Determinants of the Voluntary Provision of Public Goods: Experimental Investigations

Dissertation

zur Erlangung des akademischen Grades
„doctor rerum politicarum“ (Dr. rer. pol.)

eingereicht beim Promotionsausschuss
der Fakultät für Wirtschafts- und Sozialwissenschaften
der Ruprecht-Karls-Universität Heidelberg

von
Johannes Heiko Diederich
geboren 5. August 1981 in Wiesbaden

im August 2013

Acknowledgements

Enormous thanks goes to my wife Judith, who is the best, for her love and friendship. Besides, I greatly thank my parents Werner and Gerlinde, who always left me with freedom of choice regarding my educational course, for their devoted support. I also want to express my deep gratitude to Prof. Timo Goeschl, Ph.D. for his continuous support, close supervision, and the great working environment at the Chair of Environmental Economics in Heidelberg. Besides, I am very grateful to Prof. Dr. Andreas Lange for serving as the second referee for this dissertation, and to Dr. Israel Waichman for productive and enjoyable coauthorship. A multitude of additional persons would have to be mentioned to complete this list. Since the resulting tremendous inflation of the length of this dissertation seems prohibitively costly to both the reader and the author, this is omitted.

Table of Contents

Introduction 1

Paper 1

To Give or Not to Give: The Price of Contributing and
the Provision of Public Goods 9

Paper 2

Willingness to Pay for Voluntary Climate Action and Its Determinants:
Field-Experimental Evidence 47

Paper 3

Motivational Drivers of the Private Provision of Public Goods:
Evidence From a Large Framed Field Experiment 79

Paper 4

The Pure Group Size Effect in Linear Public Good Experiments
with Large Groups 101

Paper 5

The Effect of Ambient Noise on Cooperation in Public Good Games 145

Conclusion 167

Introduction

The private provision of public goods (PPPG) is still one of the most fascinating puzzles in economics. Underpredicted by standard economic theory but outrightly evident in empirical evidence, its presence opened doors for new methods to enter the economist's toolkit and helped birthing the nowadays more vibrant than ever field of behavioral economics.[1] The essence of the puzzle of PPPG has been the question: *Why* do people voluntarily give to provide public goods? Even with decades of research in economics and other disciplines, this question is far from being answered fundamentally such that a uniform theory of PPPG could be formulated. Thus, there are important research gaps to fill along the incremental way to provide an answer.

Closely related to the question of *causes* is the question of *determinants*: What affects voluntary giving to public goods? While the answers to the first question mostly lie in the realm of unobservable physio-psychological drivers and motivations[2], the second question asks for *observable* conditions, traits, or contexts and their effects on individuals' contributions to public goods or their propensity to give, without much focus on underlying mediators. Insights regarding the determinants of PPPG can benefit society—apart from satisfying fundamental scientific interest—along two dimensions: First, they can help advancing research about the fundamental causes of PPPG on its way to a unified theory by inspiring new potential motivations or mediators or by having implications for known ones. Second, they may in fact speak to regulators (or fundraisers) who want to encourage PPPG where efficient and who wonder about how to create a nurturing environment. While my personal interest in this research was sparked by the first of these returns, the second one cannot be denied.

[1] The interested reader is referred to Olson (1965) or any introductory microeconomic text for the neoclassical results of the theory, to, e.g., Schokkaert (2006) for a summary of empirical evidence, and to Camerer et al. (2011) and Kagel and Roth (1995) to find reviews of the development of behavioral and experimental economics.

[2] See, e.g., Kolm and Ythier (2006) for an overview.

As the reader will find out, variants of the question of *determinants* make up, for the most part, the research questions of the five articles that constitute this dissertation. Thus, while each of the five papers largely stands for itself, the articles are far from being independent from each other. One purpose of the introductory summary of the papers following below is, besides providing an overview, elucidating their interrelation by the common question of the determinants of voluntary provision.[3]

On a methodological note, the research of the five articles of this dissertation resides on the overlap of public economics, environmental economics, and behavioral/experimental economics. In particular, while public economics provides the basic theoretical framework of PPPG, environmental economics is the realm of most applications or contexts, and behavioral/experimental economics is from where I borrow the dominant method—experiments. Experiments are a natural methodological choice for targeting the determinants of giving to public goods: Among the available empirical strategies, experiments provide a particular high amount of control to the researcher, who can design appropriate experimental treatments to isolate or identify specific determinants. The five articles contained in this dissertation report on three different experiments, two of which were online experiments with subjects from the general population. In one experiment, the subject sample was representative for the Internet-using part of the German population. This experiment (Papers 1 to 3) was a "framed" field experiment (Harrison and List 2004) in which both the subject sample and the experimental stakes were "from the field". The second, "artefactual" field experiment (Paper 4) was a typical laboratory design conducted with subjects from the general population via Internet. The third experiment (Paper 5) was a standard laboratory experiment using student subjects.

Synopsis

The first article targets the probably most fundamental determinant of demand: the price. The paper presents and analyzes an online "framed" field experiment on the decision to contribute to a (real) public good which exogenously and directly varies the

[3] In providing this overview, I will largely refrain from providing references in this introduction. See the particular papers for a detailed collection of the relevant literature.

price of contributing one unit of the public good. The paper reports results for 2,440 subjects who were representative for the Internet-using population of Germany. Each subject was assigned to one of 50 different amounts of money he or she had to give up in order to provide the contribution. In contrast to earlier field-experimental results from indirect price variations using match ratios and rebate rates, the theoretical prediction of a clear negative relationship between price and the decision whether to contribute is borne out by the experimental evidence: In the analysis, my coauthor Timo Goeschl and I find a price effect that is highly significant and negative, but also small and estimate the mean elasticity of the probability to contribute as -0.31. As this finding aligns with empirical evidence that public goods exhibit lower price elasticities than private goods, the result provides one argument for research on "non-price" determinants of public good contributions. In correlating sociodemographic characteristics of the highly heterogeneous subject sample with their choices, we provide strong support for the level of education being an important determinant of public good contributions.

In the second article, we analyze the same experimental choices in the context of the particular real-world public good we used and with respect to additional, potentially context-specific determinants. The public good in the experiment was greenhouse gas emissions reductions, which do not only pose a natural opportunity to study price effects through *direct* price variation but is of interest in its own right as some voices in the public debate on climate change have been emphasizing "voluntary climate action" as a potential remedy or at least an important part of the solution strategy to climate change. In order to arrive at a willingness to pay (WTP) estimate for a voluntarily provided unit of the public good, the contribution decision subjects faced in the experiment was framed in terms of the simple dichotomous choice between the guaranteed reduction of 1 metric ton of CO_2 emissions and the randomly assigned cash amount.[4] In the statistical analysis, in order to reduce a potential bias of WTP estimates from the presence of "fat tails" in the empirical WTP distribution in our data, we jointly estimate

[4]Thus, the experimental design resembles a standard dichotomous choice valuation question in contingent valuation studies but exhibits non-hypothetical properties. By constraining the contribution to one unit, the design focuses on the "extensive margin" of giving. In contrast, many public good experiments in the literature focus on variations at the "intensive margin" without price variation: By letting subjects decide on how much of an endowment to contribute, the decision is on how many "units" of the public good to provide at a (usually implicit) fixed price.

the choice of the emissions reduction, the non-zero probability of complete indifference to the offered choice, and the endogeneity of prior knowledge about the particular public good in a three-equations mixture model. The results regarding the determinants of the contribution choice are mainly threefold: First, the analysis confirms education as the outstanding correlate among subjects' sociodemographic characteristics. Second, the joint model uncovers interesting subtleties within the submodels, suggesting that younger and male subjects demonstrate better endogenous knowledge about climate change related figures but, at the same time, are less likely to participate in any voluntary climate action at all. Third, additional controls, potentially specific to this public good, deliver significant additional determinants. Some of these additional controls come from data about the time and likely location of a subject's experimental choice within Germany, which allow us to match regional weather data as well as data on the media coverage of the issue of climate change to each subject's choice. The main result here is that higher outdoor mean temperatures during the days prior to the choice significantly increase the probability of choosing the emissions reduction. More controls come from answers to a questionnaire administered after the experimental choice. These variables suggest that higher expected benefits of emissions reductions, both personal and, more so, for future generations, as well as an awareness of a negative impact of personal lifestyle on the climate increases the propensity to choose the emissions reduction. Results regarding WTP revealed by subjects are ambiguous: Despite the use of the joint mixture model and a considerably higher bid range than previous studies on the WTP for voluntary climate action, the issue of "fat tails" renders the estimates sensitive to model assumptions. The most conservative lower-bound estimate would be €6.30 for mean and €0.30 for median WTP for 1 ton of voluntary CO_2 emissions reductions.

The third article exploits the statistical power of a dataset of 12,624 observations to (1) investigate the presence and strength of certain causal *motivations* driving the decision to contribute to a public good and (2) review some of our previous results regarding determinants. The dataset consists of the same observations of the online framed field experiment analyzed in Papers 1 and 2 (referred to as the "baseline treatment" in this paper) but also includes observations from additional subjects assigned to various treatment conditions targeted at isolating altruism, a "warm glow" of giving, image motivation, and equity concerns. Furthermore, each of the 6,312 subjects in total was observed in a second contribution choice, allowing for within-subject analysis of treatment effects besides between-subjects analysis. In the results, evidence on treatment effects is mixed and points to significant effects of framing and the sequence of presenting options. Regarding the determinants of the contribution choice in the experiment, the results confirm education as the most pronounced and robust sociodemographic determinant. In addition, the positive effect of age and being female on the propensity to give becomes marginally significant, in contrast to the findings in Papers 1 and 2 and presumably due to the statistical power of the full sample. Lastly, results largely confirm the positive causal effect of outdoor temperatures.

In the fourth article, Timo Goeschl, Israel Waichman and I follow up on the observation that many real-world public goods, such as climate change mitigation, are privately provided among large groups. Other examples such as national and international charitable organizations, open source software programs, or online encyclopedias are likewise consumed by literally millions or billions of individuals who are, at the same time, potential contributors. This observation renders *group size* a potentially important determinant of PPPG. Lab-experimental insights about an effect of group size is—for large groups, *pure* public goods, and in the standard linear environment—almost entirely based upon one piece of work by Isaac, Walker, and Williams (1994) while theory provides ambiguous predictions and empirical evidence is mixed. We present major methodological improvements to Isaac, Walker, and Williams' design and run a "long-term" online experiment with a diverse sample of the German population and group sizes of ten, 40, and 100 subjects per group. Our procedure produces data of higher quality than Isaac, Walker, and Williams' dated procedure could deliver, and the results confirm the positive relationship of group size and individual contributions, but with higher statistical confidence. Thus, individuals provide more of a pure public good in larger groups. In addition, our data provide evidence that this positive effect is also present in subjects' first-order beliefs and that subjects consistently underestimate others' contributions across group sizes. As in Papers 1 to 3, the highly heterogeneous nature of the subject sample allows to also analyze the role of several sociodemographic characteristics. In support of our previous findings, education is a significant determinant of contributions which, in contrast to our first experiment, include variations at the intensive margin. Also, age positively correlates with contributions but estimates are mostly insignificant. In contrast to our previous results, females give less in this "artefactual" field experiment. A sociodemographic attribute not elicited in Papers 1 to 3 is a rural residential environment, which positively correlates with contributions in this experiment.

In the fifth article, I take up our previous finding that certain environmental conditions (such as temperature) may determine differences in public good provision. In particular, I turn to the effects which *environmental stressors* (such as noise, air pollution, extreme temperatures, or crowding) might have on economic behavior. Especially in the controlled environment of economic experiments, this has not been subject to systematic research so far as traditionally, economics has been abstracting economic behavior from the influence of environmental stressors, much as it used to do for moods and emotions. Environmental stressors may pose relevant externalities in the economy, however, if three conditions are fulfilled: (1) The effect of environmental stressors on economic behavior is significant, (2) the ambient levels of environmental stressors are subject to (man-made) change, and (3) adaptation of humans to altered levels is imperfect or costly. In this paper, I address condition (1) and test for an effect of acute noise exposure on cooperative behavior in a standard linear public good experiment in the laboratory. I find a negative effect of noise exposure on contribution levels that is statistically significant for certain subgroups of the sample. This direction of the effect points against a mediation of the effect by physiological stress, given the available evidence on this mediator in the literature.

References

Camerer, C. F., Loewenstein, G. and Rabin, M. (2011). *Advances in Behavioral Economics*, Princeton University Press, Princeton.

Harrison, G. W. and List, J. A. (2004). Field experiments, *Journal of Economic Literature* **42**(4): 1009–

1055.

Isaac, R. M., Walker, J. M. and Williams, A. W. (1994). Group size and the voluntary provision of public goods: Experimental evidence utilizing large groups, *Journal of Public Economics* **54**(1): 1–36.

Kagel, J. H. and Roth, A. E. (1995). *The Handbook of Experimental Economics*, Princeton University Press, Princeton.

Kolm, S.-C. and Ythier, J. M. (2006). *Handbook of the Economics of Giving, Altruism and Reciprocity*, Elsevier.

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*, Vol. 124, Harvard University Press.

Schokkaert, E. (2006). The empirical analysis of transfer motives, *in* S.-C. Kolm and J. M. Ythier (eds), *Handbook of the Economics of Giving, Altruism and Reciprocity*, Vol. 1, Elsevier, pp. 127–181.

To Give or Not to Give: The Price of Contributing and the Provision of Public Goods*†

Johannes Diederich Timo Goeschl

Abstract

We examine the relationship between the price of giving and the decision to contribute in a framed field experiment ($n = 2,440$). In a departure from previous research using match rates and rebates, we vary the price of contributing to the public good directly. Treatment groups differ between subjects by the amount of money subjects have to give up in order to provide one unit of the public good. In contrast to earlier results, the theoretical prediction of a clear negative relationship between price and the decision whether to contribute is borne out by the experimental evidence. We estimate the mean elasticity of the probability to contribute as -0.31. The direct price effect is robust across specifications including sociodemographic controls for the highly heterogeneous, Internet-representative non-student sample of subjects.

Keywords: voluntary contributions; public goods; price elasticity; field experiment; online experiment

JEL Classifications: C93, D12, H41

1 Introduction

The private provision of public goods has been attracting sustained attention from economists for several decades now. A natural and recurrent question within this field has been how the price of giving to the public good affects its supply (see e.g. Vesterlund 2006). Answering this question requires observing variations in the price of giving, and relating those price variations to observed variations in giving decisions. The empirical literature, starting with Feldstein and Taylor (1976) and Feldstein and Clotfelter (1976), has been exploiting observable variations in the marginal income tax rate between households to study the price effect in settings in which contributions to public or charitable

*The authors are grateful to Andreas Lange, John List, and seminar participants at Cambridge, Heidelberg, the London School of Economics, Manchester, and the NBER Summer Institute 2012 for helpful comments. The usual waiver applies. We also thank Holger Geißler and Soreen Schroll at YouGov for cooperation, Dr. Svenja Espenhorst and Dennis Mignon at First Climate for support in acquiring EU ETS allowances, and Ruth Fieber, Christina Grimm, and Thomas Scheuerle for student assistance. Financial support by the German Science Foundation (DFG) under grant GO1604/1 is gratefully acknowledged.

†This paper is available as NBER Working Paper No. 19332. An early version of this paper is available under Diederich and Goeschl (2011).

causes are tax deductible and, therefore, subsidized.[1] More recently, the focus of empirical research in this area has shifted to gathering evidence from field experiments carried out in a fundraising context. These have provided new estimates of price elasticities of giving (Karlan and List 2007, Eckel and Grossman 2008, Karlan et al. 2011, Huck and Rasul 2011). A major benefit of field experiments is that the researcher is not restricted by given variations in marginal income tax rates. Instead, exogenous variations in the price of giving can be introduced in a controlled manner and independent of subjects' household income. The typical vehicle for such exogenous variations have been changes in the so-called "match ratio", i.e. the amount of money that some third party will contribute for every unit of money donated by the subject (Karlan and List 2007, Eckel and Grossman 2008, Karlan et al. 2011, Huck and Rasul 2011).[2] Converted into theoretically equivalent price variations, the evidence from variations in match ratios forms the basis of our current empirical understanding of the price effect.

Using variations in match ratios as a measure of the price effect in giving decisions offers a number of advantages, as it allows randomized assignment to subjects, is a familiar feature of fundraising, and is easy to implement.[3] It is also theoretically straightforward: The conversion of match ratios into theoretically equivalent price changes is simple. A 1:1 (1:2) match ratio should have the same effect as a reduction in the price by 50% (67%).

At the same time, getting at the price effect indirectly via match ratios also has drawbacks. One important drawback is the assumption implicit in the indirect approach that subjects' response to variations in match ratios can safely be interpreted as those of the theoretically equivalent price variation. The validity of this assumption has been thrown into doubt by recent experimental evidence. For example in the case of contributions to public goods, match ratios and their theoretically equivalent rebate rates give rise to systematically different behavior among potential contributors, both in the laboratory (Eckel and Grossman 2003) and in the field (Eckel and Grossman 2008).

[1]See, e.g., Peloza and Steel (2005) for a comprehensive review of empirical estimates of the price elasticity of giving.

[2]For comparative purposes, Eckel and Grossman (2008) also uses rebate rates as a vehicle.

[3]Even though one drawback is that matches and rebates only allow the experimenter to reduce the price of giving, not to increase it.

The introduction of a match leads to a greater effect on giving than the introduction of its theoretically equivalent rebate rate. Similarly, in an experiment involving private goods, Davis and Millner (2005) compare rebates, matches, and direct price variations that should be equivalent on theoretical grounds. They find that there are systematic differences in the quantity responses to these vehicles. This implies that for private goods, "match rate elasticities", i.e. price elasticities derived on the basis of variations in match ratios, are likely to be biased estimators of the price elasticity in a narrow sense. While we are not aware of comparable evidence for public goods, these results show at a minimum that the empirical equivalence of variations in match rates and in prices cannot be taken for granted.

In this paper, we present the design and report on the results of a framed field experiment that provides a *direct* measure of the price effect in a decision whether to contribute or not. The direct price effect arises out of the treatment condition in the experiment: Different subjects are randomly assigned a different amount of money that they give up if they decide to supply one unit of the public good. The observed effect on the probability to contribute therefore closely relates to the notion of the price effect from the theory of the private provision of public goods (Bergstrom et al. 1986, Andreoni 1990). The decision whether to contribute or not presents a useful first target for a study of the direct price effect: While an immediate prediction of theory is that, all else equal, the share of contributors in a population is a decreasing function of the price of contributing, a number of papers have found little field experimental support for the conjecture. Neither variations in match rates (Karlan and List 2007, Eckel and Grossman 2008, Huck and Rasul 2011) nor in rebate rates (Eckel and Grossman 2008) appear to influence subjects' decision whether to contribute.[4] Similarly, in an empirical paper exploiting variation in marginal tax rates, Smith et al. (1995) find that the rebate rate does not impact on the decision whether to contribute to a rural health care facility.

The basic idea of simply using direct price variation as a treatment in an experiment on giving is, of course, not new. For example, Andreoni and Miller (2002) and Andreoni

[4]However, there is evidence that the presence of a lead donor in itself has a significant positive impact on the probability that some positive contribution will be given. See Huck and Rasul (2011) for a careful discussion of the possible mechanisms at work.

and Vesterlund (2001) introduce, in a laboratory-based within-subject dictator game design, a direct variation in the price of giving by changing how many units of their experimental endowment a dictator has to give up in order to transfer a unit to the recipient. However, the idea has to our knowledge not been used in the context of public goods provision and in a framed field experiment (Harrison and List 2004). This enables us to control for a number of subject attributes such as age (e.g. List 2004), gender (e.g. Andreoni and Vesterlund 2001), education (e.g. Karlan 2005) and culture (Ockenfels and Weimann 1999, Brandts et al. 2004, Brosig-Koch et al. 2011) that conceivably interact with the price effect and also to check for the presence of field price censoring among subjects.

The experiment was administered to a non-student population of 2,440 subjects, employing a between-subjects design. The real public good used in the experiment was verified CO_2 emissions reductions, a natural choice since they represent a real physical contribution to a public good, are perfectly uniform and individually traceable. Subjects were randomly assigned to one of fifty treatments, with the experimental price of contributing one metric ton of emissions reductions lying between €2 and €100, depending on the treatment group.

Based on this design, we estimate a direct price effect on the probability to contribute to the public good that is negative and statistically significant: On average, increasing the price for supplying a unit of the public good by €1 decreases the probability that the individual will contribute by around 0.1%. Estimated across all price treatments, the probability to contribute has a price elasticity of -0.31. There is some evidence of non-linearity in the price effect, but the net effect is vanishingly small within the treatment range. The direct price effect therefore confirms the theoretical prediction that, all else equal, the number of contributors is a decreasing function of the price of contributing. Our data do not provide evidence for the presence of a gender, age, or a culture effect in terms of either levels or elasticities. We find, however, support for the hypothesis that the level of education has a positive role in determining contributions to public goods.[5]

The paper proceeds as follows: We explain the experimental design considerations

[5]The same design can be used to draw conclusions about individuals' willingness to pay for voluntary climate action. See Diederich and Goeschl (2013) for an implementation.

and procedures in Section 2. Section 3 presents the empirical analysis and discusses the results. Section 4 concludes.

2 Experimental design

The estimation of the direct price effect on the individual probability of contributing relies on an experimental design that manipulates the price of giving to a public good. Basic economic intuition dictates that in a sufficiently heterogeneous and large population, a higher price of giving will be associated with fewer individuals deciding in favor of contributing. The intuition can be confirmed by introducing a unit price for the public good into a variant of Andreoni's 1989, 1990 classical impurely altruistic model.[6] The experimental implementation of the intuition combines the idea of direct price variation by the experimenter (e.g. Andreoni and Miller 2002, Andreoni and Vesterlund 2001) with the idea of controlled contributions to a public good or a charity explored by Kingma (1989), Eckel and Grossman (1996), Karlan and List (2007), Eckel and Grossman (2008), Karlan et al. (2011), to name just a few. The core feature of the treatment condition consists of different units of experimental pay-off that subjects have to give up in order to contribute one unit of the real public good. The real public good are verified CO_2 emissions reductions[7] and the unit is one metric ton. The emissions reduction is realized in the form of the documented and verifiable retirement ("deletion") of an emissions allowance (EUA) under the European Union Emissions Trading Scheme (EU ETS). Retiring one EUA lowers the total ceiling of the Scheme, and hence emissions, by one ton.[8]

[6] In this model and its variants in the literature, the price of the public good is conventionally normalized to one along with the private good. We provide a formal proof of the proposition that the number of contributors in a sufficiently heterogenous population decreases in the price of contributing to the public good in the appendix to this paper.

[7] Economists have long noted that voluntary emissions reductions to mitigate climate change constitute a close empirical counterpart to a contribution in a large public goods game (e.g. Nordhaus 1993). An obvious prerequisite is that subjects agree with the economists' characterization. We come to this in the next subsection.

[8] Among several possibilities, the regulatory framework of the EU ETS, regulating the bulk of industrial CO_2 emissions across EU member states, provides the most reliable and transparent technology for real contributions to global greenhouse gas emissions reductions in an experiment. First, retiring EUAs avoids the problem of *additionality* frequently encountered for project-based carbon offsets as the total cap of the EU ETS is binding and enforced. Second, each EUA is uniquely identified by its issue number and hence individually traceable. Third, EUAs are not paper currency and have therefore no curiosity value as a tangible private commodity. Total EU emissions for the trading period 2008-2012,

Subjects are randomly assigned to one of the fifty different treatment groups, differentiated by the price of contributing. The price of contributing ranges, in increments of €2, from €2 to €100, the upper bound reflecting a current estimate of the maximum marginal cost of emissions reductions per metric ton of CO_2 (Tol 2010). Subjects only decide whether to contribute or not at the given price. They do not learn about others' choices before, during, or after the experiment.

Subjects' choices are implemented under a random incentive system (RIS) (Grether and Plott 1979, Starmer and Sugden 1991, Lee 2008) in order to limit total cost of the experiment. The RIS is between-subjects (BS)(Tversky and Kahneman 1981, Baltussen et al. 2010, Abdellaoui et al. 2011) with odds of one in fifty that the subject's choice (of either cash or contribution) was realized. On the experimental screens, the BS-RIS is framed as a lottery in which the winners' prize choices will be implemented.[9]

Like in most lab experiments, both financial incentives and public good benefits in the present design are "on the house". An alternative procedure that was considered would have involved requiring subjects to give up own money when choosing to contribute to the public good. Our choice in favor of the standard lab procedure was mainly due to questions of practicality and the cost of time and effort to the subject of transferring funds in an Internet experiment from the subject to the experimenter.[10] The latter transaction costs are equivalent to an individual minimum price on the contribution that would be unobservable and therefore out of control of the experimenter. In the literature, there is an ongoing debate on potential effects of "house money" on contributions in public good experiments.[11] Based on these results, however, there is little evidence

the relevant one for this experiment, were capped at 1.856 billion tons.

[9] Between-subjects (BS) and within-subject (WS) RIS have been subjected to examination for possible biases. While BS introduces noise and decreases risk aversion, there is less evidence of a systematic bias for simple tasks Cubitt et al. (1998), Baltussen et al. (2010). In one example, BS-RIS has been shown to affect behavior in dictator games Sefton (1992) while for ultimatum games, behavior was unaffected Bolle (1990).

[10] For example, the infrastructure of our cooperation partner is not designed to facilitate payments from subjects to the company.

[11] The evidence on a "windfall" (Keeler et al. 1985) or "house money" (Thaler and Johnson 1990) effect in public goods experiments, and if so in which direction, is mixed. While the classic finding is that with house money individuals behave less risk-averse Thaler and Johnson (1990), Clark (2002) finds no significant difference in contribution behavior in a standard voluntary contribution mechanism (VCM) in the lab. Harrison (2007) reviews Clark's analysis of the data and identifies a decrease of contributors at the extensive margin by 8% when using house money. Engel and Moffat (2012) use a panel version of the double hurdle model on the same data and find that house money increases the probability of being a "potential contributor". Carlsson et al. (2013) find in a dictator game that subjects behave more

to inform whether price elasticities would be affected by a difference in contribution probabilities, if any.

2.1 Subjects and procedures

The choice of subjects and the procedures under which the experimental design is implemented constitute a "framed field" setting.[12] The design is administered to a non-student population of 2,440 subjects drawn from the approximately 65,000 Internet panel members of the German section of YouGov and are representative for Germany's Internet using population of voting age.[13] The choice of population has some significance for an experiment that relies on economists' view of emissions reductions as public goods contributions: Irrespective of age, sex, education, or political orientation, previous surveys have concluded that German citizens overwhelmingly accept the empirical veracity of climate change and its anthropogenic cause in the form of greenhouse gas emissions (European Commission 2008). An exit questionnaire was administered to all subjects that confirmed the prior evidence.

The Internet experiment ran in two sessions in May and July 2010.[14] Session 1 lasted from May 25th to June 2nd and generated 1,640 complete observations from 1,817 invitations to the 'baseline' treatment. Session 2 lasted from July 19th to 27th and generated 800 complete observations out of 888 invitations. The recruitment of subjects followed the standard routine in which panel members are invited via an email message to proceed to the poll via a hypertext link. The introductory screen then explained, as common with the pollster's regular surveys, the thematic focus of the poll (CO_2 emissions and climate change), the expected duration (ten minutes), and the payment (in form of a lottery).[15]

generously with house money than with own money both in the lab and in the field.

[12]Following the nomenclature of Harrison and List (2004), our design falls short of a "natural field experiment" by virtue of the setting, which is familiar, but not natural, and by virtue of the awareness of subjects that their choices are being observed.

[13]We test whether our sample differs from one drawn from the general population of German voters. Using two-sided t-tests, we reject the hypothesis that the means of the socio-demographic characteristics coincide at the 1% level. Our subjects are slightly more likely to be male, younger, and educated than the average German of voting age. Income is self-reported, and therefore the lower average income in the sample is unsurprising.

[14]Prior to the experiment we ran a set of pre-tests and a pilot experiment with 200 economics students at Heidelberg University to test the online implementation and refine the set of texts and questions.

[15]The polling company usually incentivizes panel members participating in a in polls through either a

Following the introductory screen, there was a filter screen to focus on German subjects.[16] Participants then faced a sequence of 10 to 13 computer screens, depending on their decisions.[17]

On average, 49 subjects were randomly assigned to each treatment group, differentiated by the experimental price.[18] The centerpiece of the experiment were two screens, the *information screen* that set up and the *decision screen* that collected the subject's choice. The *information screen* explained three features of the experiment, (1) the choice between a cash prize in Euros and the CO_2 emissions reduction, (2) a succinct explanation of how choosing the emissions reduction results in a real, reliable, and verifiable reduction in EU CO_2 emissions through the deletion of an EUA, and (3) an explanation of the RIS with odds of 100 in every 5,000.[19] Furthermore, the text reminded subjects of the purely public nature of the contribution. Like in other field experiments on public and charitable goods, the instructions did not contain further information on what the precise public goods effects of a contribution are.[20] Instructions were kept short and simple in order to avoid well-known biases and potential misinterpretations that arise when providing subjects with potentially choice-relevant information about the public good around the time of the contribution decision (Arrow et al. 1993).

The *decision* screen of the experiment explained how the subject's choice would materialize if the subject was drawn as a winner.[21] The screen then collected the subject's choice of either the specific cash award or the real emissions reduction, which were

piece-rate reward of approximately €1 for 20 minutes expected survey time or random (lottery) prizes, e.g. in the form of shopping vouchers.

[16]Subjects of other nationalities were redirected to other surveys running at the same time.

[17]The screens required an answer for each question by entering text or choosing at least one of the options given (including "I don't know" options) before being able to proceed to the subsequent screen. This helps to prevent subjects from "rushing" through a survey.

[18]The smallest group contained 31, the largest 66 subjects (standard deviation 6.4 subjects).

[19]The number of participants implied here is due to additional treatments running at the same time.

[20]When subjects are invited to contribute to give to a liberal political organization (Karlan and List 2007, Karlan et al. 2011), a public radio station (Eckel and Grossman 2008), to a children project of an opera house (Huck and Rasul 2011), or to CO_2 emissions reductions, information about productivity should matter. Despite this, giving decisions are typically poorly informed (Krasteva and Yildirim 2013). Other authors also find that when given the opportunity, subjects take only modest effort to access additional relevant information (Berrens et al. 2004) and no more than one third of subjects have a positive willingness to pay for relevant information (Fong and Oberholzer-Gee 2011).

[21]As in other polls by the polling company, all winners would be informed via a personal email message. Cash prizes were directly credited to the subject's personal account with the polling company. A member's account balance can be converted into a variety of shopping vouchers or, having reached a threshold of € 50, wired to the member's bank account. The retirement of EUA issue numbers was verifiable through a public-sector Internet site.

Table 1: Summary statistics of subjects' sociodemographics

Variable	Description	Mean	SD	Obs.
Female	Indicator variable for gender, 1 if female	0.469	0.499	2,354
Age	Subject's age (years)	45.42	14.68	2,352
Years of education	Based on subject's stated highest educational degree	12.27	3.213	2,299
Net income	Midpoint[a] of subject's monthly household net income category (€)	2,556	1,706	1,950
Eastern Germany	Indicator variable for residence in former GDR territory	0.1895	0.392	2354

Notes: [a] In our income approximation, for the 'less than € 500' category, we assume € 450. For the two categories above € 5,000, we assume € 8,000 for compatibility with German census data. The remaining categories have widths of € 500.

presented on the screen in a randomized ordering. Subjects that chose the cash prize were automatically directed to a screen that provided them with an non-incentivized opportunity to explain their choice, which we describe in more detail below.

The experiment concluded[22] with a set of follow-up questions eliciting subjects perceptions and beliefs about EUAs and emission reductions as well as sociodemographics (age, gender, income, education, residence). Correlation of the latter variables with subjects' profiles on record with YouGov was checked. The nature of the Internet experiment also allowed us to observe when exactly subjects completed the experiment and how much time subjects spent at each screen. Table 1 presents summary statistics of the sociodemographics.

2.2 Field price censoring

A well-understood challenge created by directly varying prices in order to determine the price effect is that it can give rise to field price censoring (Harrison and List 2004). Field price censoring, henceforth FPC, arises because prices for goods within the experiment are difficult to isolate from prices of those same goods or close substitutes in the real world (Harrison et al. 2002, Cherry et al. 2004, Harrison et al. 2004). In other words, there is a possibility that subjects perceive an arbitrage opportunity introduced by the experiment, biasing the observable contribution decision. In the present experiment, subjects who would otherwise have chosen the public good contribution might choose the cash prize instead because they believe that they are able to provide an equivalent

[22]Between subjects' choices and the questionnaire, the experiment administered a second choice containing a treatment condition. This paper focuses on the independent first choice only.

CO_2 emissions reduction at a lower total cost (including time and transaction costs) than the prize offered as an alternative.[23]

Two aspects are relevant for detecting the possible presence of FPC in the experiment. First, it is relatively costly for private individuals to purchase and delete EUAs at the going spot price (€15 per metric ton at the time of the experiment)—a fact that largely excludes the possibility of FPC from perfect substitutes.[24] A subset of subjects may be aware that a variety of imperfect substitutes exist at different prices and degrees of substitutability. The alternatives range from close substitutes such as having a EUA retired through a broker[25] or purchasing an emissions offset based on a carbon reduction project[26] to more remote substitutes such as making costly changes in everyday life to reduce one's own carbon footprint.

The second issue is that the researcher should expect a high degree of heterogeneity in subjects' knowledge about these substitutes and thus, in the levels of *perceived* field prices. In fact, subjects' information status and FPC may be interrelated phenomena: uninformed subjects may have an incentive to opt for the cash prize in order to make an informed decision later.[27] In the context of the experiment, therefore, there is no single explicit field price that will censor all responses. Instead, FPC would be driven by subjects' possible perception that field opportunities are available at certain prices (Harrison et al. 2004).

[23]For our purposes, FPC is present if a subject with a reservation price for the public goods contribution r_i accepts the experiment cash prize e_i even though $r_i > e_i$ simply because the field price of an equivalent contribution in the field $\hat{f}_i$ estimated by the subject (inclusive of transaction costs) obeys $e_i > \hat{f}_i$. In cases then where $r_i > e_i > \hat{f}_i$, the experimenter may mistakenly conclude that the unobservable reservation price r_i is smaller than e_i on the basis of the subject choosing cash instead of the good and therefore systematically understate the probability to contribute. Since there is no secondary market for retired EUAs, we need not be concerned about the situation $\hat{f}_i > e_i > r_i$ in which subjects opt for the EUA despite $r_i < e_i$ in order to pocket the arbitrage margin $\hat{f}_i - e_i$.

[24]The EU ETS gives private individuals the opportunity to open an account for a fixed fee of €200. The account does not include trading, though. Obtaining EUAs in small numbers is not straightforward without an additional intermediary.

[25]At the time of the experiment, there existed only very few opportunities via the internet to commission EUA retirements, none of them in German language. One example is the UK based Carbon Retirement Ltd. (www.carbonretirement.com) with a price of around €23 per ton of CO_2 at the time of the experiment.

[26]For example, *Certified Emissions Reductions* (CER) under the United Nations Clean Development Mechanism (CDM). Being available at various grades (e.g. the "Gold Standard", www.cdmgoldstandard.org), prices exhibit significant heterogeneity. Typically, some grades of CERs were available below and above the EUA spot price at the time of the experiment.

[27]Our design prevents this effect to a certain extent since the online survey implementation allows subjects to search the internet while doing the survey, or leave the survey and take it up again later. We do not find much evidence on this behavior, though (cp. footnote 36).

To detect subjects potentially constrained by FPC without interfering with subjects' information status, we follow the strategy of a debriefing questionnaire as in Coller and Williams (1999) and Harrison et al. (2002). Our identification strategy is threefold and consists of several follow-up questions after subjects chose their desired prize. First, we gave subjects who chose the cash price the opportunity to agree to three statements following the *decision screen*. As a result, this FPC "filter" contained all subjects that did not check the first option ('*Given the two prizes, I did not want to forgo the chance of winning x Euros*'), but checked the second option ('*I believe that there is another way for me to reduce CO_2 emissions by one ton for less than x euros.*') or made a qualitatively equivalent statement in the open-ended third option ('*I had other reasons for choosing the cash prize, namely...*'). Second, we asked all subjects to estimate current EUA spot prices and the availability of EUAs to private individuals in the follow-up survey. Third, an open-ended question in the survey asked all subjects to list existing efforts to mitigate climate change. Thus, while the first and the third part of the strategy aimed at FPC from both perfect and imperfect field substitutes, part two targeted perfect substitutes only. Section 3.3 reports on several robustness checks for our results with respect to a potential bias from FPC.

3 Results and Discussion

2,440 subjects completed the experiment with a median completion time of 5 minutes.[28] A total of 382 subjects in the experiment contributed to the public good. Of the 2,058 subjects that decided not to contribute, 86 subjects expressed some form of disbelief about the payment or the real provision of the public good in answers to open-ended survey questions and were excluded from the subsequent analysis.[29] We observe contributions in each of the fifty price treatments between €2 and €100. In forty-eight treatments, the share of contributors exceeds zero at the 5% level of significance, using

[28]Average completion time was 1 hour 17 minutes. The difference between mean and median is largely driven by a small fraction of outliers (approx. 3%) in which subjects availed themselves of the opportunity to leave the survey and continue hours or days later.

[29]Results are not sensitive to their inclusion or exclusion.

a one sided t-test.[30]

The parametric analysis of subjects' discrete choice is based on a probit model. We estimate five specifications of increasing richness. The most parsimonious estimation of the direct price effect has the form

$$\Pr(Y_i = 1) = \alpha_0 + \alpha_1 P_i + \varepsilon_i$$

with $Y_i = 1$ if subject i chose the contribution to the public good and P_i denoting the cash prize offered to subject i. In several steps, in which additional controls are introduced, we arrive at the final specification of the form

$$\Pr(Y_i = 1) = \gamma_0 + \gamma_1 P_i + \gamma_2 P_i^2 + \gamma_2 N_i + \gamma_3 P_i N_i + \varepsilon_i$$

that allows for the possibility of a non-linear price effect and controls for non-price effects driven by a vector N_i of the subject's attributes as well as for interaction effects between the price of contributing and the attributes N_i.

Tables 2 and 3 report the probit coefficient estimates and the marginal effects, respectively, of the five specifications. The first two columns in both tables report on price-only specifications: Column 1 corresponds to model 1 while column 2 estimates a linear and a non-linear price effect. The second three columns augment the price-only model by including socioeconomic attributes and additional controls for experimental session, day, and daytime. Column 3 shows the coefficient estimates of the linear price model with controls for the standard suite of subjects' socioeconomic attributes. Column 4 and 5 report on the results of the final specification above, with column 4 (5) excluding (including) a possible non-linearity of the price effect.

3.1 The Direct Price Effect

Theory predicts that a higher price of contributing will be associated with a lower probability to contribute. Our data indeed confirms this prediction: The marginal effects reported in Table 3 imply that raising the price of the contribution by €1 at the sample

[30] The two prices at which contributions do not exceed zero in statistically significant way are the treatments with a price of contributing of €50 and €56.

Table 2: Probit coefficient estimates

	(1)	(2)	(3)	(4)	(5)
Price (€)	-0.0038***	-0.0223***	-0.0040***	-0.0030	-0.0022
	(0.001)	(0.004)	(0.001)	(0.002)	(0.002)
Price squared	–	0.0002***	–	–	0.0002***
		(0.000)			(0.000)
Female	–	–	0.0952	0.0834	0.0808
			(0.076)	(0.076)	(0.076)
Age	–	–	0.0037	0.0038	0.0037
			(0.003)	(0.003)	(0.003)
Years of education	–	–	0.0641***	0.0659***	0.0654***
			(0.011)	(0.011)	(0.011)
Net income (T€)	–	–	-0.0258	-0.0299	-0.0279
			(0.022)	(0.023)	(0.023)
Eastern Germany	–	–	-0.1092	-0.1239	-0.1192
			(0.095)	(0.096)	(0.097)
Price * female	–	–	–	-0.0030	-0.0034
				(0.003)	(0.003)
Price * age	–	–	–	0.0001	0.0001
				(0.000)	(0.000)
Price * years of education	–	–	–	0.0010***	0.0009**
				(0.000)	(0.000)
Price * income	–	–	–	-0.0014*	-0.0012
				(0.001)	(0.001)
Price * Eastern Germany	–	–	–	-0.0006	-0.0012
				(0.003)	(0.003)
Constant	-0.7947***	-0.4904***	-1.7739***	-1.0869***	-1.2419***
	(0.061)	(0.090)	(0.283)	(0.196)	(0.201)
Additional controls	No	No	Yes	Yes	Yes
N	2354	2354	1872	1872	1872
Log-likelihood	-1037.451	-1027.442	-786.483	-781.486	-773.769
χ^2	12.749	32.767	81.359	91.352	106.786
Pseudo R^2	0.006	0.016	0.049	0.055	0.065

Notes: Dependent variable: 1 if subject chose the contribution over the cash award. Standard errors are in parentheses. *** Significant at or below 1% ** Significant at or below 5% * Significant at or below 10%. Main effects of continuous variables in (4) and (5) are evaluated at the sample means. Additional controls include dummies for experimental session, day, and daytime.

