, finding it difficult or very difficult to live in current income, or prefer not to say. Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. Standard errors are clustered at the participant level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.17: Willingness to pay by income

	Ireland	UK	Canada	USA
A. LIVING COMFORTABLY ON CURRENT INCOME				
Stars	253.65	213.92	385.80	347.53
	[185.12 ; 322.18]	[162.96 ; 264.88]	[323.29 ; 448.32]	[280.10 ; 414.96]
Capacity	94.16	50.79	77.04	99.94
	[54.11 ; 134.20]	[19.84 ; 81.75]	[42.37 ; 111.71]	[62.07 ; 137.81]
Brand	-113.15	-106.70	-233.25	-77.30
	[-202.46 ; -23.85]	[-180.79 ; -32.62]	[-303.33 ; -163.16]	[-158.67 ; 4.08]
EE	489.87	259.37	540.90	324.69
	[317.70 ; 662.03]	[145.52 ; 373.22]	[400.91 ; 680.90]	[206.59 ; 442.78]
EE $\times$ T1	-26.61	-19.17	-65.25	-30.92
	[-262.91 ; 209.69]	[-155.86 ; 117.53]	[-246.39 ; 115.89]	[-182.26 ; 120.41]
EE $\times$ T2	-176.11	-33.18	-139.85	-91.85
	[-378.19 ; 25.97]	[-192.08 ; 125.72]	[-295.47 ; 15.78]	[-236.47 ; 52.77]
B. NOT LIVING COMFORTABLY ON CURRENT INCOME				
Stars	156.61	158.21	266.02	228.14
	[122.24 ; 190.99]	[122.00 ; 194.41]	[215.34 ; 316.70]	[187.44 ; 268.84]
Capacity	70.32	25.22	77.75	80.58
	[48.61 ; 92.03]	[4.39 ; 46.05]	[47.91 ; 107.60]	[56.29 ; 104.87]
Brand	-70.97	-48.20	-138.47	-35.23
	[-119.35 ; -22.60]	[-101.57 ; 5.17]	[-204.84 ; -72.09]	[-92.53 ; 22.08]
EE	278.27	136.33	500.83	262.40
	[208.17 ; 348.37]	[59.97 ; 212.70]	[392.01 ; 609.64]	[179.86 ; 344.93]
EE $\times$ T1	6.57	32.28	-158.39	1.19
	[-88.29 ; 101.43]	[-71.89 ; 136.46]	[-288.71 ; -28.08]	[-116.03 ; 118.41]
EE $\times$ T2	43.77	145.71	-126.03	-9.43
	[-60.02 ; 147.57]	[44.05 ; 247.38]	[-258.69 ; 6.63]	[-123.27 ; 104.41]

Notes: Panel A reports the results of mixed logit models for respondents who state they live comfortably or very comfortably on current income. Panel B presents the results for those who are copying on current income, finding it difficult or very difficult to live in current income, or prefer not to say. Energy efficiency is a dummy variable taking value 1 for the three highest levels of energy consumption (lower efficiency), and 2 for the three lowest levels of energy consumption (higher efficiency). It takes value 0 for the "neither" option like all other attributes. 95% confidence intervals in brackets.

C Appendix Chapter 4

C.1 Election Data

C.1.1 Party classifications

Our dataset contains information on the number of valid votes cast for the six main political parties in Germany — namely CDU/CSU, SPD, Greens, FDP, Die Linke and AfD — as well as the total of votes cast for the other minor parties. As mentioned in 4.3.1, our analysis primarily focuses on the vote share of incumbent and established opposition parties. For completeness, we also investigate the effect on other, non-established parties. Below is a description of the parties' classifications that have been used.

Table C.1: Classification of Parties

Category	Explanation
Established parties	These are the five main actors on the German political scene: the Social Democratic Party of Germany (SPD), the Christian Democratic Union of Germany (CDU/CSU), the Free Democratic Party (FDP), the Green Party, the Left Party (Die Linke).
Other parties	These are smaller opposition parties, many of which are not frequently represented in the federal or state parliaments. This category includes the far-right party Alternative for Germany (AfD): despite it entering the Bundestag in 2017, it was not regularly represented in parliaments over the sample period, which is why it is not classified as an established party in our analysis.

continued

Table C.1 continued

Category	Explanation
Incumbent parties	The party or coalition that was in power before a specific election. In each case, we compute the incumbent by taking the sum of the vote shares of the parties forming the coalition. For national elections, these coalitions are: SPD-Greens from 1998 to 2005; CDU-SPD from 2005 to 2009; CDU-FDP from 2009 to 2013; CDU-SPD from 2013 to 2017. For state elections, each state elects its own government and therefore the incumbent varies from state to state. *Baden-Württemberg:* CDU-FDP from 1996 to 2011; SPD-Greens from 2011 to 2016. *Bayern:* CDU from 1998 to 2008 and then from 2013 to 2018/date; CDU-FDP from 2008 to 2013. *Berlin:* CDU-SPD from 1999 to 2001 and from 2011 to 2016/date; SPD-Die Linke from 2001 to 2011. *Brandenburg:* SPD-CDU from 1999 to 2009; SPD-Die Linke from 2009 to 2018. *Bremen:* SPD-CDU from 1999 to 2007; SPD-Greens from 2007 to 2018. *Hamburg:* SPD-Greens from 1997 to 2001; CDU-FDP-Schill Partei from 2001 to 2004; CDU from 2004 to 2008; CDU-Greens from 2008 to 2011; SPD from 2011 to 2015. *Hessen:* CDU-FDP from 1999 to 2003 and from 2008 to 2013; CDU from 2003 to 2008; CDU-Greens from 2013 to 2018. *Mecklenburg-Vorpommern:* SPD-Die Linke from 1998 to 2006; SPD-CDU from 2006 to 2016. *Niedersachsen:* SPD from 1998 to 2003; CDU-FDP from 2003 to 2013; SPD-Greens from 2013 to 2017. *Nordrhein-Westfalen:* SPD-Greens from 1995 to 2005 and from 2010 to 2017; CDU-FDP from 2005 to 2010 and from 2017 to date. *Rheinland-Pfalz:* SPD-FDP from 1996 to 2006; SPD from 2006 to 2011; SPD-Greens from 2011 to 2016. *Saarland:* CDU from 1999 to 2009; CDU-FDP-Greens from 2009 to 2012; CDU-SPD from 2012 to 2017.

continued

Table C.1 continued

Category	Explanation
	Sachsen: CDU from 1999 to 2004; CDU-SPD from 2004 to 2009; CDU-FDP from 2009 to 2014. *Sachsen-Anhalt:* SPD from 1998 to 2002; CDU-FDP from 2002 to 2011; CDU-SPD-Greens from 2011 to 2016. *Schleswig-Holstein:* SPD-Greens from 1996 to 2005; CDU-SPD from 2005 to 2009; CDU-FDP from 2009 to 2012; SPD-Greens from 2012 to 2017. *Thüringen:* CDU from 1999 to 2004; CDU-SPD from 2004 to 2014; SPD-Greens-Die Linke for 2014 to 2018.
Established Opposition	The established parties that were not in power (were at the opposition) before a specific election. In each case, we compute the established opposition by taking the sum of the vote shares of all the parties. For national elections, the established opposition is represented by: CDU-FDP-Die Linke from 1998 to 2005; FDP-Greens-Die Linke from 2005 to 2009; SPD-Greens-Die Linke from 2009 to 2013; FDP-Greens-Die Linke from 2013 to date. For federal elections, each state elects its own government, so established opposition parties vary from state to state. *Baden-Württemberg:* SPD-Greens-Die Linke from 1996 to 2011; CDU-FDP-Die Linke from 2011 to 2016. *Bayern:* SPD-FDP-Greens-Die Linke from 1998 to 2008 and from 2013 to 2018; SPD-Greens-Die Linke from 2008 to 2013. *Berlin:* FDP-Greens-Die Linke from 1999 to 2001; CDU-FDP-Greens from 2001 to 2011; FDP-Greens-Die Linke from 2016 to date. *Brandenburg:* FDP-Greens-Die Linke from 1999 to 2009; CDU-FDP-Greens from 2009 to 2018. *Bremen:* FDP-Greens-Die Linke from 1999 to 2007; CDU-FDP-Die Linke from 2007 to 2018. *Hamburg:* CDU-FDP-Die Linke from 1997 to 2001; SPD-Greens-Die Linke from 2001 to 2004; SPD-FDP-Greens-Die Linke from 2004 to 2008; SPD-FDP-Die Linke from 2008 to 2011; CDU-FDP-Greens-Die Linke from 2011 to 2015.