Table 3: Marginal effects

	(1)	(2)	(3)	(4)	(5)
Price (€)	-0.0009***	-0.0054***	-0.0009***	-0.0007	-0.0005
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Price squared	–	0.0000***	–	–	0.0000***
		(0.000)			(0.000)
Female (d)	–	–	0.0223	0.0194	0.0186
			(0.018)	(0.018)	(0.018)
Age	–	–	0.0009	0.0009	0.0008
			(0.001)	(0.001)	(0.001)
Years of education	–	–	0.0150***	0.0153***	0.0150***
			(0.003)	(0.003)	(0.003)
Net income (T€)	–	–	-0.0060	-0.0069	-0.0064
			(0.005)	(0.005)	(0.005)
Eastern Germany (d)	–	–	-0.0246	-0.0275	-0.0263
			(0.021)	(0.020)	(0.020)
Price * female	–	–	–	-0.0007	-0.0008
				(0.001)	(0.001)
Price * age	–	–	–	0.0000	0.0000
				(0.000)	(0.000)
Price * years of education	–	–	–	0.0002***	0.0002**
				(0.000)	(0.000)
Price * income	–	–	–	-0.0003*	-0.0003
				(0.000)	(0.000)
Price * Eastern Germany	–	–	–	-0.0001	-0.0003
				(0.001)	(0.001)
Additional controls	No	No	Yes	Yes	Yes
N	2354	2354	1872	1872	1872
Log-likelihood	-1037.451	-1027.442	-786.483	-781.486	-773.769
χ^2	12.749	32.767	81.359	91.352	106.786
Pseudo R^2	0.006	0.016	0.049	0.055	0.065

Notes: Dependent variable: 1 if subject chose the contribution over the cash award. (d) denotes an indicator variable. Additional controls include dummies for experimental session, day, and daytime. Marginal effects are evaluated at the sample means. Standard errors are in parentheses.
*** Significant at or below 1% ** Significant at or below 5% * Significant at or below 10%

mean decreases the propensity to contribute to the public good by approximately 0.1%. The effect has the predicted negative sign and is significant at the 1% level. The effect is also robust: Comparing the magnitude of the linear price effect across specifications (columns 1, 4, and 5), the magnitude of the price effect changes only slightly when allowing for both price and non-price effects. Converting the direct price effect into a measure of elasticity, we calculate the elasticity of the probability of contributing[31] based on column 1 as -0.31 (standard error 0.09).

While in line with theoretical predictions, the evidence generated by direct price variation contrasts somewhat with the reported evidence based on indirect variation. Not all papers on the topic report on how indirect price variation impacts on the decision whether to contribute (e.g. Eckel and Grossman 2003). Those that do tend to find that variations in match rates or rebates do not have a significant impact on the share of contributors in the population. Karlan and List (2007), Karlan et al. (2011) and Eckel and Grossman (2008) conduct field experiments for political campaign organizations and public broadcasting services, respectively. Even though the experimenters offer match or rebate rates that reduce the price of giving by as much as 66%, response rates in the population do not vary systematically with the indirect price variation. Likewise, Huck and Rasul (2011) examine contributions to an educational program maintained by a large opera theater and do not find an effect on the propensity to contribute when introducing a match.[32]. Smith et al. (1995) examine contributions to rural health care facilities in Montana and do not find a significant effect of the rate of tax rebate on the decision whether to make a charitable contribution.[33]

The difference between the direct price effect on the probability to contribute and the previous evidence based on indirect price variation could be driven by several different

[31] The elasticity of probability is defined as $\eta_{\mathrm{Pr}} = \frac{\partial \Pr(Y=1)}{\partial p} \frac{p}{\Pr(Y=1)}$ where p denotes the cash prize (e.g. Miklius et al. 1976, LeClere 1992).

[32] Huck and Rasul (2011), however, find an effect of introducing a lead donor, pointing to the important confounding effect that arises when matches and lead donors are introduced simultaneously.

[33] Some observers have related this evidence to similar findings on the irrelevance of stake size on behavior in dictator games (e.g. Carpenter et al. 2005). For example, in an artefactual field experiment with an all-or-nothing design similar to ours, Bekkers (2007) exploits variations in the size of the experimental endowment, which range between €6 and €11. He finds that the probability that a subject will donate this amount to a charity is independent of the size of the endowment. A key difference to our experiment is, however, that the recipient of the donation there also receives a larger transfer while in our case, different stake sizes always results in the same physical contribution.

factors. One possibility is a bias in reporting evidence: While previous research has stressed that the drivers of whether and of how much to contribute may be different (Smith et al. 1995), some studies do not report separately how the decision whether to contribute is impacted by the variation in the indirect price of giving. Those that report on the contribution decision may do so particularly because of the surprising result that they do not find an effect. A second possibility is that experiments using indirect price variations would have found a price effect in the contribution decision at larger sample sizes. Finally, an explanation could be that indirect and direct price variation are not behaviorally equivalent when subjects decide whether to contribute to a public good (Eckel and Grossman 2003, Davis and Millner 2005, Eckel and Grossman 2008).

Before turning to possibility of FPC as a potential bias, one objection to the result that could be raised regarding the size of the direct price effect is the possibility of an anchoring effect. When subjects are poorly informed or unfamiliar with the good (Green 1992, List and Shogren 1999), higher prices offered might conceivably lead uninformed subjects to infer that the good is more valuable, prompting subjects to choose the public goods contribution. Experimental prices would therefore confound the contribution decision with the result that the true direct price effect would be even greater. To test for the possibility of such an anchoring effect, we re-estimate the model with interaction terms between price and variables that are likely to be associated with greater familiarity with the good such as subjects' confidence in their knowledge about the donation context (confidence in own estimate of the carbon "footprint" caused by personal lifestyle, confidence in own estimate of the going EUA spot price) and their education. An anchoring effect would mean that better informed subjects should be more price sensitive compared to less informed subjects, who would be more likely to base their valuation of the contribution on the cash prize offered in the experiment. We find a non-negative relationship between the propensity to provide the mitigation effort and the "information-weighted" price: Contrary to the hypothesis of the confounding price effect, more familiarity does not change the price elasticity of contributing (for the knowledge variables) or even decreases it (for education, see columns 4 and 5 in Table 3). This resonates with experimental findings that price elasticity does not systematically vary with uncertainty about good characteristics Heffetz and Shaya (2009).

3.2 Non-price controls

There are a number of non-price attributes of subjects that have been examined in the literature as determinants of contributing and that conceivably interact with the price of contributing. These attributes include mainly gender, age, education, income, and 'culture'. Column 3 in Tables 2 and 3 reports the estimated effect of non-price attributes on the probability to contribute while columns 4 and 5 report on the estimated interaction effects.

List (2004) succinctly sums up much of the experimental evidence on the socioeconomic drivers of a failure to contribute in public goods games in his dictum of "young, selfish, and male". In the present experiment, *females* seem to be more inclined to opt for the public good contribution across all tested specifications, but the effect is not significant. The result is in line with the currently equivocal evidence on gender effects in public goods settings where the evidence on gender differences is less clear-cut than its behavioral salience in areas such as risk taking or competition (see Croson and Gneezy 2009, and references therein). As Andreoni and Vesterlund (2001) point out, however, the lack of a level effect in social dilemmas may mask interaction effects: In a laboratory setting, they find male subjects to be more altruistic than female subjects when the price of giving is low, and vice versa. We therefore test for a possible price-gender interaction term to allow for elasticities to differ between men and women. The estimates in columns 4 and 5, however, yield no evidence for a gender effect in the present setting.

Like gender, *age* has attracted increasing attention as a determinant of behavior in public goods settings (Harbaugh and Krause 2000, List 2004). List (2004) and Carpenter et al. (2008), for example, find that social preferences increase with age in laboratory public goods games and charitable donations experiments. Also, like gender, the age effect is consistently positive but insignificant in all model specifications. Again, we test for a possible interaction effect with the price of contributing, but do not uncover a significant relationship.

In contrast to gender and age, *education* stands out as highly significant across all specifications. As the results in Table 3 show, subjects' propensity to contribute increases by as much as 1% for every year spent in education. Education also stands out for an interaction effect with the price of contributing: Additional years of education are associated with a higher probability of contributing at higher prices. Education therefore makes subjects decision to contribute less price elastic.

Both the presence and strength of the education effect are interesting. Many papers studying pro-social behavior do not report on the educational status of participants. Notable exceptions are List (2004), Karlan (2005), and Bekkers (2007): In three field experiments measuring social preferences reported by List (2004), education is either insignificant or weakly associated with higher contributions. In an experimental study in the context of a Peruvian microcredit program, Karlan (2005) finds that educational attainment is a determinant of observed behavior in a number of archetypical strategic situations such as the trust game, but is not associated with a greater willingness to contribute in public goods games. Bekkers (2007) studies dictator behavior in a survey-based, anonymous, all-or-nothing version of the game. There, educational status is binary (with or without a university degree) and a high status is associated with a significantly elevated probability of donating.

Pro-social behavior may be acquired through education, but the strong relationship observed in the data may also arise from a different source. One plausible explanation could be that education and the perception of benefits from public goods provision are positively correlated, as is the case for climate policy benefits in the U.S. as survey data indicates (Borick et al. 2011). However, there is less evidence of this type of correlation in EU countries: 89% among those with a high-school degree or less and 92% of those with tertiary education regard climate change as at least "a fairly serious problem" (European Commission 2008). The strong education effect may also be explained by the specific public good used in the experiment: Emission reductions have long-run public good characteristics in a complex climate system. Patience and cognitive ability are therefore likely to matter. A number of empirical studies link cognitive ability and its proxy, education, with lower discount rates when assessing future costs and benefits and with overall stronger forward-looking behavior by individuals (Bettinger and

Slonim 2007, Kirby et al. 2005, Parker and Fischhoff 2005). Other studies emphasize the lower cognitive cost to abler individuals of making decisions in complex settings (Peters et al. 2006). Against the background of self-reported income, another explanation is that education is a possible alternative measure of income and wealth. Since both tend to be positively correlated with cognitive ability (Banks and Oldfield 2007), this provides an additional causal channel through which education could enter as a significant explanatory variable.

The effect of *income* is insignificant in every model specification and the interaction effect borderline significant at the 10% level in one specification. While surprising in the context of the tax rebate literature (Auten et al. 2002), income elasticities of contribution close to zero have also been reported in a field experiment on charitable contributions by Eckel and Grossman (2008). However, the authors warn against overinterpreting the result due to the aggregate nature of their income data. In the present experiment, income data is indeed available on an individual level. At the same time, caution is advised as income is self-reported and therefore subject to potential biases, and 482 subjects are excluded that did not report their income. Data speaks against multicollinearity of income and education as explanation for the persistent insignificance of the one and strong significance of the other. The correlation coefficient with education is positive at 0.29, but excluding education from the regression as a robustness check fails to give rise to a significant income effect.

Previous research has stressed the role of culture as a potential determinant (or not) of contribution decisions in public goods. While some experiments fail to find evidence for cultural difference (e.g. Brandts et al. 2004), two experiments on contribution behavior conducted in Germany (Ockenfels and Weimann 1999, Brosig-Koch et al. 2011) find significant and highly persistent differences between East-German and West-German residents regarding their behavior in a so-called “solidarity game”. We test for the presence of significant differences in the contribution decision between subjects located in East and West Germany both in terms of level and in terms of an interaction with the price effect. In both cases, there is no evidence for a significant effect of the place of residence on the contribution decision when considering all subjects.[34]

3.3 Field price censoring

As pointed out earlier, one potential drawback of varying the price of contributing directly and in the field is the possibility of field price censoring (FPC) among subjects. If present, FPC has the potential of biasing results. In the limit, e.g. in the context of highly familiar goods, the presence and magnitude of the direct price effect could conceivably hinge entirely on the fact that subjects know or believe that they can provide the public good more cheaply outside the experiment.

To identify subjects possibly affected by FPC, we draw on the FPC "filter" statements described in Section 2.2 as well as on answers to the follow-up questions on EUAs and on efforts for climate change mitigation. A common problem in debriefing questionnaires that are not payoff-relevant is that, while easily implemented, they are not immune to contamination through strategic behavior or ex post rationalization (Corrigan and Rousu 2008). In the context of the FPC identification strategy pursued here, both a subject's "filter" statements and his or her estimate of the EUA spot price may be endogenous to the preceding choice whether to contribute or not at the given price. The conservative strategy we adopt here is to use these answers to identify the observations that are *potentially* subject to FPC and test in four different ways whether their inclusion causes a bias in the overall price effect. Previewing the results, the available evidence points against a substantive bias in the price effect on account of omitted FPC. In three of four different approaches, the coefficient estimates for the price effect are not affected, in one case they are affected marginally.

Table 4 summarizes subjects' FPC "filter" statements and identifies 511 (25.9%) of 1,973 cash choosing subjects who declare, by not checking statement 1 but checking statement 2, that at the given experimental price, they would make a contribution, but chose not to because they believe they can make the same contribution to the public good at a lower price elsewhere.[35] The question now is whether the inclusion of these

[34] However, excluding younger age groups who would have spent their formative years after German reunification, we find a location effect: Subjects resident in Eastern Germany have a 5% lower probability of contributing, with the highest significance ($p = 0.047$) for those aged 33 years or more.

[35] Among the 1,973 cash choosing subjects, 276 gave an open-ended answer in own words without

Table 4: FPC "filter": Joint distribution of subjects' statements about their choice of cash

"Given the two prizes, I did not want to forgo the chance of winning x euros"	"I assume that there is another possibility for me to reduce CO_2 emissions by 1 ton for less than x euros" 0	1	Total
0	18	511	529
1	1,321	123	1,444
Total	1,339	634	1,973

Note: x denotes the cash prize the subject was assigned to

subjects bias the estimate of the price effect in column 1 of Tables 2 and 3? If FPC played a role, the estimated coefficient of price on the contribution decision in the full sample would be plausibly biased towards zero since a rational agent making those statements would always choose cash, irrespective of the price.

Column (1) in Table 5 reports that the price coefficient of the reduced sample that excludes the 511 potentially affected subjects does not differ significantly from the coefficient of the full sample. The regression replicates the significantly negative price effect on the decision to contribute in the full sample (cp. column 1 in Table 2) and compares it to that in the reduced sample. The coefficient of the interaction term is insignificant ($p = 0.69$). Naturally, the overall probability of choosing the reduction is significantly higher if one excludes cash choosing subjects, leading to a significantly positive coefficient on the dummy for the reduced sample. We obtain a price elasticity of probability of -0.33 (standard error 0.089) if computed for the reduced sample only, compared to -0.31 (standard error 0.09) derived for the full sample.

Another way of utilizing the "filter" statements is to assume that all subjects identified by the statements were indeed subject to FPC and then recode their choice from choosing cash to choosing the reduction. Column (2) compares the original and the recoded sample the same way column 1 does for the reduced sample. Again, a significant difference in the coefficients on cash prize cannot be established. The evidence based on the "filter" statements thus points against a significant bias from FPC.

Columns (3) and (4) of Table 5 present the results of the second part of the strategy

checking one of the two statements. 258 answers provided paraphrases of the given statements and could therefore be reassigned. 249 of them implied an actual comparison of benefits and costs of the prizes (statement 1), 9 answers corresponded to a preferred opportunity outside the experiment given the choice (statement 2).

Table 5: Robustness of the price effect to field price censoring

	(1)	(2)	(3)	(4)
Cash prize	-0.0038***	-0.0038***	-0.0038***	-0.0042***
	(0.001)	(0.001)	(0.001)	(0.001)
Reduced sample	0.2024**	–	0.1161	–
	(0.090)		(0.092)	
Reduced sample * cash prize	-0.0006	–	-0.0004	–
	(0.002)		(0.002)	
Recoded sample	–	0.6557***	–	–
		(0.081)		
Recoded sample * cash prize	–	0.0005	–	–
		(0.001)		
EUA estimate below	–	–	–	-0.5297***
				(0.148)
EUA estimate below * cash prize	–	–	–	0.0048**
				(0.002)
Constant	-0.7960***	-0.7960***	-0.7960***	-0.6799***
	(0.061)	(0.061)	(0.061)	(0.069)
N	4199.000	4710.000	3714.000	2355.000
Log-likelihood	-1970.881	-2594.222	-1698.694	-1027.371
χ^2	41.701	312.406	28.654	33.265
Pseudo R^2	0.010	0.057	0.008	0.016

Notes: Probit coefficient estimates. Standard errors in parentheses. Dependent variable: 1 if subject chose the contribution over the cash award. Independent variables: '*Reduced sample*' is 0 if the observation belongs to the full sample and 1 if the observation belongs to the sample excluding subjects that are potentially affected by FPC according to the "filter" statements (column 1) or EUA price estimates (column 3). '*Recoded sample*' is 0 if the observation belongs to the original sample and 1 if the observation belongs to the sample with recoded choices according to the FPC "filter" statements. '*EUA estimate below*' is an indicator variable and 1 if the observation is potentially affected by FPC according to subject's EUA price estimate. *** Significant at or below 1% ** Significant at or below 5% * Significant at or below 10%

Table 6: Subjects' EUA price estimates

Survey question		Freq.	Rel. freq.	Cum.
"What is your estimate	Below 2	100	4.25	4.25
of the current market price	2 to below 10	110	4.67	8.92
(in EUR) for 1 ton of CO_2	10 to below 20	328	13.93	22.85
in the EU emissions trading	20 to below 30	240	10.19	33.04
system?"	30 to below 50	213	9.04	42.08
	50	286	12.14	54.22
	Above 50 to below 100	496	21.06	63.14
	100	355	15.07	78.21
	Above 100 to below 1,000	215	9.13	87.35
	1,000 to below 10,000	210	8.92	96.26
	10,000 and more	88	3.74	100.00

Notes: Continuous variable (open-ended question).

to detect FPC. This part specifically targets FPC from the potential availability of a *perfect* substitute and is based on subjects' estimates of the going EUA spot price elicited in the ex-post questionnaire.[36] Table 6 gives a detailed summary of this variable. About 74% of subjects gave an estimate within the range of the randomly assigned experimental prices (€ 2 to € 100) while the median subject gave an estimate of € 50, close to the experimental mean and median. Thus, most subjects do not seem to be well informed about the field price (about € 15 at the time of the experiment). Comparing assigned experimental cash prizes and estimated field prices, we identify 996 subjects who estimated an EUA price below the cash prize amount they were assigned to. 1,359 subjects gave an EUA price estimate greater or equal to the cash prize. If subjects implicitly or explicitly took their perception of a field price into account when pondering their contribution decision, and not vice versa, then the choice of subjects who anticipate an EUA price below the experimental price may be affected by FPC.[37]

As before, we compare the unconditional price coefficient of the full sample with

[36] Evidence for endogenous information acquisition during the experiment, e.g. by searching the Internet for EUA spot prices, comes from a careful examination of the "time stamps" of each screen in each individual experiment. The time stamp measures the exact time at which the subject moved on to the next screen. As information collection requires time for targeted search, search activity should be associated with time delay at screens that ask for relevant information relative to other screens. We impose ambitious assumptions on how quickly a subject can collect the information: For example, subjects would need to find EUA prices and information on annual per capita emissions on the Internet in under 2 minutes. We find no more than 1.4% of subjects with time delays that would be consistent with information collection. In addition, these candidates do not exhibit above average accuracy on the factual questions in the experiment. On this basis, we conclude that endogenous information acquisition does not play a role in explaining the results and confirm results by Berrens et al. (2004) and Fong and Oberholzer-Gee (2011). Importantly, this result also means that a potential field price censoring is not a product of endogenous information acquisition by subjects during the experiment, but can at most be generated by differences in information prior to the experiment.

[37] To a rational agent, the choice would also depend on perceived transaction costs.

that of a reduced sample. This time, the reduced sample excludes subjects potentially affected by FPC due to their EUA price estimate given afterwards. Column (3) in Table 5 reports on the results. Again, the price coefficient of the reduced sample is not significantly different from that of the full sample. The corresponding elasticity of probability for the reduced sample is −0.29 (standard error 0.095).

For the final column (4) of Table 5, we split the original full sample into two subsamples, one consisting of subjects whose EUA price estimate exceeds the cash prize and the other of those whose estimates are below the cash prize. Column (4) reports on the results of a direct comparison of contribution choices between the two subsamples with respect to price. The results show that, first, controlling for cash prize, subjects who estimate an EUA price below their cash prize are significantly less likely to contribute than those who estimate a spot price above their cash prize. Second, the contribution choice of the former group is not significantly correlated with price: the interaction term is significantly positive and, regarding magnitude, offsets the significantly negative main effect.[38] The observed effects in this split-sample case are as one would expect them to arise from FPC. The test using a split-sample approach is weak, however, as the result can equally well be generated by reasons other than FPC: First, given the distributions of the cash prize and the price estimate variables, there are only few observations for low prices where the estimate undercuts the experimental price. This inflates the variance at low prices for this group and may prevent detection of a significant price effect. Second, the price estimate reported by the subject may itself not be independent of the choice that the subject has taken. These competing hypotheses cannot be tested against each other, given the data. However, even if there is a FPC bias, it is both small and reinforcing the price effect.

In the third and final part of the detection strategy for FPC, we qualitatively analyzed the answers to the open-ended question on subjects' existing efforts to mitigate climate change. Most comments related to behavioral changes or investments into energy saving measures. None of the subjects mentioned any type of carbon offset or certificate. We take this as further evidence that close substitutes and their field prices did not play a

[38]Performing a separate regression for the reduced sample gives an insignificant effect of the price.

role in determining subjects' contribution choices.

4 Conclusion

The relationship between the price of giving a the public good and its private provision is a natural subject of interest to economists. Empirical opportunities in the form of exogenous variations in marginal tax rates (for tax deductible contributions), laboratory experiments, and field experiments have provided the basis for important insights into how variations in rebate rates and match ratios affect the probability that individuals will choose to contribute and how much they contribute if they do. Among the many results of this indirect approach to price variation, one finding is that the decision whether to contribute appears to be largely immune to variations in match or rebate rates. While variations in rebate rates and match ratios can be converted into theoretically equivalent price variations, recent experimental evidence has thrown into doubt whether this theoretical equivalence also implies behavioral equivalence. Using price elasticities derived on the basis of their theoretically equivalent match rate or rebate rate elasticities may therefore be problematic.

This paper presents field experimental evidence from an alternative approach to examining the relationship between the price of giving and public goods provision, namely through *direct* price variation: We compare across thousands of subjects how the decision whether to contribute varies with the amount of money that subjects have to give up in order to provide one unit of the public good. The theoretical prediction of a clear negative relationship between price and public goods provision is borne out by our experimental data. There is a negative and robust direct price effect on the probability whether to contribute. We estimate its mean elasticity across the treatment range as -0.31. The direct price effect is robust with respect to a range of controls and with respect to the potential problem of field price censoring. This provides strong evidence that in the present case, making contributing cheaper through, for example, public subsidies has only a modest impact on the probability to contribute.

Among subjects' socioeconomic attributes that we use as controls, education stands out as a key determinant of the decision whether to contribute. Keeping in mind the possible limitations of self-reported income data and the lack of an established education-social preference channel in the literature, the role of education could be due to both cognitive and income or wealth effects. For gender and age, on the other hand, the literature provides reasons for expecting a significant role, but both effects fail to materialize in the experiment.

Given the difference between the evidence on the contribution decision by direct and indirect price variation, an obvious next research step is to directly compare match rates, rebate rates, and direct price changes in the context of public goods, preferably in a field setting. This would be important both in order to confirm independently the nature of the direct price effect and to quantify the differences between these approaches in terms of direction and magnitude.

References

Abdellaoui, M., Baillon, A., Placido, L. and Wakker, P. P. (2011). The rich domain of uncertainty: Source functions and their experimental implementation, *The American Economic Review* **101**: 695–723.

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and ricardian equivalence, *Journal of Political Economy* **97**(6): 1447–1458.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving, *Economic Journal* **100**(401): 464–477.

Andreoni, J. and Miller, J. (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism, *Econometrica* **70**(2): 737–753.

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

Arrow, K. J., Solow, R., Portney, P. R., Leamer, E. E., Radner, R. and Schuman, H. (1993). Report of the NOAA panel on contingent valuation, *Technical report*. Washington D.C.

Auten, G. E., Sieg, H. and Clotfelter, C. T. (2002). Charitable giving, income, and taxes: An analysis of panel data, *The American Economic Review* **92**(1): 371–382.

Baltussen, G., Post, T., Van den Assem, M. J. and Wakker, P. P. (2010). Random incentive systems in a dynamic choice experiment. Working Paper.

Banks, J. and Oldfield, Z. (2007). Understanding pensions: Cognitive function, numerical ability and retirement saving, *Fiscal Studies* **28**(2): 143–170.

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

Bergstrom, T., Blume, L. and Varian, H. (1986). On the private provision of public goods, *Journal of Public Economics* **29**(1): 25–49.

Berrens, R. P., Bohara, A. K., Jenkins-Smith, H. C., Silva, C. L. and Weimer, D. L. (2004). Information and effort in contingent valuation surveys: Application to global climate change using national internet samples, *Journal of Environmental Economics and Management* **47**(2): 331–363.

Bettinger, E. and Slonim, R. (2007). Patience among children, *Journal of Public Economics* **91**(1-2): 343–363.

Bolle, F. (1990). High reward experiments without high expenditure for the experimenter?, *Journal of Economic Psychology* **11**(2): 157–167.

Borick, C. P., Lachapelle, E. and Rabe, B. G. (2011). Climate compared: Public opinion on climate change in the United States and Canada, *Brookings Institution Issues in Governance Studies* (39).

Brandts, J., Saijo, T. and Schram, A. (2004). How universal is behavior? A four country comparison of spite and cooperation in voluntary contribution mechanisms, *Public Choice* **119**(3-4): 381–424.

Brosig-Koch, J., Helbach, C., Ockenfels, A. and Weimann, J. (2011). Still different after all these years: Solidarity behavior in east and west germany, *Journal of Public Economics* **95**(11-12): 1373–1376.

Carlsson, F., He, H. and Martinsson, P. (2013). Easy come, easy go, *Experimental Economics* **16**(2): 190–207.

Carpenter, J., Connolly, C. and Myers, C. (2008). Altruistic behavior in a representative dictator experiment, *Experimental Economics* **11**(3): 282–298.

Carpenter, J., Verhoogen, E. and Burks, S. (2005). The effect of stakes in distribution experiments, *Economics Letters* **86**(3): 393–398.

Cherry, T., Frykblom, P., Shogren, J., List, J. and Sullivan, M. (2004). Laboratory testbeds and non-market valuation: The case of bidding behavior in a second-price auction with an outside option, *Environmental and Resource Economics* **29**(3): 285–294.

Clark, J. (2002). House money effects in public good experiments, *Experimental Economics* **5**(3): 223–231.

Coller, M. and Williams, M. B. (1999). Eliciting individual discount rates, *Experimental Economics* **2**(2): 107–127.

Corrigan, J. R. and Rousu, M. C. (2008). Testing whether field auction experiments are demand revealing in practice, *Journal of Agricultural and Resource Economics* **33**(2): 290–301.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences, *Journal of Economic Literature* **47**(2): 448–474.

Cubitt, R., Starmer, C. and Sugden, R. (1998). Dynamic choice and the common ratio effect: An experimental investigation, *Economic Journal* **108**(450): 1362–1380.

Davis, D. D. and Millner, E. L. (2005). Rebates, matches, and consumer behavior, *Southern Economic Journal* **72**(2): 410–421.

Diederich, J. and Goeschl, T. (2011). Giving in a large economy: Price vs. non-price effects in a field experiment, *Discussion Paper No. 514* . Department of Economics, Heidelberg University.

Diederich, J. and Goeschl, T. (2013). Willingness to pay for voluntary climate action and its determinants: Field-experimental evidence, *Environmental and Resource Economics* . DOI 10.1007/s10640-013-9686-3.

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

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

Eckel, C. and Grossman, P. (2008). Subsidizing charitable contributions: A natural field experiment comparing matching and rebate subsidies, *Experimental Economics* **11**(3): 234–252.

Engel, C. and Moffat, P. G. (2012). Estimation of the house money effect using hurdle models, *Max Planck Institute for Research on Collective Goods Working Paper* .

European Commission (2008). Europeans' attitudes towards climate change, *Special Eurobarometer* **300**. European Parliament / European Commission.

Feldstein, M. and Clotfelter, C. (1976). Tax incentives and charitable contributions in the united states: A microeconometric analysis, *Journal of Public Economics* **5**(1-2): 1–26.

Feldstein, M. and Taylor, A. (1976). The income tax and charitable contributions, *Econometrica* **44**(6): 1201–1222.

Fong, C. M. and Oberholzer-Gee, F. (2011). Truth in giving: Experimental evidence on the welfare effects of informed giving to the poor, *Journal of Public Economics* **95**(5-6): 436–444.

Green, D. P. (1992). The price elasticity of mass preferences, *The American Political Science Review* **86**(1): 128–148.

Grether, D. M. and Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon, *The American Economic Review* **69**(4): 623–638.

Harbaugh, W. T. and Krause, K. (2000). Children's altruism in public good and dictator experiments, *Economic Inquiry* **38**(1): 95–109.

Harrison, G. W. (2007). House money effects in public good experiments: Comment, *Experimental Economics* **10**(4): 429–437.

Harrison, G. W., Harstad, R. M. and Rutström, E. E. (2004). Experimental methods and elicitation of values, *Experimental Economics* **7**(2): 123–140.

Harrison, G. W., Lau, M. I. and Williams, M. B. (2002). Estimating individual discount rates in denmark: A field experiment, *American Economic Review* **92**(5): 1606–1617.

Harrison, G. W. and List, J. A. (2004). Field experiments, *Journal of Economic Literature* **42**(4): 1009–1055.

Heffetz, O. and Shaya, M. (2009). How large are non-budget-constraint effects of prices on demand?, *American Economic Journal: Applied Economics* **1**: 170–199.

Huck, S. and Rasul, I. (2011). Matched fundraising: Evidence from a natural field experiment, *Journal of Public Economics* **95**: 351–362.

Karlan, D. and List, J. A. (2006). Does price matter in charitable giving? evidence from a large-scale natural field experiment, *National Bureau of Economic Research Working Paper Series* **No. 12338**.

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

Karlan, D., List, J. A. and Shafir, E. (2011). Small matches and charitable giving: Evidence from a natural field experiment, *Journal of Public Economics* **95**(5-6): 344–350.

Karlan, D. S. (2005). Using experimental economics to measure social capital and predict financial decisions, *American Economic Review* **95**(5): 1688–1699.

Keeler, J. P., James, W. L. and Abdel-Ghany, M. (1985). The relative size of windfall income and the permanent income hypothesis, *Journal of Business & Economic Statistics* **3**(3): 209–215.

Kingma, B. R. (1989). An accurate measurement of the crowd-out effect, income effect, and price effect for charitable contributions, *The Journal of Political Economy* **97**(5): 1197–1207.

Kirby, K. N., Winston, G. C. and Santiesteban, M. (2005). Impatience and grades: Delay-discount rates correlate negatively with college GPA, *Learning and Individual Differences* **15**(3): 213–222.

Krasteva, S. and Yildirim, H. (2013). (un)informed charitable giving, *Journal of Public Economics* **106**(0): 14–26.

LeClere, M. J. (1992). The interpretation of coefficients in models with qualitative dependent variables, *Decision Sciences* **23**(3): 770–776.

Lee, J. (2008). The effect of the background risk in a simple chance improving decision model, *Journal of Risk and Uncertainty* **36**(1): 19–41.

List, J. A. (2004). Young, selfish and male: Field evidence of social preferences, *Economic Journal* **114**(492): 121–149.

List, J. A. and Shogren, Jason, F. (1999). Price information and bidding behavior in repeated second-price auctions, *American Journal of Agricultural Economics* **81**(4): 942–949.

Miklius, W., Casavant, K. L. and Garrod, P. V. (1976). Estimation of demand for transportation of agricultural commodities, *American Journal of Agricultural Economics* **58**(2): 217–223.

Nordhaus, W. D. (1993). Reflections on the economics of climate change, *The Journal of Economic Perspectives* **7**(4): 11–25.

Ockenfels, A. and Weimann, J. (1999). Types and patterns: An experimental East-West-German comparison of cooperation and solidarity, *Journal of Public Economics* **71**(2): 275–287.

Parker, A. M. and Fischhoff, B. (2005). Decision-making competence: External validation through an individual-differences approach, *Journal of Behavioral Decision Making* **18**(1): 1–27.

Peloza, J. and Steel, P. (2005). The price elasticities of charitable contributions: A meta-analysis, *Journal of Public Policy & Marketing* **24**(2): 260–272.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C., Mazzocco, K. and Dickert, S. (2006). Numeracy and decision making, *Psychological Science* **17**(5): 407–413.

Sefton, M. (1992). Incentives in simple bargaining games, *Journal of Economic Psychology* **13**(2): 263–276.

Smith, V. H., Kehoe, M. R. and Cremer, M. E. (1995). The private provision of public goods: Altruism and voluntary giving, *Journal of Public Economics* **58**(1): 107–126.

Starmer, C. and Sugden, R. (1991). Does the random-lottery incentive system elicit true preferences? An experimental investigation, *American Economic Review* **81**(4): 971–978.

Thaler, R. H. and Johnson, E. J. (1990). Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice, *Management Science* **36**(6): 643–660.

Tol, R. S. J. (2010). The economic impact of climate change, *Perspektiven der Wirtschaftspolitik* **11**: 13–37.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice, *Science* **211**(4481): 453–458.

Vesterlund, L. (2006). Why do people give, *in* W. W. Powell and R. Steinberg (eds), *The nonprofit sector: a research handbook*, 2. ed. edn, Yale Univ. Press, New Haven, pp. 168–190.

Appendix

A Formal proof of the direct price effect at the extensive margin of contributions

We introduce a unit price for the public good into a variant of Andreoni's (1989, 1990) classical impurely altruistic model in order to guide the intuition for the effects of a direct price change and of non-price factors at the extensive margin. Assume n individuals who derive utility from the amount of private numéraire x, the level of a public good G, and their own contributions to the public good of size g_i ("warm glow"). Let preferences also depend on a vector of individual-specific characteristics, $\boldsymbol{\theta}_i$. Thus, we write the utility function as

$$U_i = U\left(x_i, \delta_i G, g_i; \boldsymbol{\theta}_i\right)$$

where $\delta_i \in [0,1]$ denotes heterogeneous perceptions about the value of the public good (Karlan and List 2006). Another interpretation of δ_i is incomplete information about the benefits produced by the public good. In our case, δ_i represents any heterogeneous beliefs about the size of climatic changes and thus the benefits generated by the total provision of emissions reductions.

Let the public good be measured in units which individuals can "purchase" and provide at price p. Total provision is the sum of individual provisions, $G = \sum_{i=1}^{n} g_i$. Also define $G_{-i} = \sum_{j \neq i} g_i$. Individuals are endowed with wealth w_i and thus maximize utility subject to their budget constraint,

$$\max_{x_i, g_i} U\left(x_i, \delta_i G, g_i; \boldsymbol{\theta}_i\right)$$

$$\text{s.t. } x_i + pg_i = w_i \tag{1}$$

$$G = G_{-i} + g_i \tag{2}$$

$$g_i \geq 0 \ . \tag{3}$$

Substituting for g_i, the problem reduces to

$$\max_{x_i, G} U\left(x_i, \delta_i G, G - G_{-i}; \boldsymbol{\theta}_i\right)$$

$$\text{s.t. } x_i + pG = w_i + pG_{-i}$$

$$G \geq G_{-i} \ .$$

We assume U to be strictly quasi-concave and increasing in the first three arguments. Thus, if we ignore the inequality constraint for a moment, this resembles an ordinary consumer choice problem. The demand function for G solving the problem can thus be written as

$$f\left(p, w_i + pG_{-i}, G_{-i}, \delta_i; \boldsymbol{\theta}_i\right) \ .$$

The third argument in f is the warm glow argument. Taking into account the inequality constraint (3), demand for the public good is

$$G = \max\left\{f(p, w_i + pG_{-i}, G_{-i}, \delta_i; \boldsymbol{\theta}_i), G_{-i}\right\} \ .$$

In order to derive first-order effects at the extensive margin, we take the inverse of f with respect to the second argument, $w_i + pG_{-i}$ and add pg_i to both sides. Solving for g_i gives

$$g_i = \frac{1}{p}\left[w_i - f^{-1}\left(p, G, G_{-i}, \delta_i; \boldsymbol{\theta}_i\right)\right] + G \ .$$

Given (3), the condition to provide a strictly positive amount of public good is

$$w_i > f^{-1}(p, G, G_{-i}, \delta_i; \boldsymbol{\theta}_i) - pG \ .$$

Let w_i^* denote the threshold level of wealth at which individual i switches from non-contribution to contribution. Here, (3) holds with equality and thus, $G = G_{-i}$. It follows that

$$w_i^* = f^{-1}(p, G_{-i}, \delta_i; \boldsymbol{\theta}_i) - pG_{-i} \tag{4}$$

Note that the third argument of f^{-1} drops out since at $g_i = 0$ the individual does not derive any utility from warm glow. Also note that w_i^* is not identical for all individuals because of δ_i and $\boldsymbol{\theta}_i$.

We are now interested in how the set of contributors changes if certain parameters change. From

(4) it follows that

$$\frac{\partial w_i^*}{\partial p} = f_p^{-1} - G_{-i} > 0$$

if we assume normality for both goods.[39] Thus, an increase in price *ceteris paribus* increases the threshold level of wealth for individual i, which makes it less likely that individual i will contribute. Similarly, normality of both goods implies that[40]

$$\frac{\partial w_i^*}{\partial \delta_i} = f_\delta^{-1} < 0 \ .$$

Intuitively, if individual i's perceived benefits from the public good provision increase then it is more likely that i will provide a strictly positive amount of the public good. With regard to individual characteristics, we have already demonstrated that w_i^* depends on $\boldsymbol{\theta}_i$.

B Instructions (translation of experimental screens into English)

B.1 Welcome screen

Dear participants,

we would like to invite you to participate in two lotteries and to answer some questions about CO_2-emissions and climate change.

Your participation will take approximately ten minutes. In the lotteries, you have the chance to win points worth up to a three-digit amount in euros.

As usual, all your information will be treated confidentially.

B.2 Citizenship screen

Of which countries do you hold citizenship?

In case you hold multiple citizenships, please tick all [of the following citizenships] which apply!

[...]

B.3 Information Screen

In the following two lotteries, you may choose between two different prizes. These are:

A cash prize in points

or

[39]Note that normality implies that any increase in wealth will always go in consumption of *both* goods.

[40]Note that an increase of δ_i in f^{-1} *ceteris paribus* implies lower demand for x, hence $f_\delta^{-1} < 0$.

the reduction of carbon dioxide (CO_2) emissions by 1 ton

How will the reduction of the CO_2 emissions take place? We will make use of a reliable opportunity provided by the EU emissions trading system: We will purchase and delete an *EU emissions allowance* for you. Within the EU, emissions allowances are needed by power plants and other large installations in order to be allowed to emit CO_2. Since there is only a fixed overall amount of allowances in place, deleted allowances are no longer available to facilitate emissions.

Emissions in Germany and other EU countries decrease from one deleted allowance by exactly one ton.

Because of the way in which CO_2 mixes in the air, it does not matter for the effect on the climate where on the globe CO_2 emissions are reduced. What counts is only total emissions worldwide. In the lotteries, 100 winners will be randomly selected out of about 5,000 participants. The following two lotteries may differ in the prizes offered as well as in the payoff procedures.

B.4 Decision Screen

In this lottery, you have the choice between the two prizes listed below.

- If you choose the cash amount and win, then the corresponding amount of points will be automatically credited to your points account within the next few days. All winners will receive a short notification email for that.
- The deletion of emissions allowances will, in this lottery, take place as a collective order for all winners: For every winner who has chosen the emissions reduction, one additional allowance will be deleted. Winners will receive a short notification email containing a hyperlink through which they can reliably verify the deletion on Heidelberg University webpages.