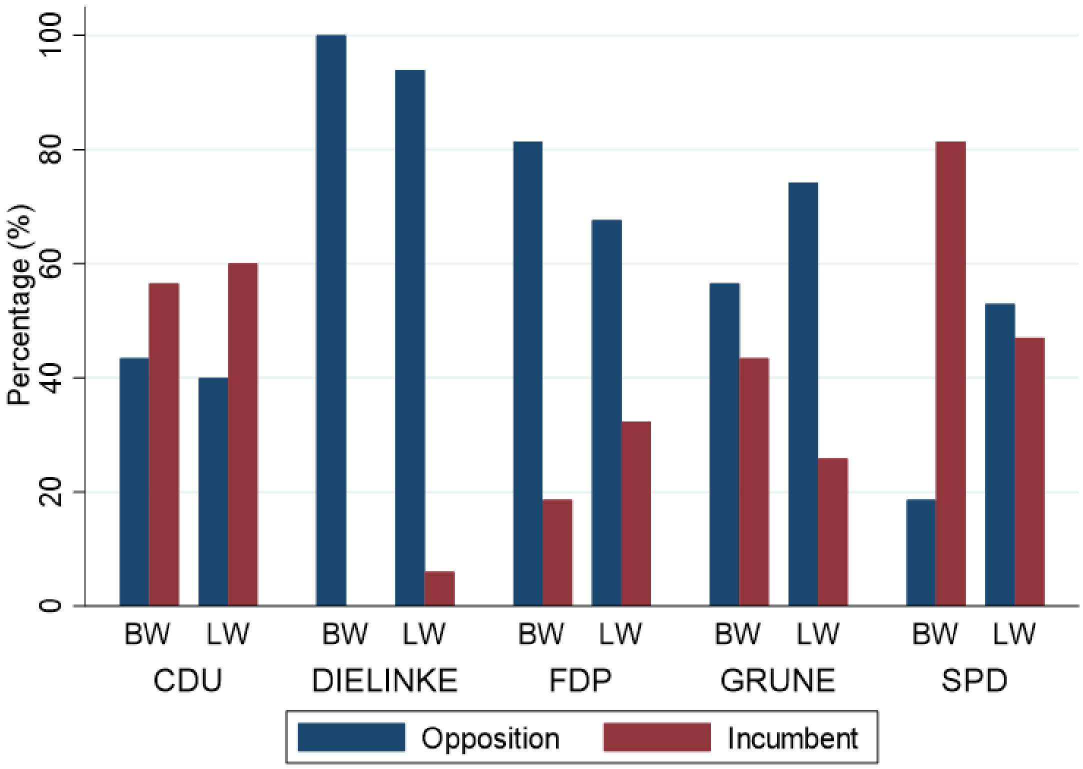
continued

Table C.1 continued

Category	Explanation
	Hessen: SPD-Greens-Die Linke from 1999 to 2003 and from 2008 to 2013; SPD-FDP-Greens-Die Linke from 2003 to 2008; SPD-FDP-Die Linke from 2013 to 2018. *Mecklenburg-Vorpommern:* CDU-FDP-Greens from 1998 to 2006; FDP-Greens-Die Linke from 2006 to 2016. *Niedersachsen:* CDU-FDP-Greens-Die Linke from 1998 to 2003; SPD-Greens-Die Linke from 2003 to 2013; CDU-FDP-Die Linke from 2013 to 2017. *Nordrhein-Westfalen:* CDU-FDP-Die Linke from 1995 to 2005 and from 2010 to 2017; SPD-Greens-Die Linke from 2005 to 2010 and from 2017 to date. *Rheinland-Pfalz:* CDU-Greens-Die Linke from 1996 to 2006 and from 2011 to 2016; CDU-FDP-Greens-Die Linke from 2006 to 2011. *Saarland:* SPD-FDP-Greens-Die Linke from 1999 to 2009; SPD-Die Linke from 2009 to 2012; FDP-Greens-Die Linke from 2012 to 2017. *Sachsen:*: SPD-FDP-Greens-Die Linke from 1999 to 2004; FDP-Greens-Die Linke from 2004 to 2009; SPD-Greens-Die Linke from 2009 to 2014. *Sachsen-Anhalt:* CDU-FDP-Greens-Die Linke from 1998 to 2002; SPD-Greens-Die Linke from 2002 to 2011; FDP-Die Linke from 2011 to 2016. *Schleswig-Holstein:* CDU-FDP-Die Linke from 1996 to 2005; FDP-Greens-Die Linke from 2005 to 2009; SPD-Greens-Die Linke from 2009 to 2012; CDU-FDP-Die Linke from 2012 to 2017. *Thüringen:* SPD-FDP-Greens-Die Linke from 1999 to 2004; FDP-Greens-Die Linke from 2004 to 2014; CDU-FDP from 2014 to 2018.

Figure C.1 offers a graphical representation of Table C.1, showing the frenquency with which each of the main established parties was the incumbent or the opposition in our sample period of 2000-2018, distinguishing between federal (BW) and state (LW) elections.

Figure C.1: Frequency of Incumbency and Opposition by Party and Election Type (2000–2018)



Notes: This graph shows how frequently each of the major parties has been the incumbent or opposition party in federal elections (BW) and state elections (LW). For example, the SPD was an incumbent party in 80% of the federal elections in our sample and in around 50% of the state elections.

C.1.2 Election dates

We investigate outcomes from 82 election, five national (*Bundestagswahl*, BW) and 67 state (*Landtagswahl*, LW), in 2000-2018 the period. The national parliament (*Bundestag*) is elected for a four-year term, while the federal ones (*Landtag*) remain in power for four or five years depending on the state. Below the election dates for both bodies are listed.

Table C.2: Date and Type of Elections

Date	Type
National Elections (*Bundestag*) - All States	
September 22, 2002	National (*Bundestag*)
September 18, 2005	National (*Bundestag*). Early election.
September 27, 2009	National (*Bundestag*)
September 22, 2013	National (*Bundestag*)
September 24, 2017	National (*Bundestag*)
Baden-Württemberg	
March 25, 2001	State (*Landtag*)
March 26, 2006	State (*Landtag*)
March 27, 2011	State (*Landtag*)
March 13, 2016	State (*Landtag*)
Bayern	
September 21, 2003	State (*Landtag*)
September 28, 2008	State (*Landtag*)
September 15, 2013	State (*Landtag*)
October 14, 2018	State (*Landtag*)
Berlin	
October 21, 2001	State (*Landtag*)
September 17, 2006	State (*Landtag*)
September 18, 2011	State (*Landtag*)
September 18, 2016	State (*Landtag*)
Brandenburg	
September 19, 2004	State (*Landtag*)
September 27, 2009	State (*Landtag*)

continued

Table C.2 continued

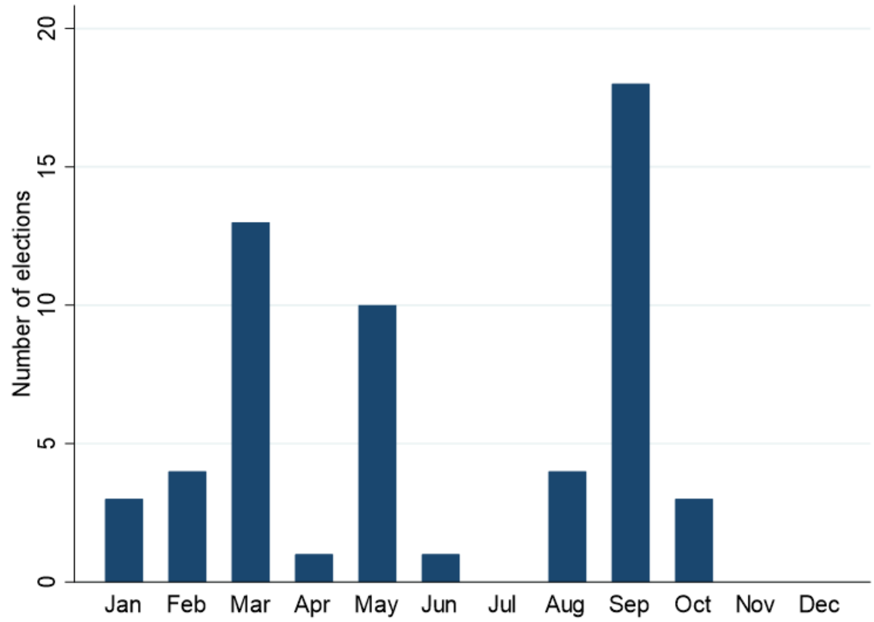
Date	Type
September 14, 2014	State (*Landtag*)
Bremen	
May 25, 2003	State (*Landtag*)
May 13, 2007	State (*Landtag*)
May 22, 2011	State (*Landtag*)
May 10, 2015	State (*Landtag*)
Hamburg	
September 23, 2001	State (*Landtag*)
February 29, 2004	State (*Landtag*). Early election.
February 27, 2008	State (*Landtag*)
February 20, 2011	State (*Landtag*)
February 15, 2015	State (*Landtag*)
Hessen	
February 2, 2003	State (*Landtag*)
January 27, 2008	State (*Landtag*)
September 22, 2013	State (*Landtag*)
October 28, 2018	State (*Landtag*)
Mecklenburg-Vorpommern	
September 22, 2002	State (*Landtag*)
September 17, 2006	State (*Landtag*)
September 4, 2011	State (*Landtag*)
September 4, 2016	State (*Landtag*)
Niedersachsen	
February 2, 2003	State (*Landtag*)
January 27, 2008	State (*Landtag*)
January 20, 2013	State (*Landtag*)
October 15, 2017	State (*Landtag*)
Nordrhein-Westfalen	
May 14, 2000	State (*Landtag*)
May 22, 2005	State (*Landtag*)
May 09, 2010	State (*Landtag*)
May 13, 2012	State (*Landtag*). Early election.

continued

Table C.2 continued

Date	Type
May 14, 2017	State (*Landtag*)
Rheinland-Pfalz	
March 25, 2001	State (*Landtag*)
March 26, 2006	State (*Landtag*)
March 27, 2011	State (*Landtag*)
March 13, 2016	State (*Landtag*)
Saarland	
September 5, 2004	State (*Landtag*)
August 30, 2009	State (*Landtag*)
March 25, 2012	State (*Landtag*). Early election.
March 26, 2017	State (*Landtag*)
Sachsen	
September 19, 2004	State (*Landtag*)
August 30, 2009	State (*Landtag*)
August 31, 2014	State (*Landtag*)
Sachsen-Anhalt	
April 21, 2002	State (*Landtag*)
March 26, 2006	State (*Landtag*)
March 20, 2011	State (*Landtag*).
March 13, 2016	State (*Landtag*)
Schleswig-Holstein	
February 27, 2000	State (*Landtag*)
February 20, 2005	State (*Landtag*)
September 27, 2009	State (*Landtag*)
May 6, 2012	State (*Landtag*). Early election.
May 7, 2017.	State (*Landtag*)
Thüringen	
June 13, 2004	State (*Landtag*)
August 30, 2009	State (*Landtag*)
September 14, 2014	State (*Landtag*).

Figure C.2: Distribution of elections by calendar month



C.2 Voting by Mail as a Measurement Error Problem

A challenge for our empirical analysis is that some votes are cast by mail. The share of mail voters has steadily increased over the sample period: in federal elections, it increased from 13.4% of all votes in 2002 to 28.6% in 2017 (Bundeswahlleiter, 2017). The possibility of voting by mail adds two potential problems to our estimation strategy. First, for most counties, voting data are not available separately by means of voting, at the ballot box or by mail. Therefore, the overall voting data at the county level that we use are a combination of votes cast at the ballots in the county on the election day and votes mailed by residents of that county, potentially at anytime in the month before the election day and from anywhere that they might have been at that time. Both these facts make it likely for us to assign an incorrect level of air pollution to those votes sent by mail.

The fact that we potentially assign an incorrect pollution level to a share of voters is akin to a measurement error problem, as the true exposure when making voting decisions is different from the exposure we assign, namely the level of PM10 on election day. Let the level of PM10 on election day be $PM10_{it}$ with mean μ_{PM10} and standard deviation σ_{PM10} and the share of mail voters be $\alpha \in (0, 1)$. We assume that mail voters only vote on one particular day, on which the level of PM10 is Q_{it}, with mean μ_Q and standard deviation σ_Q. $PM10$, Q and the outcome y are within-transformed, i.e. they represent the residuals after differencing out county and election date fixed effects. The indices i and t stand for county and election date, respectively. Assume that the true relationship is

$$y_{it} = \beta_0 + \beta_1((1-\alpha)PM10_{it} + \alpha Q_{it}) + \varepsilon_{it}, \tag{C.1}$$

and that the model is otherwise correctly specified, $cov(PM10_{it}, \varepsilon_{it}) = cov(Q_{it}, \varepsilon_{it}) = 0$. In Equation (C.1), the true exposure on the day when people cast their vote is a weighted average of the concentration of PM10 on the election day and the day on which the mail voters made their voting decisions. By contrast, in our analysis we assign to each voter the level of PM10 on the election day and estimate

$$y_{it} = \gamma_0 + \gamma_1 PM10_{it} + \eta_{it}. \tag{C.2}$$

The estimate we obtain from estimating Equation (C.2) is

$$\gamma_1 = \beta_1 \left[(1-\alpha) + \alpha \frac{cov(PM10_{it}, Q_{it})}{Var(PM10_{it})}\right] = \beta_1 \left[(1-\alpha) + \alpha\delta\right]. \tag{C.3}$$

The bias resulting from measurement error is the the term in the square brackets. The bias is a function of the share of mail votes α and the serial correlation in the level of PM10, δ. Measurement error attenuates the estimates as long as

$$\frac{\alpha - 1}{\alpha} < \delta < 1. \tag{C.4}$$

Whether δ lies within this range is an empirical question. Note that δ is equivalent to a coefficient of a regression of PM10 several days before the election on PM10 on election day. In Table C.3, we run this regression and control for county and election date fixed effects. The results suggest that δ is positive and small: for most lags the coefficient is around 0.25 and for the lag $t-25$ the coefficient is zero. The estimates for δ being in the range $[0, 0.27]$ means that we have attenuation bias.