Please choose now, which prize you want if you are drawn as winner:

( ) The reduction of CO_2 emissions by 1 ton through the deletion of one EU emissions allowance

( ) 46 euro in bonus points[41] in bonus points

B.5 FPC filter question

Please give some particulars about the reasons why you personally chose the euro amount. In order to do so, please tick all statements that apply to you. Please answer spontaneously.

( ) Given the two prizes, I don't not want to forgo the chance of winning 46 euros.

[41] Example amount. The order in which the two prizes appeared was randomized.

( ) I assume that there is another way for me to abate 1 ton of CO_2 emissions for less than 46 euros.

( ) I decided in favor of the euro amount for other/further reasons, namely: ____

B.6 Transition to lottery 2 and second lottery

On the next page, there will be the description of our second lottery.

[...]

B.7 Introduction follow-up questions

Thank you for your evaluations. On the following pages, we would like to ask you some concluding questions.

B.8 Follow-up questions (screen 1)

What is your estimate of the current market price for one ton of CO_2 in the EU emissions trading system?

____ euros

How sure are you about your estimate?

( ) I know the price

( ) Very sure

( ) Somewhat sure

( ) Somewhat unsure

( ) Very unsure

( ) I don't know

B.9 Follow-up questions (screen 2)

In this lottery, EU emission allowances are bought and deleted by the organizer. Do you think that there exists a possibility for you personally to buy and delete EU emissions allowances?

( ) Yes

( ) Somewhat yes

( ) Somewhat no

( ) No

( ) I don't know

Do you think that you will personally benefit from positive effects of reduced CO_2 emissions (for example from the mitigation of climate change)?

( ) [Same answer options as above]

Do you think that future generations in Germany (for example your children and grandchildren) will benefit if climate change mitigating CO_2 emissions reductions are undertaken in the present time?

( ) [Same answer options as above]

Do you think that your personal behavior or lifestyle has contributed or is contributing to climate change?

( ) [Same answer options as above]

B.10 Follow-up questions (screen 3)

What is your estimate of the yearly CO_2 emissions caused by your lifestyle?

____ tons

How sure are you about your estimate?

( ) I've had the emissions calculated

( ) Very sure

( ) Somewhat sure

( ) Somewhat unsure

( ) Very unsure

( ) I don't know

B.11 Follow-up questions (screen 4)

Do you consciously act in a climate-protecting way? If yes, please list some forms of behavior, decisions and measures through which you have consciously contributed or are contributing to the reduction of CO_2 or other greenhouse gases (in keywords).

B.12 Enquiry of socio-demographic information (if not or only partially on record)

Please state your gender.

( ) Male

( ) Female

In what year were you born? ___

How many children under 18 live in your household? ___

B.13 Enquiry of socio-demographic information if not on record

What is your highest educational degree?

( ) Still in school

( ) Special-needs school

( ) Elementary secondary school (*Hauptschule*, 9th grade)

( ) Polytechnic school of the GDR (10th grade)

( ) Highschool (*Realschule*, 10th grade)

( ) Advanced technical college entrance qualification (*Fachhochschulreife*)

( ) A-levels (*Abitur*, 12th or 13th grade)

( ) Advanced technical college - *Fachhochschule* (Diplom (*FH*), Bachelor, Master)

( ) University degree (*Diplom*, *Magister*, Bachelor, Master)

( ) Ph.D.

( ) Dropout

( ) No specification

What is the overall net income of the household that you live in?

( ) under EUR 500

( ) from EUR 500 up to EUR 1000

( ) from EUR 1000 up to EUR 1500

( ) from EUR 1500 up to EUR 2000

( ) from EUR 2000 up to EUR 2500

( ) from EUR 2500 up to EUR 3000

( ) from EUR 3000 up to EUR 3500

( ) from EUR 3500 up to EUR 4000

( ) from EUR 4000 up to EUR 4500

( ) from EUR 4500 up to EUR 5000

( ) from EUR 5000 up to EUR 10000

( ) EUR 10000 and more

( ) no specification

B.14 Closing screen

Dear participant,

Thank you very much for your participation in this survey. If you are one of the winners, we will contact you by e-mail shortly.

Willingness to Pay for Voluntary Climate Action and Its Determinants: Field-Experimental Evidence*†

Johannes Diederich   Timo Goeschl

Abstract

The determinants of individual, voluntary climate action (VCA) in combating climate change and its potential scale are frequently debated in public but largely underresearched. We provide estimates of the willingness to individually reduce EU greenhouse gas emissions by one ton, using the European Union Emissions Trading Scheme. Estimates are derived from an online field experiment with a large, highly heterogenous, and Internet-representative sample of voting-aged Germans. Jointly estimating willingness to pay (WTP), non-indifference to VCA, and prior knowledge, we uncover important determinants of preferences for VCA, such as education, the information structure among the population, and exogenous environmental conditions.

Keywords: climate change, EU ETS, field experiment, online experiment, public goods, voluntary contributions, voluntary climate action, willingness to pay

JEL-Classifications: C93, Q51, Q54

"Each and every one of us can make changes in the way we live our lives and become part of the solution [to climate change]." – *Al Gore*

1 Introduction

Voluntary climate action (VCA) offers members of the general population the opportunity to individually provide additional reductions in greenhouse gas (GHG) emissions above and beyond those implemented by their governments. Within the climate policy

*The authors thank conference participants at the CESifo Summer Institute (Venice 2010), at the annual meeting of the European Association of Environmental and Resource Economists (Rome 2011), at the annual meeting of the Verein für Socialpolitik (Frankfurt 2011), at the Climate Economics and Law Conference (Bern 2011) and at the Summer Meeting of the Association of Environmental and Resource Economists (Asheville, NC 2012), seminar participants at Heidelberg, and the handling editor at Environmental Resource Economics and an anonymous referee for helpful comments. We also thank Ruth Fieber, Christina Grimm, and Thomas Scheuerle for student assistance and Dr. Svenja Espenhorst and Dennis Mignon at First Climate for support in acquiring EU ETS allowances. Financial support by the German Science Foundation (DFG) under grant GO1604/1 is gratefully acknowledged.

†This paper has been published in Environmental and Resource Economics (2013), DOI 10.1007/s10640-013-9686-3. The final publication is available at link.springer.com

debate, a rhetoric has evolved among commentators, climate researchers, and government bodies that attributes a high potential to such voluntary action (e.g. Gore and Guggenheim 2006, Pachauri 2007, European Commission 2011). VCA, it is sometimes argued, might alleviate the need for coercive measures of emissions reductions by governments. The demand for VCA and the factors that determine it are therefore key empirical questions.

A limited number of studies have investigated the demand for VCA, typically with a focus on estimating central measures of willingness to pay (WTP). These studies estimate mean WTP to be €25 (Brouwer et al. 2008), £24 (MacKerron et al. 2009), or €12 (Löschel et al. 2013) per ton of abated carbon (CO_2) emissions. Such numbers point to the possibility that, given the opportunity, voluntary behavior might give rise to substantial GHG emissions reductions. Additional evidence regarding the presence and determinants of VCA would help to substantiate this possibility, and various opportunities for conceptual and methodological improvements present themselves. One opportunity lies in mitigating the potential hypothetical bias of the numbers reported. With the exception of Löschel et al. (2013), existing estimates are derived using stated preferences methods and thus may overstate WTP (e.g. Cummings et al. 1995, Carlsson and Martinsson 2001) or bias covariate effects. Secondly, all existing studies are constrained by comparably small samples ($N < 350$) consisting of a specific subgroup of the general population: Frequent fliers passing through a specific airport (Brouwer et al. 2008), young adults with higher education (MacKerron et al. 2009), or residents of a specific city (Löschel et al. 2013). Thirdly, participants faced bid prices for emissions reductions that mostly fall in the neighborhood of current offset or permit prices. Estimates of true marginal abatement costs, however, are up to one order of magnitude higher (e.g. Tol 2010). The fourth opportunity lies in that the existing studies have not focused on covariates or accounted for the likely presence of indifferent respondents in a way suitable for estimating covariate effects.

The present paper embarks on providing the first study of preferences for VCA and a rich set of covariates based on both non-hypothetical choices and a comparably large sample from the general population. Specifically, we report on the WTP for the voluntary abatement of one ton of CO_2 emissions through the retirement of an emissions allowance (EUA) under the European Union Emissions Trading Scheme (EU-ETS).[1] The basic design of the "framed" field experiment (Harrison and List 2004) is a closed-ended single-bounded valuation question implemented under a random incentive system (Grether and Plott 1979, Starmer and Sugden 1991, Lee 2008): Experimental subjects indicate their preference between, on the one hand, a randomly drawn cash award and, on the other, the EUA retirement. The cash prize presented to the subject is the outcome of an equiprobable draw from prizes between €2 and €100 in steps of €2, the upper bound reflecting an economically meaningful maximum abatement cost for one ton of CO_2 emissions (Tol 1999, 2009, 2010). We use between-subjects randomization (Tversky and Kahneman 1981, Baltussen et al. 2010, Abdellaoui et al. 2011) with odds of 1 in 50 that a subject's choice of either cash or emissions reductions is realized. The choice is taken by an Internet-representative sample of 2,440 Germans of voting age drawn from a population of approximately 65,000 panel members of an Internet polling firm.

Any field study aimed at investigating the public's demand for VCA must take into account that VCA corresponds to a unilateral, private provision of a global public good, GHG emissions reductions (Nordhaus 1993). Therefore, individual behavior will likely be affected by typical behavioral patterns of voluntary giving. First and foremost, some people will not be willing to pay anything for voluntary emissions reductions, not because they do not value climate change mitigation but because their behavior is determined by strategic "free riding" or the perceived marginality of their individual contribution, for example. Other people will be willing to pay more due to the altruistic component of a contribution. This contribution character of the valued good has not always been sufficiently accounted for by the existing VCA literature. Indeed, our data suggests a significant number of subjects with zero WTP as well as a robust share of contributors even at highest prices. Since our focus is on covariate effects, we explicitly model in-

[1] EUAs are the vehicle of choice for facilitating credible GHG emissions reductions in an experiment as a binding cap of the EU ETS avoids problems of additionality that are often encountered for project-based offsets (e.g. Certified Emissions Reductions (CER) under the Clean Development Mechanism of the Kyoto Protocol). Retirement (or officially, "deletion") of EUAs is an option available to all trading account holders in the EU ETS. As a result, a retired EUA reduces the total amount of GHG emissions in the European Union by one ton of CO_2. The EUAs used in the experiment are Phase II emission allowances with a market price of around €15 apiece at the time of the experiment.

difference (or nonparticipation) in the econometric analysis using a mixture model (e.g. McFadden 1994, Hanemann and Kriström 1995, Haab 1995, Kriström 1997) to prevent parameter estimates to be biased if participation and WTP are determined by the same set of covariates (Haab 1999, Werner 1999). Both indifference to the offered environmental change and WTP can also be determined by a subject's endogenous information status about or experience with the good. Empirical work suggests that the knowledge among the general population about the causes and functional relationships of climate change (Ungar 2000, Lorenzoni and Pidgeon 2006, Lorenzoni et al. 2007, Sterman and Sweeney 2007) and the logic underpinning climate policy (Sinn 2008) varies considerably. Follow-up survey results in our experiment support this picture. We exploit this heterogeneity and jointly model WTP, participation, and endogenous information in the econometric analysis of covariate effects.[2]

Our key results can be summarized as follows. First, we identify a variety of significant drivers and correlates of VCA in our experiment, which points to important heterogeneities regarding VCA among the population. For example, measures of educational status turn out to be a key predictor of VCA: Years of schooling correlate with WTP both directly and indirectly through the information submodel. Similarly, WTP positively correlates with stated perceptions of both selfish and altruistic benefits from today's emissions reductions as well as perceptions of a lifestyle-related responsibility for climate change. Policy-makers interested in harnessing VCA may find it useful to be aware of these heterogeneities. A second key result is that the empirical reality of VCA is likely to be subtle: Estimated coefficients and WTP are sensitive to accounting for nonparticipation and endogenous information in the econometric model, necessitating the use of joint modeling. Such modeling reveals, for example, that both subjects' knowledge about climate change and their likelihood of being indifferent with respect to VCA differs by age and gender: The typical young male is more likely to be informed about climate change and also more likely to indicate indifference towards VCA. Third, WTP appears to be influenced by unexpected exogenous drivers: We find evidence that ambient temperatures around the time of the experiment positively affect

[2]Thanks to an anonymous referee and the editor for insightful comments about the presence of nonparticipation and endogenous experience.

WTP. Fourth, the empirical distribution of WTP uncovered in our experiment suggests that future research will have to widen the bid range even beyond seemingly reasonable upper bounds in order to cover the tails of WTP. Our experiment extends the upper bound of experimental prices well beyond those in the existing literature and yet, like some of the previous studies, finds significant demand at the maximum bid price. This renders central measures of the estimated WTP distribution sensitive to the assumed utility model and error distribution. While mitigated somewhat by the joint mixture model or the use of nonparametric estimators, ours and similar existing WTP estimates should be interpreted cautiously. The most conservative lower-bound estimate in our data would be €6.30 for mean and €0.30 for median WTP for one ton of voluntary CO_2 emissions reductions.

The paper proceeds as follows: We describe the experimental design, protocol, and data in the following section. We then outline the econometric analysis in section 3 and present and discuss the results in section 4. Section 5 concludes.

2 Experimental design and data

The basic design choices of the experiment are intended to address a number of issues that arise in the context of assessing WTP in the field. First, the design combines the advantages of a standard dichotomous choice format (e.g. Lusk and Hudson 2004, Shogren 2006)—such as short administration time, limited cognitive load, and a familiar decision situation—with incentivized choices in order to alleviate the potential hypothetical bias of stated preferences methods (e.g. Cummings et al. 1995, Harrison 2006, Harrison and Rutström 2008). Secondly, the field experimental design combined with the comparably large sample, that is representative with respect to sex, age, and region of residence for the Internet-using population of Germany, enhances external validity. Thirdly, the design excludes to the greatest extent possible confounding public or private good attributes associated with the experimental good by employing, on the one hand, a website-based certification procedure of the EUA retirements, and by reminding, on the other hand, the subjects of the spatial indifference of local reductions for a global effect on the climate. If subjects received EUA retirement certificates in hardcopy, for example, it would plausibly increase WTP not because of the GHG emissions reduction, but because of the curiosity dimension of the good or because of private co-benefits derived from an increased visibility of the decision to others. Fourthly, with a focus on endogenous, "homegrown" values (Cummings et al. 1995), we do not provide any exogenous information regarding the issue of climate change or the employed metric for emissions reductions.[3] This allows us to investigate endogenous knowledge as a driver of WTP in the joint estimation (Cameron and Englin 1997) and parallels the problem policy makers would face when promoting VCA, at least in the short run. While this is not a necessary design choice,[4] providing "unbiased full information" (Arrow et al. 1993, Munro and Hanley 1999) would be difficult given the complexity of the issue on the one hand and the requirement of a particularly low cognitive load in an experiment that runs online and with members of the general population on the other. The "snap shot" character of the design extends to the point that our experiment, like others, elicits demand given the current market equilibria for (im)perfect substitutes for the experimental good and given existing national or international climate policies.[5] Thus, some of a subject's demand may already be met, and the results need to be interpreted in this light.

2.1 Procedures

The experiment was administered using the infrastructure of a large Internet polling company. The recruitment of subjects followed the standard routine of our cooperation partner in which panel members are invited via an email message to proceed to the survey via a hypertext link. The introductory screen explained the thematic focus of the survey, the expected duration of the survey (ten minutes), and the use of the random

[3]Note that the field nature of the experiment allows subjects to collect additional information while the experiment is in progress (e.g. by consulting Internet resources on the side in a separate browser window). In the data on subjects' speed of progress in the experiment, however, we do not find much evidence for simultaneous endogenous information acquisition—an observation that has also been made in the literature (Berrens et al. 2004).

[4]Among the existing studies on VCA, the amount of information provided by the researchers differs. While Brouwer et al. (2008) appear to be silent about causes of climate change and used metrics, MacKerron et al. (2009) provide a minor amount of information on metrics, and Löschel et al. (2013) provide information on both climate change and metrics (based on the IPCC report).

[5]Thanks to an anonymous referee for pointing this out.

incentive system with a prize worth up to a three-digit Euro figure.[6] Following the invitation screen, there was a filter screen to focus on German subjects of voting age.[7] Participants then saw 10 to 13 computer screens asking for 16 to 19 choices or answers, depending on their decisions.[8] Median completion time was approximately five minutes.[9]

Subjects' valuation decision was collected using two screens, one that introduced the good to be valued and set up the choice (subsequently called "information screen") and one that explained the payment procedures and collected the choice (subsequently called "decision screen"). The information screen explained three features of the experiment: (1) the trade-off between a cash prize and the CO_2 emissions reduction, including a succinct explanation of how the deletion of an EUA reliably reduces EU carbon emissions, (2) the public good character of the emissions reduction and (3) the random incentive system with odds of 100 in 5,000.[10] The decision screen explained the consequences if the subject was drawn as a winner, and elicited the subject's choice. Subjects that chose the cash amount received the award through their personal account at the polling company[11] while those who chose the emissions reductions could verify that the emissions reduction had been carried out through authoritative certification presented on a university website. The two prize alternatives were presented in random order, including the randomly determined subject-specific cash amount, and subjects had to check the preferred option.

The experiment concluded with a set of screens containing follow-up questions on the knowledge, attitudes, and beliefs about climate change, EUAs, and the metrics used. For subjects that had chosen the cash prize, a screen testing for field price censoring was inserted. Finally, the survey collected socio-demographic variables.

[6]These design criteria would have been familiar to panel members from previous polls as they decided on whether to proceed. The polling firm would regularly incentivize polls through either a piece-rate reward of approximately €1 for 20min expected survey time or random (lottery) prizes, e.g. in the form of a shopping vouchers.

[7]Subjects of other nationalities were redirected to other surveys running at the same time.

[8]Some screens and questions were due to a second valuation question posed after the independent first one reported here. For a translation of the relevant experimental screens, see Diederich and Goeschl (2011b). Screen shots are available from the authors upon request.

[9]Mean completion time was 1 hour 18 minutes. This is driven by a small fraction of surveys (about 4%) in which subjects availed themselves of the opportunity to leave the survey and continue hours or days later.

[10]The number of participants implied here is due to additional treatments running at the same time.

[11]A panel member can convert his or her account balance into cash as soon as a threshold of €50 is reached.

The Internet experiment ran in two sessions in May and July 2010. Session 1 lasted from May 25th to June 2nd and generated 1,640 complete[12] observations from 1,817 invitations. Session 2 lasted from July 19th to 27th and generated 800 complete observations out of 888 invitations. In the pooled sample, answers to the open-ended questions revealed 85 subjects who either objected to the EU ETS as a proper method to reduce emissions or said they distrusted the experiment itself. Following the usual procedure in the literature, these observations were excluded from the subsequent analysis.[13] The experiment was preceded by a set of pre-tests and a pilot experiment with 200 economics students in order to test the online implementation and refine the set of texts and questions.

2.2 Data

Table 1 reports descriptive statistics of experimental variables and matched environmental controls. Table 2 compares the sample means of key socio-demographic characteristics with census data. While showing considerable variation, all compared demographics of the Internet-representative sample turn out to be statistically different from those of the general German population. As one might expect, the average Internet user is more likely to be male, younger, and educated, and lives with more children. Regarding household income, both very low and very high income categories are slightly underrepresented: While mean income in the census data is higher,[14] the difference reverses if one drops incomes above €5,000.

Subjects' stated views regarding climate change in Table 1 are in line with a characterization that citizens are generally concerned about climate change and have some understanding about the physical inertia of the climate problem, but also differ in their knowledge about the metrics involved. A majority accepts that their lifestyle is contributing to climate change and understands that current emissions reductions do not benefit themselves but instead constitute an intertemporal benefit transfer to future gen-

[12]We count an observation as complete if the subject saw the final dismissing screen. All screens required an answer for each question by entering text or choosing at least one of the options (including "I don't know" options) before being able to proceed to the next screen.

[13]Results presented are not sensitive to inclusion or exclusion of these observations.

[14]Income categories above €5,000 were checked by 6% in our sample, while census data indicate a share of around 19%.

Table 1: Descriptive statistics of experimental variables and matched controls

Variable	Description	Mean	S.d.	Min.	Max.
A. Socio-demographic characteristics					
Female	1 if female	0.469	0.499	0	1
Age	Years	45.42	14.68	18	89
Children in HH	Number of household members below age 18	0.466	0.846	0	6
Education	Years based on highest educational degree	12.27	3.213	9	22
Income	Monthly household net income[a] (Euros)	2,556	1,706	450	8,000
> 1 citizenships	1 if has citizenship besides German	0.017	0.129	0	1
B. Climate change attitudes and beliefs					
Personal benefits	Degree of agreement to personal benefits from effects of carbon emissions reductions[b]	2.367	0.990	1	4
Future benefits	Degree of agreement to benefits for following generations from today's emissions reductions[b]	2.902	0.967	1	4
Lifestyle impact	Degree of agreement that personal lifestyle has contributed to climate change[b]	2.761	0.951	1	4
Footprint estimate	Estimate of yearly CO_2 emissions from lifestyle (metric tons)	3,021[c]	15,340	0	100,000
Footprint est. confidence	Confidence in own footprint estimate, 1 if at least "rather sure"	0.075	0.263	0	1
EUA price estimate	Estimate of current EUA spot price (Euros)	1,656[d]	10,306	0	100,000
Price est. confidence	Confidence in own EUA price estimate, 1 if at least "rather sure"	0.106	0.308	0	1
EUA availability	Believes that EUAs would be personally available for purchase somewhere else (1 if at least "rather yes")	0.197	0.398	0	1
C. Matched environmental controls					
Media coverage	Number of hits in a climate change related keywords search in German print and online media[e,f]	136.9	28.13	69.5	160
Temperature	Mean ambient air temperature in subject's region of residence[f,g] (°C)	15.1	4.186	8.05	25.8

Notes: [a] In our income approximation from subjects' reported income categories, for the "less than €500" category, we assume €450. For the two categories above €5,000, we assume €8,000 for compatibility with German census data. The remaining categories have a width of €500 each. [b] Answer categories: 1="no", 2="rather no", 3="rather yes", 4="yes" [c] Median is 10 [d] Median is 50 [e] Keywords used: '*climate change*', '*climate protection*', '*global warming*', '*carbon dioxide*', 'CO_2'. Database: LexisNexis [f] The variable is the moving 2-day average of the daily values of the day at which the subject took the experiment and the day before [g] Source: German National Meteorological Service (DWD)

Table 2: Socio-demographics: sample vs. census

Variable	Mean values		T-test (two-sided)
	Experimental sample	German census data	
Female	0.469	0.521	$p < 0.01$
	(0.499)	(0.500)	
Age	45.43	50.05	$p < 0.01$
	(14.68)	(18.31)	
Children in HH	0.466	≈0.340	$p < 0.01$
	(0.846)	(≈0.900)	
Education	12.27	11.02	$p < 0.01$
	(3.214)	(3.01)	
Income	2,556	4,057	$p < 0.01$
	(1,705)	(1,170)	
Income $\leq 5,000$	2,205	2,150	$p < 0.05$
	(1,030)	(1,300)	

Notes: Standard errors in parantheses. Source: Federal Statistical Office (Destatis), Mikrozensus 2008, 2009, EVS 2008 and own computations

erations. The evidence on the understanding of the metrics is mixed: While the median subject provides a surprisingly close estimate to the yearly per-capita carbon emissions in Germany (about 10 tons), a number of subjects has difficulties in giving a reasonable estimate, and only 7.5% feel at least somewhat certain about their guess. A similar pattern arises for estimates of spot prices of EU emissions allowances (about €15 at the time of the experiment).

Environmental controls were matched to subjects using data from a print and online media database (LexisNexis) and from the National Meteorological Service (DWD). The Germany-wide media coverage can be matched to subjects by experimental day while the temperature data can be matched by both experimental day and region of residence (Bundesland). Both variables reflect the 2-day moving average of daily values of the day at which the subject decided to start the experimental survey and the day before. In order to verify the robustness of the media coverage variable, we used two mutually exclusive sets of keywords who turned out to be highly correlated (correlation coefficient 0.81).

Descriptive results regarding subjects' valuation choices can be summarized as follows. In total, 382 (16.2%) of 2,354 subjects chose the emissions reduction through the retirement of an EUA. 1,972 (83.8%) chose the cash amount. Despite a bid range that is considerably larger than in previous studies, offered prices do not cover the tails of the

WTP distribution: At the lowest two bids (€2 and €4), still about two third of subjects are not willing to take the reduction and thus, reveal an even lower WTP or complete indifference to the choice. At the highest bid of €100, still about one sixth of subjects choose the reduction and thus reveal a WTP even larger. Note that in general, a low price elasticity appears not unusual for voluntary contributions to public goods (Green 1992, Diederich and Goeschl 2011a).

3 Analysis

The econometric analysis jointly models (1) the WTP decision whether to contribute the emissions reduction, (2) the participation decision whether the individual is indifferent to the offered choice, and (3) the endogenously determined knowledge regarding the valued good. To model the WTP decision, we employ a version of the classic Bishop-Heberlein model, which is a frequently used model for dichotomous choice data in contingent valuation. Thus, subject i's probability to choose the emissions reduction instead of the money can be expressed as

$$\Pr{}_i\,(\text{choice is emissions reduction}) = \begin{matrix} 1 - G_\varepsilon\left(-Z_i'\alpha + \beta \ln t_i\right) & \text{if } t_i < y_i \\ 0 & \text{if } t_i \geq y_i \end{matrix}\,. \tag{1}$$

where G_ε is the cdf of the error term of the utility difference[15], t_i denotes the offered cash amount, y_i is income, Z_i is a vector of covariates, and (α, β) is the parameter vector. The Bishop-Heberlein model has two advantages in our case. First, it bounds WTP from below at zero, a necessary assumption for a mixture model with a spike at zero, without further need of truncation.[16] Second, it assumes constant marginal utility of

[15]Shown is the formulation as a random utility model (RUM) censored from above by income (Hanemann and Kanninen 1999). The formulation as a expenditure difference model (Cameron 1988) is analogue (Haab 1999, Hanemann and Kanninen 1999).

[16]Our data provides little evidence on the presence of negative WTP. When asked about the reason for choosing the cash prize, only two of 1,972 subjects expressed some disutility from the emissions reduction, with harmful consequences for the economy as one of the main arguments. However, since subjects were not forced to use this open-ended answer option, this number may understate the true share. By contrast, 71% of subjects expect some benefits for future generations from the reduction, 45% expect some personal benefits, and 72% give examples of personal climate-friendly behavior.

income which appears reasonable for our data.[17]

In order to account for indifference, choice probabilities in the experiment become

$$\begin{aligned} \widetilde{\Pr}_i\,(\text{choice is emissions reduction}) &= (1-\gamma_i)\,\Pr{}_i\,(\text{choice is emissions reduction}) \\ \widetilde{\Pr}_i\,(\text{choice is money}) &= \gamma_i + (1-\gamma_i)\,\Pr{}_i\,(\text{choice is money}) \end{aligned} \tag{2}$$

where $(1-\gamma_i)$ denotes subject i's probability of participation, i.e., of being not indifferent but having a positive WTP (Haab 1999, Hanemann and Kanninen 1999). In order to estimate $(1-\gamma_i)$, we follow two complementary approaches. One is to make $(1-\gamma_i)$ a function of covariates,

$$(1-\gamma_i) = Q_i'\theta + \zeta_i, \tag{3}$$

and to identify participation for each subject through an indicator variable, POSWTP. In order to classify subjects, we analyze answers to two open-ended survey questions. One asked whether the subject intentionally behaves climate protecting and solicited examples of individual behavior or measures aimed at mitigating climate change. The other question asked for reasons for choosing the money instead of the emissions reduction. Coding of these answers was done conservatively such that unclear answers were treated as participation, POSWTP=1. The other approach is to assume $(1-\gamma_i) = (1-\overline{\gamma})$ as constant across individuals and exogenously given by the share of subjects who chose the reduction in a separate treatment with 39 subjects which was identical to the one described here except that subjects faced an alternative cash prize of €0.[18] Since there are several other plausible reasons for choosing cash besides indifference in this case (e.g. "protest voting" due to disappointment from being assigned a zero cash prize) we interpret this share as a lower bound of $(1-\overline{\gamma})$.[19]

[17] Performing a linear grid search using a single-equation Box-Cox model, we find maxima of the log-likelihood function around $\lambda = 1$.

[18] 21 of 39 subjects (46.15%) facing a zero cash prize chose the emissions reduction.

[19] Other approaches used in the literature are to estimate $(1-\gamma)$ endogenously if no information about individual indifference is available (An and Ayala 1996, Haab 1999) or to utilize a follow-up question on indifference without estimating it as a function of covariates (Kriström 1997). The downside of both approaches is that if the assumed distribution of WTP and the true distribution of $(1-\gamma)$ depend on the same covariates, then treating $(1-\gamma)$ as constant or treating both equations as independent can bias the estimates of covariate effects on WTP (Werner 1999) and makes them more sensitive to the assumed distribution for the error terms (Haab 1999). If we estimate $(1-\gamma)$ endogenously using (2), we obtain participation probability estimates of around 70% for the lognormal, 60% for the log-logistic, and about

For the information submodel, we follow Cameron and Englin (1997) in assuming that the error terms of the WTP equation (1) and the information equation,

$$\text{INF}_i^{\text{obj,subj}} = R_i'\delta + \eta_i, \tag{4}$$

are correlated. In addition, we assume information and participation to be correlated. In the estimation results, we subsequently compare two possible proxies for information in our data as alternative dependant variables. The first, INF^{obj}, provides a measure of the "objective" knowledge related to carbon emissions and is constructed from the standardized deviation variables of subjects' EUA price and carbon footprint estimates. The second, INF^{subj}, represents "subjective" knowledge and is constructed from subjects' standardized self-assessed quality of both estimates. Both variables are distributed approximately normally.

4 Results

4.1 Estimation results

If we assume ε_i, ζ_i, and η_i to be multivariate normally distributed with correlations ρ and ν, then the individual log-likelihood can be written as

$$\begin{aligned}
\log L_i \quad = \quad & \ln\left[(1/\eta_i)\phi\left((\text{INF}_i^{\text{obj,subj}} - R_i'\delta)/\eta_i\right)\right] \\
& + \text{POSWTP}_i I_i \ln\left[\Phi\left(P_i\right)\Phi\left(W_i\right)\right] \\
& + \text{POSWTP}_i (1 - I_i)\ln\left[\Phi\left(P_i\right)\left(1 - \Phi\left(W_i\right)\right)\right] \qquad (5) \\
& + (1 - \text{POSWTP}_i)\ln\left[1 - \Phi\left(P_i\right)\right]
\end{aligned}$$

50% for the Weibull version of (2).

where

$$
\begin{aligned}
P_i &= \frac{Q_i'\theta - \nu\left(\left(\mathrm{INF}_i^{\mathrm{obj,subj}} - R_i'\delta\right)/\eta_i\right)}{\left(1-\nu^2\right)^{0.5}}, \\
W_i &= \frac{\left(Z_i'\alpha - \beta \ln t_i\right) - \rho\left(\left(\mathrm{INF}_i^{\mathrm{obj,subj}} - R_i'\delta\right)/\eta_i\right)}{\left(1-\rho^2\right)^{0.5}}
\end{aligned}
$$

and I_i is the subject's discrete prize choice with $I_i = 1$ if the subject chooses the reduction over the offered cash amount t_i. The log-likelihood function $\log L = \sum_{i=1}^{N} \log L_i$ can then be maximized with respect to the coefficient vectors $(\alpha, \beta, \delta, \theta)$ and the parameter values (η, ν, ρ).

If $(1-\gamma_i) = (1-\overline{\gamma})$, then individual log-likelihood of the resulting two-equations model becomes

$$
\begin{aligned}
\log L_i = \; & \ln\left[(1/\eta_i)\phi\left(\left(\mathrm{INF}_i^{\mathrm{obj,subj}} - R_i'\delta\right)/\eta_i\right)\right] \\
& + I_i \ln\left[(1-\overline{\gamma})\Phi\left(W_i\right)\right] \\
& + (1-I_i)\ln\left[(1-\overline{\gamma})\left(1-\Phi\left(W_i\right)\right) + \overline{\gamma}\right]
\end{aligned} \tag{6}
$$

and $\sum_{i=1}^{N} \log L_i$ is maximized with respect to the coefficient vectors (α, β, δ) and parameter values (η, ρ).

Table 3: Parameter estimates

	Standard censored lognormal	Joint censored lognormal-normal-normal, subjective information	Joint censored lognormal-normal-normal, objective information	Joint censored lognormal-normal, with exog. part., subjective inf.	Joint censored lognormal-normal, with exog. part., objective inf.
A. WTP submodel					
Cash prize amount	-0.2012***	-0.2037***	-0.2106***	-0.3327***	-0.3592***
	(0.039)	(0.043)	(0.044)	(0.075)	(0.078)
Female	0.0650	0.0422	0.0303	0.0175	-0.0150
	(0.077)	(0.085)	(0.086)	(0.135)	(0.138)
Age	0.0008	0.0021	0.0015	0.0030	0.0023
	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)
Children in HH	-0.0348	-0.0703	-0.0754	-0.0726	-0.0876
	(0.047)	(0.051)	(0.052)	(0.075)	(0.078)
Education	0.0495***	0.0575***	0.0587***	0.0833***	0.0845***
	(0.011)	(0.012)	(0.012)	(0.020)	(0.020)
Personal benefits	0.1489***	0.1368**	0.1358**	0.2575***	0.2458***
	(0.051)	(0.057)	(0.058)	(0.086)	(0.089)
Future benefits	0.2467***	0.2256***	0.2145***	0.3095***	0.3130***
	(0.059)	(0.066)	(0.067)	(0.093)	(0.095)
Lifestyle impact	0.1504***	0.1332**	0.1208**	0.2737***	0.2683***
	(0.051)	(0.057)	(0.057)	(0.086)	(0.088)
EUA availability	0.0025	-0.0292	-0.0554	-0.0797	-0.0790
	(0.089)	(0.098)	(0.099)	(0.157)	(0.160)
Price est. precision	0.1779***	–	–	–	–
	(0.042)				
Price est. confidence	0.3195**	–	–	–	–
	(0.128)				
Footprint est. precision	-0.0220	–	–	–	–
	(0.056)				
Footprint est. confidence	-0.5699***	–	–	–	–
	(0.164)				
Media coverage	0.0000	-0.0009	-0.0011	-0.0034	-0.0036
	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
Ambient temperature	0.0185**	0.0204**	0.0196*	0.0300*	0.0338**
	(0.009)	(0.010)	(0.010)	(0.016)	(0.017)

Continued on next page

Table 3 – *Continued from previous page*

	Standard censored lognormal	Joint censored lognormal-normal-normal, subjective information	Joint censored lognormal-normal-normal, objective information	Joint censored lognormal-normal, with exog. part., subjective inf.	Joint censored lognormal-normal, with exog. part., objective inf.
Constant	-2.9099***	-2.5286***	-2.3676***	-2.5931***	-2.4328***
	(0.402)	(0.443)	(0.445)	(0.666)	(0.677)
B. Participation submodel					
Female	–	0.3028***	0.2861***	–	–
		(0.089)	(0.090)		
Age	–	0.0071**	0.0069**	–	–
		(0.003)	(0.003)		
Children in HH	–	0.0918*	0.0787	–	–
		(0.052)	(0.053)		
Education	–	-0.0196	-0.0196	–	–
		(0.013)	(0.013)		
Personal benefits	–	0.2347***	0.2413***	–	–
		(0.060)	(0.060)		
Future benefits	–	0.1011*	0.0941*	–	–
		(0.055)	(0.056)		
Lifestyle impact	–	0.2950***	0.3189***	–	–
		(0.052)	(0.053)		
EUA availability	–	-0.1325	-0.1219	–	–
		(0.105)	(0.105)		
Media coverage	–	-0.0035**	-0.0029*	–	–
		(0.002)	(0.002)		
Ambient temperature	–	-0.0209**	-0.0185*	–	–
		(0.011)	(0.011)		
Constant	–	0.0556	-0.1141	–	–
		(0.426)	(0.426)		
$(1-\gamma)$	1	0.8421	0.8392	0.4615	0.4615
	(–)	(0.133)	(0.137)	(–)	(–)
C. Information submodel					
Female	–	-0.1401***	-0.0791**	-0.1407***	-0.0793**
		(0.045)	(0.031)	(0.045)	(0.031)
Age	–	-0.0057***	-0.0040***	-0.0057***	-0.0040***

Continued on next page

Table 3 – *Continued from previous page*

	Standard censored lognormal	Joint censored lognormal-normal-normal, subjective information	Joint censored lognormal-normal-normal, objective information	Joint censored lognormal-normal, with exog. part., subjective inf.	Joint censored lognormal-normal, with exog. part., objective inf.
		(0.002)	(0.001)	(0.002)	(0.001)
Children in HH	–	0.0884***	0.0253	0.0887***	0.0254
		(0.026)	(0.018)	(0.026)	(0.018)
Education	–	0.0188***	0.0092**	0.0190***	0.0093**
		(0.007)	(0.005)	(0.007)	(0.005)
Income	–	0.0000	0.0000	0.0000	0.0000
		(0.000)	(0.000)	(0.000)	(0.000)
> 1 citizenships	–	0.2694	-0.0588	0.2628	-0.0624
		(0.169)	(0.116)	(0.169)	(0.116)
Media coverage	–	-0.0019**	0.0003	-0.0020**	0.0003
		(0.001)	(0.001)	(0.001)	(0.001)
Constant	–	0.3384**	-0.7869***	0.3394**	-0.7864***
		(0.163)	(0.112)	(0.163)	(0.112)
ρ	–	0.1076***	0.1270***	0.1234*	0.2078***
		(0.041)	(0.038)	(0.063)	(0.067)
ν	–	-0.0650	-0.0323	–	–
		(0.043)	(0.042)		
N	1842	1597	1552	1597	1552
Log-likelihood	-757.462	-3225.815	-2531.033	-2664.652	-1983.074
χ^2	219.989	124.219	116.112	114.448	105.675
Pseudo R^2	0.127	–	–	–	–

Notes: Standard errors in parentheses. *** Significant at or below 1 percent. ** Significant at or below 5 percent. * Significant at or below 10 percent. Controls for experimental date, region of residence, or Eastern Germany are not included in the regressions reported here. If included they would not yield further significances and would not change WTP coefficient estimates much. See Diederich and Goeschl (2011b) for specifications that include some of these controls.

Table 3 presents the estimation results. The first column corresponds to the standard lognormal single-equation model (1). The second and third columns report results of the joint three-equations model (5), and the last two columns report results of the joint two-equations model (6) with exogenous participation probability. In all five columns, the significance levels of the estimated parameters rarely change between models and specifications. At the same time, there is reason to suspect the presence of a bias in the single equation model, indicated by differences in the magnitudes of the significant coefficients estimates between the single and the joint models. In addition, the significance of the EUA price and footprint estimates in the single equation model justifies employing a joint model that endogenizes prior information among subjects. Across the joint models, the choice of the proxy for participation clearly matters as a comparison of the coefficient estimates of the two and the three-equation model shows. This points to a strong impact of the participation rate $(1 - \gamma)$ on estimates. Finally, comparing the two specifications for each of the joint models, coefficient estimates are largely invariant with respect to the information proxy used: Subjective (left column) and objective (right column) measures of subjects' knowledge return broadly similar estimates, except for the coefficient estimates in the information submodel itself.

The estimation results in Table 3 return signs for the significant variables that are within expectations. Starting with the WTP submodel, the price variable has the desired negative effect and is highly significant, irrespective of the model. Among the socio-demographic variables, education stands out as a highly significant correlate of the choice of the emissions reduction. Among attitudinal variables, the expectation of benefiting future generations shows a higher correlation with the propensity to choose the EUA than the expectation of personal benefits or the acknowledgement of personal negative contributions to climate change. Finally, the two matched environmental controls deliver an unexpected effect: While a larger number of recent news items related to climate change has no direct effect, higher mean temperatures in subjects' regions of residence around the time of the experiment are associated with a higher propensity to opt for the GHG emissions reduction.[20] Moving on to the participation submodel,

[20]The effect is robust to including maximum instead of mean temperatures at the ten percent level of significance.

we observe important subtleties regarding the determinants of the prize choice: Female subjects, older subjects, and, in part, subjects with more children in the household are more likely to indicate non-indifference to VCA. In addition, attitudinal variables reverse their role in the participation submodel compared to the WTP submodel: Personal benefits and acknowledged lifestyle impacts are much stronger correlated with participation than are future benefits. Both recent media coverage as well as temperatures decrease statements of participation in VCA. At around 84%, the predicted average probability of participation in the three-equations-model is close to the observed share of subjects who indicate non-indifference to VCA. Finally, in the information submodel, female and older subjects are both subjectively and objectively less informed. Subjects with more children in the household feel better informed, but do not differ significantly in objective terms. As expected, education is a significant driver of endogenous information status, thus adding to the direct effect of education on WTP through ρ. Interestingly, higher recent media coverage has a slightly negative effect on the subjective assessment of information. The correlation of the endogenous information status, both subjective and objective, and WTP is always highly significant. In contrast, endogenous information status is not correlated with indifference, as ν is insignificant.

4.2 Willingness to Pay

The specific experimental design permits interpretation of the welfare measure elicited as both WTP and willingness to accept (WTA) since either of the two implied reference points can be defended.[21] We subsequently follow the literature and report WTP.

[21] First, the experimental task may be interpreted as eliciting minimum WTA to forgo the emissions reduction. To see this, denote the vector of public goods with and without the emissions reduction as q^1 and q^0, respectively, and let the corresponding element of q be given in terms of abatement. Then, the equivalent variation $v(p, q^1, y) = v(p, q^0, y + E)$ defines WTA, where p is the vector of prices for private goods. Second, however, the converse of perceiving the the experimental choice as a purchasing decision appears equally if not more plausible. In this case, maximum WTP is given by the compensation variation, $v(p, q^1, y' - C) = v(p, q^0, y')$, with income $y' = y + t$. Note that most of the reasons believed to create the frequent disparity between WTP and WTA do not apply in our case (Hanemann 1999). First, the notion of a "loss" à la Tversky and Kahnemann and thus, loss aversion is ambiguous and depends on the reference point used as described above. Second, we use a closed-ended and paid elicitation format, which has been suggested to minimize the strategic incentives to understate WTP and overstate WTA found for open-ended, unpaid elicitation formats and paid auctions. Third, when the change in q is small, a WTP/WTA disparity due to low elasticities of substitution between public and private good and income elasticities larger than unity should play a negligible role. One potential bias in this context is the possible presence of a "windfall" (Keeler et al. 1985) or "house money" (Thaler and Johnson 1990)

Table 4 reports WTP computations and illustrates the impact of covariates on mean and median WTP estimates. Beginning with the covariates, the main results of the first five columns can be summarized as follows. First, the joint models, which allow for indifferent subjects, feature considerably higher mean and median WTP estimates than the single-equation model.[22] The strong impact of accounting for indifference is also emphasized by substantial differences in estimates between the three-equations model and the two-equations model with exogenous participation rate. In contrast to the observed sensitivity to the participation rate, WTP estimates of the joint models are highly robust to employing either the subjective or the objective proxy for knowledge. Second, calibrating the vector of covariates using values from the census (Table 2) gives mean and median WTP estimates of about 77% of the values for the experimental sample. Thus, we would expect estimates for a truly representative sample to result in numbers that are about one quarter below ours. Third, changes in climate change attitudes and education affect WTP considerably. For example, the calibration to reflect a well-informed "realist", who acknowledges future benefits and personal lifestyle impacts but not personal benefits from today's reductions, more than doubles median WTP compared to the sample mean.

Turning to the absolute level of WTP, the differences between mean and median estimates are considerable in all versions of the models. This points to the fact that despite censoring by income and even though the joint models allow for indifference, the effect of the "fat tails" (Boyle et al. 1988) on both sides of the empirical WTP distribution persists. Thus, mean estimates are sensitive to the assumed model of utility, the imposed distribution of the error terms, and the bounds implied by non-negativity and income restrictions. The challenges of distributional and model assumptions also affect the median estimates since the empirical median falls below the lowest bid price (Hanemann and Kanninen 1999).[23] The estimated levels, particularly of mean WTP,

effect common to all experiments in which subjects always gain irrespective of their choice. Evidence on the presence, direction, and scale of a potential bias in public good situations is inconclusive, however (Harrison 2007).

[22]Running separate regressions including either the participation or the information submodel shows that the changes in WTP estimates are almost entirely due to allowing for indifference.

[23]For example, mean (median) WTP estimates in the first column at the sample mean are €127.40 (€0.22) if errors are assumed log-logistic and €139.36 (€0.22) if errors are distributed Weibull.

Table 4: Mean and median WTP

	Standard censored lognormal	Joint censored lognormal-normal-normal, subjective information	Joint censored lognormal-normal-normal, objective information	Joint censored lognormal-normal with exog. part., subjective inf.	Joint censored lognormal-normal with exog. part., objective inf.	Turnbull LB / UB
At sample mean	€ 109.82	€ 135.69	€ 132.48	€ 189.26	€ 173.76	€ 6.31 / €45.69
	[€ 0.18]	[€ 0.34]	[€ 0.41]	[€ 10.39]	[€ 12.03]	–
At German census average	€ 85.47	€ 105.50	€ 102.59	€ 148.02	€ 136.38	–
	[€ 0.14]	[€ 0.27]	[€ 0.32]	[€ 8.16]	[€ 9.50]	–
At example calibrations of variables:[a]						
"Enthusiast"	€ 371.07	€ 394.73	€ 370.98	€ 818.21	€ 751.97	€ 22.20 / € 73.18[b]
(Personal benefits=4, Future benefits=4, Lifestyle impact=4)	[€ 5.95]	[€ 7.77]	[€ 7.35]	[€ 283.16]	[€ 283.16]	–
"Enthusiast" with university degree	€ 546.04	€ 608.52	€ 582.32	€ 1,219.78	€ 1,150.56	–
	[€ 24.32]	[€ 39.11]	[€ 36.24]	[€ 1,188.96]	[€ 930.13]	–
"Realist"	€ 178.40	€ 205.40	€ 191.77	€ 317.75	€ 298.70	€ 16.87 / € 22.30[b]
(Personal benefits=1, Future benefits=4, Lifestyle impact=4)	[€ 0.65]	[€ 1.04]	[€ 1.06]	[€ 27.79]	[€ 31.03]	
"Realist" with university degree	€ 288.90	€ 350.32	€ 333.46	€ 596.56	€ 569.06	–
	[€ 2.64]	[€ 5.21]	[€ 5.23]	[€ 116.67]	[€ 119.41]	–

Notes: Mean WTP is calculated as $\int_0^{y^*} \Phi(\bar{Z}'\hat{\alpha} - \hat{\beta}\ln t)\mathrm{d}t$ where y^* is the median monthly income of the sample (EUR 2.250) or the German census (EUR 1.750), respectively. Median WTP is calculated as $\min(y, \exp(\hat{\alpha}/\hat{\beta}))$ and reported in square brackets (Hanemann and Kanninen 1999). [a] Values of explanatory variables not made explicit are mean values of the sample, $\bar{Z}$. [b] Values are derived by pooling answer categories 1 and 2 as well as 3 and 4 of the three attitude variables therefore need to be interpreted with caution. Observing that (1) the exogenous participation probability of 46% in the two-equation model can be interpreted as a lower bound of participation in the sample and (2) the conservative coding of POSWTP for the three-equation model can be interpreted as an upper bound, the two models may be plausibly viewed as estimating an upper and lower bound for median WTP in our data. Regarding mean WTP, the last column of Table 4 reports estimates from the Turnbull Distribution-Free Estimator (Turnbull 1976, Carson et al. 1994) which has been suggested as a conservative approach for mean WTP in the presence of "fat tails" (Haab and McConnell 1997, 2002). Similar to our parametric model, the Turnbull allows for nonparticipation but indifferent subjects do not need to be uniquely identified.[24] In its most conservative lower bound version, the Turnbull exlusively relies on the assumption of non-negativity and the information that the WTP of a subject who chooses the reduction is not less than the alternative cash prize. Taken together, we would suggest a mean WTP of €6.31 and a median WTP between €0.30 and €12 as the most conservative and best available estimates for central measures of WTP in our sample. The Turnbull estimator has limited power to quantify covariate effects, however, as it does so by simply confining the estimator to subsamples that exhibit the desired configuration. If we compute the Turnbull for our subsamples of "enthusiasts" and "realists" (defined by including the "rather yes" and "rather no" categories to increase subsample size), lower bound estimates increase to about €22 and €17, respectively.

4.3 Discussion

In this section, we compare our findings to the existing literature and discuss two potential limitations of the preference elicitation. To begin with the former, Table 5 provides a comparison of covariate results with the VCA literature. In addition, results can be compared to some extent with two related strands: (1) papers on preferences for mandatory, collective climate policies (CCP) and (2) papers on voluntary contributions to public goods in general. Table 5 includes a selection of the former category[25] while the latter

[24] The assumption here is that subjects with $v(p, q^0, y_i'; z_i) = v(p, q^1, y_i'; z_i)$ choose cash at all prices $t_i > 0$.

[25] The selection of papers is based on comparability of the valued scenario and availability of covariates. See Johnson and Nemet (2010) for a more comprehensive survey of the growing literature on WTP for

has been discussed to some extent elsewhere (Diederich and Goeschl 2011a). The present study benefits from a larger set of covariates than most other studies on VCA or CCP. Among the socio-demographic variables, education stands out as the most frequent and unanimously positive driver. In contrast to our findings, income is positively correlated in most studies where available. A possible explanation for the ambigious results for gender and age may be countervailing effects in both variables that are specific to climate change. For example, the delayed arrival of benefits from emissions reductions may militate against older subjects contributing who in general have been found to give more (List 2004). The second panel in Table 3 reports on stated climate change attitudes. Making the variables comparable across the literature involves some imprecise adjustments, such as pooling the expectations of personal and future generation benefits. The almost equivocal finding is that of a positive correlation between WTP and benefit expectations as well as the acknowledgement of personal responsibility for climate change. The latter may not only arise from concerns of justice (Konow 2003) or offset motives (Kotchen 2009) but could also be driven by an "outrage" premium for human-made environmental damages (Bulte et al. 2005). A novel element in the present study is the matching of exogenous data of environmental conditions at the time and location of the valuation choice. While this allows to establish causality, it limits comparability to previous findings which are based on respondents' statements. If the effect uncovered in our results is not a general effect in public good provision but rather context-specific, a possible explanation may be a heuristic shortcut: subjects might associate lower GHG emissions with lower temperatures, making emissions reductions—without further reasoning—appear instantaneously more desirable. Without further evidence, however, such reasoning is entirely speculative.

The large range of estimates for central measures of WTP in our data and their sensitivity puts existing estimates of the VCA literature with comparable survey designs into perspective and warrants cautious interpretation of the available welfare measures for VCA. In comparison, we would expect our estimates in Table 4 to be lower than those of the two contingent valuation studies on VCA as they are based on stated pref-

CCP.

erences and respondents which—in light of the covariate results—will probably display an above-average WTP. Regarding results based on the Turnbull lower bound estimator, our data indeed suggests a lower mean WTP than found by Brouwer et al. (2008) at €25/tCO_2 by employing the Turnbull. Our parametric mean WTP estimates, however, considerably exceed the RUM based estimates around £24/tCO_2 by MacKerron et al. (2009). One plausible reason for this is that the "fat tail" in MacKerron's et al. data is at a lower maximum bid (£20) than ours (€100). Moreover, MacKerron et al. (2009) employ an unresticted linear RUM which allows for negative WTP. Among the existing estimates, our results most closely correspond to Löschel et al. (2013) who nonparametrically calculate a mean WTP at €12/tCO_2 (median at €0) from observed demand in a variant of the Becker-DeGroot-Marshak mechanism.

The literature as well as our design suggest two qualifications for our results. First, field price censoring (FPC) can arise in valuation experiments because prices for goods within the experiment cannot easily be isolated from prices "in the field" (Harrison et al. 2004, Cherry et al. 2004). As a result, there are circumstances when the experimentally observable WTP is censored at the level of the field price as subjects avail themselves of arbitrage opportunities. Careful examination of the data leads us to conclude that FPC is an unlikely source of bias in the present experiment (Diederich and Goeschl 2011a). Additional evidence to the analysis provided in Diederich and Goeschl (2011a) comes from answers to the post-experimental survey in which only 6.2 (13.5) percent of subjects confidently (tentatively) believe that they personally have access to the EUA market. Any remaining effect of FPC on covariate estimates will be at least partly accounted for by the information submodel while for WTP, estimates would be downward biased. Second, the number of sceptics about the reduction technology that is identified in the ex-post survey and excluded from the sample may be a lower bound as some subjects may be reluctant to mention their dislike about the employed reduction technology or other reservations in their answers.[26] Since they probably show up as participants in the data who place a positive value on climate change mitigation but choose cash due to unidentified scepticism, WTP estimates would be biased downward. A potential effect on

[26]Thanks to an anonymous referee for pointing this out.

Table 5: Covariate effects on WTP for voluntary climate action and for collective climate policies

	WTP for VCA (1tCO_2)				WTP for CCP							
	This study	Brouwer et al. (2008)	MacKerron et al. (2009)	Löschel et al. (2013)	Berrens et al. (2004)[a]	Viscusi and Zeckhauser (2006)[b]	Solomon and Johnson (2009)	Cai et al. (2010)	Carlsson et al. (2010)[c]	Lee et al. (2010)	Akter and Bennett (2011)[d]	Kaczan et al. (n.d.)
A. Socio-demographic characteristics												
Female	○	○	+	–	–	–	○	+	–	n.a.	n.a.	○
Age	○	○	n.a.	–	○	n.a.	n.a. / ○	n.a	–	○	n.a.	○
Children	○	○ / n.a.	○	○	n.a.	n.a.	n.a.	n.a.	○	n.a.	n.a.	○
Education	+	○ / n.a.	n.a.	+	+	n.a.	n.a.	n.a.	+	+	+	+
Income	○	+	○	+	+	n.a.	+	n.a.	+	+	n.a.	–
B. Climate change attitudes and beliefs												
Personal or future benefits	+	+[e]	n.a.	+[f]	n.a.	n.a.	+[f]	+	n.a. / ○	n.a.	+	n.a.
Lifestyle impact or personal responsibility	+	+	n.a.	+[g]	n.a.	n.a.	n.a.	+[h]	+[i]	n.a.	n.a.	n.a.
C. Environmental controls												
Media attention	○	n.a.	n.a.	+[j]	n.a.	n.a.	n.a.	n.a.	n.a	n.a.	+[k]	n.a.
Temperature	+	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	+[l]

Notes: +: positive effect, significant at 10 percent level or less –: negative effect, significant at 10 percent level or less ○: insignificant effect n.a.: variable or estimate not available [a] Based on the "pooled" model of Table 6. [b] Based on "Gas tax remedy till 2100" in Table IV. [c] Based on the unconditional WTP results for Sweden. [d] Based on stated WTP to support the proposed Australian carbon trading scheme. [e] Based on perceived effectiveness of a carbon tax. [f] Based on stated general concerns about climate change. [g] Based on stated importance of personal activity against climate change. [h] Based on the finding that WTP rises when a policy assigns larger cost share to groups or countries that are believed to be more responsible for climate change. [i] Based on the general statement that humans have affected the temperature increase. [j] Based on self-reported influence through media reports. [k] Respondents stated to have watched Al Gores's "An Inconvenient Truth". [l] Based on respondents' stated perception of generally rising temperatures.

covariate estimates would again be partially accounted for by the information submodel.

5 Conclusions

Individual, unilateral action to reduce GHG emissions, it has been suggested, could play an important part in the endeavor to tackle climate change. Empirical estimates of the public's willingness to engage in VCA are sparse, however, and its key drivers are not well understood. In this paper, we provide non-hypothetical estimates of the preferences for VCA by giving subjects the costly opportunity to reduce GHG emissions by one ton through the retirement of an EU ETS emissions allowance. To do so, we run a dichotomous choice online valuation experiment with a sample of voting-aged Germans. In contrast to prior studies, our sample is large, highly heterogeneous and Internet-representative, and subjects face a considerably larger range of bid prices. Furthermore, we explicitly take into account the public nature of the good and the voluntary contribution character of the valuation exercise and allow for zero WTP in our econometric analysis through a mixture model. Consistent with the approach to elicit "homegrown values" (Cummings et al. 1995), we focus on endogenous information status and model it jointly with WTP and the proxy for participation in VCA activities.

In the joint estimation, the main correlates of WTP in our experiment are subjects' education—both directly and indirectly through the endogenous information status—, their perceived benefits from emissions reductions—with a greater weight of altruistic compared to selfish benefits—, the acknowledgement of personal responsibility for climate change, and matched outdoor temperatures around the time and place of the experiment. Regarding subjects indication of not being generally indifferent towards VCA, the likelihood of participation is higher for females and increases with age, with stated benefit expectations and with perceived personal responsibility for climate change. Regarding the proxies for information status about climate change, young males with higher education are both subjectively and objectively better informed about climate change related facts. In addition, the subjective measure of information correlates positively with the number of children in the household and negatively with recent coverage of climate change in German media. Finally, subjects' endogenous information status and WTP highly significantly correlate.

Central measures of the estimated WTP distribution vary considerably with the identified drivers. When calibrated with German census means of covariates, we obtain estimates of about 77% of the estimates for our sample. In turn, a calibration to "realistic" expectations of benefits and personal lifestyle impact considerably increases WTP estimates compared to the sample mean. In absolute terms, both mean and median estimates are sensitive to model assumptions despite a much larger bid range than previous studies. If one wanted to arrive at a welfare measure based on the data we obtained, we would suggest a mean WTP of €6.30 and a median WTP of €0.30 for an individual voluntary contribution of one ton of GHG emissions reductions as conservative lower-bound estimates.

References

Abdellaoui, M., Baillon, A., Placido, L. and Wakker, P. P. (2011). The rich domain of uncertainty: Source functions and their experimental implementation, *The American Economic Review* **101**: 695–723.

Akter, S. and Bennett, J. (2011). Household perceptions of climate change and preferences for mitigation action: the case of the Carbon Pollution Reduction Scheme in Australia, *Climatic Change* **109**(3-4): 417–436.

An, Y. and Ayala, R. A. (1996). A mixture model of willingess to pay distributions, *Working Paper* .

Arrow, K. J., Solow, R., Portney, P. R., Leamer, E. E., Radner, R. and Schuman, H. (1993). Report of the NOAA panel on contingent valuation, *Technical report*. Washington D.C.

Baltussen, G., Post, T., Van den Assem, M. J. and Wakker, P. P. (2010). Random incentive systems in a dynamic choice experiment. Working Paper.

Berrens, R. P., Bohara, A. K., Jenkins-Smith, H. C., Silva, C. L. and Weimer, D. L. (2004). Information and effort in contingent valuation surveys: Application to global climate change using national internet samples, *Journal of Environmental Economics and Management* **47**(2): 331–363.

Boyle, K. J., Welsh, M. P. and Bishop, R. C. (1988). Validation of empirical measures of welfare change: Comment, *Land Economics* **64**(1): 94–98.

Brouwer, R., Brander, L. and Van Beukering, P. (2008). "A convenient truth": Air travel passengers' willingness to pay to offset their CO2 emissions, *Climatic Change* **90**(3): 299–313.

Bulte, E., Gerking, S., List, J. A. and de Zeeuw, A. (2005). The effect of varying the causes of environmental problems on stated WTP values: Evidence from a field study, *Journal of Environmental Economics and Management* **49**(2): 330–342.

Cai, B., Cameron, T. and Gerdes, G. (2010). Distributional preferences and the incidence of costs and benefits in climate change policy, *Environmental and Resource Economics* **46**(4): 429–458.

Cameron, T. A. (1988). A new paradigm for valuing non-market goods using referendum data: Maximum likelihood estimation by censored logistic regression, *Journal of Environmental Economics and Management* **15**(3): 355–379.

Cameron, T. A. and Englin, J. (1997). Respondent experience and contingent valuation of environmental goods, *Journal of Environmental Economics and Management* **33**(3): 296–313.

Carlsson, F., Kataria, M., Krupnick, A., Lampi, E., Lofgren, A., Qin, P., Chung, S. and Sterner, T. (2010). Paying for mitigation: A multiple country study, *Resources For the Future, Discussion Papers* .

Carlsson, F. and Martinsson, P. (2001). Do hypothetical and actual marginal willingness to pay differ in choice experiments?: Application to the valuation of the environment, *Journal of Environmental Economics and Management* **41**(2): 179–192.

Carson, R. T., Wilks, L. and Imber, D. (1994). Valuing the preservation of Australia's Kakadu conservation zone, *Oxford Economic Papers* **46**: 727–749.

Cherry, T., Frykblom, P., Shogren, J., List, J. and Sullivan, M. (2004). Laboratory testbeds and nonmarket valuation: The case of bidding behavior in a second-price auction with an outside option, *Environmental and Resource Economics* **29**(3): 285–294.

Cummings, R. G., Harrison, G. W. and Rutström, E. E. (1995). Homegrown values and hypothetical surveys: Is the dichotomous choice approach incentive-compatible?, *The American Economic Review* **85**(1): 260–266.

Diederich, J. and Goeschl, T. (2011a). Giving in a large economy: Price vs. non-price effects in a field experiment, *Discussion Paper No. 514* . Department of Economics, Heidelberg University.

Diederich, J. and Goeschl, T. (2011b). Willingness to pay for individual greenhouse gas emissions reductions: Evidence from a large field experiment, *Discussion Paper No. 517* . Department of Economics, Heidelberg University.

European Commission (2011). You control climate change. Available at: http://ec.europa.eu/clima/sites/campaign/index.htm [Accessed September 23, 2011].

Gore, A. and Guggenheim, D. (2006). *An Inconvenient Truth: A Global Warning*, Paramount Pictures. Documentary Movie.

Green, D. P. (1992). The price elasticity of mass preferences, *The American Political Science Review* **86**(1): 128–148.

Grether, D. M. and Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon, *The American Economic Review* **69**(4): 623–638.

Haab, T. C. (1995). *The Impact of Nonparticipants on Nonmarket Valuation Techniques*, PhD thesis.

Haab, T. C. (1999). Nonparticipation or misspecification? The impacts of nonparticipation on dichotomous choice contingent valuation, *Environmental and Resource Economics* **14**(4): 443–461.

Haab, T. C. and McConnell, K. E. (1997). Referendum models and negative willingness to pay: Alternative solutions, *Journal of Environmental Economics and Management* **32**(2): 251–270.

Haab, T. C. and McConnell, K. E. (2002). *Valuing environmental and natural resources: The econometrics of non-market valuation*, Edward Elgar Publishing, Northampton.

Hanemann, W. M. (1999). The economic theory of WTP and WTA, *in* I. J. Bateman and K. G. Willis (eds), *Valuing Environmental Preferences. Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*, Oxford University Press, Oxford, pp. 42–96.

Hanemann, W. M. and Kanninen, B. (1999). *The Statistical Analysis of Discrete-Response CV Data*, Oxford University Press, Oxford, chapter 11, pp. 302–441.

Hanemann, W. M. and Kriström, B. (1995). Preference uncertainty, optimal designs and spikes, *in* P.-O. Johansson, B. Kriström and K.-G. Mäler (eds), *Current issues in environmental economics*, Manchester University Press, Manchester, pp. 58–77.

Harrison, G. W. (2006). Experimental evidence on alternative environmental valuation methods, *Environmental and Resource Economics* **34**(1): 125–162.

Harrison, G. W. (2007). House money effects in public good experiments: Comment, *Experimental Economics* **10**(4): 429–437.

Harrison, G. W., Harstad, R. M. and Rutström, E. E. (2004). Experimental methods and elicitation of values, *Experimental Economics* **7**(2): 123–140.

Harrison, G. W. and List, J. A. (2004). Field experiments, *Journal of Economic Literature* **42**(4): 1009–1055.

Harrison, G. W. and Rutström, E. E. (2008). Experimental evidence on the existence of hypothetical bias in value elicitation methods, *in* C. Plott and V. Smith (eds), *Handbook of Experimental Economics Results*, Vol. 1, North Holland, Amsterdam, chapter 81, pp. 752–767.

Johnson, E. and Nemet, G. F. (2010). Willingness to pay for climate policy: A review of estimates, *La Follette School Working Paper No. 2010-011* . La Follette School of Public Affairs at the University of Wisconsin-Madison.

Kaczan, D., MacDonald, D. H., Morrison, M. and Hatfield-Dodds, S. (n.d.). Willingness to pay to reduce the risk of severe climate change: Evidence from Australia, *Working Paper* . Reviewed in Johnson and Nemet (2010).

Keeler, J. P., James, W. L. and Abdel-Ghany, M. (1985). The relative size of windfall income and the permanent income hypothesis, *Journal of Business & Economic Statistics* **3**(3): 209–215.

Konow, J. (2003). Which is the fairest one of all? A positive analysis of justice theories, *Journal of Economic Literature* **41**(4): 1188–1239.

Kotchen, M. J. (2009). Voluntary provision of public goods for bads: A theory of environmental offsets, *Economic Journal* **119**(537): 883–899.

Kriström, B. (1997). Spike models in contingent valuation, *American Journal of Agricultural Economics* **79**(3): 1013–1023.

Lee, J. (2008). The effect of the background risk in a simple chance improving decision model, *Journal of Risk and Uncertainty* **36**(1): 19–41.

Lee, J.-S., Yoo, S.-H. and Kwak, S.-J. (2010). Public's willingness to pay for preventing climate change, *Applied Economics Letters* **17**(4-6): 619–622.

List, J. A. (2004). Young, selfish and male: Field evidence of social preferences, *Economic Journal* **114**(492): 121–149.

Lorenzoni, I., Nicholson-Cole, S. and Whitmarsh, L. (2007). Barriers perceived to engaging with climate change among the UK public and their policy implications, *Global Environmental Change* **17**(3-4): 445–459.

Lorenzoni, I. and Pidgeon, N. (2006). Public views on climate change: European and USA perspectives, *Climatic Change* **77**(1): 73–95.

Löschel, A., Sturm, B. and Vogt, C. (2013). The demand for climate protection—empirical evidence from Germany, *Economics Letters* **118**(3): 415–418.

Lusk, J. L. and Hudson, D. (2004). Willingness-to-pay estimates and their relevance to agribusiness decision making, *Review of Agricultural Economics* **26**(2): 152–169.

MacKerron, G. J., Egerton, C., Gaskell, C., Parpia, A. and Mourato, S. (2009). Willingness to pay for carbon offset certification and co-benefits among (high) flying young adults in the UK, *Energy Policy* **37**(4): 1372–1381.

McFadden, D. (1994). Contingent valuation and social choice, *American Journal of Agricultural Economics* **76**(4): 689–708.

Munro, A. and Hanley, N. (1999). Information, uncertainty, and contingent valuation, *in* I. Bateman, K. G. Willis and K. J. Arrow (eds), *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*, Oxford University Press, Oxford, chapter 9, pp. 258–279.

Nordhaus, W. D. (1993). Reflections on the economics of climate change, *The Journal of Economic Perspectives* **7**(4): 11–25.

Pachauri, R. (2007). Conference of the Parties to the UNFCCC serving as the meeting of the Parties to the Kyoto Protocol (COP/MOP), opening ceremony 12 December 2007 – WMO/UNEP Intergovernmental Panel on Climate Change. Video presentation, available at: http://www.un.org/webcast/unfccc/2007/index.asp [Accessed September 23, 2011].

Shogren, J. F. (2006). Experimental methods and valuation, *in* K. G. Mäler and J. R. Vincent (eds), *Valuing Environmental Changes*, Vol. 2 of *Handbook of Environmental Economics*, Elsevier, pp. 969–1027.

Sinn, H.-W. (2008). Public policies against global warming: A supply side approach, *International Tax and Public Finance* **15**(4): 360–394.

Solomon, B. D. and Johnson, N. H. (2009). Valuing climate protection through willingness to pay for biomass ethanol, *Ecological Economics* **68**(7): 2137–2144.

Starmer, C. and Sugden, R. (1991). Does the random-lottery incentive system elicit true preferences? An experimental investigation, *American Economic Review* **81**(4): 971–978.

Sterman, J. and Sweeney, L. (2007). Understanding public complacency about climate change: Adults' mental models of climate change violate conservation of matter, *Climatic Change* **80**(3): 213–238.

Thaler, R. H. and Johnson, E. J. (1990). Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice, *Management Science* **36**(6): 643–660.

Tol, R. S. J. (1999). The marginal costs of greenhouse gas emissions, *Energy Journal* **20**(1): 61–81.

Tol, R. S. J. (2009). The economic effects of climate change, *Journal of Economic Perspectives* **23**(2): 29–51.

Tol, R. S. J. (2010). The economic impact of climate change, *Perspektiven der Wirtschaftspolitik* **11**: 13–37.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data, *Journal of the Royal Statistical Society. Series B (Methodological)* **38**(3): 290–295.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice, *Science* **211**(4481): 453–458.

Ungar, S. (2000). Knowledge, ignorance and the popular culture: Climate change versus the ozone hole, *Public Understanding of Science* **9**(3): 297–312.

Viscusi, W. and Zeckhauser, R. (2006). The perception and valuation of the risks of climate change: A rational and behavioral blend, *Climatic Change* **77**(1): 151–177.

Werner, M. (1999). Allowing for zeros in dichotomous-choice contingent-valuation models, *Journal of Business & Economic Statistics* **17**(4): 479–486.

Motivational Drivers of the Private Provision of Public Goods: Evidence From a Large Framed Field Experiment

Johannes Diederich Timo Goeschl

Abstract

Disentangling the motivational drivers of individuals is frequently regarded a key step in reconciling theory and empirical evidence on the voluntary provision of public goods. We present results of a large online field experiments with 12,624 contribution choices by members of the Internet-using German population. Subjects are assigned to six treatments targeted at motivations such as altruism, "warm glow", image motivation, or equity concerns. While evidence on treatment effects is mixed, the data point to significant effects of framing and the sequence of presenting options. Exploiting variations within the highly heterogeneous sample, the results confirm previous results from a subset of the data on sociodemographics and exogenous environmental conditions as determinants of subjects' choices and add additional evidence that females and older subjects are more inclined to give to the public good.

1 Introduction

The question of the underlying drivers of the private provision of public goods has spawned a still ongoing discussion in the literature. Since both empirically and experimentally, consumers voluntarily contribute to public goods at higher levels than the standard theory predicts, several extensions of and alternatives to the traditional theoretical formulation of the problem have been discussed. Most extensions concern additional sources of utility over and above the "purely altruistic" setup, in which, besides consumption, the overall level of the public good is the only argument in the utility function, no matter how it is provided (Samuelson 1954, Bergstrom et al. 1986).

Among the additional drivers discussed, "warm glow", image motivation, and moral norms have received a considerable amount of attention. Commonly, the *"warm glow"* of giving is defined as any private utility gains from the act of giving itself regardless of its impact on the aggregate level of the public good (Cornes and Sandler 1984, Andreoni 1989, 1990). Also, it is assumed to increase in the size of one's contribution. *Image motivation*, in turn, is characterized by (1) the visibility of the contribution or pro-social action and (2) by some (perceived) norm causing external (social) approval or sanctioning (Akerlof 1980, Hollander 1990, Benabou and Tirole 2006, Ariely et al. 2009). We may therefore also call image motivation a *social* norm. In contrast, a norm where sanctioning takes place internally, i.e. within the subject, may be called a *moral* norm (Brekke et al. 2003, Konow 2003). One may view the particular moral norm considered in the present paper as a norm of distributive fairness or a concern for equity.

This paper reports results from five additional treatment groups in the experiment reported in our articles "To Give or Not to Give: The Price of Contributing and the Provision of Public Goods" and "Willingness to Pay for Voluntary Climate Action and Its Determinants: Field-Experimental Evidence".[1] Treatments were designed to disentangle, in a uniform, controlled procedure, the presence and strength of the four motivational drivers mentioned above in the propensity to contribute a 1 ton carbon emissions reduction to the global pure public good of climate change mitigation. In total, the data set contains 12,624 contribution choices made by 6,312 subjects.

Main results regarding a successful disentanglement of the targeted motivations through the treatments are ambiguous. On the one hand, we find that when removing altruism from subjects' set of motivations, contribution probabilities do not differ, while when adding image motivation, there is weak evidence for an increase. Both findings are plausible and in line with theory. On the other hand, several findings point to difficulties of the experimental design to successfully operate: First, there is evidence for strong framing effects as two theoretically equivalent treatments delivered significantly different contribution probabilities from slight changes in the information given. Second, results from an ex-post control questions suggest that the treatments removing altruism from subjects' set of motivations have not been well understood and thus may have failed. Third, we find in a treatment combining the removal of altruism and the activation of image motivation that contribution probabilities increase relatively strongly which is difficult to reconcile with the result of the two single treatments mentioned above.

In addition to treatment effects, the set of available covariates provides an opportu-

[1]Both papers are part of this dissertation. The latter is also available under Diederich and Goeschl (2013).

nity to check previous results on determinants of contributions based on a subsample of the full sample. Results confirm previous findings on sociodemographics such as a highly significant correlation with education. Due to the larger statistical power, results add evidence for a positive correlation of age and being female with the probability to contribute to previously insignificant estimates. Lastly, additional evidence for a positive causal effect of outdoor temperatures, matched to subjects' choices by experimental day and region of residence, on the probability to contribute can be provided.

2 Theoretical framework

The following presents a simple linear model to illustrate the treatment design. The model is based on the standard impurely altruistic public good model (Andreoni 1989, 1990) and also draws from Benabou and Tirole (2006). Let utility be additively separable and expressed as

$$\begin{aligned} U_i =& v_i^x x_i + v_i^G G + v_i^g g_i \\ &+ \eta \left[\gamma_i^G E_{-i} \left(v_i^G | g_i, \mathbf{p} \right) - \gamma_i^g E_{-i} \left(v_i^g | g_i, \mathbf{p} \right) \right] \end{aligned} \tag{1}$$

where x_i denotes i's consumption of a monetary numéraire, g_i is i's contribution to the public good, and $G = G_{-i} + g_i$ with G_{-i} being contributions by others. The vector $\mathbf{v}_i = \left(v_i^x, v_i^G, v_i^g\right)$ denotes the individual's valuations for money, the overall level of public good, and own contributions, respectively. Thus, the second (third) term in (1) is the altruistic ("warm glow") component of utility. The last term represents image motivation where $\eta \in [0, 1]$ is a parameter of the visibility of the act of contributing to others, and the vector $\boldsymbol{\gamma}_i = (\gamma_i^G, \gamma_i^g)$ denotes individual i's concerns for being perceived by others as altruistic and for being perceived as *not* interested into "warm glow" (the latter corresponding to the "private reward" in Benabou and Tirole 2006). We assume that the way expectations on $\mathbf{v}_i$ are formed is common knowledge. Let the individual be endowed with wealth w_i and maximize utility with respect to a budget constraint $w_i = x_i + \mathbf{p}\mathbf{y}_i$ where $\mathbf{p}$ denotes an M-dimensional vector of prices representing M available technologies which convert the numéraire into a unit of public good. Lastly, $g_i = \sum_{j=1}^{M} y_{ij}$.[2]

The first-order derivative of (1) with respect to contribution technology j is

$$\frac{\partial U}{\partial y_{ij}} = -\,p_j v_i^x + v_i^G + v_i^g + \eta \left[\gamma_i^G \frac{\partial E_{-i}\left(v_i^G | \sum_{j=1}^{M} y_{ij}, \mathbf{p}\right)}{\partial y_{ij}} - \gamma_i^g \frac{\partial E_{-i}\left(v_i^g | \sum_{j=1}^{M} y_{ij}, \mathbf{p}\right)}{\partial y_{ij}} \right]. \qquad (2)$$

Obviously, the solution of the simple linear model will not be interior but fully depend on the relative weight of the additive motivational components. In our experimental design, we investigate whether a subject is willing to contribute *one* more unit to the public good. That is, whether utility is increasing in y_{ij} at the current level of contributions.[3]

Note that the image term in (2) comprises of two effects: the change in others' posterior expectation of i's altruism and the change in their posterior expectation of i's interest in personal reward from giving. Both effects should be expected to be positive from an increase of giving (Benabou and Tirole 2006). Hence the net effect of increasing visibility η is unclear ex ante.

Differentiating between the four different components of utility is straightforward formally. In order to isolate "warm glow" one needs to ensure anonymity, $\eta = 0$, and exogenously keep G invariant to g_i.[4] In this case, net marginal utility, $v_i^g - p_j v_i^x$, is either increasing or decreasing in y_i which determines i's experimental choice. This will

[2] The model would owe realism a potential but, for our purposes here, unnecessary extension. One property of the specific public good used in the experiment is that contributions will *offset* concurrent negative contributions since carbon emissions are a by-product of consumption in most economies (Kotchen 2009). Drawing from Kotchen's model, we could write *net contributions* as

$$g_i = \sum_{j=1}^{M} y_{ij} - \delta x_i, \ \delta \in [0, 1]$$

where we assume a linear externality of private consumption on the public good, δx_i. Thus, i's direct contributions now (partly) offset or exceed the harm done through consumption. Note that G_{-i} may now include some initial level of the public good provided by nature, and that the individual can boost G also by reducing consumption. As an additional extension, one could assume δ to be individual-specific and to represent the individual's *perception* of the impact of consumption. We follow up on this issue through questions in the post-experimental survey.

[3] The possibility of field price censoring has been discussed at length in the article "To Give or Not to Give: The Price of Contributing and the Provision of Public Goods" (this dissertation).

[4] Note that v_i^g will, as a residual category, capture every source of utility that only depends on the size of personal giving. This is "warm glow" by definition but may also relate to other motivations that one would rather categorize into more complex moral reasoning (Brekke et al. 2003).

be the condition of our experimental "Warm glow" treatment. If we let G vary with g_i, this will correspond to the "Baseline" condition where both the "warm glow" and the altruistic components are active. In the "Image" condition, $\eta > 0$. Comparison to the Baseline will reveal whether the net effect of the image term is positive or negative.

Note that the production technology is constant across treatment conditions so far. Different technologies can have different (perceived) by-products of producing the unit of public good, however. It is clear that differences in attributes of the otherwise identical good may affect utility of contributors (Hanley et al. 1998). We make use of this by offering, in one treatment, another contribution technology which differs from production via EUAs particularly with respect to equity effects due to the geographic region in which the contribution would be produced (in a developing country instead of within the EU). One attribute of the technology made explicit to subjects was that production in a developing country would generate positive side benefits to the local population and environment.[5] In the model, we therefore assume for simplicity that the production technologies of G differ only with respect to their impact on equity. Without explicit functional specification, we may add a technology-dependent concern for equity, Q, to utility. (1) becomes

$$U_i = U_i \left(v_i^x x_i + v_i^G G + v_i^g g_i + \eta R(\boldsymbol{\gamma}_i, g_i, \mathbf{p}); Q(\mathbf{y}) \right) \tag{3}$$

where $R(\cdot)$ denotes the image term in (1). Thus, Q plays only a role for treatment effects when the technology changes.

3 Treatment Design

Inspired by the model above, the treatment design of the experiment (incompletely) varies three factors: (1) whether the choice of the emissions reduction actually has an impact on total emissions, (2) whether the choice is to some extent visible to others, and (3) whether the choice has different distributional impacts (Table 1). This gave six treat-

[5] Following up on footnote 2, one explicit attribute of the baseline EUAs compared to the alternative technology was that domestic production would contribute to emissions reductions in the region where subjects' personal negative contributions have occurred.

Table 1: Partial three-factor design of treatments

	Factor 1: Impact on total contributions	
Factor 2 (Factor 3):	Has impact	Has no impact
Private contribution (domestic)	Base/EUA	WG
Visible contribution (domestic)	Image	WGI
Private contribution (developing country)	CER	–

Notes: Base: Baseline treatment. WG: Warm Glow treatment. WGI: Warm-Glow-Image treatment.

Table 2: Two-stages counterfactual design of the experiment

Stage	Experimental group								Total
	1	2	3	4	5	6	7	8	
1	Base	WG	Base	Image	Base	WGI	EUA	CER	
2	WG	Base	Image	Base	WGI	Base	CER	EUA	
# of subjects:	779	778	784	796	792	798	788	797	6,312

Notes: Base: Baseline treatment. WG: Warm Glow treatment. WGI: Warm-Glow-Image treatment.

ments in total. Treatments were administered to subjects in a two-stages counterfactual design in order to allow for both between-subjects and within-subjects comparison of behavior (Table 2).[6] One reason for adding a within-subject component in design was to exploit any coherence of treatment effects within subjects and to immunize treatment effects against any arbitrariness in valuation or constructed preferences that could manifest in between-subjects comparison (Ariely et al. 2003, Hanley et al. 2009).

At the beginning of the experiment[7] subjects were informed that they would participate in two consecutive lotteries. Following the first prize choice and, if applicable, the FPC "filter" screen, subjects made a second choice based on another version on the *decision screen* that administered a treatment condition by containing a different description of how the public good would be provided, as described below. In two treatments (EUA and CER), also a second *information screen* was shown that differed in

[6]Thus, the full sample of the Baseline treatment (excluding sceptics) analyzed in the two other papers on this experiment consists of the pooled first-stage choices of experimental groups 1, 3, and 5.