Table C.3: Regression of lagged PM10 on contemporaneous PM10

Dependent variable:	$PM10_{t-5}$	$PM10_{t-10}$	$PM10_{t-15}$	$PM10_{t-20}$	$PM10_{t-25}$	$PM10_{t-30}$
PM10 ($10\mu g/m^3$)	0.2847***	0.2284***	0.2700***	0.2791***	-0.0006	0.2358***
	(0.026)	(0.032)	(0.031)	(0.025)	(0.039)	(0.029)
Mean dep. var.	2.14	2.20	2.18	1.88	2.41	2.49
R^2	0.825	0.800	0.783	0.703	0.666	0.790
N	2770	2765	2759	2758	2756	2760
Controls						
County FE	✓	✓	✓	✓	✓	✓
El. Date FE	✓	✓	✓	✓	✓	✓

Notes: This table displays the results of separate OLS regressions of the level of PM10 (in $10\mu g/m^3$) on $t-s$ days before the election on the day of election on the level of PM10 on election day. Standard errors clustered at the county level are displayed in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Based on Equation (C.3), we can also quantify the attenuation bias. Suppose the share of mail votes is $\alpha = 0.2$ and the regression coefficient in Table C.3 is 0.25. In that case, the term in square brackets equals 0.85, which means that our estimates are 15% lower than the true effect because we assign the incorrect level of PM10 to mail votes.

The model can also be generalized to mail voting being up to S days before the election $s = 1, \dots, S$. Assuming that $cov(Q_{i,t-s}, PM10_{it}) = 0 \ \forall \ s$, the estimate becomes

$$\gamma_1 = \beta_1 \left[\left(1 - \sum_{s=1}^{S} \alpha_s\right) + \sum_{s=1}^{S} \alpha_s \frac{cov(PM10_{it}, Q_{i,t-s})}{Var(PM10_{it})} \right] = \beta_1 \left[\left(1 - \sum_{s=1}^{S} \alpha_s\right) + \sum_{s=1}^{S} \alpha_s \delta_s \right], \tag{C.5}$$

with α_s being the share of voters who vote by mail s days before the election. The results suggest that the most coefficients δ_s are around 0.25, which means that we likely have attenuation bias and our estimates represent a lower bound to the true effect.

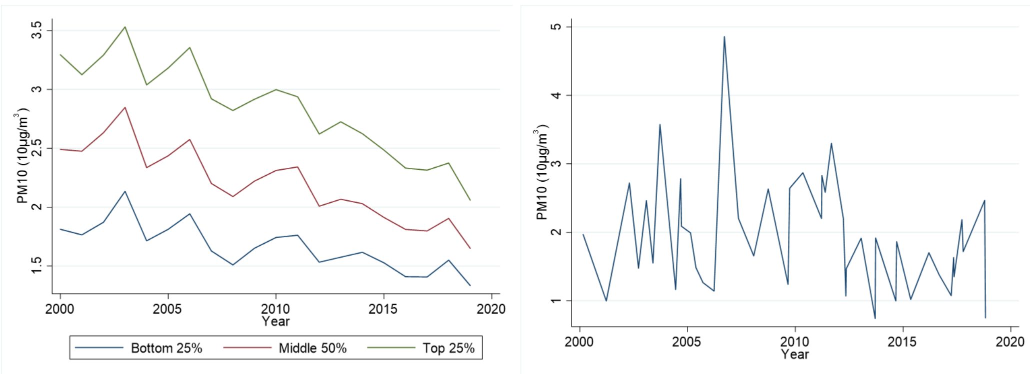
When we instrument for PM10 on the election day with wind directions on the same day, we can eliminate one part of the attenuation bias, namely the part governed by δ. Wind directions on the election day are plausibly orthogonal to pollution levels several days or weeks before, such that $\delta = 0$. However, the IV estimation cannot fully eliminate the attenuation bias, which in large parts is governed by $(1-\alpha)$, the share of people voting on the election day.

C.3 Additional Checks

C.3.1 More on the Variation of PM10

Figures C.3-C.6 illustrate the variation of particulate matter within and across counties. Figure C.3 shows the variation in average levels of PM10. Panel A reports the year-on-year variation dividing measuring stations into three groups: stations in the bottom 25% of average levels of PM10 across years, those in the middle 50%, and those in the top 25%. In each of these groups, the level of PM10 varies from year to year, but the variation follows the same decreasing trajectory in all three groups. Panel B reports average levels of PM10 across election dates.

Figure C.3: Fluctuation in Average PM10

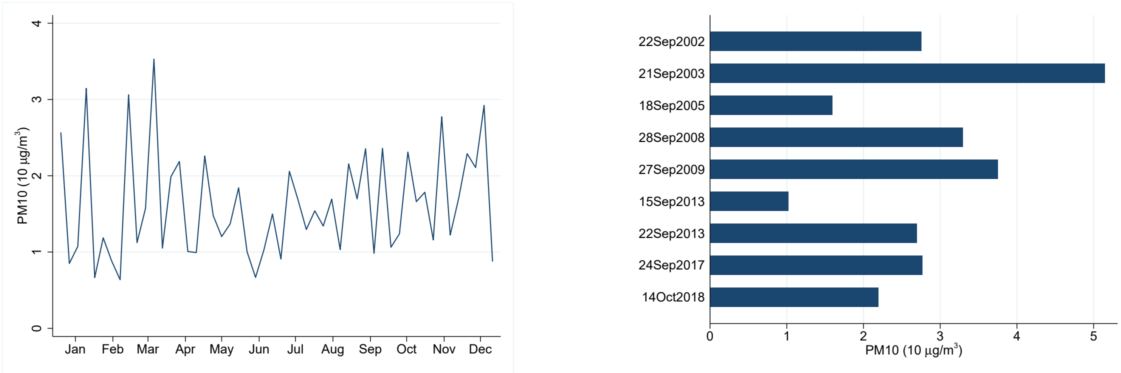


(a) Average PM10 by year (b) Average PM10 by election date

Notes: This graph displays the yearly average level of PM10 in three types of counties: 1) counties in the lowest 25% in terms of average pollution across election dates, 2) counties between the 25-th and 75-th percentile in this distribution, and 3) counties in the top 25% of this distribution.