[7]The reader is referred to the two articles "To Give or Not to Give: The Price of Contributing and the Provision of Public Goods" (this dissertation) and "Willingness to Pay for Voluntary Climate Action and Its Determinants: Field-Experimental Evidence" (this dissertation and Diederich and Goeschl 2013) for a full account of the experimental procedure.

wording. If subjects opted for the cash prize in their second choice, a second FPC "filter" screen identical to that used in the first choice appeared. The following describes the particular design and wording on the *decision screen* and, partially, the *information screen* for each treatment condition.[8]

Baseline treatment. The Baseline treatment most closely corresponded to the situation known from laboratory public good experiments, but in a framed field experiment (Harrison and List 2004) with a real public good and non-student subjects. In particular, subjects' choices were completely private information and affected total contributions. Instructions on the *decision screen* described that winners would be notified via email and that the deletion of the EUA, if they chose this prize, would be verifiable on a Heidelberg University[9] web page via a web link embedded in the notification email.

Warm Glow treatment. Differentiating "warm glow" from altruism and other motives is not a straightforward task. Most of the experimental designs in the laboratory exogenously vary the individual opportunity cost of contributions and the marginal value of the public good to separate between altruism and "warm glow" (Andreoni 1993, Palfrey and Prisbrey 1997, Goeree et al. 2002, Eckel et al. 2005). However, the marginal benefit of a real public good is fixed. We therefore used a variant of a design by Crumpler and Grossman (2008) which mimics a complete crowding out of contributions.[10] In their experiment, subjects were informed that the charity they select would receive \$10 from the experimenter. Subjects were then endowed with \$10 and asked to indicate how much of their endowment they would like to pass to the charity. Instructions stated that "The amount contributed by the proctor to your selected charity WILL be reduced by however much you pass to your selected charity. *Your selected charity will receive neither more nor less than \$10.*" Having made sure that subjects understood the procedure, the authors find a stable average contribution rate of around 20% among 150 subjects. We

[8] See the appendix for screenshots of these two experimental screen for each treatment (in German).

[9] This mentioning of our home institution was the only hint to the identity of the experimenters in the experiment and intended to increase credibility of the deletion confirmations to subjects. Confirmation certificates of the deletion were official transaction protocols by the German Emissions Trading Authority (DEHSt).

[10] Hence, this design corresponds to holding G constant in the model.

adapted this design by stating on the *decision screen* of the Warm Glow treatment:

> "In this lottery, a certain number of emission allowances will definitely be bought and deleted. The emission allowance offered to you today is part of these allowances. This means that regardless of your choice, the number of allowances to be deleted will not change. However, you have the opportunity to personally contribute to this emission reduction. You can do so by foregoing the cash prize and selecting the emission reduction instead."

One limitation of the design by Crumpler and Grossman (2008) is that it may allow for an experimenter demand effect or a desire to give to the experimenter. The fact that the experimenters' identity was much less clear in our field setting mitigates this problem: First, there was no personal interaction. Second, it was much less clear whose financial burden would be reduced from contributing.

Image treatment. In order to boost visibility of a winners' contribution compared to the Baseline treatment, the decision screen described that the reduction-choosing winner would be personally contacted by a staff member via email to arrange the EUA deletion and to ask for the consent to publish the winners' name on a section of YouGov's website dedicated to this purpose. This procedure increased visibility of the subject's pro-social choice while cash choices remained private. The more personal interaction for deletion contrasts the anonymous aggregate deletion procedure announced in the Baseline treatment. In order to account for potentially higher demands for data privacy in an environment such as the Internet, subjects were informed that their names would be published only with first name, the first letter of the surname, and city.

When designing the Image treatment, we deliberately decided not to increase visibility by issuing personalized certificates (hard copy or electronic) that confirm the deletion (e.g Löschel et al. 2010). The reason is that such certification can generate additional private utility even if not shown to others. Thus, this option seemed unlikely to increase visibility without activating additional sources of utility. Instead, subjects in the Image treatment were informed, just like in the Baseline treatment, that the deletion would be verifiable on a Heidelberg University website. In implementing this, winners of the

Image treatment were assigned a single EUA number in the notification email and could verify that "their" EUA number fell within a range of EUA numbers for which a single official deletion confirmation was provided on a Heidelberg University webpage.

Another feature in treatment design of the Image treatment comes out of the theoretical model. Since the size of the net image effect in eq. (2) is likely to depend on $\mathbf{p}$, we "wash out" these second order effects by concealing the price a winning contributor has faced. Therefore, instructions noted that the alternative cash prize of a winner would not be made public and that other participants may face different trade-offs.

Warm Glow Image treatment (WGI). This treatment exactly combined the instructions of the Warm Glow and the Image treatments. Thus, we added publicity to the barebone "warm glow" motivation to contribute in the Warm Glow treatment. The text of the *decision screen* made clear which informations would be published, thus subjects could conclude that the information on whether their EUA was pre-bought or not would not be disclosed.

EUA and CER treatments. These treatments were designed to facilitate a change in abatement technologies. Thus, the CER treatment offered a Certified Emissions Reduction (CER) based on the Clean Development Mechanism (CDM) of the Kyoto Protocol instead of an EUA. The CER was of the "Gold Standard" quality.[11] Instructions explicitly mentioned the differences between both technologies along two dimensions: (1) region of abatement and (2) region of investment. We expected both to give rise to specific equity concerns or moral considerations. While the former can trigger a polluter-pays motivation in favor of domestic abatement through an EUA, the latter can trigger distributional concerns due to the side-benefits of Gold Standard CERs which require the carbon offset project to benefit the local community and local environment in a developing country.[12] Both motivations can be found in anecdotal evidence.[13] For the experimental implementation, presentation of the two technologies required not only

[11] http://www.cdmgoldstandard.org

[12] Thus, the difference between both technologies may be interpreted as a normative conflict between equality and equity (Konow 2003, Nikiforakis et al. 2012).

[13] For example, Carbon Retirement, a commercial UK based service for deleting EUAs, advertises with the feature of domestic abatement for moral reasons.

two different *decision screens* but also two different *information screens*. For the EUA treatment, both screens closely corresponded to the Baseline treatment, with minor differences (see the results below).

4 Results

Before turning to treatment effects, we compare (independent) behavior in the first lottery of subjects in the pooled Baseline groups 1, 3, and 5 with that of subjects in the EUA group 7. Since both treatments correspond to the same combination of factors (Table 1), behavior should not differ. The result is, however, that contribution choices of the EUA treatment significantly exceed those in the Baseline treatment ($p = 0.01$ for a two-sided Mann-Whitney U test and student's t test), by about 3.8%. Thus, minor changes in framing of the screens had a remarkably significant impact: The *information* and *decision screens* of both theoretically identical treatments differed only slightly at three places in the text. First, in presenting the two options in headlines on the *information screen*, the option of "reduction of carbon (CO_2) emissions by 1 ton" added "within the European Union" in the EUA treatment. Second, within the text below, it stated that this would reduce "domestic emissions in Germany and other EU countries, to which your personal energy use contributes" instead of only saying that this would reduce "emissions in Germany and other EU countries" as in the Baseline treatment. Third, the *decision screen* in the baseline noted that the deletion of EU allowances would take place through a "joint order" for all winners choosing this option. The first two differences were intended to help contrasting the EUA treatment from the CER treatment, the third difference was intended to help contrasting the Warm Glow and Image treatments from the Baseline, which was not necessary in the EUA treatment. Since all three differences occurred simultaneously, we cannot further differentiate between the possible causes of this framing effect. The following will therefore present results for the EUA and CER treatments separately from those of the other treatments.

In the analysis, choices in the first lottery, which are completely independent from each other, can be directly compared between-subjects. To exploit both choices each subject made, analysis through panel regressions will account for between- and within subject effects. Table 3 reports results of Probit regressions of the choice of the emissions reduction comparing the Baseline treatment with the treatments Warm Glow, Image, and WGI. Columns (1) to (3) are Probit regressions reporting between-subjects differences of choices in the first lottery. Columns (4) to (6) account for the panel structure of the two-lottery counterfactual design and report results of random-effects Probit regressions. Columns (4) to (6) thus also take into account within-subject treatment effects as well as a “time” trend. Some specifications control for subjects’ characteristics and for matched exogenous environmental conditions used in previous analyses. Table 4 restates descriptions of these variables and provides summary statistics for the full experimental sample.

Coefficient estimates of treatment effects are positive throughout. However, estimates are mostly insignificant with the exception of the WGI treatment (marginal effect up to 3.5%) and, for one specification only, the Image treatment (marginal effect 2.8%).

The insignificant difference between Baseline and Warm Glow treatments would imply that altruism was a negligible motivational component in subjects’ choices. However, evidence for a lack of understanding of the treatment design of the two “warm glow” treatments comes from a control question in which we asked subjects in treatment groups 1, 2, 5, and 6 for the number of lotteries (“both”, “one”, or “none of the lotteries”) in which a winner’s prize choice would influence the actual amount of emissions reductions. Overall, only 13.8% chose the correct answer (“one lottery”).

A positive effect of visibility of the contribution, implied marginally by the results, is in line with the theoretical expectations and indicates that subjects expectation of being perceived as altruistic when observed in the act of contributing by others (the first term in eq. (2)) seems to dominate.

In light of the results of the Warm Glow and Image treatments, the comparably large positive effect of the WGI treatment appears counterintuitive, however. This is even more so as the “warm glow” component of the treatment has apparently not been well understood as the control question mentioned above indicates. Again, framing effects could be the reason for this otherwise inconclusive effect: In exactly combining the

Table 3: Probit coefficient estimates comparing Baseline, Warm Glow, Image, and WGI treatments. Dependent variable: choice of emission reduction.

	Between-subjects			Between- and within-subjects		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment:						
Warm Glow	0.065	0.055	0.057	0.027	0.106	0.125
	(0.061)	(0.069)	(0.070)	(0.088)	(0.103)	(0.106)
Image	0.069	0.112*	0.101	0.074	0.051	0.081
	(0.060)	(0.067)	(0.068)	(0.087)	(0.102)	(0.105)
WGI	0.132**	0.124*	0.118*	0.157*	0.118	0.146
	(0.059)	(0.067)	(0.068)	(0.085)	(0.100)	(0.103)
Lottery 2	–	–	–	0.578***	0.445***	0.451***
				(0.055)	(0.065)	(0.067)
Cash prize	–	-0.004***	-0.004***	–	-0.020***	-0.021***
		(0.001)	(0.001)		(0.001)	(0.002)
Female	–	0.084*	0.081	–	0.311**	0.298**
		(0.050)	(0.051)		(0.127)	(0.130)
Age	–	0.003*	0.003*	–	0.004	0.005
		(0.002)	(0.002)		(0.004)	(0.005)
Education	–	0.051***	0.052***	–	0.139***	0.143***
		(0.007)	(0.008)		(0.021)	(0.021)
Income	–	-0.011	-0.013	–	-0.034	-0.038
		(0.015)	(0.015)		(0.038)	(0.039)
Ambient temperature	–	–	0.009	–	–	0.034**
			(0.006)			(0.017)
Media coverage	–	–	-0.001	–	–	0.001
			(0.001)			(0.002)
Constant	-0.985***	-1.595***	-1.661***	-4.190***	-5.179***	-5.974***
	(0.031)	(0.138)	(0.236)	(0.117)	(0.367)	(0.625)
N	4727	3866	3763	9454	7732	7526
Log-likelihood	-2179.660	-1737.092	-1691.138	-3866.061	-3053.855	-2963.011
χ^2	5.389	75.290	81.157	114.634	290.534	283.337
D.f.	3	8	10	4	9	11
Pseudo R^2	0.001	0.021	0.023			
AIC	4367.319	3492.184	3404.276	7744.122	6129.711	5952.022
BIC	4393.164	3548.524	3472.838	7787.047	6206.195	6042.062

Notes: Columns (1)-(3) are Probit coefficient estimates of the choice of emission reduction in the first lottery. Columns (4)-(6) are random-effects Probit estimates of choices in the full panel, reported as marginal effects at the sample means. Standard errors are shown in parentheses. Stars indicate significance levels (* 10%, ** 5%, *** 1%).

Table 4: Summary statistics of subjects' characteristics ($N = 6,312$)

Variable	Description	Mean	SD	Min	Max
Sociodemographic characteristics					
Female	1 if female	0.475	0.499	0	1
Age	Subject's age (years)	45.73	14.57	18	91
Years of education	Years based on subject's stated highest educational degree	12.26	3.222	9	22
Income	Midpoint[a] of subject's reported monthly household net income category (Euros)	2,508	1,681	450	8,000
Environmental controls					
Ambient temperature	Mean ambient outdoor temperature in subject's region of residence[b] (°C)	15.08	4.089	8.05	25.8
Media attention	Number of hits for a climate change related keyword search[c] in German print and online media[b]	135.7	29.34	69.5	160

Notes: [a] In our income approximation, for the 'less than €500' category, we assume €450. For the two categories above €5,000, we assume €8,000 for compatibility with German census data. The remaining categories have widths of €500. [b] The variable is the moving 2-day average of the daily values of the day at which the subject took the experiment and the day before [c] Keywords used: '*climate change*', '*climate protection*', '*global warming*', '*carbon dioxide*', 'CO_2'. Database: LexisNexis [d] Answer categories 1=disagree, 2=tend to disagree, 3=tend to agree, 4=agree [e] Median is 10 [f] Median is 50

wording of the two other treatments, the WGI treatment contained the largest amount of text of all treatments. Potentially, an apparently more detailed description, compared to the Baseline, may have increased subjects' trust in the procedure, or made the option more interesting from an hedonic point of view.

In contrast to the weak evidence on treatments, the results show a highly significant effect of repetition. Estimated at the margin, subjects are up to 57.8% more likely to choose the emissions reduction in the second lottery.

Estimates of other covariate effects confirm previous analyses of the Baseline subsample but also add additional evidence based on the increased statistical power of the full sample. Results for the effects of price, education, and income corroborate the estimates in our article "To Give or Not to Give: The Price of Contributing and the Provision of Public Goods" (this dissertation) in sign, size, and significance. In addition, the higher statistical power of the full sample reveals some evidence for positive correlations of being female and older with the choice of the contribution (marginal effect up to 31% for being female and up to 0.1% per year for age). This finding is in accordance with some previous works, while others have found insignificant effects.[14]

[14] See the original article for references.

In exploiting the statistical power of the full sample, there is also evidence in support of the effects of matched environmental controls analyzed in our article “Willingness to Pay for Voluntary Climate Action and Its Determinants: Field-Experimental Evidence” (this dissertation and Diederich and Goeschl 2013): Higher ambient outdoor temperatures in a subject’s region of residence around the time the subject took part in the experiment caused a higher probability of choosing the emissions reduction. The effect is significant in the panel regression only, however (marginal effect 3%).

Finally, Table 5 reports the same model specifications in comparing EUA and CER treatments. Estimates of the treatment effect are ambiguous with some weak within-subject evidence for a preferring the CER through an offset project in a developing country (marginal effect 12%). Other estimates are qualitatively similar but statistically weaker than in Table 3 due to smaller size of the subsample.

5 Conclusion

Following up on previous analyses, this paper presented additional results of treatments targeted at contribution motives known from the literature such as “warm glow”, image motivation, and equity concerns based on a framed field experiment on giving to a real and global public good: climate change mitigation. Results regarding the disentanglement of these motives are ambiguous: On the one hand we find an insignificant presence of altruism compared to a “warm glow” of giving and a slightly positive image motivation, which is in line with theoretical predictions.[15] On the other hand, these results do not align with our finding of a relatively strong positive effect when simply combining these two treatments, i.e., removing altruism and adding image motivation in a combined treatment. Also, the two altruism-removing “warm glow” treatments were apparently not well understood by subjects. Besides, there is evidence for framing effects such that slight changes in wording had significant impacts. Other than treatment effects, a strong result is that subjects are much more inclined to choose the emissions reduction if presented with a second chance. Regarding covariate effects, results of the

[15] While our own model does not predict signs, the model by Ribar and Wilhelm (2002), for example, predicts a dominance of “warm glow” and zero impact of altruism in very large groups.

Table 5: Probit coefficient estimates comparing EUA and CER treatment. Baseline: EUA treatment. Dependent variable: choice of emission reduction.

	Between-subjects			Between- and within-subjects		
	(1)	(2)	(3)	(4)	(5)	(6)
CER treatment	-0.022	-0.057	-0.050	0.124*	0.070	0.085
	(0.072)	(0.082)	(0.083)	(0.075)	(0.085)	(0.087)
Lottery 2	–	–	–	0.615***	0.563***	0.550***
				(0.080)	(0.091)	(0.092)
Cash prize	–	-0.001	-0.001	–	-0.012***	-0.011***
		(0.001)	(0.001)		(0.002)	(0.002)
Female	–	0.020	0.010	–	0.056	0.056
		(0.087)	(0.088)		(0.166)	(0.169)
Age	–	0.008**	0.007**	–	0.011*	0.010
		(0.003)	(0.003)		(0.006)	(0.006)
Education	–	0.045***	0.045***	–	0.092***	0.090***
		(0.013)	(0.013)		(0.026)	(0.026)
Income	–	0.031	0.032	–	0.077	0.074
		(0.025)	(0.026)		(0.050)	(0.051)
Ambient temperature	–	–	0.002	–	–	-0.018
			(0.011)			(0.022)
Media coverage	–	–	-0.000	–	–	-0.003
			(0.002)			(0.003)
Constant	-0.840***	-1.777***	-1.730***	-2.675***	-3.864***	-3.092***
	(0.051)	(0.245)	(0.416)	(0.190)	(0.518)	(0.821)
N	1585.000	1256.000	1230.000	3170.000	2512.000	2460.000
Log-likelihood	-787.516	-605.400	-596.333	-1492.148	-1171.425	-1146.591
χ^2	0.091	25.989	24.333	62.349	90.271	84.913
D.f.	1.000	6.000	8.000	2.000	7.000	9.000
Pseudo R^2	0.000	0.021	0.020			
AIC	1579.031	1224.800	1210.666	2992.296	2360.851	2315.182
BIC	1589.768	1260.749	1256.699	3016.542	2413.310	2379.069

Notes: Columns (1)-(3) are Probit coefficient estimates of the choice of emission reduction in the first lottery. Columns (4)-(6) are random-effects Probit estimates of choices in the full panel, reported as marginal effects at the sample means. Standard errors are shown in parentheses. Stars indicate significance levels (* 10%, ** 5%, *** 1%).

previous analyses are confirmed. In addition, the higher statistical power of the present sample allows us to find evidence for a positive effect of age and being female on the likelihood of contribution the emissions reduction.

References

Akerlof, G. A. (1980). A theory of social custom, of which unemployment may be one consequence, *Quarterly Journal of Economics* **94**(4): 749–775.

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and ricardian equivalence, *Journal of Political Economy* **97**(6): 1447–1458.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving, *Economic Journal* **100**(401): 464–477.

Andreoni, J. (1993). An experimental test of the public-goods crowding-out hypothesis, *The American Economic Review* **83**(5): 1317–1327.

Ariely, D., Bracha, A. and Meier, S. (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially, *American Economic Review* **99**(1): 544–555.

Ariely, D., Loewenstein, G. and Prelec, D. (2003). "Coherent arbitrariness": Stable demand curves without stable preferences, *The Quarterly Journal of Economics* **118**(1): 73–105.

Benabou, R. and Tirole, J. (2006). Incentives and prosocial behavior, *American Economic Review* **96**(5): 1652–1678.

Bergstrom, T., Blume, L. and Varian, H. (1986). On the private provision of public goods, *Journal of Public Economics* **29**(1): 25–49.

Brekke, K. A., Kverndokk, S. and Nyborg, K. (2003). An economic model of moral motivation, *Journal of Public Economics* **87**(9-10): 1967–1983.

Cornes, R. and Sandler, T. (1984). Easy riders, joint production, and public goods, *Economic Journal* **94**(375): 580–598.

Crumpler, H. and Grossman, P. J. (2008). An experimental test of warm glow giving, *Journal of Public Economics* **92**(5-6): 1011–1021.

Diederich, J. and Goeschl, T. (2013). Willingness to pay for voluntary climate action and its determinants: Field-experimental evidence, *Environmental and Resource Economics* . DOI 10.1007/s10640-013-9686-3.

Eckel, C. C., Grossman, P. J. and Johnston, R. M. (2005). An experimental test of the crowding out hypothesis, *Journal of Public Economics* **89**(8): 1543–1560.

Goeree, J. K., Holt, C. A. and Laury, S. K. (2002). Private costs and public benefits: unraveling the effects of altruism and noisy behavior, *Journal of Public Economics* **83**(2): 255–276.

Hanley, N., Kriström, B. and Shogren, J. F. (2009). Coherent arbitrariness: On value uncertainty for environmental goods, *Land Economics* **85**(1): 41–50.

Hanley, N., Wright, R. and Adamowicz, V. (1998). Using choice experiments to value the environment, *Environmental and Resource Economics* **11**(3): 413–428.

Harrison, G. W. and List, J. A. (2004). Field experiments, *Journal of Economic Literature* **42**(4): 1009–1055.

Hollander, H. (1990). A social exchange approach to voluntary cooperation, *American Economic Review* **80**(5): 1157–1167.

Konow, J. (2003). Which is the fairest one of all? A positive analysis of justice theories, *Journal of Economic Literature* **41**(4): 1188–1239.

Kotchen, M. J. (2009). Voluntary provision of public goods for bads: A theory of environmental offsets, *Economic Journal* **119**(537): 883–899.

Löschel, A., Sturm, B. and Vogt, C. (2010). The demand for climate protection—an empirical assessment for Germany, *ZEW Discussion Papers* (10-068). Centre for European Economic Research (ZEW), Mannheim.

Nikiforakis, N., Noussair, C. N. and Wilkening, T. (2012). Normative conflict and feuds: The limits of self-enforcement, *Journal of Public Economics* **96**(910): 797–807.

Palfrey, T. R. and Prisbrey, J. E. (1997). Anomalous behavior in public goods experiments: How much and why?, *The American Economic Review* **87**(5): 829–846.

Ribar, D. C. and Wilhelm, M. O. (2002). Altruistic and joy-of-giving motivations in charitable behavior, *Journal of Political Economy* **110**(2): 425–457.

Samuelson, P. A. (1954). The pure theory of public expenditure, *The Review of Economics and Statistics* **36**(4): 387–389.

Appendix

Information and *decision screens*

This section contains screenshots of the actual screens used in the experiment (in German). Translations of the screens belonging to the Baseline treatment can be found in the appendix to the article "To Give or Not to Give: The Price of Contributing and the Provision of Public Goods" (this dissertation). Translations of the other treatments can be mostly found in the text above.

Figure 1: *Information screen* of the Baseline, Warm Glow, Image, and WGI treatments

Figure 2: *Decision screen*, Baseline treatment

Figure 3: *Decision screen*, Warm Glow treatment

YouGov What the world thinks

20%

In dieser Verlosung haben Sie die Wahl zwischen den beiden unten stehenden Gewinnmöglichkeiten:

- Falls Sie den Geldbetrag wählen und gewinnen, erhalten Sie in den nächsten Tagen die entsprechenden Punkte automatisch auf Ihrem Punktekonto gutgeschrieben. Alle Gewinner erhalten dazu eine Benachrichtigungs-E-Mail.

- Falls Sie die Senkung der CO2-Emissionen wählen und gewinnen, werden Sie von einem unserer Mitarbeiter in den nächsten Tagen persönlich per E-Mail kontaktiert. Der Mitarbeiter wird die Löschung der von Ihnen gewonnenen Berechtigung für Sie durchführen, die Sie dann über einen Weblink auf den Internetseiten der Universität Heidelberg zuverlässig nachvollziehen können. Außerdem würden wir Ihren Klimaschutzbeitrag gern als Dankeschön auf der Gewinnerseite des Panelportals veröffentlichen. Der Mitarbeiter wird Sie dazu nach Ihrem Einverständnis fragen, das selbstverständlich freiwillig ist. Alle übrigen Angaben bleiben natürlich streng vertraulich.

Bitte wählen Sie nun aus, welchen Preis Sie in dieser Verlosung möchten, falls Sie als Gewinner gezogen werden:

○ 86 Euro in Form von Bonuspunkten

○ Die Senkung der CO2-Emissionen um 1 Tonne durch Löschen einer EU Emissionsberechtigung

Figure 4: *Decision screen*, Image treatment

YouGov What the world thinks

20%

In dieser Verlosung haben Sie die Wahl zwischen den beiden unten stehenden Gewinnmöglichkeiten.

- Falls Sie den Geldbetrag wählen und gewinnen, erhalten Sie in den nächsten Tagen die entsprechenden Punkte automatisch auf Ihrem Punktekonto gutgeschrieben. Alle Gewinner erhalten dazu eine kurze Benachrichtigungs-E-Mail.

- Falls Sie die Senkung der CO2-Emissionen wählen und gewinnen, werden Sie von einem unserer Mitarbeiter in den nächsten Tagen persönlich per E-Mail kontaktiert. Der Mitarbeiter wird die Löschung der von Ihnen gewonnenen Berechtigung für Sie durchführen, die Sie dann über einen Weblink auf den Internetseiten der Universität Heidelberg zuverlässig nachvollziehen können. Außerdem würden wir Ihren Klimaschutzbeitrag gern als Dankeschön auf der Gewinnerseite des Panelportals veröffentlichen. Der Mitarbeiter wird Sie dazu nach Ihrem Einverständnis fragen, das selbstverständlich freiwillig ist. Alle übrigen Angaben bleiben natürlich streng vertraulich.

Bitte beachten Sie noch: Für diese Verlosung wird eine bestimmte Anzahl an Emissionsberechtigungen auf jeden Fall gekauft und gelöscht. Dazu zählt auch die Berechtigung, die wir Ihnen anbieten. Das heißt, egal, welche Wahl Sie treffen: Die Anzahl der gelöschten Berechtigungen ändert sich nicht mehr. Sie haben jedoch die Möglichkeit, sich an dieser Senkung von Emissionen persönlich zu beteiligen. Dies können Sie tun, indem Sie auf den Geldbetrag als Gewinn verzichten und stattdessen die Senkung der CO2-Emissionen wählen.

Bitte wählen Sie nun aus, welchen Preis Sie in dieser Verlosung möchten, falls Sie als Gewinner gezogen werden:

○ Die Senkung der CO2-Emissionen um 1 Tonne durch Löschen einer EU Emissionsberechtigung

○ 64 Euro in Form von Bonuspunkten

Figure 5: *Decision screen*, WGI treatment

Figure 6: *Information screen* of the EUA treatment

Figure 7: *Decision screen*, EUA treatment


YouGov What the world thinks

18%

Im Folgenden informieren wir Sie über eine der beiden Verlosungen. Sie dürfen in dieser Verlosung zwischen zwei verschiedenen Gewinnen wählen. Diese sind in dieser Verlosung:

100 Euro in Punkten

oder

Die Senkung der Kohlendioxid (CO2) - Emissionen um 1 Tonne durch ein Klimaschutzprojekt in einem Entwicklungsland.

Wie funktioniert die Senkung der Emissionen in dieser Verlosung? Wir verwenden dazu eine zuverlässige Möglichkeit. Wir unterstützen für Sie ein zertifiziertes lokales *Klimaschutzprojekt in einem Entwicklungsland*, z.B. für Biogas, Windkraft oder Energieeffizienz. Dadurch wird in dem Entwicklungsland 1 Tonne CO2 vermieden. Dabei wählen wir ausschließlich Projekte des höchstmöglichen Gütesiegels, dem CDM Gold Standard. Solche Projekte müssen so durchgeführt werden, dass sie der Bevölkerung vor Ort zu Gute kommen (z.B. durch Beschäftigung von Arbeitskräften vor Ort) und besonders umweltschonend durchgeführt werden.

Durch die Art und Weise, wie sich CO2 in der Luft verteilt, macht es für die Wirkung auf das Klima keinen Unterschied, wo auf der Welt CO2-Emissionen gesenkt werden. Es zählen nur die Gesamtemissionen weltweit.

Bei den Verlosungen werden insgesamt 100 Gewinner aus etwa 5.000 Teilnehmern zufällig ausgewählt.

Figure 8: *Information screen* of the CER treatment


YouGov What the world thinks

20%

Sollten Sie zu den Gewinnern gehören, werden Sie in den nächsten Tagen per E-Mail benachrichtigt.

- Falls Sie den Geldbetrag wählen, werden die entsprechenden Punkte dann auf Ihrem Punktekonto gutgeschrieben.
- Falls Sie die Senkung der CO2-Emissionen wählen, erhalten Sie mit der Benachrichtigungs-E-Mail einen Weblink, über den Sie den Beitrag zum Klimaschutzprojekt auf den Internetseiten der Universität Heidelberg zuverlässig nachvollziehen können.

Bitte wählen Sie nun aus, welchen Preis Sie in dieser Verlosung möchten, falls Sie als Gewinner gezogen werden:

○ Die Senkung der CO2-Emissionen um 1 Tonne durch ein Klimaschutzprojekt in einem Entwicklungsland

○ 88 Euro in Form von Bonuspunkten

Figure 9: *Decision screen*, CER treatment

The Pure Group Size Effect in Linear Public Good Experiments with Large Groups*

Johannes Diederich Timo Goeschl Israel Waichman

Abstract

Many pure public goods are privately provided among large groups. Existing evidence about the behavior of large groups in linear public good experiments is sparse and debatable while theory provides ambiguous predictions and empirical results are mixed. We compare individual contribution levels for experimental group sizes of 10, 40, and 100 in a between-subjects design and find a positive "pure" effect of group size that is statistically more pronounced than previously established by Isaac, Walker, and Williams (1994). Differences in group sizes also manifest in subjects' first-order beliefs, in crowding-out of contributions by others' giving, and in subjects' tendency to stick to their first-round decision. Our design of the long-term Internet experiment provides an improved methodology to overcome typical administrative and budgetary challenges in running large-group, multi-period experiments. Our experiment also benefits from a diverse sample of the German population revealing gender, education, and the urbanity of the place of residence as main correlates to the private provision of the experimental public good.

1 Introduction

Many public goods are privately provided on a global scale with literally millions of consumers and potential suppliers. Examples include web resources such as Wikipedia, charities such as the Red Cross, and environmental public goods such as climate change mitigation. Most experimental research on the private provision of public goods has focused on comparably small groups, playing a public good game in a laboratory. Typically, these groups consist of two to ten subjects of a convenience sample of students. The immediate question about the external validity of this approach is threefold: First, to what extent do results obtained for students hold for humans in general? Second, to what extent do results obtained in a small group setting hold for large groups? And

*We thank seminar participants in Heidelberg and Mannheim for helpful comments, and Brigitte Galiger and Hans Baumann for student assistance. We also thank the people at Lightspeed Research for cooperation. Funding by the German Science Foundation (DFG) under grant GO1604/1 is gratefully acknowledged.

third, to what extent do results obtained from laboratory public goods hold for (the framing of) real-world public goods?

While a multitude of papers has been addressing the first and the third question, the literature is surprisingly silent about the second one. Among the thousands of published replications of the standard linear public good experiment (henceforth called linear PGE or LPGE), we are aware of only one study that explicitly compares behavior of small and large[1] groups (Isaac, Walker, and Williams 1994, henceforth also IWW). The most likely reasons for this lack of evidence are methodological. In general, researchers face two challenges in running a PGE with large groups (Isaac et al. 1994, Ledyard 1995). One is to facilitate simultaneous play when group sizes exceed typical lab sizes. The other one is funding since for parameter values of typical small-group PGEs, maximum attainable as well as expected average payoffs rapidly increase with group size.[2] IWW overcame the first challenge by what they termed "multiple session" experiments in which each round lasts several days so that subjects can sequentially access the lab and submit their decision. IWW overcame the second challenge by paying, instead of money, extra-credit points improving students' grades which, on the downside, made final payment almost dichotomous. The result IWW found in their experiments was that for certain values of experimental parameters, larger groups provide more of the public good while for other values, they could not identify a significant effect of the groups size.[3] Besides this experimental finding, theoretical and empirical work on the effect of numbers provide ambiguous findings regarding contribution levels.

In this paper, we report on a LPGE played with groups of the unexceptional size of 10 as well as of the large sizes of 40 and 100 subjects. Our results clearly support the pres-

[1]In an experimental context, we use the term "large" as referring to group sizes exceeding $N = 20$, which roughly corresponds to the capacity of a typical experimental laboratory. Notably, there have been several contributions to the literature that compare different sizes of small groups with $N < 20$.

[2]For a standard LPGE, the increase is quadratic in N if we assume perfectly identical subjects.

[3]A tendency to find the result counterintuitive at first can result from a confusion of terminology. A "pure" effect of numbers (IWW) implies that the public good in question is also *pure*. To the experimentalist, this requires the marginal returns from a contribution to the public good to be constant across experimental group sizes. Another effect of group size arises for non-excludable but rival goods, i.e. *common* goods. For common goods, an increase in group size causes a decrease in the marginal per-capita return (MPCR) of contributions since consumption of the good becomes more congested. Sometimes, the intuition of a negative effect of group size on contributions is based on this increase in the marginality of one's contribution (Olson 1965). However, this is not a "pure group size effect" as IWW call it. In the experimental literature, the significant effect of MPCR on contributions is a well-established result (Ledyard 1995).

ence of a positive effect of *pure* group size on individual contributions. At the same time, we offer an improved methodology for running experiments that require simultaneous interaction with large groups. In particular, we resolve several important shortcomings in IWW's design. First, administering the experiment online leads to an extremely low rate of "defaulting" subjects, i.e. subjects who fail to participate in individual rounds and thus, trigger an automatic individual decision. Second, using a regionally highly dispersed sample largely excludes uncontrolled communication between subjects during multi-day rounds. Third, paying monetary earnings to induce experimental preferences ensures continuous incentives. Apart from these improvements, we are the first to administer a LPGE with large groups to a heterogeneous sample of subjects from the general population, which also allows to exploit considerable variation in demographic characteristics.

In addition to establishing a positive effect of group size, we obtain results from subjects' stated first-order beliefs, from the analysis of individual behavior over time, and from correlating demographic information. Regarding beliefs, we find evidence that the observed positive group size effect on contributions is also present in first-order beliefs about others' contributions. However, the predicted share of contributors among fellow group members is inversely related to group size, an effect which we do not observe in actual behavior. In our regression results, observed previous-round contributions by others crowd-in individual contributions for small groups. However, the crowding-in is insignificant for groups with $N = 40$ and becomes a significant crowding-out for large groups of $N = 100$ as well as for small groups of $N = 10$ with high endowments. Overall, observed cooperative behavior is always higher than the believed behavior of others. Regarding individual behavior over time, we find that the share of subjects who display a "sticky" contribution pattern is larger in larger groups. Exploiting sociodemographic data of our non-student subject sample, we find significantly higher contributions for male subjects, for subjects holding a higher educational degree, and for subjects living in a more rural residential environment. These effects do not significantly differ with group size.

The remainder of this paper proceeds as follows: Section 2 provides a review of the relevant theoretical, empirical, and experimental literature. Section 3 describes the experimental design and the collected data. Section 4 reports on the analysis and results. Section 5 discusses the main finding with respect to explanatory hypotheses and future research.

2 Literature

The literature on a "pure" effect of group size on public good provision is very limited and provides ambiguous predictions and evidence. The widespread premise among economists that contributions will decrease with group size traces back to Olson (1965) who wrote "the larger a group is the farther it will fall short of providing an optimal supply of any collective good".[4]

2.1 Theoretical literature

Economic theory provides ambiguous predictions for the direction of an effect of group size when the public good is pure and linear. One needs to be careful, however, to differentiate between those types of models where predictions carry over to the environment of a standard linear public good experiment and those types of models where they do not. Consider the (neo-)classical "purely altruistic" model, in which individuals care, besides consumption, about the aggregate level of public good in the economy only. For an experimental environment with a linear monetary public good (e.g. an experimental "group account"), the unanimous prediction for this type of preferences is to contribute zero if MPCR < 1. This prediction is independent of group size and due to free riding behavior. The predictive power of the classical model is therefore limited in a linear public good experiment with varying group size. If an interior solution was possible, however, like for many real-world public goods, the purely altruistic model always predicts a negative effect of group size on individual contributions, both for identical individuals (Chamberlin 1974, McGuire 1974) and for individuals who differ in wealth or preferences (Andreoni 1988), while total giving will rise. The intuition for

[4] As pointed out by many scholars (see, e.g., Esteban 2001) and noted in footnote 3, Olson's proposition implied a non-excludable but rival public good.

this result fully parallels the intuition for the famous crowding-out result in the purely altruistic model (Warr 1983, Bernheim 1986, Bergstrom et al. 1986) that has inspired so much behavioral research: If one cares only about the level of public good provided in the economy, adding more contributors will partially crowd out one's own contribution. Thus, more potential contributors increases free riding, which thus is the driving force of the negative effect of group size in the purely altruistic model.

The stark predictions of the purely altruistic model would be mitigated somewhat if one adds "warm glow" (Andreoni 1989, 1990). Ribar and Wilhelm (2002) show that under a reasonable set of assumptions, such "joy-of-giving motives" will crowd out altruism as the motive for giving in an economy of infinite size. Hence, contributions would be invariant to additional contributors in the limit. In a linear public good experiment, joy-of-giving motives may be important to explain the typically observed above-Nash behavior. However, since the altruistic component of motivation predicts zero contributions for all group sizes in a linear experimental environment with MPCR < 1, composition of the altruistic and the "warm glow" component will not change in the number of recipients. Thus, also an "impure" (Andreoni 1989, 1990) public good model will not predict a pure group size effect for a linear PGE.

An important step towards modeling group size effects is the model of "congested altruism" by Andreoni (2007). In his model, individuals explicitly care about the benefits to others generated by the contribution to the public good.[5] However, utility is assumed not to increase proportionally to the increase of total social benefits from larger numbers but is discounted as the size of the economy grows. Thus, the individuals derive utility from a combination of caring about *total* surplus and caring about *average* surplus generated by the contribution. As a result, the overall predicted effect of group size is ambiguous: On the one hand, larger groups imply that the marginal contribution will create more social benefits, providing an incentive to increase contributions. On the other hand, larger groups imply that a smaller contribution is needed to achieve

[5] Note that there is potential for confusion due to the terminology used in the literature. Traditionally, the models in which the individual only cares about aggregate provision have been called "purely altruistic", as explained above, although these types of preferences are perfectly consistent with a purely selfish desire to consume the public good. In contrast, in Andreoni (2007)'s model, agents are explicitly interested in the benefits to others. Intuitively, Andreoni calls this "altruism" although it is different from the meaning of altruism in traditional models.

the same average surplus, providing an incentive to decrease contributions. Calibrating a utility function with experimental results, Andreoni postulates that the latter effect would dominate. In particular, average giving would decrease while total giving would increase. Note that the decreasing effect from group size is not due to increased strategic free riding, as in the purely altruistic model, but results from a decrease in the value of social benefits to the contributor. Thus, this type of preferences may very well carry over to an experimental linear public good environment.

2.2 Experimental literature

Experimental evidence on the pure effect of numbers in a linear public good environment is extremely limited for large groups. This is especially surprising as the topic has been of key interest to experimentalists—inspired by Olson's conjecture—, especially in the early days of public good experiments (Ledyard 1995). To the best of our knowledge, only IWW and, recently, Weimann et al. (2012), which we both briefly discuss below, have investigated the effect of group size in linear public good experiments with groups of $N > 20$. However, there are other papers reporting on large-group linear public good experiments who do not compare group sizes. A meta analysis on linear public good experiments (Zelmer 2003) lists McCorkle and Watts (1996), which is a simple one shot experiment consisting of a single large group, and we are also aware of Marwell and Ames (1981)[6] and Rondeau et al. (2005).[7] A larger set of papers tested for the effect of group size in linear public good experiments with "small" ($N < 20$) groups. The seminal paper is Isaac and Walker (1988) who find no significant effect of numbers comparing groups of 4 and 10 subjects. In her meta-study, Zelmer (2003) finds a weakly positive effect of group size across small-group studies. Other experimental public good papers reporting on large groups focused on the provision point mechanism not linear environments, especially among the early literature (Chamberlin 1978, Marwell and Ames 1979, 1981, Brewer and Kramer 1986).[8] Sweeney (1973) tested Olson's group size conjecture through

[6] Among other treatments, Marwell and Ames (1981) consider a linear environment with a verly low MPCR of 0.0275.

[7] In Rondeau et al. (2005), subjects know that they are part of a large group but not the exact size of 32.

[8] In many of the early one-shot experiments, deception was used to make subject believe that they are part of a large group (Marwell and Ames 1979, 1981, Brewer and Kramer 1986).