Figure C.4 illustrates our identifying variation based on the example of Munich. Panel a) shows the fluctuation in the level of PM10 across Sundays in a given year, in this case in 2016. The level of PM10 fluctuates considerably around an annual mean of around $18\mu g/m^3$. The fluctuations are larger in winter than in summer, although even in winter pollution levels can be low. Panel b) illustrates the the identifying variation, namely the fluctuation in PM10 within the same city across election dates. The level of PM10 fluctuates considerably across election dates. In the regressions, we condition on election date fixed effects, which absorb fluctuations that are common to all counties in Germany, not just Munich.

Figure C.4: Example for Variation of Pollution within a County: Munich



(a) Pollution in Munich on all Sundays in 2016

(b) Pollution in Munich across Election Dates

Notes: This graph illustrates the variation of PM10 in Munich, a large city in the south of Germany. Panel a) shows the level of PM10 on all Sundays in 2016. Panel b) illustrates the identifying variation. It displays the level of PM10 on all the election dates in our sample period.

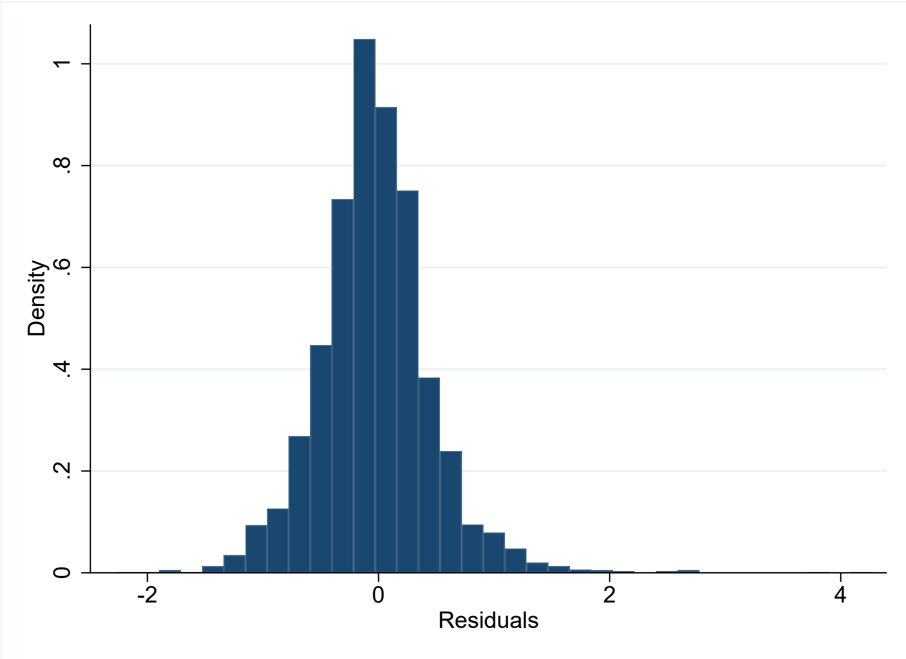
Figure C.5 illustrates the extent of identifying variation that is left after conditioning on fixed effects and controls. Each panel displays the residuals of PM10 after conditioning on county and election date fixed effects (Panel a) and additionally conditioning on weather controls (Panel b). Both graphs show that our estimation relies on a significant degree of identifying variation, even after controlling for weather.

Whereas Figure C.5 shows the amount of identifying variation in the sample, Figure C.6 shows how the identifying variation varies across counties. The figure displays the distribution of the within-county variation in PM10 after conditioning on election date fixed effects. Most counties have a within-standard deviation between 0.5 and 1, although there are some positive outliers, i.e. counties with a particularly high within-county standard deviation of PM10.

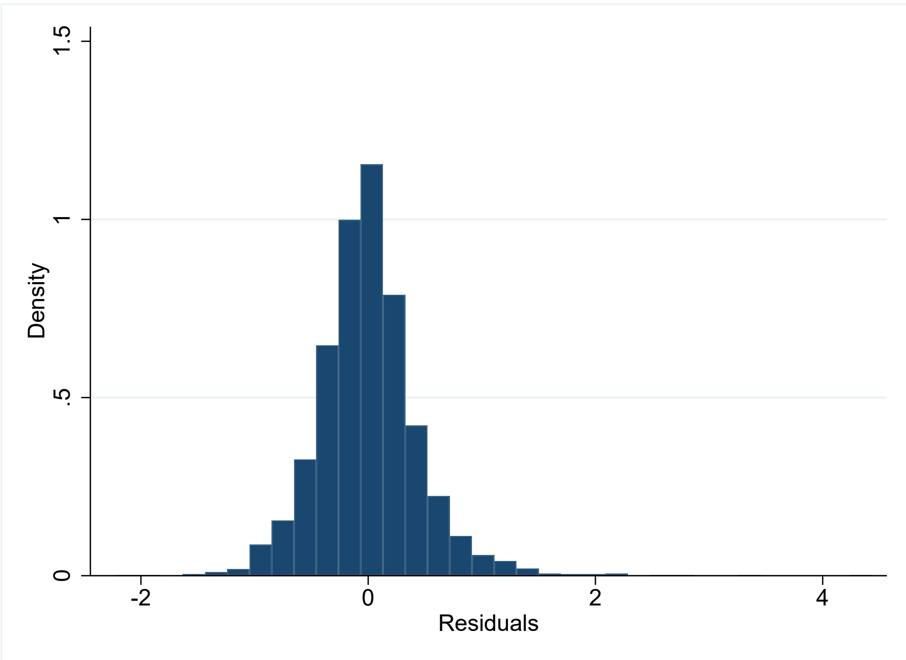
C.3.2 States with territorial reforms

Three states underwent substantial territorial reforms during our sample period: in 2007 Sachsen-Anhalt moved from 24 to 14 counties, in 2008 Sachsen moved from 29 to 13 counties, and in 2011 Mecklenburg-Vorpommern moved from 18 to 8 counties. The goal of these reforms was to reduce the number of counties within a state, which was achieved by including counties into other existing ones, merging counties to form a brand new entity, or counties being dissolved with their territory being redistributed between various