Table 1: Number of experimental groups in Isaac, Walker, and Williams (1994)

	Group size			
	$N = 4$	$N = 10$	$N = 40$	$N = 100$
MPCR = 0.30	17	16	6	3
MPCR = 0.75	10	10	6	3
MPCR = 0.03	–	–	6	–

another design than actually varying group size and is thus not comparable.

2.2.1 Isaac, Walker, and Williams (1994)

IWW ran a series of long-term, or as they called it, a "multiple session" experiments with non-monetary incentives to systematically investigate the effect of group size for linear pure public goods in large groups. Table 1 lists the treatments they report. IWW compared groups of 4, 10, 40, and 100 subjects, using MPCR values of 0.3 and 0.75 and, partly, 0.03. Their design aimed at providing a feasible methodology to experimental economists to cope with the challenges of running large-groups, multi-round experiments that require simultaneous interaction. Each experimental round lasted for two or three days. During this time the student subjects could use on-campus computer facilities at their own convenience to enter their round decisions. This part of the design overcame the "physical constraint" of limited lab space. Another innovation with respect to this constraint was to complement the long-term experiments by "muti-site" experiments in which they connected two experimental labs to run a few regular short term lab experiments with groups of 40. To overcome the "financial constraint", subjects' earnings were converted into extra credit points to improve their final grades by up to three points on a 100-point scale. This procedure was thoroughly tested by comparing behavior between long-term extra-credit experiments and standard short-term cash experiments, mostly for group sizes of four and ten.

IWW's findings have surprised many economists who would have expected a negative effect of group size on contributions. IWW found significantly higher individual contributions for the large groups of 40 and 100 compared to the two small groups of four and ten. In particular, cooperation was sustained over time with only slight if any decline for the large groups while the small groups exhibited the usual declining pattern over time. This finding held for the MPCR of 0.3. For the MPCR of 0.75, IWW found no statistically significant difference in average contributions across the four group sizes. Instead, all four treatments showed only a slight if any decline compared to the first-round level.

Despite their careful design, IWW's method leaves room for several important improvements. First, chances are high that subjects' decisions were not independent from each other as they were all volunteers from undergrad microeconomic theory classes at (mostly) the same university. This is especially problematic for the large group treatments since in this case, chances are high that one's peers are in the same group (and subjects know that), and MPCR values of 0.3 or 0.75 require only three or two other members of the group to cooperate in order to outperform the free rider's earnings. Thus, communication between rounds could be an important alternative explanation to group size for increased cooperation in IWW's results. Second, IWW report a significant amount of "default decisions". The design of a long-term experiment requires to specify a default rule in case a subject does not participate in a particular round. This is especially important in a public goods experiment where groups size is to be kept constant. The default rule IWW used was to allocate all endowment to a subject's private account, thus interpreting non-participation as a decision to free ride. It is obvious that this rule conditions a subject's costs of participating in a round on the decision whether to free ride or not and thus, may strongly bias incentives in the case of a long-term experiment where subjects repeatedly need to physically access some facility. In IWW's data, default rates range around 25-45% for the large groups of 100 and between 15% and 60% for the small groups of four. For the groups of 100, this means that about half of all free riding decisions were done through defaulting. Note that average contribution levels did mostly not significantly differ in the comparison to short-term, cash experiments, which points against a bias for the small groups. Third, translating monetary earnings into extra-credit points made payment almost dichotomous. According to IWW's description, a unique letter grade (e.g. B+) typically comprised of 3-4 points of the 100-point scale. Thus, at most one level change of the letter grade (e.g. from B+ to A-, or from

B- to B) could be achieved from earnings in the experiment.

2.2.2 Weimann, Brosig-Koch, Hennig-Schmidt, Keser, and Stahr (2012)

In a recent working paper, Weimann et al. (2012) report on playing the linear public good game with groups of 60 and 100 by connecting four experimental labs at German universities through the Internet. Live video surveillance of all four labs increases credibility of the procedure to subjects. While this "connected lab" approach improves upon IWW, who only connected two labs and did not provide video surveillance, Weimann et al. (2012) do not provide a solution to the financial constraint as they focus very low MPCR values (0.02 and 0.04). The contribution behavior they find compares well to IWW's treatment with groups of 40 and MPCR = 0.03 and resembles the usual declining pattern of standard small-group lab experiments. Comparing the groups of 60 and 100, they find weak evidence for a positive effect of group size.

2.3 Empirical literature

Empirical evidence on the effect of group size for pure public goods is sparse and ambiguous. Goetze et al. (1993) analyze data from 137 public broadcasting stations and use the number of (potential) receivers of the program as the group size variable. Controlling for income, they find smaller average contributions for larger audiences while total contributions increase. Haan and Kooreman (2002) analyze data from 166 small companies who were provided biweekly with a box of candy bars by a firm run by business students. The candy bars were a public good insofar as in each firm, the box containing the candy as well as the container collecting the payment were freely accessible and payment was on a voluntary basis. Using the number of sales as proxy for the size of the company, the authors find tentative evidence that average contributions to the provision of the box increase with group size. The third empirical paper we are aware of is Zhang and Zhu (2011) who exploit a natural experiment with the Chinese Wikipedia website: In 2005, access to the website was blocked in mainland China which reduced the size of the group of potential beneficiaries and contributors significantly. The authors find that this significantly reduced individual contributions. Note that all three real-world public goods analyzed in the empirical studies can be interpreted as both pure and linear public goods.

3 Experimental design and procedures

3.1 Design

The experiment on which we report is a standard linear public good game using the voluntary contribution mechanism (VCM).[9] Thus, individual payoffs are calculated as

$$\pi_i = y - g_i + a \left(g_i + \sum_{j \neq i}^{N} g_j \right) \tag{1}$$

where y denotes subjects' identical endowment, g_i denotes subject i's contribution to the public good, and a is the marginal per capita return (MPCR): the return to each member of the group, including the donor, from a marginal contribution. Thus, if $1/N < a < 1$ then the environment of the linear public good experiment mimics a social dilemma in which free riding ($g_i = 0$) is the dominant strategy to the rational player but full contribution ($g_i = y$) would be socially optimal.

In order to overcome the physical and financial constraints and to improve upon IWW's design after 20 years, we introduced three features in running a large-groups public good experiment. One was to administer the experiment online. In particular, we ran a long-term Internet experiment in which each experimental round lasted for exactly three days. This has several advantages. First, handling large groups and communicating with many subjects via the Internet can be easier than in-person. Second, participation costs to subjects are extremely small, both in general and for each round as subjects simply click a link in an email, log in, and enter their round decision instead of physically come to a lab. We expected low participation costs to encourage low default rates. Further support for going online comes from recent evidence that long-term online experiments produce similar results to short-term lab experiments in social dilemma situations (Normann et al. Forth.). The second novel feature was to recruit a more

[9] E.g. Isaac and Walker (1988), Fehr and Gächter (2000).

general sample of the German population than a convenience sample of students. Apart from the fact that to our knowledge no paper has investigated the behavior of "normals" (Wilson 2007) in large-group public good games, this has mainly three advantages: First, the regional dispersion of subjects would make interaction extremely unlikely. Second, it enhances external validity of the results. Third, higher diversity in sociodemographic characteristics would allow to link some of the results to these idiosyncatics. Subjects were recruited from the German panel of an international online polling company. This guaranteed that subjects were either used to participate in online surveys or at least able to do so. The third feature of the design was to return to monetary earnings for the subjects. To do so, we made use of the polling firm's infrastructure to pay their survey participants: The pollster compensates panel members with "points" where one point is worth €0.01.[10] Panel members can cash points by transferring their monetary equivalent to a PayPal account where they can use it for purchases or wire the balance to a regular bank account. Members can also use the points directly to acquire shopping vouchers (e.g. for amazon.de or Karstadt, one of Germany's major department store chains) or donate to a charity (e.g. UNICEF). There is a threshold for panel members to be able to use the points which is an account balance of 550 points. Using the pollster's infrastructure has a major advantage that needs to be emphasized: The usual payment for participating in online surveys varies considerably, depending on the polling company and the length of the survey.[11] From that, we expected the induced preferences to be sustainable over a wider range of prospective earnings (very low money for the small groups but relatively large money for the large groups) than it would be possible in a lab experiment with students. In particular, we endowed subjects with the equivalent of €0.40 each round with one randomly determined round being payoff-relevant. This endowment allowed to remain within a reasonable research budget despite the large groups. Similar to IWW's tests of extra-credit rewards, we test whether the induced incentives work by comparing, for the group size of ten, behavior with a treatment

[10]Members earn points through all kinds of activities with the pollsters's panel, predominantly answering surveys. Survey topics of these types of pollsters are very broad and range from politics and consumer preferences to lifestyle issues.

[11]The pollster at hand advised us with earnings of €5-10 per subject for the proposed experiment. Other players in the market provide much lower rates, e.g. €1 for 20min.

Table 2: Experimental treatments (between-subjects design)

	Treatment			
	10L	40L	100L	10H
Group size	$N = 10$	$N = 40$	$N = 100$	$N = 10$
Endowment	40 points	40 points	40 points	1,000 points
Number of groups	16	5	5	5

Notes: 1 point = €0.01. MPCR is 0.3 in all treatments.

where endowment is substantially higher (€10).

Parameter values in the experiment largely follow IWW. First, we chose a MPCR of 0.3, which is the value at which IWW identified a group size effect. Second, we chose group sizes of 10, 40, and 100 subjects but left out IWWs' smallest group of 4 since very small groups have been extensively studied in the literature already. Third, we adopted IWWs' default rule of allocating all endowment to the private account if a subject did not participate in a given round. Fourth, all subjects were assigned to one group size only and participated in the experiment once (between-subjects treatment design). We deviated from IWWs' parameters in that we chose five instead of three groups for the largest group size of 100 to increase statistical power. Furthermore, we decided to confine the online experiment to seven instead of ten rounds since we anticipated that an effect of group size could be established within the first seven rounds already, and we were not sure about the persistence of online panel members used to one-time surveys in a long-term interaction. Table 2 summarizes the treatments and number of independent observations in the experiment.

3.2 Procedures

The experiment took place in two sessions in July 2012 and in September 2012. The July session consisted of six groups of Treatment 10L and all five groups of Treatment 10H (see Table 2). The September session consisted of 10 groups of Treatment 10L and the treatments 40L and 100L.

In order to recruit subjects, our cooperation partner screened a random selection of panel members by asking for their consent to participate in what was presented to them as an "interactive survey" in which final payment would depend on own answers as well as the answers of other participants and which would run in seven rounds, each round lasting

2-3 days. With the agreement, subjects were required to provide a valid email address, phone number, and ZIP code ("PLZ" in German). Also, the screener survey asked for basic sociodemographic characteristics. About two weeks after screening, consenting panel members received an email inviting them to the first round of the "interactive survey" and providing them with personal login data and the link to the experimental website, *www.interaktive-umfrage.de* (translates "www.interactive-survey.de").

The experimental website consisted of only three pages[12], thus minimizing cognitive load and time demands. The first page was shown after log in and only in the first round of the game. This screen asked subjects to download and read the experimental instructions and emphasized that only by reading the instructions, subjects would understand how to influence their earnings. The screen provided a link to download the instructions in PDF format, consisting of two pages of written text (in German) using 12pt script size. Every subsequent screen in the experiment showed a link to review the instructions in the upper right corner. On the second screen, subjects saw the history of previous rounds in the upper part of the screen, were provided with a button to access the payoff calculator in the middle part of the screen, and were asked to enter their decision for the current round in the lower part of the screen. In addition, the screen reminded subjects of the size of the group they were part of. After entering a decision, the website presented the third screen, which asked subjects for their estimates of (1) how many of the other members of their group would contribute at least one point to the group account and (2) how many points on average these contributing members would allocate. These questions were not paid.[13] Having answered these two questions, subjects saw a dismissing screen. Subjects would also see the dismissing screen if they would log in a second time after having entered their answers but before the end of the round.

Experimental instructions[14] described the recurring question in each round of how to divide the endowment between two "alternatives": a "private account" and a "group

[12]Screenshots (in German) can be found in the appendix.

[13]There is evidence that unincentivized prediction questions do not bias behavior in a linear public good experiment (Gächter and Renner 2010). In addition, since the time between decisions would usually consist of three days, we were not too worried that the questions would influence decisions as answers to the prediction questions were not part of the history shown to subjects.

[14]Original instructions (in German) can be found in the appendix.

account". We decided to follow the literature which uses this framing since, on the one hand, this framing seems sufficiently general while, on the other hand, it retains some control about what a sample drawn from the general population will associate with the experimental task. The latter might not hold anymore with an even more abstract framing. Instructions verbally explained the payoff from the group account through both the per-capita effect from giving (i.e., via the MPCR) and by illustrating the per-capita effect with a multiplication of the group account balance by a certain factor and an equal split of the result.[15] In addition, instructions gave the formula to calculate personal earnings, expressed in simple words. Instructions were pre-tested using a small convenience sample of members of the general population.

The payoff calculator familiarized subjects with the decision space of the linear public good game. We felt this option to be preferable over control questions in order to keep the amount of text as small as possible in an online survey with members of the general population. Also, control questions may have the disadvantage of providing subjects with reference allocations. Instead of reading text, subjects could playfully enter possible allocations of themselves and the average co-member of their group and learn about the earnings.

The experimental protocol was designed to keep default rates as low as possible. During the morning of the first day of each round, subjects received an email that announced the new round and contained the personal login data and the link to the website. In the first round, this email announced the start of the experiment as described previously. Subjects who would not enter an allocation during the first (second) day received an automatic reminder email at 2:00 a.m. on the second (third) day sent by the experimental website. In addition, and to avoid misallocations of automatic reminder emails to spam folders, student assistants sent out personalized emails to non-participating subjects on the second day at around 5:00 p.m. At about the same time on the third day, student assistants began to call subjects who had still not participated by phone and politely reminded them of the approaching end of the round. If necessary, they were prepared to elicit an allocation on the phone, which in fact almost never happened. If a subject

[15]This corresponds to the notation of MPCR as k/N. The factor is $k = \{3; 12; 30\}$ for $N = \{10; 40; 100\}$, respectively (Isaac et al. 1994).

did not participate until 11:55 p.m. of the third day, the website assigned the default decision and started the next round.

In round *one*, the procedure on the third day was slightly different. Instead of phoning, student assistants sent out another personalized reminder email during the late afternoon. During the evening, invited panel members who had not completed the first round were then discarded and experimental groups were re-matched.[16]

After completion of all seven rounds, the random number fixing the payoff-relevant round was determined by the online polling company who had not seen the data at this point. After calculating and transmitting earnings, points were credited to the members' accounts.

3.3 Sample characteristics

Our cooperation partner provided us with 163 (1,239) screened panel members for the July (September) session, of which some suffered from obviously invalid email addresses or phone numbers. Subsequently, 150 (1,162) members were invited to the experiment. 42 (308) invited members in July (September) did not complete round one. In order to fit group sizes and treatments, two of the no-shows were assigned to groups despite their absence while the remaining were removed. In addition, 54 subjects in September were dismissed despite their completion of round one in order to fit group sizes and treatments. Dismissed subjects were paid a flat fee of 500 points. This left us with a total of 110 (800) experimental subjects assigned to treatments and groups as shown in Table 2. The average earning from the public good game was 772 points with a range of [34 points; 2,483 points] in July and 487 points with a range of [27 points; 1,825 points] in September. In addition, subjects received a fixed participation fee of 200 points.

Table 3 summarizes sociodemographic characteristics which show considerable heterogeneity. About half of the sample is female. Among the age categories, people above age 65 are somewhat underrepresented. With respect to education, more than half of the sample consists of workers or employees who completed a vocational training or professional school. About 30% of the sample have some form of higher education (college,

[16] Our partner oversampled panel members during screening, assuming a response rate of about 75% to the first-round invitations to the experiment.

Table 3: Demographic characteristics

Variable	Categories	Freq.	Percent
Female		453	49.78
Age	18-24	89	9.78
	25-34	146	16.04
	35-44	240	26.37
	45-54	221	24.29
	55-64	192	21.10
	65 and older	22	2.42
Education	Vocational training (Lehre/Berufsfachschule)	402	44.18
	Professional school / tertiary college (Fachschule)	61	6.70
	Master of crafts (Meister) / technician	51	5.60
	College degree (Fachhochschule)	112	12.31
	University degree	146	16.04
	Other degree or training	26	2.86
	No degree or training	38	4.18
	Student at college or university	51	5.60
	Apprentice or trainee	10	1.10
	Student in school	3	0.33
	Ph.D.	10	1.10
Region I	North Germany	281	30.88
	Central Germany	397	43.63
	South Germany	232	25.49
Region II	West Germany	668	73.41
	East Germany (former GDR)	178	19.56
	Berlin	64	7.03
Big city	1 if lives in city with population $> 500,000$	177	19.45
	0 else	733	80.55
Small town / rural	1 if ZIP code is unique to town	448	64.18
	0 if town has multiple ZIP codes	250	35.82
	Missing	212	-

university, Ph.D.). About 7% are students in various levels of their education. Subjects' regions of residence are highly dispersed across Germany (Figure 1). We exploit the regional data about federal states (Bundeslaender) and ZIP codes along several dimensions. The variable "region I" proxies for latitude. "Region II" corresponds to former East and West Germany. Since the state of Berlin was itself divided, we treat it separately. "Big city" is an indicator variable for all subjects whose ZIP code is part of 14 German cities with a population larger than 500,000.[17] In contrast, "small town" is an indicator variable for all subjects who live in a community that consists of only one ZIP code.

Table 4 gives a more detailed impression of the regional dispersion by summarizing

[17]The variable largely excludes subjects who live in suburban areas of these cities, however.

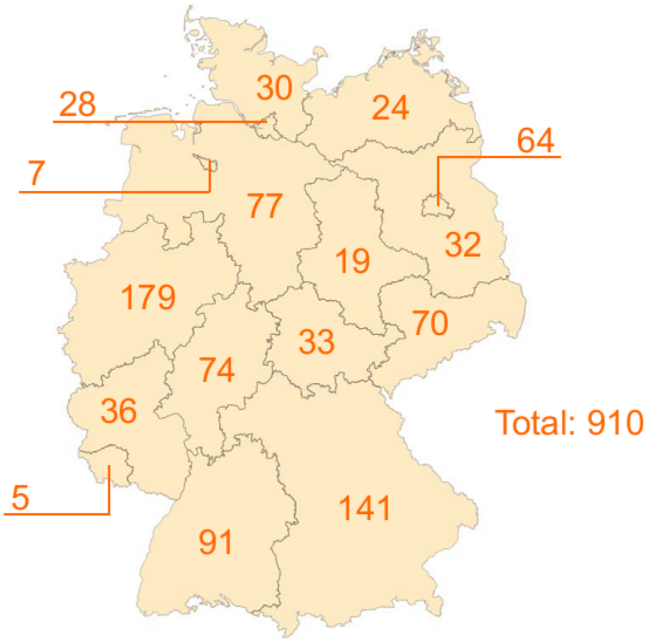


Figure 1: Distribution of subjects across German federal states (Bundeslaender)

Table 4: Distribution of subjects across ZIP codes

# of subjects per ZIP code	ZIP codes	
	Frequency	Percent
1	705	87.91
2	89	11.10
3	6	0.75
4	1	0.12
5	1	0.12
Total	802	100.00

the distribution of subjects across ZIP codes. In total, our subjects live in 802 different ZIP codes. In only eight of them, three or more subjects reside within the same code. Thus, personal interaction of subjects is extremely unlikely.

4 Results

4.1 Data quality

Figure 2 gives an overview about the data by depicting mean contributions of each independent observation, i.e. each experimental group. In treatments 10L and 10H, we observe the the typical declining pattern over time for some groups while for other groups, contributions stay about the same. One group in Treatment 10L exhibits exceptionally high contribution rates. In Treatment 40L, most groups exhibit a slightly declining path, while in Treatment 100L, contributions almost appear reversed U-shaped. Naturally, the variance of group means is much lower for large groups.

Figure 3 illustrates that our design succeeded in keeping default rates at low levels. The graph shows for each round and treatment the percentage of subjects who did not enter a round decision. Default rates manifest at 4.5% on average, slightly rise over time, and never exceed 10%. This is much lower than what IWW obtained. We conclude that a significant bias caused by the default rule is very unlikely in our data.

Result 1 *Default rates are very low, with 4.5% of subjects defaulting on average in each round.*

Figure 4 compares mean contributions for the two treatments with identical group size of 10 but endowments differing by the factor 25. Contribution schedules of both treatments are strikingly similar. Testing for differences does not establish significance for any round, using the two-sided Fligner-Policello robust rank order test (Fligner and Policello 1981, Feltovich 2005).[18] If we removed the outlier group of Treatment 10L, the 10L line would shift downwards and would almost coincide with the 10H line. These observations lead us to conclude that for our subject sample, the stakes employed in the 10L, 40L, and 100L treatments suffice in order to induce experimental preferences.

Result 2 *Mean contributions in the treatments with lower and higher endowment do not significantly differ.*

Compared to the literature, the typical declining pattern of contributions in groups of size $N = 10$ appears slightly less pronounced than usual. One reason can be that we only observe seven rounds. In experiments with ten rounds, it has been found that much of the decline takes place after round seven (e.g., Isaac and Walker 1988, Fehr and Gächter 2000, List 2004) which is usually attributed to the "end game effect". Another

[18] This result also holds for the standard Mann-Whitney-Wilcoxon test or the t-test. The robust rank order test is the more appropriate test for non-normal populations in our case since it does not assume equal variances.

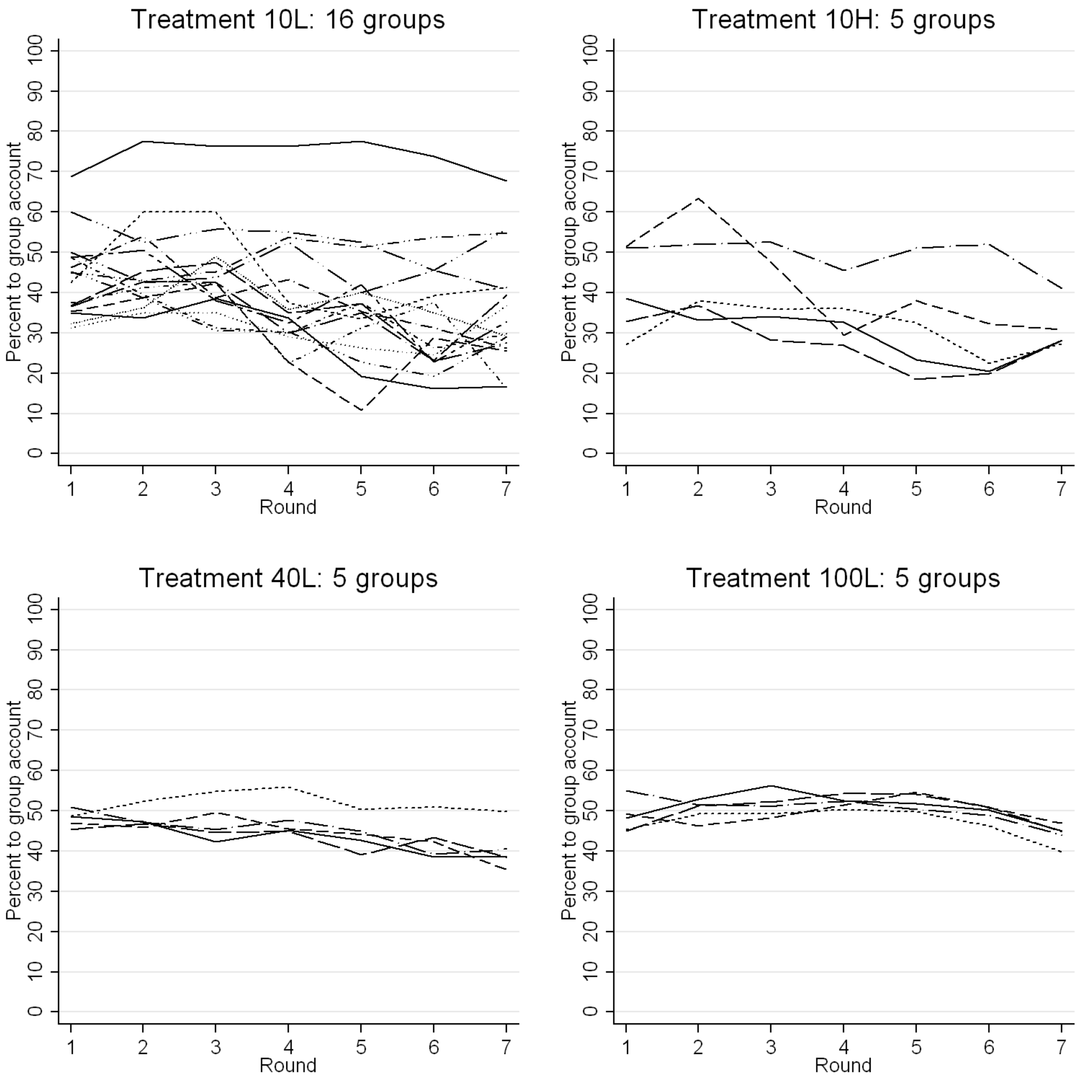


Figure 2: Mean contributions (in percent of endowment) of experimental groups by rounds and treatments

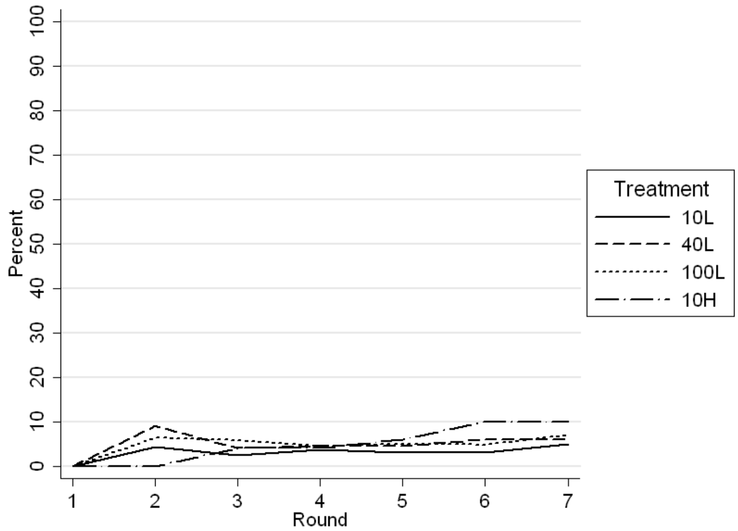


Figure 3: Default rates by rounds and treatments

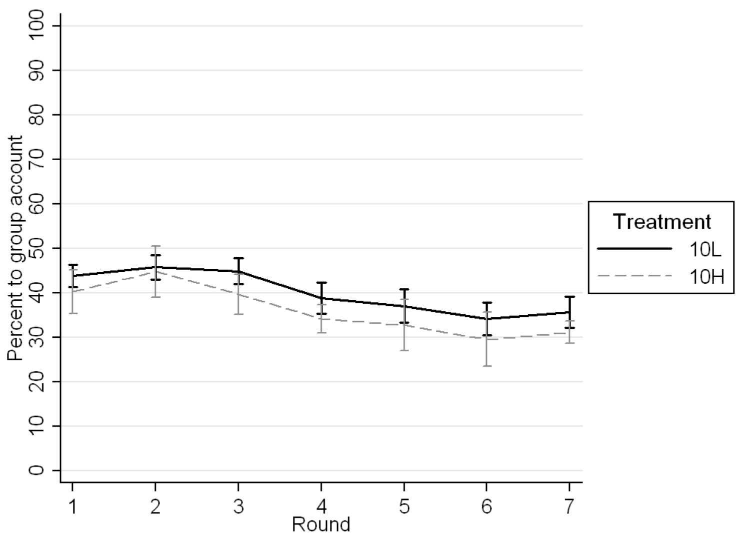


Figure 4: Mean contributions (in percent of endowment) of low and high stake groups with group size $N = 10$. Error bars denote standard errors.

reason might be our subject sample. Intuitively, non-students may be less able to cope with the abstract environment of an experimental game and therefore may adapt slower on average. It has been shown for one-shot public good games (Gächter et al. 2004) as well as for repeated public good games (Belot et al. 2010) that non-students contribute more than student subjects. However, we are not aware of existing results regarding differences in the speed of convergence to the Nash equilibrium. We return to this point in Section 4.3 on individual behavior. Regarding a potential bias introduced by slower adaptation of subjects or higher noise in behavior, the (between-subjects) treatment effects for group size should be unaffected or, if anything, biased downwards.

4.2 Group size effect

Figure 5 compares mean contributions in the three group size treatments. Clearly, contribution levels are strictly ordered for group size. Regarding dynamics, the groups of size $N = 10$ show an overall downward trend while groups of size $N = 40$ stay rather constant at first and begin to slightly decline after round four. In contrast, the group size of $N = 100$ shows a slight upward trend before declining after round five. Table 5 reports test results from a roundwise comparison of group means between treatments. The group size effect is significant in all rounds between sizes $N = 10$ and $N = 100$. Comparing group sizes $N = 10$ and $N = 40$ as well as $N = 40$ and $N = 100$, the effect is significant in most rounds, especially in rounds four to six.

Result 3 *There is strong evidence for a positive effect of group size on contributions to a pure public good in a standard linear PGE, using a MPCR of 0.3.*

4.3 Individual behavior

One meaningful way of summarizing individual behavior in large-group experiments is to compare histograms of the contributions across rounds (IWW). Figure 6 compares behavior in the first and in the final round, differentiated by group sizes. From the graph, we can observe several stylized facts. In round one, equal split of the endowment is the clear mode of the distribution for all treatments, with almost 40% choosing this

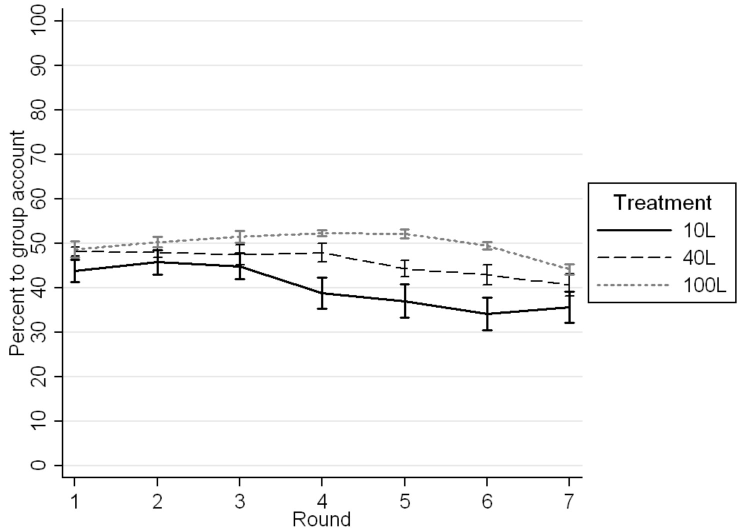


Figure 5: Mean contributions (in percent of endowment) for group sizes $N = 10$, $N = 40$, and $N = 100$. Error bars denote standard errors.

Table 5: Significance of group size effect

Round	Mean contribution in % of endowment			Two-sided Fligner-Policello robust rank order test		
	10L	40L	100L	100L vs. 10L	40L vs. 10L	100L vs.40L
1	43.73	48.09	48.58	Ù=-1.792 *	Ù=-1.772 *	Ù=0.275
	(2.56)	(0.92)	(1.81)			
2	45.70	47.90	50.22	Ù=-2.046 **	Ù=-1.615 *	Ù=-1.136
	(2.80)	(1.15)	(1.13)			
3	44.80	47.35	51.45	Ù=-3.040 **	Ù=-1.396	Ù=-1.448 *
	(2.88)	(2.21)	(1.39)			
4	38.73	47.83	52.22	Ù=-2.673 **	Ù=-2.638 **	Ù=-1.531 *
	(3.56)	(2.07)	(0.67)			
5	37.00	44.25	52.12	Ù=-4.472 ***	Ù=-2.469 **	Ù=-5.493 **
	(3.82)	(1.83)	(0.97)			
6	34.09	42.88	49.43	Ù=-4.121 ***	Ù=-2.202 **	Ù=-1.531 *
	(3.74)	(2.21)	(0.88)			
7	35.59	40.58	44.11	Ù=-2.647 **	Ù=-1.310	Ù=-1.322
	(3.53)	(2.45)	(1.18)			

Notes: Mean contribution is the average of group means. Standard errors in parentheses are clustered at group means. The Fligner-Policello nonparametrically tests the equality of the distributions of group means. Stars indicate significance levels (* 10%, ** 5%, *** 1%).

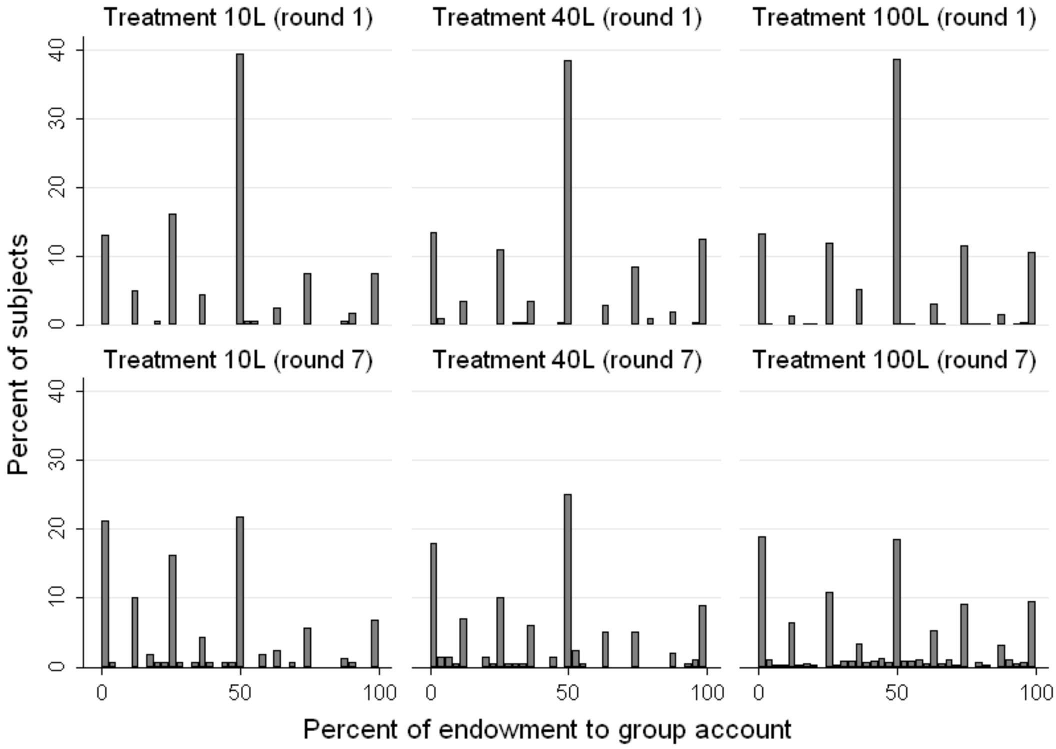


Figure 6: Individual behavior: Distribution of contributions (in percent of endowment) in rounds one and seven, by treatments

allocation. Zero contribution is chosen by almost 15% of subjects while full cooperation is chosen by up to about 10% of subjects. The remaining subjects show a preference for dividing endowment along intervals of 5 or 10 points (12.5% or 25% of endowment). In round seven, the fraction of equal split allocations has dropped considerably to about 20% but remains modal. The share of full free riders has slightly increased while the share of full cooperators has slightly decreased. Overall, many more combinations in between the 5-point and 10-point intervals are chosen, with some accumulation close to full cooperation and free riding. If we test for differences, the distributions of individual allocations in rounds one and seven differ significantly at the 10% level for all treatments. Between treatments, test results are largely in line with Table 5.

Another way to identify individual patterns within the data is to investigate the subject-specific variation in contributions. Panel A of Table 6 summarizes the distribution of the 15% quantile[19] of subjects who largely stick to a certain allocation over all seven rounds. We observe that the overall share of "sticky" subjects is larger for group sizes $N = 40$ and $N = 100$ at around 16% compared to 11% for group size $N = 10$. This can be due to the feedback (the results from previous rounds shown to subjects) being less volatile in larger groups. Conditional on the "sticky" subsample, about 13% of decisions in all three treatments are free riding, about 32% are choosing the equal split of endowment, and about 25% are full cooperation. If we compare this to the distribution of decisions in the full sample (panel B of Table 6), the most striking difference is a considerably higher frequency of full cooperation among "sticky" subjects. The same effect can be found to some extent for equal split but not for free riding.

Result 4 *Subjects tendency to stick to their initial allocation over all seven rounds is higher for larger groups.*

Result 5 *The share of full cooperation among "sticky" subjects is more than three times the overall share of full cooperation. Thus, cooperators are more likely to be "sticky" than other types of players. This result holds across group sizes.*

At the other end of the distribution of standard deviations, we observe a number of subjects with very high variation in contributions. Categorizing the individual contribution patterns of the top 7.5% quantile of standard deviation, we identify roughly 40% of them with a decreasing pattern over time but only one individual with an increasing pattern. 45% show a pattern that can be interpreted as "pulsing" behavior (IWW). About 12% display patterns of some other sort.

4.4 First-order beliefs

We analyze subjects' beliefs about others' behavior along two dimensions. First, we compare beliefs between treatments. Second, we compare beliefs with observed behavior.

In Figure 7, the three figures on the right-hand side (RHS) of each panel depict subjects' answers to the two belief elicitation questions: in Panel (a), it gives subjects'

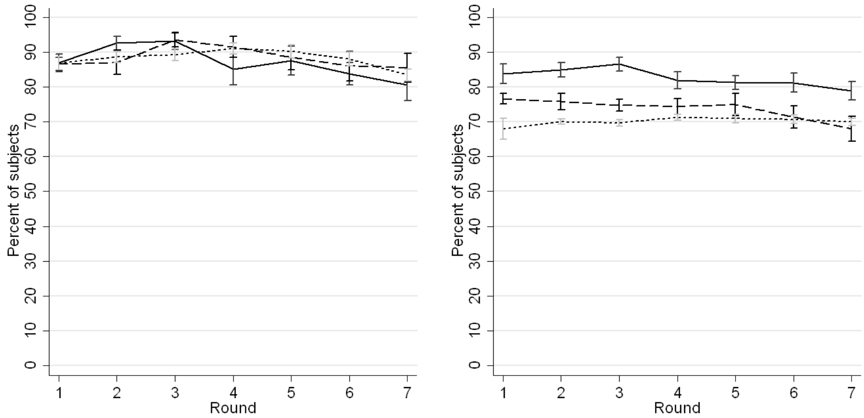
[19] The 15% quantile of subjects with the lowest variation in contributions seemed a plausible choice as it consists of 9.9% with zero variation and about 5.1% with SD< 5.15% of endowment.

Table 6: Distribution of individual decisions conditional on low variation in contributions and in the full sample

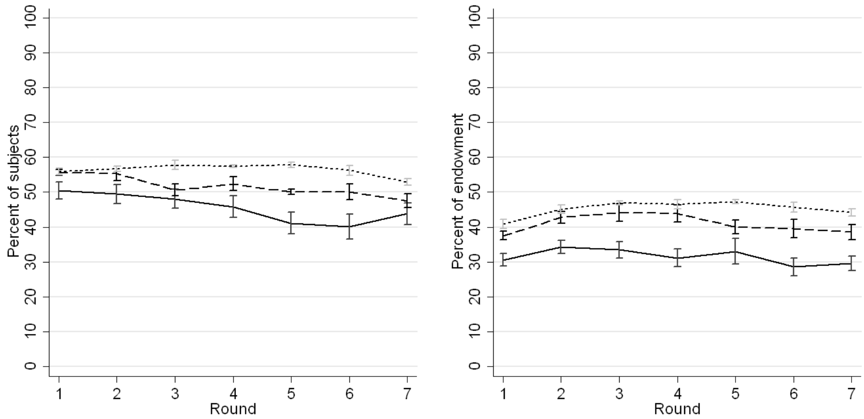
Contribution to group account	Percent of decisions			
	10L	40L	100L	Pooled
A. Decisions by subjects with low variation (n=952)[a]				
Free riding (0%)	11.9	8.3	14.3	12.6
Equal split (50%)	31.8	29.0	33.1	32.4
Full cooperation (100%)	33.3	25.8	24.8	26.6
Other allocations	23.0	36.9	27.8	28.4
Total conditional	100.0	100.0	100.0	100.0
Total unconditional	11.25	15.50	15.80	14.94
B. Full set of decisions (n=6,370)				
Free riding (0%)	13.0	11.6	11.8	12.0
Equal split (50%)	25.9	28.3	23.3	24.9
Full cooperation (100%)	6.8	8.5	9.5	8.8
Other allocations	54.3	51.6	55.4	54.3
Total all decisions	100.0	100.0	100.0	100.0

Notes: [a] The 15% quantile of subjects with the lowest variation in contributions consists of 9.9% with zero s.d. and about 5.1% with SD< 5.15% of endowment.

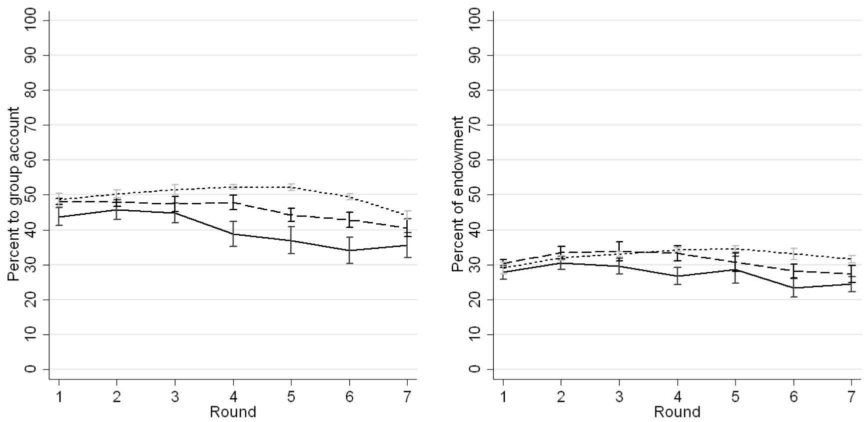
beliefs about the number of contributors in their group (question 1), in Panel (b), it shows the predicted average contributions of those who would contribute, and in Panel (c), it depicts predicted unconditional average contributions, obtained from multiplying the answers to questions 1 and 2. The predicted share of contributors is in fact decreasing with group size. The differences between Treatment 10L and the large group treatments are statistically significant in all but one round using the Fligner-Policello robust rank order test. This includes the first round. Thus, expectations already reflect differences in group sizes before subjects get any feedback from previous rounds. In contrast, we do not find significant differences in the share of free riders for actual behavior (LHS of Panel (a)). The RHS of Panel (b) shows that beliefs about conditional average contributions are increasing with group size. Differences between the large group treatments and the 10L treatment are highly significant for all rounds. The LHS of Panel (b) shows that we also observe the group size effect in the conditional sample. Regarding computed beliefs about unconditional average contributions, results are less pronounced. Comparing each of the large group treatments with the 10L treatment yields a significant effect only after round three.



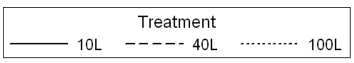
(a) Percentage of contributors in group (question 1)



(b) Percent of endowment allocated to group account, conditional on being a contributor (question 2)



(c) Percent of endowment allocated to group account, unconditional (question 1 x question 2)



(d) Legend

Figure 7: Observed behavior (left) vs. first-order beliefs (right)

Result 6 *There is evidence of group size effects in subjects' first-order beliefs. In particular, the predicted contributions of contributors increase with group size, which we also find in observed behavior. In contrast, the predicted share of contributors is inversely correlated with group size while there is no significant effect in actual behavior.*

The presence of the group size effects in first-order beliefs suggests that the observed effect is not merely some artifact of behavior since it appears robust to a mental process of forming expectations about others' behavior. The fact that we follow a between-subjects design reinforces this conclusion. Thus, people do not only give more in large groups in a linear public good experiment, they also state that they expect others to do so. Consequently, the group size effect may be consistent with conditional cooperative or reciprocal behavior. They also expect relatively fewer people to contribute in larger groups which we do not observe, however.

Two potential limitations apply to these observations. First, an endogeneity problem would arise if subjects did not process the belief elicitation questions correctly but used their own decision as a shortcut to give an answer. Evidence from the following result below points somewhat against this behavior. Second, due to our design we cannot control for the possibility that the conditional nature of the second question has not been fully understood by all subjects. In this case, answers by these subjects represent unconditional contributions so that the presence of an overall positive group size effect in the RHS of panel (c) in Figure 7 would be actually underestimated.

Comparing predicted and observed behavior in each panel of Figure 7, we find that actual behavior almost unanimously exceeds predictions in all rounds and for each treatment.[20] We also investigate the possibility that the conditionality of the second question has not been understood by subjects. In this case, one needs to interpret the answers to the second question as predictions of the unconditional group average and compare the RHS of Panel (b) with the LHS of Panel (c). This can be seen in Figure 8. Observed mean contributions still marginally exceed beliefs in all rounds, and differences

[20] Observed values are significantly higher than predicted values at mostly 5%, sometimes 1%, and sometimes 10% significance using the Wilcoxon matched-pairs signed-ranks test. An exception is Treatment 10L in Panel (a) in which the predicted share is not significantly different from the observed one in rounds 1, 4, 6, and 7.

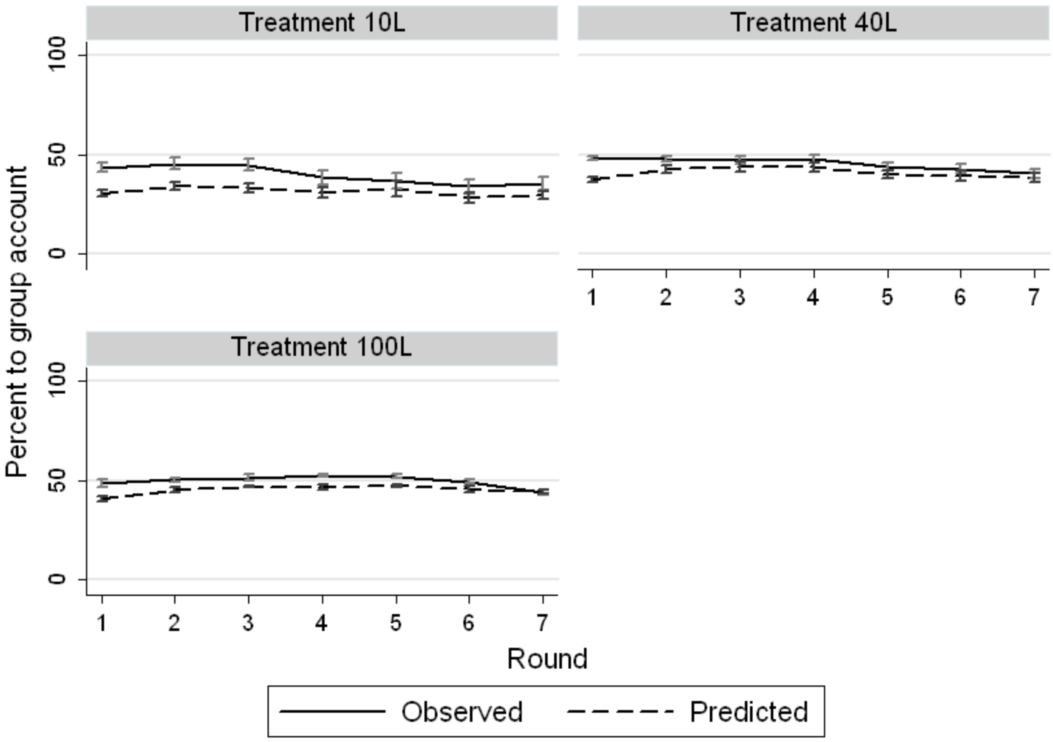


Figure 8: Observed average contributions vs. predicted conditional contributions, by treatment

are significant for most rounds.[21]

Result 7 *Subjects regularly underestimate the probability and extent of contributions by other group members for both small and large groups.*

This result reminds of the common observation that people think of their own behavior as above-average. In the literature on public good games, both deviations of beliefs compared to observed behavior have been found. For example, Thöni et al. (2012) find lower beliefs for a non-student sample and Gächter and Renner (2010) for a student sample. Others have found the opposite effect (Thöni et al. 2012, p.639).

[21]Significance levels are mostly 5% and sometimes at 1% or 10%. For rounds three to seven in Treatment 40L, round seven in Treatment 100L, and round five in Treatment 10H equality cannot be rejected.

4.5 Regression analysis

Regressions of individual contribution behavior provide further support for the pure group size effect as well as insights about the effect of covariates. For the basic model we use

$$g_i = \beta_0 + \beta_1 T_i + \beta_2 G_{-i,r-1} + \beta_3 B_i + \beta_4 G_{-i,r-1} T_i + \beta_5 B_i T_i + \beta_6 Z_i + \beta_8 R_i \qquad (2)$$

where g_i denotes subject i's contribution to the group account in a given round, and T_i denotes i's experimental treatment (10L, 40L, 100L, or 10H). The lagged variable $G_{-i,r-1}$ is part the history shown to subjects and denotes total contributions by other group members in the previous round. B_i is the vector of answers to the two belief elicitation questions. As before, to facilitate comparability across treatments, we measure $G_{-i,r-1}$ in percent of total group endowment, the first element of B_i in percent of group size, and the second element of B_i in percent of individual endowment. Finally, Z_i is a vector of sociodemographic characteristics, and R_i denotes round dummies.[22]

Table 7 reports coefficient estimates of OLS regressions for five different specifications. Specification (1) only includes the treatment dummies, besides dummy controls for rounds. Specification (2) consists of specification (1) plus the full set of sociodemographic characteristics. Specification (2) thus contains all available variables that are fully exogenous to the contribution choice. Specification (3) consists of specification (1) plus the experimental variables. Specification (4) reports the full set of variables. In addition, specification (5) contains interaction terms of the experimental variables with treatments.[23]

[22]Random effects GLS panel regressions with rounds as the time variable yield very similar results to the ones reported below. Apparently, controlling for rounds through dummies is more standard in the literature, however.

[23]We also tested interaction of the treatments with the demographic variables and round dummies. These results do not deliver additional meaningful significant results and are therefore not included in Table 7.

Table 7: OLS regressions. Dependent variable: Percentage of endowment contributed to group account

	(1)	(2)	(3)	(4)	(5)
A. Experimental variables					
Treatment:					
40L	5.6009*	7.3198*	2.9023	2.6086	15.3161
	(3.234)	(3.939)	(2.324)	(2.347)	(15.364)
100L	9.7812***	10.1732***	6.0805**	5.2327**	31.6540***
	(2.951)	(3.601)	(2.677)	(2.543)	(10.392)
10H	-3.9755	-5.2241	0.9328	1.4229	14.3801*
	(4.748)	(4.385)	(2.126)	(2.324)	(8.163)
Others' contributions (lagged)	–	–	0.1779	0.1788	0.3224***
			(0.129)	(0.112)	(0.092)
Belief % contributors	–	–	0.2134***	0.1844***	0.1787***
			(0.026)	(0.029)	(0.056)
Belief % contributions	–	–	0.4726***	0.4963***	0.4913***
			(0.053)	(0.045)	(0.092)
Others' contrib. (lagged) × ...					
40L	–	–	–	–	-0.3394
					(0.342)
100L	–	–	–	–	-0.5307***
					(0.167)
10H	–	–	–	–	-0.5483***
					(0.131)
Belief % contributors × ...					
40L	–	–	–	–	0.0090
					(0.057)
100L	–	–	–	–	0.0005
					(0.069)
10H	–	–	–	–	0.0705
					(0.120)
Belief % contributions × ...					
40L	–	–	–	–	0.0405
					(0.099)
100L	–	–	–	–	-0.0184
					(0.112)
10H	–	–	–	–	0.0928
					(0.146)
B. Demographic variables					
Female	–	-4.3781***	–	-3.7521*	-3.8123*
		(1.343)		(2.070)	(2.061)
Age:					
25-34	–	7.7380**	–	4.7482*	4.8432**
		(3.487)		(2.406)	(2.247)
35-44	–	3.3364	–	3.2801	3.3741
		(3.851)		(3.830)	(3.744)
45-54	–	4.4989	–	3.5503	3.5967
		(4.126)		(3.280)	(3.135)
55-64	–	4.0601	–	3.1711	3.1310
		(4.515)		(3.378)	(3.258)
65 and older	–	12.2621	–	7.4505	7.4544
		(8.841)		(6.127)	(5.940)
Education:					
Professional school	–	3.7312	–	2.4003	2.2137
		(3.716)		(4.373)	(4.299)
Master of crafts	–	3.1782	–	-0.3666	-0.3489
		(4.205)		(3.557)	(3.470)

Continued on next page

Table 7 – *Continued from previous page*

	(1)	(2)	(3)	(4)	(5)
College degree	–	4.7907*	–	0.9490	0.9303
		(2.752)		(2.133)	(2.132)
University degree	–	6.7378**	–	2.0347	1.7786
		(2.674)		(2.861)	(2.868)
Other degree or training	–	-7.0496**	–	-7.5202**	-7.6114**
		(3.441)		(2.898)	(2.842)
No degree or training	–	2.7568	–	4.4807	4.4834
		(3.272)		(2.799)	(2.714)
Student at college / univ.	–	3.4143	–	0.9818	0.9155
		(5.422)		(5.639)	(5.524)
Apprentice / trainee	–	2.2138	–	4.8963	5.0601
		(6.574)		(5.068)	(4.959)
Student in school	–	12.3849**	–	15.7820**	15.7142**
		(5.371)		(6.921)	(6.739)
Ph.D.	–	23.2421***	–	16.8054***	16.4940***
		(5.139)		(5.104)	(4.858)
Region I:					
Central Germany	–	0.7721	–	1.0214	0.8861
		(1.456)		(1.589)	(1.617)
South Germany	–	0.3665	–	0.9461	0.7865
		(1.560)		(1.453)	(1.541)
Region II:					
East Germany	–	-0.0605	–	-0.3903	-0.3687
		(1.916)		(1.491)	(1.498)
Berlin	–	0.1245	–	1.6105	1.4767
		(2.816)		(2.852)	(2.879)
Big city (>500,000)	–	1.1233	–	3.9818	3.6411
		(3.188)		(2.853)	(2.945)
Small town / rural	–	4.2146*	–	4.5555*	4.3211*
		(2.253)		(2.387)	(2.392)
Constant	42.2247***	32.9797***	3.8445	-2.1270	-7.6067
	(2.898)	(5.381)	(4.736)	(6.183)	(5.797)
N	6370	4886	5099	3907	3907
R^2	0.028	0.057	0.251	0.274	0.278
# of clusters	31	31	31	31	31

Notes: Reported are coefficients. Standard errors are clustered at the group level and shown in parentheses. Stars indicate significance levels (* 10%, ** 5%, *** 1%). The baseline is Treatment 10L, age 18-24, vocational training, North and West Germany, and round 2 (to allow for lagged variables). All model specifications also include controls for round indicating a highly significant negative time trend.

Clearly, the positive effect of group size on contributions manifests in each model specification. For $N = 40$ however, the effect is significant in specifications (1) and (2) only, when experimental variables are excluded. Differences in stake sizes, represented by the coefficient on the 10H treatment, are not significant, although specification (5) starts to pick up some effect that could be due to the selection of round two as the baseline (cp. Figure 4). Overall, these results provide additional support for our findings in Sections 4.1 and 4.2.

When included, experimental variables correlate with the contribution choices in mostly expected ways but also uncover unexpected results. First, the effect of the sum of contributions by other members of the group, as implicitly contained in the history of total contributions shown to subjects, is insignificant in specifications (3) and (4). However, differentiating for treatments in specification (5) reveals a significant positive effect in the 10L treatment and a negative effect in the 100L and 10H treatments. Thus, we observe a crowding-in of exogenous provision for small groups but a crowding-out for large groups and small groups with large stakes. Second, as one would expect, the two types of elicited beliefs strongly and positively correlate with the contribution. Thus, the evidence for average behavior to be consistent with conditional cooperation (Fischbacher et al. 2001) is mixed.[24]

Result 8 *Stated first-order beliefs about others are positively correlated with individual contributions. However, observed actual contributions by others crowd-out contributions in large groups whereas they crowd-in contributions in small groups.*

Result 9 *Evidence from regressions for an effect of stakes is mixed: While there is essentially no effect on contribution levels, differences can be found for crowding behavior, with crowding in for small-stake groups and crowding out for large-stake groups.*

Several sociodemographic characteristics correlate with contribution choices. First, females give less than males in our sample. This finding compares to varying findings in the literature where, e.g., Thöni et al. (2012) also find a negative effect among Danish subjects from the general population. In contrast, field studies using real-world public goods found positive correlations (e.g. List 2004). Second, the effect of age is largely insignificant: only the category consisting of subjects around the age of 30 contributes

[24]A possible regressor that may provide further insights but is omitted in our model is a lagged variable for subjects' own contributions in previous rounds. Including contributions of the previous round in columns (3) and (4) results in a highly significant coefficient while the R^2 increases to about 0.45. Significance levels of the other experimental variables do not change. However, coefficient estimates of all variables, including of demographic variables, decrease considerably since the variable necessarily controls for effects that have already been endogenously present in the round before. Thus, the interpretation of coefficients of those variables that are fixed across rounds (treatments and demographics) now gets a marginal component: a significantly positive coefficient can be interpreted as the effect being increasing compared to the previous round and vice versa. If one wanted to include lagged own contributions, ideally one would jointly estimate first round contributions without lagged variables and the remaining round including lagged variables.

significantly more than the baseline which consists of subjects around 20. Both findings are somewhat in contrast to previous studies which typically find, if any, a positive effect of age on altruism and cooperation (e.g., Carpenter et al. 2008, List 2004, Thöni et al. 2012). Third, level and type of education seems to be correlated with contributions: Subjects with certain higher educational degrees give more, so do students in school but not trainees or university students. Note that we cannot exclude the effect to partially reflect an effect of unobserved income. We have, however, previously found evidence for a positive effect of education in giving to a public good in Diederich and Goeschl (2011) and Diederich and Goeschl (2013) and in the literature, weakly, in List (2004). Fourth, subjects living in small towns or rural areas contribute significantly more while other regional variables are insignificant. Not part of the model of Table 7 are interaction effects of demographic variables with group size treatments. When included, we do not find further meaningful significant effects.

Result 10 *Individual contributions to the group account are higher for subjects who are male, of age around 30, have higher educational degrees, and reside in small towns or rural areas. Interaction terms do not suggest significant differences of these effects for different group sizes.*

5 Discussion and concluding remarks

Our results provide strong support for the presence of a positive effect of "pure" group size on contributions in linear public goods experiments when the marginal return from contributing is 0.3. We add to the existing literature in several respects. First, we provide an additional piece of evidence given ambiguous theoretical predictions, ambiguous empirical findings, and only one previous experimental study on large ($N > 20$) groups. Second, we improve on several methodological drawbacks of the existing experimental evidence. In particular, the long-term online experiment reported here benefits from extremely low default rates, an independent sample of subjects, and monetary earnings (instead of grade improvements). To the researcher, our design provides an improved framework for running simultaneous large-group experiments that helps in overcoming the "physical constraint" of standard lab sizes as well as "financial constraints" (IWW). Third, we are the first to test for the effect of group size using a sample from the general population, which allows to correlate a rich set of covariates with observed behavior.

By virtue of the experimental design, we can exclude many possible explanations for the positive effect of group size but not all. The standard linear public good game we use holds constant the incentives for strategic free riding, which is the driver of a group size effect in the classical "purely altruistic" model, across group sizes. Consequently, the observed group size effect must be due to a non-standard preference component.[25] Among those, the design of the basic linear public good game precludes many possible drivers known from the literature such as image motivation, punishment, or communication. Likewise, a potential "warm glow" component of utility, i.e. utility that is derived from the mere act of giving, will be invariant to changes in the group size by definition. Based on the available set of hypotheses known to us from the literature, we are thus left with three possible explanations for which we can unfortunately not differentiate further given our design.

First, our findings are consistent with the presence of other regarding preferences, particularly with Andreoni (2007)'s theory of "congested altruism". If this is the case then our results suggest that the group size-dependent discounting is low such that care about *aggregate* surplus dominates care about *average* surplus in preferences. Thus, contrary to Andreoni's findings, an increase in numbers has an overall positive effect in our data. In a sense, subjects would perceive an increase in group size as an improved opportunity of a "welfare-generating machine". Note that this explanation would also be consistent with the presence of a group size effect in first-order beliefs and with signaling behavior through "pulsing" contributions.

Second, our findings are consistent with the hypothesis that the public good environment, though linear by design, is not represented linearly in subjects' preferences. Instead, an argument of this type assumes that subjects care about some minimum level of earnings (e.g. of the size of initial endowment) more than about earnings above this level. Thus, loss aversion is at the heart of this hypothesis. An variant of the argument

[25]Thanks to Christoph Vanberg for pointing this out.

can be found in IWW who define a measure of how much the $N-1$ other group members need to contribute on average in order to restore the earnings of subject i, who makes a marginal contribution, to the level of free riding (i.e. the endowment). Obviously, this measure is decreasing in group size. Thus, chances of "hedging" the initial endowment are higher for larger groups. Another variant of this argument would state that in larger groups, it is more likely for a loss-averse subject to encounter the required number[26] of like-minded subjects who will also contribute and thus permit earnings above the endowment for everyone. Note that implicit in all versions of this argument is that it assumes the presence of something like a *perceived provision point* in subjects' preferences. Now, it is well known from threshold public good games that cooperation is more likely to succeed in environments with provision points and the public good is more likely to be provided compared to linear environments. Thus, an implicit or perceived provision point may cause a positive effect of group size since average contributions necessary to achieve the perceived threshold always decrease in group size. Again, this explanation would be consistent with the presence of a group size effect in first-order beliefs and with signaling behavior through "pulsing" contributions.

Third, it is also possible that the positive group size effect is due to intuitive, non-strategic behavior: It is not necessary for subjects who only fully understand the rules of the game to also grasp the strategic environment or to be motivated by others' well-being in order to be influenced by salient differences between group sizes. For example, a simple computation of maximum possible earnings from cooperation will produce a much larger number for $N = 100$ compared to $N = 10$ while both minimum earnings and the free rider's earnings are constant across group sizes.[27] Thus, a myopic cost benefit analysis may lead to higher contributions in large groups (IWW). We can also interpret such behavior in the context of dual process theories, originally developed in social psychology (Sloman 1996, Kahneman 2003). According to these theories decisions are made either through a fast system based on *intuition* (also called system 1) or a slow system based on *reasoning* (also called system 2). The hypothesis is then that

[26]Obviously, for a MPCR of 0.3 the required number is four.

[27]Minimum attainable payoff is when $m_i = z$ and $m_j = 0\ \forall j \neq i$, maximum attainable payoff is when $m_i = 0$ and $m_j = z\ \forall j \neq i$. Minimum payoff is always 12 points, the free rider's payoff is always 40 points, and maximum payoffs are 148 points (10L), 508 points (40L), and 1,228 points (100L).

subjects use, for whatever reason, intuition rather than reasoning when confronted with the decision environment of our experiment. Note that "pulsing" behavior which we partly observe in our data would not be consistent with this explanation if it is meant to signal cooperation. In contrast, the result of a group size effect in first-order beliefs is consistent with this explanation since subjects using intuition may simply extrapolate their own behavior to others. Future research is necessary to test these hypotheses and to further investigate the drivers of the positive effect of group size confirmed in our results.

References

Andreoni, J. (1988). Privately provided public goods in a large economy: The limits of altruism, *Journal of Public Economics* **35**(1): 57–73.

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and ricardian equivalence, *Journal of Political Economy* **97**(6): 1447–1458.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving, *Economic Journal* **100**(401): 464–477.

Andreoni, J. (2007). Giving gifts to groups: How altruism depends on the number of recipients, *Journal of Public Economics* **91**(9): 1731–1749.

Belot, M., Miller, L. and Duch, R. (2010). Who should be called to the lab? A comprehensive comparison of students and non-students in classic experimental games, *Nuffield Centre for Experimental Social Sciences Discussion Paper Series* **2010-001**.

Bergstrom, T., Blume, L. and Varian, H. (1986). On the private provision of public goods, *Journal of Public Economics* **29**(1): 25–49.

Bernheim, B. D. (1986). On the voluntary and involuntary provision of public goods, *The American Economic Review* **76**(4): 789–793.

Brewer, M. B. and Kramer, R. M. (1986). Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing, *Journal of Personality and Social Psychology* **50**(3): 543–549.

Carpenter, J., Connolly, C. and Myers, C. (2008). Altruistic behavior in a representative dictator experiment, *Experimental Economics* **11**(3): 282–298.

Chamberlin, J. (1974). Provision of collective goods as a function of group size, *The American Political Science Review* **68**(2): 707–716.

Chamberlin, J. R. (1978). The logic of collective action: Some experimental results, *Behavioral Science* **23**(5): 441–445.

Diederich, J. and Goeschl, T. (2011). Giving in a large economy: Price vs. non-price effects in a field experiment, *Discussion Paper No. 514* . Department of Economics, Heidelberg University.

Diederich, J. and Goeschl, T. (2013). Willingness to pay for voluntary climate action and its determinants: Field-experimental evidence, *Environmental and Resource Economics* . DOI 10.1007/s10640-013-9686-3.

Esteban, J. (2001). Collective action and the group size paradox, *American Political Science Review* **95**(3): 663–672.

Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments, *The American Economic Review* **90**(4): 980–994.

Feltovich, N. (2005). Critical values for the robust rank-order test, *Communications in Statistics - Simulation and Computation* **34**(3): 525–547.

Fischbacher, U., Gächter, S. and Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment, *Economics Letters* **71**(3): 397–404.

Fligner, M. A. and Policello, G. E. (1981). Robust rank procedures for the Behrens-Fisher problem, *Journal of the American Statistical Association* **76**(373): 162–168.

Gächter, S., Herrmann, B. and Thöni, C. (2004). Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence, *Journal of Economic Behavior & Organization* **55**(4): 505–531.

Gächter, S. and Renner, E. (2010). The effects of (incentivized) belief elicitation in public goods experiments, *Experimental Economics* **13**(3): 364–377.

Goetze, L., Glover, T. F. and Biswas, B. (1993). The effects of group size and income on contributions to the corporation for public broadcasting, *Public Choice* **77**(2): 407–414.

Haan, M. and Kooreman, P. (2002). Free riding and the provision of candy bars, *Journal of Public Economics* **83**(2): 277–291.

Isaac, R. M. and Walker, J. M. (1988). Group size effects in public goods provision: The voluntary contributions mechanism, *The Quarterly Journal of Economics* **103**(1): 179–199.

Isaac, R. M., Walker, J. M. and Williams, A. W. (1994). Group size and the voluntary provision of public goods: Experimental evidence utilizing large groups, *Journal of Public Economics* **54**(1): 1–36.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality, *American Psychologist* **58**(9): 697.

Ledyard, J. O. (1995). Public goods: A survey of experimental research, *in* J. H. Kagel and A. E. Roth (eds), *The handbook of experimental economics*, Princeton University Press, Princeton, pp. 111–194.

List, J. A. (2004). Young, selfish and male: Field evidence of social preferences, *Economic Journal* **114**(492): 121–149.

Marwell, G. and Ames, R. E. (1979). Experiments on the provision of public goods. I. resources, interest, group size, and the free-rider problem, *American Journal of Sociology* **84**(6): 1335–1360.

Marwell, G. and Ames, R. E. (1981). Economists free ride, does anyone else?: Experiments on the provision of public goods, IV, *Journal of Public Economics* **15**(3): 295–310.

McCorkle, S. and Watts, M. (1996). Free riding indexes for Ukrainian economics teachers, *The Journal of Economic Education* **27**(3): 233–237.

McGuire, M. (1974). Group size, group homogeneity, and the aggregate provision of a pure public good under Cournot behavior, *Public Choice* **18**(1): 107–126.

Normann, H. T., Requate, T. and Waichman, I. (Forth.). Do short-term laboratory experiments provide valid descriptions of long-term economic interactions? A study of cournot markets, *Experimental Economics* .

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*, Vol. 124, Harvard University Press.

Ribar, D. C. and Wilhelm, M. O. (2002). Altruistic and joy-of-giving motivations in charitable behavior, *Journal of Political Economy* **110**(2): 425–457.

Rondeau, D., Poe, G. L. and Schulze, W. D. (2005). VCM or PPM? A comparison of the performance of two voluntary public goods mechanisms, *Journal of Public Economics* **89**(8): 1581–1592.

Sloman, S. A. (1996). The empirical case for two systems of reasoning, *Psychological Bulletin* **119**(1): 3–22.

Sweeney, J. W. J. (1973). An experimental investigation of the free-rider problem, *Social Science Research* **2**(3): 277–292.

Thöni, C., Tyran, J.-R. and Wengström, E. (2012). Microfoundations of social capital, *Journal of Public Economics* **96**(78): 635–643.

Warr, P. G. (1983). The private provision of a public good is independent of the distribution of income, *Economics Letters* **13**(2-3): 207–211.

Weimann, J., Brosig-Koch, J., Hennig-Schmidt, H., Keser, C. and Stahr, C. (2012). Public-good experiments with large groups, *University of Magdeburg Working Paper* **No. 9/2012**.

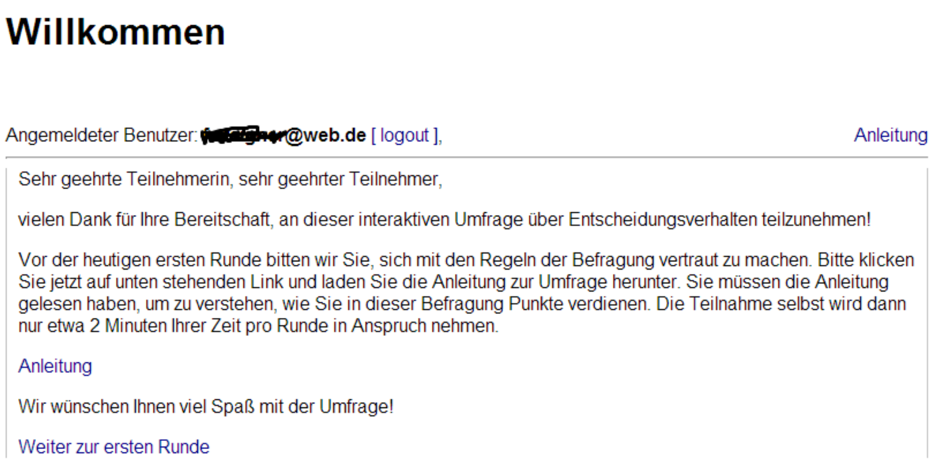
Wilson, R. K. (2007). Political economy and experiments, *The Political Economist* **XIV**(1): 1–7.

Zelmer, J. (2003). Linear public goods experiments: A meta-analysis, *Experimental Economics* **6**(3): 299–310.

Zhang, X. and Zhu, F. (2011). Group size and incentives to contribute: A natural experiment at Chinese Wikipedia, *American Economic Review* **101**(4): 1601–1615.

Appendix

A Experimental website



Willkommen

Angemeldeter Benutzer: [redacted]@web.de [logout], Anleitung

Sehr geehrte Teilnehmerin, sehr geehrter Teilnehmer,

vielen Dank für Ihre Bereitschaft, an dieser interaktiven Umfrage über Entscheidungsverhalten teilzunehmen!

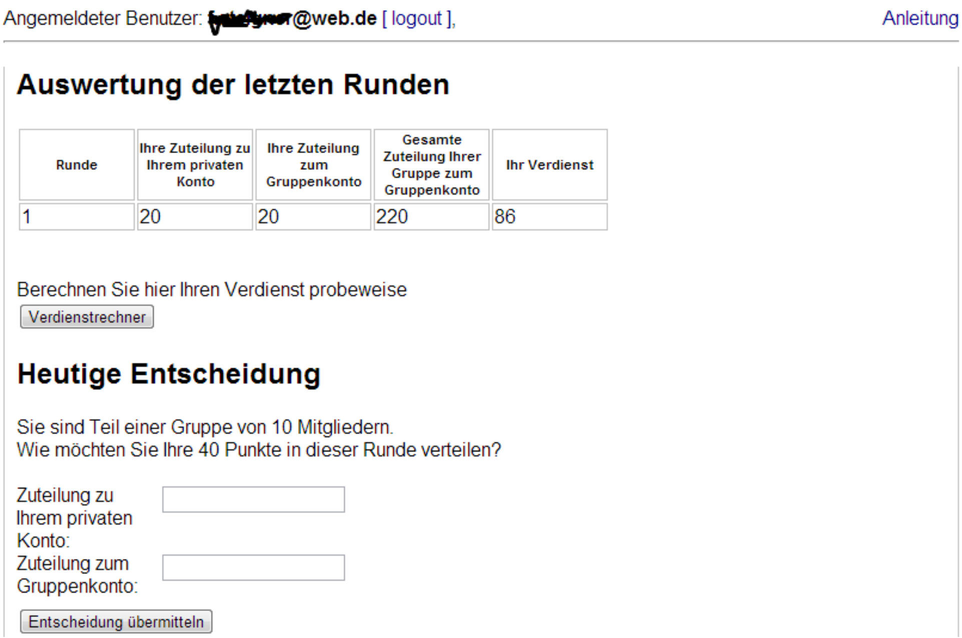
Vor der heutigen ersten Runde bitten wir Sie, sich mit den Regeln der Befragung vertraut zu machen. Bitte klicken Sie jetzt auf unten stehenden Link und laden Sie die Anleitung zur Umfrage herunter. Sie müssen die Anleitung gelesen haben, um zu verstehen, wie Sie in dieser Befragung Punkte verdienen. Die Teilnahme selbst wird dann nur etwa 2 Minuten Ihrer Zeit pro Runde in Anspruch nehmen.

Anleitung

Wir wünschen Ihnen viel Spaß mit der Umfrage!

Weiter zur ersten Runde

Screen 1 (Welcome), shown in round 1 only



Angemeldeter Benutzer: [redacted]@web.de [logout], Anleitung

Auswertung der letzten Runden

Runde	Ihre Zuteilung zu Ihrem privaten Konto	Ihre Zuteilung zum Gruppenkonto	Gesamte Zuteilung Ihrer Gruppe zum Gruppenkonto	Ihr Verdienst
1	20	20	220	86

Berechnen Sie hier Ihren Verdienst probeweise

Verdienstrechner

Heutige Entscheidung

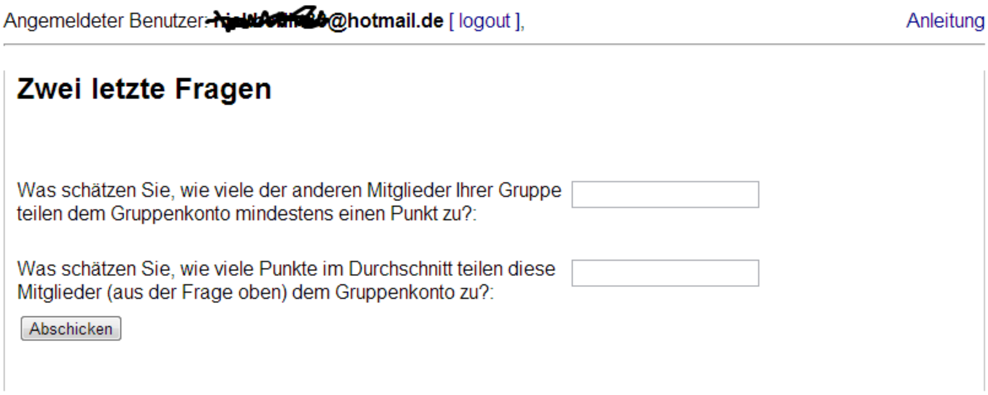
Sie sind Teil einer Gruppe von 10 Mitgliedern.
Wie möchten Sie Ihre 40 Punkte in dieser Runde verteilen?

Zuteilung zu Ihrem privaten Konto:

Zuteilung zum Gruppenkonto:

Entscheidung übermitteln

Screen 2 (Decision), shown here for round 2



Angemeldeter Benutzer: [redacted]@hotmail.de [logout], Anleitung

Zwei letzte Fragen

Was schätzen Sie, wie viele der anderen Mitglieder Ihrer Gruppe teilen dem Gruppenkonto mindestens einen Punkt zu?:

Was schätzen Sie, wie viele Punkte im Durchschnitt teilen diese Mitglieder (aus der Frage oben) dem Gruppenkonto zu?:

Abschicken

Screen 3 (Belief elicitation)

Verdienstrechner

Meine Zuteilung zu meinem privaten Konto: 23

Meine Zuteilung zum Gruppenkonto: 17

Durchschnittliche Zuteilung eines jeden der anderen Gruppenmitglieder zum Gruppenkonto: 20

Berechnen

Ihre Zuteilung zu Ihrem privaten Konto	Ihre Zuteilung zum Gruppen-Konto	Gesamte Zuteilung Ihrer Gruppe zum Gruppen-Konto	Auszahlung aus dem Gruppen-Konto	Ihr Verdienst der Runde
10	30	1020	306	316
9	31	1021	306	315
20	20	2000	600	620
19	21	2001	600	619
23	17	1997	599	622

Payoff calculator

B Experimental instructions

Presented here is the original wording of the instructions for the 10L treatment. Instructions for the other treatments differed in the group size stated in the text.

Anleitung
für die interaktive Umfrage
in Zusammenarbeit mit MySurvey

Bitte lesen Sie diese Anleitung sorgfältig durch, bevor Sie mit der Umfrage beginnen! Nur so wissen Sie, wie Sie Ihren Verdienst durch Ihre Antworten beeinflussen.

Wie viele Punkte Sie durch diese interaktive Umfrage verdienen, hängt sowohl von Ihren eigenen Entscheidungen beim Beantworten ab, als auch von den Entscheidungen anderer Teilnehmer. Zusätzlich zum variablen Verdienst erhalten Sie in jeden Fall 200 Punkte als Dankeschön für Ihre Teilnahme. Alle Punkte werden Ihnen nach Abschluss aller 7 Runden der Befragung gutgeschrieben.

Ihre Angaben während der Befragung werden vollständig anonym behandelt. Sowohl die von Ihnen angegebene E-Mail-Adresse als auch die Telefonnummer werden nach Abschluss der Befragung wieder gelöscht.

IHRE ENTSCHEIDUNGSSITUATION

Sie wurden einer Gruppe von 10 MySurvey-Mitgliedern zugeordnet, die wie Sie an dieser Befragung teilnehmen. Alle Mitglieder Ihrer Gruppe lesen die gleiche Anleitung wie Sie.

In jeder der 7 Runden der Befragung werden Sie gebeten, die gleiche Frage zu beantworten: Sie bekommen 40 MySurvey-Punkte zur Verfügung gestellt – wie möchten Sie diese auf zwei Alternativen verteilen? Die beiden Alternativen sind ein privates Punktekonto und ein gemeinsames Konto Ihrer Gruppe. Die anderen 9 Teilnehmerinnen und Teilnehmer in Ihrer Gruppe beantworten die Frage gleichzeitig mit Ihnen.

Die beiden Konten bedeuten dabei folgendes: Jeder Punkt, den Sie Ihrem privaten Konto zuteilen, erhöht Ihren eigenen Punkteverdienst um 1 Punkt (d.h. Sie behalten den Punkt einfach). Jeder Punkt, den Sie dem Gruppenkonto zuordnen, erhöht den Punkteverdienst bei allen Mitgliedern der Gruppe (also auch bei Ihnen) um 0,3 Punkte. Ebenso erhöht jeder Punkt, den ein anderes Mitglied dem Gruppenkonto zuteilt, den Verdienst aller 10 Mitglieder (inkl. Ihnen) um 0,3 Punkte. (Dies geschieht, indem die von allen eingezahlten Punkte auf dem Gruppenkonto mit 3 multipliziert werden, bevor sie gleichmäßig auf alle 10 Mitglieder verteilt werden.)

Die oben beschriebenen Regeln lassen sich in einer kurzen Formel zusammenfassen:
Ihr persönlicher Verdienst = [Ihre Zuteilung zum privaten Konto] + 0,3 x [Summe aller Beiträge zum Gruppenkonto in der Gruppe].

ABLAUF DER RUNDEN

Vor jeder neuen Runde der Umfrage erhalten Sie Informationen darüber, wie die vorhergehenden Runden ausgegangen sind. Sie können nachschauen, wie viele Punkte Sie Ihrem persönlichen Konto und wie viele Sie dem Gruppenkonto zugeteilt haben. Außerdem erfahren Sie, wie viele Punkte in Ihrer Gruppe insgesamt dem Gruppenkonto zugeteilt wurden. Aus diesen Angaben wird Ihr Verdienst für die jeweilige Runde berechnet und angezeigt.

Seite **1** von **2**

Für jede Runde haben Sie drei Tage Zeit, sich einzuloggen und Ihre Antworten einzutragen. Nach Abschluss aller 7 Runden wird eine Runde der Befragung zufällig ausgelost. Nur das Ergebnis dieser einen Runde gilt für Ihren endgültigen Verdienst für Sie und die anderen Mitglieder Ihrer Gruppe! **Jede Runde ist also gleich wichtig und könnte für Ihren MySurvey-Punkteverdienst die Entscheidende sein.**

DER ENTSCHEIDUNGSBILDSCHIRM

Der Entscheidungsbildschirm ist der Hauptbildschirms in jeder Runde der Befragung.