Figure C.5: Residuals of PM10



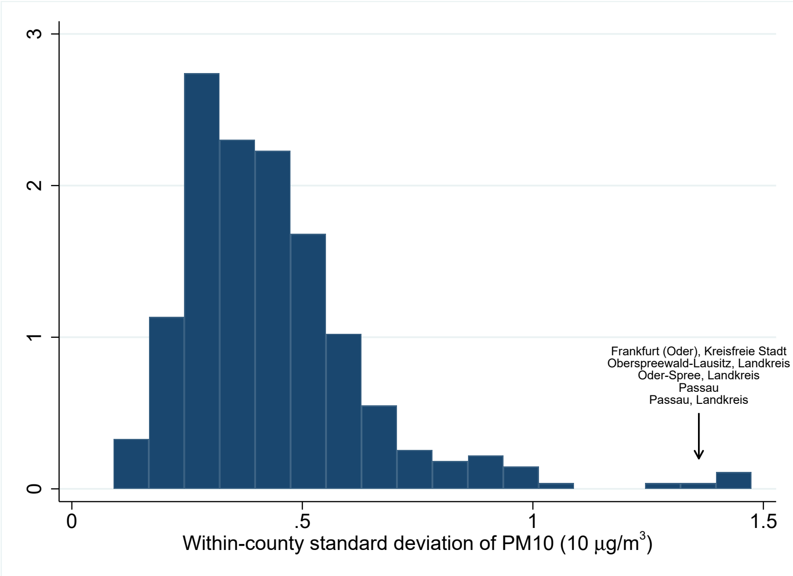
(a) Residuals of PM10 Conditional on Two-way FE



(b) Residuals of PM10 Conditional on Two-way FE and Weather Controls

Notes: This graph displays the distribution of the residuals of PM10 after conditioning on two-way fixed effects (Panel a), and after conditioning on two-way fixed effects and weather controls (Panel b). One unit equals $10\mu g/m^3$.

Figure C.6: Histogram of within standard deviations



Notes: This graph displays the distribution of the within-county standard deviations of PM10 after conditioning on election date fixed effects.

existing or newly formed counties.

It is practice in the literature to "reconstruct" the new territorial entity for the entire estimation period. The challenge for our analysis is that county-level data pre- and post-election are not readily comparable. In order to build a panel dataset for counties in these states, we require information on voting results *before the reform* based on the county definition *after the reform*. We construct this data based on municipality-level voting data, which we obtained from the statistical offices of the three states along with correspondence files that link municipalities and counties.

All of the results presented in the main analysis of this paper use voting and socio-economic observations for these three states thus created. However, we also decided to conduct a robustness check employing the same specifications reported in Table 4.2 excluding the three states that saw counties' reforms from the estimation sample.

The results are presented in Table C.4. The sample size decreases by 90 observations, but the coefficients remain quantitatively and qualitatively unaltered.

C.3.3 States fixed effects

In Germany many important political decisions are made at the state level. This means that neighbouring counties belonging to different states can experience relatively similar levels of air pollution while being subject to considerably different economic and environmental regulations. For this reason, we re-estimate our main models additionally including a state fixed effect. The results, displayed in Table C.5, are identical to those reported in Table 4.2.

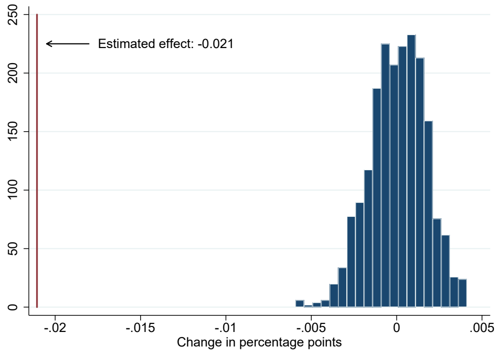
C.3.4 Balancing tests

A potential concern with our identification strategy is that pollution levels may be driven by economic or shocks. To test whether economic factors are systematically related to changes in pollution, we regress three variables – population, GDP per capita, and the employment rate – on the level of PM10 and control for fixed effects and weather controls. The results, shown in Table C.6, do not point to a systematic relationship between pollution and any of these three variables. In none of the cases do we find significant effects. Although this is no proof of the absence of omitted variables, we view these results as one piece of evidence along with the placebo tests, permutation tests, and instrumental variable strategy.

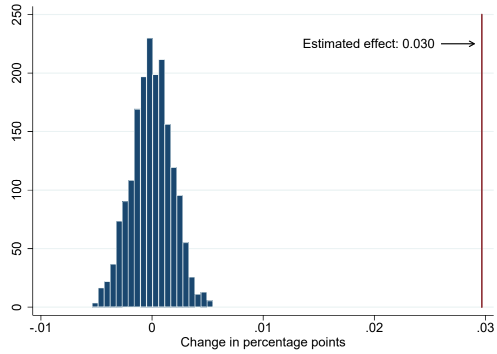
C.3.5 Permutation tests

To corroborate our identification strategy — and to exclude that our estimates are the result of fitting noise — we perform permutation tests. Within each county, we randomly re-shuffle the level of PM10 across election dates. For example, instead of the pollution level in Munich on the day of the state election in 2018, the procedure assigns the pollution level of the federal election in 2005 or the level on the day of some other election. For each outcome, we perform 500 permutations and regress the outcome on PM10, two-way fixed effects, and weather controls. In all three panels, our estimates based on the true levels of pollution (Panel B of Table 4.2) are far away from the distribution of placebo estimates. In none of the three permutation tests could we find a single placebo estimate that is more extreme than our estimate. These findings clearly reject the notion that our estimates are the result of fitting noise.

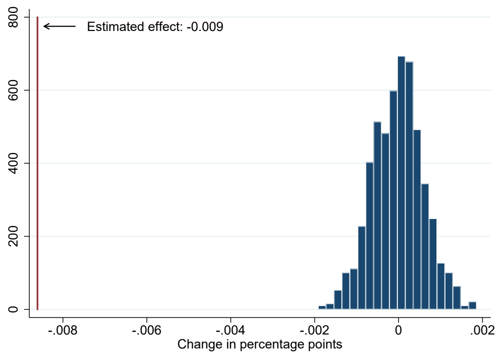
Figure C.7: Permutation Tests



(a) Vote Share of Incumbent Parties



(b) Vote Share of Established Opposition



(c) Vote Share of Other Parties

Notes: Each panel displays the distribution of 500 placebo estimates of a regression of a vote share on the level of PM10, controlling for election date and county fixed effects as well as weather controls. In each permutation, the level of PM10 was randomly re-shuffled within a county. The vertical lines indicate the estimates based on the true pollution levels, as shown in Panel B of Table 4.2.

Table C.4: Removing States with county reforms

Outcome:	Vote Share of Incumbent Parties (1)	Vote Share of Established Opposition Parties (2)	Vote Share of Other Parties (3)	Turnout (4)
A. Without controls				
PM10 ($10\mu g/m^3$)	-0.0207***	0.0263***	-0.0056***	0.0014*
	(0.003)	(0.004)	(0.001)	(0.001)
Mean dep. var.	0.48	0.42	0.10	0.69
R^2	0.575	0.682	0.901	0.962
N	2680	2680	2680	2680
Controls				
County FE	✓	✓	✓	✓
El. Date FE	✓	✓	✓	✓
B. With controls				
PM10 ($10\mu g/m^3$)	-0.0219***	0.0284***	-0.0064***	0.0012
	(0.003)	(0.004)	(0.001)	(0.001)
Mean dep. var.	0.48	0.42	0.10	0.69
R^2	0.605	0.703	0.907	0.963
N	2680	2680	2680	2680
Controls				
County FE	✓	✓	✓	✓
El. Date FE	✓	✓	✓	✓
Weather	✓	✓	✓	✓
Ozone	✓	✓	✓	✓
Demographics	✓	✓	✓	✓
El. Type FE	✓	✓	✓	✓
Turnout	✓	✓	✓	

Notes: This table displays the results of the same OLS regressions presented in Table 4.2, excluding those states that experienced reforms of counties in the sample period (i.e. Mecklenburg-Vorpommern, Sachsen and Sachsen-Anhalt). Standard errors clustered at the county level are displayed in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.5: Including States fixed effects

Outcome:	Vote Share of Incumbent Parties (1)	Vote Share of Established Opposition Parties (2)	Vote Share of Other Parties (3)	Turnout (4)
A. Without controls				
PM10 ($10\mu g/m^3$)	-0.0198***	0.0278***	-0.0080***	0.0012
	(0.003)	(0.004)	(0.002)	(0.001)
Mean dep. var.	0.48	0.42	0.10	0.69
R^2	0.576	0.685	0.893	0.961
N	2770	2770	2770	2770
Controls				
County FE	✓	✓	✓	✓
State FE	✓	✓	✓	✓
El. Date FE	✓	✓	✓	✓
B. With controls				
PM10 ($10\mu g/m^3$)	-0.0205***	0.0291***	-0.0087***	0.0010
	(0.003)	(0.004)	(0.002)	(0.001)
Mean dep. var.	0.48	0.42	0.10	0.69
R^2	0.604	0.705	0.902	0.961
N	2770	2770	2770	2770
Controls				
County FE	✓	✓	✓	✓
State FE	✓	✓	✓	✓
El. Date FE	✓	✓	✓	✓
Weather	✓	✓	✓	✓
Ozone	✓	✓	✓	✓
Demographics	✓	✓	✓	✓
El. Type FE	✓	✓	✓	✓
Turnout	✓	✓	✓	

Notes: This table displays the results of the same OLS regressions presented in Table 4.2, additionally including states fixed effects. Standard errors clustered at the county level are displayed in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.6: Balancing Test: Does Pollution Predict Economic Outcomes?

Outcome:	Total Population	GDP per capita	Employment rate
A. No controls			
PM10 ($10\mu g/m^3$)	-499.423	-56.688	-0.002
	(656.322)	(137.329)	(0.001)
Mean dep. var.	214509.95	31128.08	0.76
R^2	0.998	0.955	0.986
N	2770	2770	2770
Controls			
County FE	✓	✓	✓
El. Date FE	✓	✓	✓
B. With controls			
PM10 ($10\mu g/m^3$)	-665.484	-51.529	-0.002
	(632.854)	(146.427)	(0.001)
Mean dep. var.	214509.95	31128.08	0.76
R^2	0.998	0.955	0.987
N	2770	2770	2770
Controls			
County FE	✓	✓	✓
El. Date FE	✓	✓	✓
Weather	✓	✓	✓
Ozone	✓	✓	✓

Notes: This table displays the results of an OLS regression of the outcomes listed at the top on the air concentration of PM10 (in $10\mu g/m^3$) and the controls listed at the bottom. Standard errors clustered at the county level are displayed in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.