Im oberen Teil des Bildschirms sehen Sie die Auswertung vergangener Runden wie in obigem Abschnitt „Ablauf der Runden“ beschrieben.

Durch Klicken auf die Schaltfläche „Verdienstrechner“ öffnet sich ein neues Fenster, in dem Sie den Verdienst vor Ihrer Entscheidung probeweise berechnen können. Der Verdienstrechner ist im nächsten Abschnitt beschrieben.

Im unteren Teil des Bildschirms treffen Sie Ihre Wahl zur Aufteilung der 40 Punkte für die aktuelle Runde. Wenn Sie in eines der Felder eine Zahl eingeben, wird der Rest Ihrer 40 Punkte automatisch in das andere Feld eingetragen.

Durch Klick auf den Link „Anleitung“ oben rechts im Bildschirm können Sie dieses Dokument erneut einsehen, herunterladen oder ausdrucken.

DIE SCHALTFLÄCHE „VERDIENSTRECHNER“

Sie haben hier die Möglichkeit, verschiedene Aufteilungen auszuprobieren. Tragen Sie dazu eine Punkteverteilung in die ersten beiden Zeilen ein. Damit Ihr Verdienst ermittelt werden kann, müssen Sie außerdem eine Vermutung über das Verhalten der anderen Mitglieder für die aktuelle Runde anstellen. Tragen Sie dazu in Zeile 3 ein, wie viele Punkte jedes andere Mitglied Ihrer Gruppe im Durchschnitt dem Gruppenkonto zuteilen wird (also ein Wert zwischen 0 und 40 Punkte).

Durch Klicken auf „Berechnen“ wird aus diesen Angaben Ihr persönlicher Verdienst berechnet und in die Tabelle im unteren Teil des Bildschirms eingetragen. Sie können so mehrere Kombinationen der Angaben ausprobieren und miteinander vergleichen.

ABSCHLIESSENDE FRAGEN

Nachdem Sie Ihre Entscheidung in einer Runde getroffen haben, werden Ihnen auf dem nachfolgenden Bildschirm noch zwei kurze Fragen bezüglich Ihrer Einschätzung zum Ergebnis gestellt. Danach ist die Runde zu Ende. Sie werden per Email eingeladen, wenn die nächste Runde beginnt.

KONTAKT

Sollten Sie Rückfragen haben, erreichen Sie die Betreuer der Umfrage durch einen Klick auf den Link "Kontakt" am oberen Ende der Umfrage-Bildschirme. Dorthin können Sie auch gerne Kommentare oder Anregungen bezüglich dieser Befragung schicken.

Seite **2** von **2**

The Effect of Ambient Noise on Cooperation in Public Good Games*

Johannes Diederich

Abstract

Environmental stressors such as noise, pollution, extreme temperatures, or crowding can pose relevant externalities in the economy if certain conditions are met. This paper presents experimental evidence that exposure to acute ambient noise decreases cooperative behavior in a standard linear public good game.

1 Introduction

Traditionally, economics has been abstracting economic behavior from the influence of environmental stressors[1] on humans, much as it used to do for the influence of moods and emotions (Kirchsteiger et al. 2006). In particular, researchers have assumed either that individual preferences are invariant to any impact of environmental stress or that the potential effect on preferences is stochastic and negligible. Environmental stressors may become relevant to economics, however, if three conditions are fulfilled: (1) The effect of environmental stressors on economic behavior is significant, (2) the ambient levels of environmental stressors are subject to (permanent) change (Rabin 1998),[2] and (3) adaptation of humans to altered levels is imperfect or costly.

As a first step to this agenda, this paper addresses condition (1) and tests for an aftereffect of acute noise exposure on cooperative behavior in a standard linear public good experiment. One reason for investigating noise among the available set of environmental stressors is that it has been suggested to be the most important stressor for conditions of environmental overload (Moser 1988). To my knowledge, neither noise

*Thanks to Magdalena Buckert and various colleagues for helpful discussions. Funding by the German Science Foundation under grant GO1604/1 is gratefully acknowledged.

[1]Noise, air pollution, extreme temperatures, and crowding have been typically subsumed under environmental stressors (Evans 1984).

[2]For example, the populations of industrialized countries have been subject to considerable variations in ambient levels of noise, pollution, and crowding for the past two centuries, especially in urban areas.

nor other environmental stressors have been subject to systematic research in economic experiments so far. While I am therefore not aware of previous results regarding an effect of ambient noise on cooperative behavior in social dilemmas, such as the voluntary provision of public goods, the psychological literature reports that acute exposure to noise reduces helping behavior, both in laboratory (Glass and Singer 1972, Sherrod and Downs 1974) and in field experiments (Page 1977, Moser 1988). Helping behavior and cooperative behavior in social dilemmas are related in that altruism has been suggested to be a major motivational driver for behavior in both situations (e.g. Andreoni 2006). Thus, assuming that exposure to noise affects the altruistic component of motivation, the immediate hypothesis for the question of an effect of noise in a public good experiment is that observed cooperation will *decline*.

The literature also provides evidence for the opposite hypothesis, however, if one takes into account possible physiological pathways mediating the potential effect of noise on cooperation. Traditionally, the effects of environmental stressors have been linked to *physiological* stress. Thus, one immediately plausible mechanism would be that noise produces physiological stress, and that physiological stress affects cooperative behavior or altruism. Regarding the first part of this pathway, the early literature on environmental stressors, which focused on *behavioral* measures of physiological stress, found that noise decreases frustration tolerance and attention (Glass and Singer 1972, Sherrod and Downs 1974, Page 1977). Evidence on a link of noise and *physiological* measures of stress, which are typically elevations in the cardiovascular and neuroendocrine systems, is more rare. A meta study by Dickerson and Kemeny (2004) on the effect of various psychological stressors on the stress hormone cortisol does not find a significant effect for noise. However, as the authors note, the sample size is small in the case of noise (n=6). A result where physiological stress from noise could be identified is the "effort-by-stress tradeoff" (Tafalla and Evans 1997, Evans and Cohen 2004): Noise increases norepinephrine and cortisol levels in a laboratory setting if accompanied with high effort to complete a simultaneous task. At the same time, performance in the task is unaffected by noise. If completing the task is associated with low effort, however, no physiological indication of stress under noise can be observed, while performance is significantly worse. Thus, physiological stress seems to be traded-off for effort and produced only if needed to compensate for the increased psychological demands from noise in order to maintain performance. Evidence for the effort-by-stress tradeoff was also found in field studies on job demand and occupational noise (Evans and Cohen 2004). A second result where noise was found to cause physiological stress was in the case of chronic exposure: In studies with children, chronic exposure to aircraft or traffic noise increased cortisol and epinephrine as well as blood pressure levels, adversely affected psychophysiological, cognitive, motivational, and affective indices of stress, and led to decreased persistence in problem solving tasks (Evans et al. 1995, Bullinger et al. 1999, Evans and Cohen 2004).[3]

Regarding the second part of the hypothesized link between noise, physiological stress, and cooperation/altruism, evidence is even more sparse. In an early study, Dovidio and Morris (1975) show that a high-stress condition leads to increased helping behavior towards others who share the same stressful situation. They observed less helping behavior, however, if the potential recipient was in a dissimilar and less stressful situation.[4] To my knowledge, there is only one paper investigating behavior in standard economic games: von Dawans et al. (2012) show that physiological stress, induced by a standardized laboratory stressor, is associated with increased pro-social and altruistic behavior in versions of the trust game and the dictator game. Taking together this evidence with that on the link between noise and physiological stress described above, there is reason to expect a *positive* effect of noise exposure on public good contributions if one assumes that physiological stress mediates the effect.

In the experiment reported in this paper, treated subjects were exposed to a conglomerate of typical urban sounds directly before playing a standard linear public good game in groups of four players. The data suggest a *negative* effect of acute noise exposure on public good contributions, compared to an unexposed control group of subjects, and would thus support the first of the two aforementioned hypotheses. The effect is statistically significant for certain subgroups among the student subjects, specifically,

[3]Some of these papers experimentally exploited airport openings that corresponded to natural experiments.

[4]These results found for male subjects were confirmed for female participants by Hayden et al. (1984)

Bachelor and Master students, but not for students pursuing other degrees. The effect is not significantly different between men and women. While it is beyond the scope of this paper to further disentangle the motivational drivers or physiological mediators of the noise effect, the present study collects some survey-based controls for chronic, acute pre-experimental, and acute intra-experimental stress as well as for subjects' history of chronic and acute pre-experimental noise exposure in order to account for adaptation or multiple-stressor effects. Most of these variables, in particular chronic stress and noise history, do not significantly interact with the observed effect of noise. Likewise, there is no significant evidence for an effort-by-stress trade-off based on these survey measures. Evidence on an effect from the presence of multiple stressors is mixed, and the possibility of endogeneity inherent to ex post survey measures cannot be excluded.

The paper proceeds as follows. Section 2 presents the experimental design. Section 3 presents the results, and Section 4 concludes.

2 Method

Participants Participants were 63 female and 49 male students at Heidelberg University, recruited via ORSEE (Greiner 2004). 64 subjects were randomly assigned to the noise treatment, 48 subjects to the control treatment. The share of females is somewhat higher in the control treatment (64.6% vs. 50%), which turns out not to be significantly different using two-sided tests. Half of the subjects majors in an academic subject of the social sciences (economics, political science, sociology) while the other half majors in subjects belonging to the humanities, sciences, or another area. Subjects of the social sciences turn out to be unequally distributed across treatments, with a share of 62.5% in the noise treatment but only 33.3% in the control group (Table 1). Likewise, differences exist with respect to the degree pursued. The unequal distributions are likely to be due to the sample size, as the procedure of matching treatment condition and experimental sessions was not systematic. The differences can be a problem for simple tests of the treatment effect if economics students display systematically different behavior, or if academic maturity matters, and thus calls for an additional regression analysis. Mean

Table 1: Share of students in the social sciences and pursued degrees, by treatments

	Major belongs to social sciences	Degree pursued			
		Bachelor	Master	Staatsexamen[a]	Other, or no answer
Noise	62.5%	70.3%	6.3%	18.8%	4.7%
Control	33.3%	43.8%	22.9%	29.2%	4.2%

Notes: [a] The *Staatsexamen* is a degree in Germany issued by the government, not the university. For examples, teachers, lawyers, and medical doctors graduate via *Staatsexamen*.

earnings were €8.40 per subject (SD €1.01), including a fixed show-up fee of €3.

Noise treatment In the first part of the experiment, subjects who received the noise treatment were exposed to 25min of noise at about 65dB(A) on average with bursts at about 75dB(A) on average.[5] The noise was administered through speakers[6] (Glass and Singer 1972, Sherrod and Downs 1974, Tafalla and Evans 1997) placed such that noise levels were fairly equal across cubicles in the lab. Loudness was measured several times during exposure.[7] The noise was a mixture of typical urban sounds at varying levels (such as road traffic, aircraft sound, a drilling jackhammer[8], a ringing cellphone, and passing people engaged in a indiscernible chat) which was interrupted by random bursts of electronic static.[9] During exposure, subjects had to perform a paid proofreading task on four magazine or newspaper articles on various unrelated topics. Payment in this part was conditioned on performance in the task. The control group performed the same task for the same time without being exposed to noise. The typical noise level in the laboratory in sessions in the control treatment was about 45dB(A).

Administering the noise during an unrelated effortful task prior to the public good game is guided by two findings of the literature. First, there is stronger evidence of behavioral aftereffects than of simultaneous effects for noise and other environmental stressors. In Glass and Singer (1972) and Sherrod and Downs (1974), for example,

[5] Following the guidelines of the German Association of Otolaryngologists, the noise would not exceed 85dB(A).

[6] Bose Companion 2.

[7] Trotec BS15.

[8] Referential note: “Builders drilling sound” recorded by Koops.

[9] Cohen (1980) reviews several papers using different types of noise. Results suggest that besides the uncontrollability of the noise, the unpredictability matters for an effect of intermittent noise samples while variations in the intensity of the noise matters for an effect of continuous noise samples. The noise administered in the present experiment combines both types using unpredictable, interrupting bursts and variations in intensity of the continuous parts.

subjects were confronted with an opportunity to display helping behavior after being exposed to noise during which they were working on some other task. While performance in the task did not significantly differ between the treatment and the control group, differences manifested in the behavior afterwards. This suggests that subjects may adapt during exposure.[10] Glass and Singer (1972) and many replications in the literature (see Evans and Cohen 2004) found similar behavioral aftereffects for other environmental stressors and for different tasks following exposure, for instance, persistence in a puzzle task or performance in a proofreading task. However, the dominance of aftereffects does not preclude to also find simultaneous effects: For the relevant case of helping behavior, for example, Page (1977) identifies differences during acute exposure. The second reason for having subjects solve a cognitively demanding task during noise exposure is to allow an effort-by-stress tradeoff to manifest, if present. One way to interpret aftereffects is in the context of behavior that occurs after a daily routine in an exposed environment, e.g., changes in home life from a noisy working environment (Evans and Cohen 2004).

Public good game The public good game was played directly after the proofreading task, in groups of four subjects and for ten periods. The game was of the standard linear type using the voluntary contribution mechanism (VCM). The choice was framed as distributing the endowment of 20 "points" of each round (worth €0.30) between a "private account" and a "group account", the latter yielding a marginal per capita return (MPCR) of 0.4. Matching was constant across rounds (partner treatment) and earnings of each round added to the final payoff. Instructions included a table illustrating a subject's round earnings for various combinations of own and others' contributions to the group account.[11]

Questionnaire The experiment concluded with a questionnaire collecting measures of subjective effort in the proofreading task, perceived stress, history of noise exposure, and some demographics. The question in which subjects have to rank their effort put in

[10]This is in line with others' findings regarding the effect of simultaneous noise on cognitive performance (e.g. Hancock and Pierce 1985, Hygge and Knez 2001).

[11]See the appendix for the exact wording of the instructions (in German, translation to English available upon request from the author).

the proofreading task serves as a control for the presence of physiological stress during noise exposure (Tafalla and Evans 1997). To control for background levels of chronic and pre-experimental stress, the questionnaire used the 4-item version of the Perceived Stress Scale (PSS) by Cohen et al. (1983) as well as a simple question asking for perceived stress on the same day prior to the experiment, respectively. Questions regarding noise asked how bothered the subject felt by the experimental noise (Sherrod and Downs 1974) as well as a subjective ranking of the noise exposure on the same day prior to the experiment, during the past year, and during childhood. Although subjective in nature, collecting some long-term history of noise exposure can serve as a control for “adaptation level shifts” through which noise might be perceived less annoying if a subject has a history of chronic exposure (Berglund et al. 1975, Evans and Cohen 2004). Likewise, the questions about perceived stress and noise exposure right before the experiment can serve as controls for the presence of multiple stressors which could increase the effect of noise due to “diminished coping with multiple stressors” (Evans and Cohen 2004).

Protocol The experiment consisted of seven experimental sessions, conducted in Spring 2013 in the experimental laboratory of the Department of Economics at Heidelberg University, Germany. Each session was scheduled for 1 hour and 15 minutes including seating and payment, which turned out to be about the time each session took. After seating, subjects read the instructions (in hardcopy) for both parts of the experiment in private. The experimenter afterwards paraphrased the important parts of the instructions in a standardized way, with an emphasis on the understanding of the public good game, including the payoff table. Subjects had the opportunity to ask questions in private. When all questions were answered, the texts for the proofreading task were distributed in hard copy and the experiment started. All other elements of the experiment were programmed and conducted with the software z-Tree (Fischbacher 2007). Subjects were not informed about their earnings in the proofreading task before finishing the end of the experiment. After finishing the questionnaire, subjects were paid and dismissed.

3 Results

3.1 Nonparametric treatment effects

Before investigating an aftereffect in the public good game, it is meaningful to evaluate the reception of the noise treatment during exposure. In the proofreading task, subjects' earnings (which are perfectly collinear to their performance) turn out to be significantly lower under noise exposure ($p < 0.10$ in a two-sided Fligner-Policello Robust Rank Order Test, Fligner and Policello 1981). However, part of the difference can be attributed to a few subjects[12] who missed to enter and confirm the number of mistakes they had counted into their computer terminals before time ran out, thus earning zero money in the task.[13] Interestingly, these instances only occurred in the noise treatment, despite identical oral warnings prior to beginning the task as well as 60-100 seconds before time was up. If these subjects are excluded, the difference in performance between the treatment group and the control group is insignificant ($p = 0.57$). As mentioned before, both results, unaffected performance and decreased performance under noise, have been found for comparable tasks in the literature. At the same time, subjects report to experience the administered noise as bothersome and disruptive. In the ex post questionnaire, a majority of treated subjects (61%) chooses one of the upper three categories on a scale of six answer possibilities to describe their feelings about the noise during the proofreading task. The six answer categories were presented as ranging from "not bothersome at all" to "extremely bothersome". Both results together would suggest the presence of an effort-by-stress trade-off among those subjects who are unaffected in performance (i.e. do not miss the timeout) and rate the noise as bothersome, thus the literature would suggest elevated levels of physiological stress among these subjects to prevent performance losses under noise.

Result 1 *Subjects performance in the proofreading task is not significantly different under noise exposure. However, some treated subjects miss to pay sufficient attention to the time constraint under noise. The majority of treated subjects rates the noise as bothersome.*

[12] 8 out of 112.

[13] The upper right corner of the computer screen displayed a count down for the 25 minutes (in seconds).

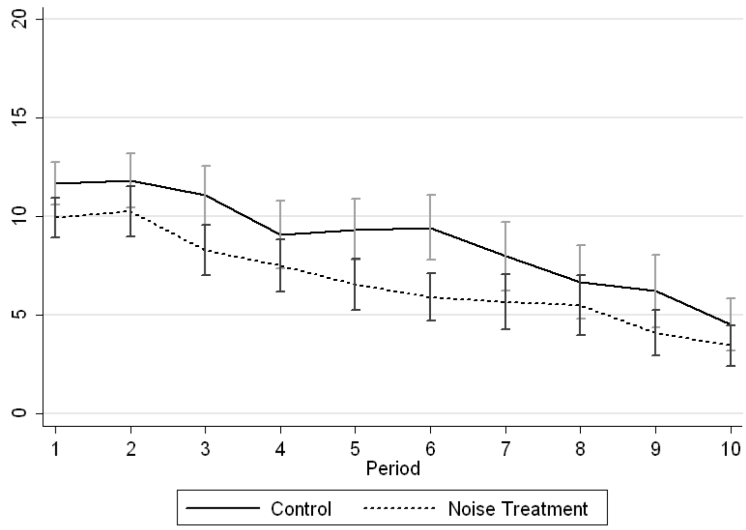


Figure 1: Mean contributions (in experimental points) of subjects in the control group and the treatment group. Error bars denote standard errors which are clustered at experimental groups ($N = 28$).

Turning to the aftereffect of noise in the public good game, Figure 1 depicts mean contributions for the control group and the treatment group. Clearly, average contributions of those exposed to noise prior to the game seem to be consistently below those of unexposed subjects. The effect is almost never significant, however, using two-sided tests, as Table 2 shows. Plausibly, the number of independent observations (28 experimental groups) is too small to identify a roundwise effect. Pooling group means of all periods would lead to highly significant two-sided test results ($p < 0.002$).

Result 2 *There is tentative nonparametric evidence that public good contribution levels in the noise treatment are consistently below those in the control treatment.*

3.2 Regression results

In order to exploit the controls obtained by the questionnaire, and in order to control for the unequal distributions between the treatment and the control group found for some demographics, this section presents regression results. Table 3 reports summary statistics for variables elicited in the questionnaire. The first three variables will be assumed to proxy for physiological stress, either acute (if subjects rate their day before

Table 2: Significance of the treatment effect

		Period				
		1	2	3	4	5
Mean contribution (points)	Noise	9.94	10.26	8.28	7.50	6.55
		(1.00)	(1.28)	(1.27)	(1.34)	(1.32)
	Control	11.67	11.83	11.06	9.08	9.33
		(1.08)	(1.38)	(1.51)	(1.73)	(1.56)
Two-sided (one-sided) Fligner-Policello test		Ù=1.364 (*)	Ù=0.875	Ù=1.448 (*)	Ù=0.657	Ù=1.519 (*)
		Period				
		6	7	8	9	10
Mean contribution (points)	Noise	5.89	5.65	5.48	5.48	3.44
		(1.20)	(1.41)	(1.51)	(1.16)	(1.03)
	Control	9.44	7.98	6.67	6.21	4.52
		(1.67)	(1.73)	(1.86)	(1.83)	(1.32)
Two-sided (one-sided) Fligner-Policello test		Ù=2.006** (***)	Ù=1.229	Ù=0.708	Ù=0.812	Ù=0.774

Notes: Mean contribution is the average of the means of the experimental public good groups. Standard errors in parentheses are clustered at the group level. For periods 2-10, the Fligner-Policello tests the equality of the distributions of the group means. For period 1, the Fligner-Policello tests the equality of the distributions of the individual contributions. Stars indicate significance levels (* 10%, ** 5%, *** 1%).

the experiment as stressful or if they rate their effort in the proofreading task as high inducing an effort-by-stress tradeoff) or chronic (as measured by the 4-item PSS). The next three variables represent subjects' self-rated history of noise exposure on the experimental day prior to the experiment, over the past year, and during childhood at the place they lived most of the time. All but the PSS scale are dummy variables which are constructed from subjects' answers on a 6-item rating scale. The second set of variables are demographic characteristics.

Testing for systematic differences in the distributions of the answer ratings between the treatment group and the control group yields some evidence for endogeneity of the questionnaire responses. While for most variables, significant differences cannot be established (using a two-sided Fligner-Policello test), for the two questions on noise exposure during the past year and during childhood, subjects in the noise treatment report significantly lower exposures compared to the control group. The potential presence of a bias from endogeneity needs to be taken into account when interpreting regression

Table 3: Summary statistics of questionnaire variables

Variable	Mean	SD	Min	Max
Pre-experimental stress	.5714	.4971	0	1
Effort in part I	.6696	.4725	0	1
Chronic stress (4-item PSS)	6.009	3.036	0	15
Pre-experimental noise exposure	.6429	.4813	0	1
Chronic noise exposure (past year)	.4911	.5022	0	1
Noise exposure in childhood	.5446	.5002	0	1
Female	.5625	.4983	0	1
Age	22.13	2.625	18	33
Major belongs to the social sciences	.5	.5022	0	1
Pursues Bachelor's degree	.5893	.4942	0	1
Pursues Master's degree	.1339	.3421	0	1
Pursues Staatsexamen	.2321	.4241	0	1
Pursues other degree / no answer	.0446	.2074	0	1

results.

The econometric specification used in the regressions is

$$\begin{aligned} C_{i,t} = \quad & \beta_0 + \beta_1 P_{i,t} + \beta_2 T_{i,t} + \beta_3 S_{i,t} + \delta_1 (T_{i,t} \times S_{i,t}) \\ & + \beta_4 H_{i,t} + \delta_1 (T_{i,t} \times H_{i,t}) + \beta_5 D_{i,t} + \delta_1 (T_{i,t} \times D_{i,t}) \\ & + \beta_6 \sum_{-i,t-1} C_{-i,t-1} + \beta_7 C_{i,t-1} \end{aligned}$$

where $C_{i,t}$ denotes subject i's contributions in period t, $P_{i,t}$ denotes a vector of dummy variables for the experimental periods, $T_{i,t}$ is the dummy indicating the treatment, $S_{i,t}$ and $H_{i,t}$ are the vectors of stress-related and noise-related controls as described above, $D_{i,t}$ is the vector of demographic controls, and the last two terms represent others' and i's own contributions to the group account in the previous period.

Table 4 reports coefficient estimates from OLS regressions.[14,15] The first specification only includes the treatment variable, besides the period dummies. Columns (2) to (4) each include, in addition, one of the vectors containing variables related to either stress, noise history, or demographics as well as their interactions with the treatment variable.[16] Column (5) presents the full model. Column (6) corresponds to column (5),

[14]Running random effects GLS panel regressions with period as the time variable yields very similar results with the same significance levels for all variables.

[15]Note that due to the small sample size, Table 4 also marks significance levels of below 20% (with one star in parentheses) for illustrative purposes, in order to point out variables that might plausibly qualify for significance in a larger sample.

[16]Among the demographic controls, I leave out age due to its small variation in the student sample. If there are effects, I expect them to be better reflected by the pursued academic degrees which plausibly

but, in addition, includes the experimental feedback variables for the previous period. Coefficient estimates for the period dummies and the constant, which are not reported in Table 4 but included in each specification, confirm a highly significant negative time trend, as suggested by Figure 1.

Column (1) confirms the insignificant negative treatment effect suggested by the nonparametric tests in Section 3.1. Controlling for self-reported variables related to stress in column (2), the results show significantly higher contributions among those subjects who both report a stressful day and were exposed to the noise. The limited statistical power of the sample prevents to clearly disentangle whether the interaction effect is due to a negative main effect of noise among the self-reported unstressed that is absent among the stressed, or whether the effect is due to a negative main effect of stress among the control group that is absent among the treated. Taking together the evidence from all specifications in Table 4 and from additional regressions, there is evidence for both. In an attempt to explain this finding, it appears that both a negative effect of noise as well as of stress is intuitive, but the significant absence of the negative effect if both conditions meet appears counterintuitive. In particular, the negative effect of pre-experimental stress would be consistent with "diminished coping from multiple stressors" (Evans and Cohen 2004) but this is inconsistent with un-diminished or even better coping under pre-experimental stress and noise. A potential explanation for this finding could be endogeneity of the questionnaire answers, however: If subjects felt tempted to ex post report a stressful day as an excuse for having given less, then the positive interaction effect would imply that in the noise treatment, subjects are simply less tempted to use this excuse and, at the same time, give less because of the noise exposure. Turning to the effort in the proofreading task, there is no evidence that treated subjects reporting higher efforts behave different, potentially from an effort-by-stress tradeoff, as indicated by the insignificant interaction effect.[17] However, all subjects across treatments who report higher effort also tend to give more to the public good. This could be a manifestation

proxy for maturity among students.

[17] In a future analysis of the data, a better way to test for a potential effect of an effort-by-stress tradeoff given the available variables would be to test of different behavior of subjects who (1) were in the treatment group, (2) reported over-average levels of effort, (3) reported the noise as bothersome, and (4) did not miss the timeout.

Table 4: OLS coefficient estimates of contributions to the group account

	(1)	(2)	(3)	(4)	(5)	(6)
Noise	-2.070	-3.784	-0.638	-5.199*	-5.919*	-2.316*
	(1.789)	(2.997)	(2.417)	(2.736)	(3.304)	(1.308)
Pre-exper. stress	–	-1.846(*)	–	–	-1.932(*)	-1.529***
		(1.238)			(1.429)	(0.551)
. . . * Noise	–	3.489*	–	–	2.917(*)	2.106***
		(1.771)			(2.150)	(0.704)
Effort in part I	–	2.719*	–	–	2.833(*)	0.728(*)
		(1.467)			(1.957)	(0.486)
. . . * Noise	–	-1.969	–	–	-1.608	-0.250
		(1.991)			(2.623)	(0.786)
Chronic stress (PSS)	–	-0.076	–	–	-0.088	0.000
		(0.218)			(0.283)	(0.110)
. . . * Noise	–	0.158	–	–	0.200	0.077
		(0.326)			(0.367)	(0.140)
Pre-exp. noise expos.	–	–	-1.527	–	-0.206	0.116
			(1.560)		(2.279)	(0.891)
. . . * Noise	–	–	1.765	–	0.961	-0.212
			(1.802)		(2.497)	(1.079)
Chronic noise expos.	–	–	3.397	–	3.236(*)	0.161
			(2.617)		(2.432)	(0.554)
. . . * Noise	–	–	-2.519	–	-0.773	0.783
			(2.783)		(2.762)	(0.705)
Noisy childhood	–	–	0.272	–	0.677	0.496
			(1.361)		(1.710)	(0.610)
. . . * Noise	–	–	-1.410	–	-1.455	-0.602
			(1.782)		(2.073)	(0.761)
Female	–	–	–	-0.906	-1.164	0.435
				(1.157)	(1.456)	(0.759)
. . . * Noise	–	–	–	0.449	-0.758	-0.368
				(1.543)	(2.179)	(0.891)
Major of the Soc. Sciences	–	–	–	-2.494(*)	-3.501*	-1.341(*)
				(1.858)	(1.844)	(0.817)
. . . * Noise	–	–	–	2.184	3.519(*)	1.437(*)
				(2.545)	(2.431)	(1.040)
Master student	–	–	–	-1.697	-2.918**	-0.961
				(1.822)	(1.383)	(0.916)
Staatsex. stud.	–	–	–	-3.384(*)	-3.526(*)	-1.094
				(2.510)	(2.372)	(0.858)
Other stud./n.a.	–	–	–	-0.906	-2.186	-0.850
				(1.743)	(1.774)	(1.313)
Master * Noise	–	–	–	-1.453	0.106	-0.004
				(2.193)	(2.216)	(1.019)
St.-ex. * Noise	–	–	–	6.916**	7.538**	1.520
				(3.112)	(2.920)	(1.167)
Other/n.a. * Noise	–	–	–	6.368**	9.664***	3.324**
				(2.405)	(2.888)	(1.577)
Others' contributions	–	–	–	–	–	0.138***
						(0.018)
Own contribution	–	–	–	–	–	0.465***
						(0.047)
N	1120	1120	1120	1120	1120	1008
# of clusters	28	28	28	28	28	28
R^2	0.113	0.140	0.134	0.171	0.227	0.597

Notes: (*) significant at the 20 % level, * at 10 %, ** at 5 %, *** at 1 %. Standard errors are clustered at the group level and shown in parentheses. All regressions include dummy controls for experimental periods and a constant. In columns (4)-(6), Bachelor students are the baseline.

of an experimenter demand effect. The measure of chronic stress, the PSS, does never significantly correlate with either contributions or noise. Altogether, the evidence on an effect of stress in the experiment is mixed and supports careful interpretation of simple ex post survey measures.

Result 3 *Survey measures of chronic stress insignificantly correlate with the both contributions to the public good as well as the effect of noise. Survey measures of acute pre-experimental or intra-experimental stress provide mixed evidence, and contamination of questionnaire answers by endogeneity, e.g. from ex-post rationalization or an experimenter demand effect, cannot be excluded.*

Column (3) reveals no significant main or interaction effects of the various types of self-reported noise exposure in the past. This points against a measurable presence of “adaptation level shifts” (Evans and Cohen 2004).

Result 4 *Regression results do not reveal any effects of the various types of self-reported past noise exposure on contributions or on the effect of noise on contributions.*

Employing demographic controls in column (4) delivers a significantly negative effect of the noise treatment for the baseline of Bachelor students. The effect persists for Master students, who do not significantly differ from the baseline, but is offset by a positive interaction effect of about the same magnitude for Staatsexamen students and for the residual category of other degrees and non-responding students. In addition, Staatsexamen students show tentative evidence for a negative main effect on contributions. As one would expect, there is (tentative) evidence that students with a major in the social sciences (mostly economics) give less. Male and female subjects do not significantly differ in giving behavior or reception of the treatment in the present sample, as one might have expected from findings by Epstein and Karlin (1975) for the effect of crowding.

Result 5 *Controlling for demographics, noise exposure significantly reduces contributions among subjects pursuing a Bachelor’s or Master’s degree, but not for other subjects in the sample. No such differences can be established between male and female subjects.*

All effects described above generally persist if we test the full model in column (5), with slight changes in the significance levels of some variables. Also, the full model adds (weak) evidence that the significantly negative effect of noise is less received by students of the social sciences, and that Master students give less in general. The latter effect can plausibly related to experience or age. The final column (6) conditions contribution choices on what happened in the previous period. Both experimental feedback variables are highly significant in explaining behavior in the current period. Including them affects significance levels of some variables, partly heavily, but does not alter coefficient signs.[18]

4 Conclusions

Despite a considerable body of literature in other disciplines, for example, environmental psychology, the effects of environmental stressors have not yet been in the focus of economists. And rightly so, if environmental stressors are, on aggregate, irrelevant to economic behavior and decision making. Environmental stressors pose an externality and source of market failure, however, if their effect is significant, if changes in their levels are permanent and man-made, and if adaptation is imperfect or costly. This paper provides evidence for the first of these three conditions. My results suggest a significant adverse effect of the exposure to acute ambient noise on the extent of voluntary giving to a linear laboratory public good.

Among the limited related literature, this result confirms findings of a negative effect of noise on helping behavior (Glass and Singer 1972, Sherrod and Downs 1974, Page 1977, Moser 1988). The finding supports altruism as the affected motivational transmitter, as altruism is regarded a common motivation for both helping behavior and public good contributions (Andreoni 2006). In contrast, the result appears not in line with the hypothesis that physiological stress is the dominant physiological pathway of the noise effect, since the (sparse) literature suggests a positive effect of physiological stress on pro-social behavior (Dovidio and Morris 1975, von Dawans et al. 2012). However, it is beyond the scope of this paper to clearly identify the mediator causing the observed effect, and alternative hypotheses do exist. For example, risk preferences or cognitive processing capacities could equally plausible be the affected motivational drivers of public

[18] Again, random effects GLS panel regressions with period as the time variable yield very similar results.

good giving if subjects perceive the group account as the more risky or more cognitively demanding option.[19] The same argument may hold for the effect on helping behavior reported in the literature. Likewise, alternative hypotheses besides physiological stress exist for the physiological transmission of the effect. For example, Page (1977) mentions stimulus overload, distraction, or escape and avoidance behavior as possible explanations for the negative effect on helping behavior. These could also be plausibly related to selfish behavior in cases of social dilemmas. Disentangling these various explanations for physiological or motivational causes of the effect is left for future research.

The effect of noise found in the present data is an effect of acute, short-term noise exposure. The extent to which the result may generalize to an environment of chronic noise exposure, such as urban areas or the vicinity to an airport, depends on adaptation, as mentioned before. Research on chronic effects naturally poses more difficulties as the researcher mostly relies on either empirical data or natural experiments. Regarding adaptation costs, the literature provides some evidence on the costs of adaptation to permanently increased noise levels. For example, chronic noise was found to adversely affect auditory discrimination in children and thus, reading acquisition, presumably due to a habituation of adaptation strategies to cope with chronic exposure (Cohen et al. 1973, Evans et al. 1995, Evans and Cohen 2004). If adaptation to chronically increased levels incurs significant costs, man-made changes in ambient levels of environmental stressors such as noise would be sources of relevant externalities in the economy, however.

References

Andreoni, J. (2006). Philanthropy, *in* S.-C. Kolm and J. M. Ythier (eds), *Handbook on the Economics of Giving, Reciprocity and Altruism*, Vol. 2, Elsevier, pp. 1201–1269.

Berglund, B., Berglund, U. and Lindvall, T. (1975). A study of response criteria in populations exposed to aircraft noise, *Journal of Sound and Vibration* **41**(1): 33–39.

Bullinger, M., Hygge, S., Evans, G. W., Meis, M. and Mackensen, S. v. (1999). The psychological cost of aircraft noise for children, *Zentralblatt für Hygiene und Umweltmedizin* **202**(2-4): 127–138.

[19] Not entirely related to this point, von Dawans et al. (2012) find unaltered propensities to take risks for higher levels of physiological stress.

Cohen, S. (1980). Aftereffects of stress on human performance and social behavior: A review of research and theory, *Psychological Bulletin* **88**(1): 82–108.

Cohen, S., Glass, D. C. and Singer, J. E. (1973). Apartment noise, auditory discrimination, and reading ability in children, *Journal of Experimental Social Psychology* **9**(5): 407–422.

Cohen, S., Kamarck, T. and Mermelstein, R. (1983). A global measure of perceived stress, *Journal of Health and Social Behavior* **24**(4): 385–396.

Dickerson, S. S. and Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research, *Psychological Bulletin* **130**(3): 355–391.

Dovidio, J. F. and Morris, W. N. (1975). Effects of stress and commonality of fate on helping behavior, *Journal of Personality and Social Psychology* **31**(1): 145–149.

Epstein, Y. M. and Karlin, R. A. (1975). Effects of acute experimental crowding, *Journal of Applied Social Psychology* **5**(1): 34–53.

Evans, G. W. (1984). *Environmental stress*, Cambridge University Press, Cambridge.

Evans, G. W. and Cohen, S. (2004). Environmental stress, *in* C. Spielberger (ed.), *Encyclopedia of Applied Psychology*, Vol. 1, Elsevier, pp. 815–824.

Evans, G. W., Hygge, S. and Bullinger, M. (1995). Chronic noise and psychological stress, *Psychological Science* **6**(6): 333–338.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments, *Experimental Economics* **10**(2): 171–178.

Fligner, M. A. and Policello, G. E. (1981). Robust rank procedures for the Behrens-Fisher problem, *Journal of the American Statistical Association* **76**(373): 162–168.

Glass, D. and Singer, J. (1972). *Urban Stress: Experiments on Noise and Social Stressors*, Academic Press, New York.

Greiner, B. (2004). An online recruitment system for economic experiments, *Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht* **63**: 79–93.

Hancock, P. A. and Pierce, J. O. (1985). Combined effects of heat and noise on human performance: A review, *American Industrial Hygiene Association Journal* **46**(10): 555–566.

Hayden, S. R., Jackson, T. T. and Guydish, J. (1984). Helping behavior of females: Effects of stress and commonality of fate, *The Journal of Psychology* **117**(2): 233–237.

Hygge, S. and Knez, I. (2001). Effects of noise, heat and indoor lighting on cognitive performance and self-reported affect, *Journal of Environmental Psychology* **21**(3): 291–299.

Kirchsteiger, G., Rigotti, L. and Rustichini, A. (2006). Your morals might be your moods, *Journal of Economic Behavior & Organization* **59**(2): 155–172.

Moser, G. (1988). Urban stress and helping behavior: Effects of environmental overload and noise on behavior, *Journal of Environmental Psychology* **8**(4): 287–298.

Page, R. A. (1977). Noise and helping behavior, *Environment and Behavior* **9**(3): 311–334.

Rabin, M. (1998). Psychology and economics, *Journal of Economic Literature* **36**(1): 11–46.

Sherrod, D. R. and Downs, R. (1974). Environmental determinants of altruism: The effects of stimulus overload and perceived control on helping, *Journal of Experimental Social Psychology* **10**(5): 468–479.

Tafalla, R. J. and Evans, G. W. (1997). Noise, physiology, and human performance: The potential role of effort, *Journal of Occupational Health Psychology* **2**(2): 148–155.

von Dawans, B., Fischbacher, U., Kirschbaum, C., Fehr, E. and Heinrichs, M. (2012). The social dimension of stress reactivity: Acute stress increases prosocial behavior in humans, *Psychological Science* **23**(6): 651–660.

Appendix

Experimental instructions

[*Original wording. Translation to English available upon request from the author.*]

Willkommen zu diesem Experiment! Bitte **lesen Sie diese Anleitung sorgfältig** durch. Nur so wissen Sie, **wie Sie Ihren Verdienst durch Ihre Entscheidungen und Antworten beeinflussen**. Wenn Sie den Instruktionen folgen, können Sie insgesamt einen angemessenen Geldbetrag verdienen.

Dieses Experiment besteht aus **zwei Teilen** und einem abschließenden **Fragebogen**. Alle Teilnehmer nehmen am gleichen Experiment teil und lesen die gleiche Anleitung wie Sie.

Ab jetzt und bis zum Ende des Experiments ist es Ihnen **untersagt, mit anderen Teilnehmern zu kommunizieren**. Wenn Sie während des Experiments eine **Frage** haben, **heben Sie einfach die Hand**.

Alle Ihre Angaben werden **vertraulich** behandelt und sind **anonym**. Das einzige Mal, wo Ihr Namen und Ihre Adresse benötigt wird, ist auf der Empfangsbestätigung / Quittung der Auszahlung, die Sie **bitte jetzt ausfüllen**. Tragen Sie dort auch die Computernummer Ihres Sitzplatzes ein, damit

Ihnen Ihr Verdienst nach dem Experiment zugeordnet werden kann. Das Feld für den **Betrag bitte freilassen**.

Um die **wissenschaftliche Verwertbarkeit** dieses Experiments zu fördern, treffen Sie alle Ihre Entscheidungen am besten frei von Mutmaßungen über den wissenschaftlichen Gegenstand dieses Experiments. Vielen Dank.

Auszahlungen

Während des Experiments spielen Sie um Punkte. Nach dem Experiment werden die Punkte, die Sie verdient haben, in Euro umgerechnet. Dabei entspricht **1 Punkt = 1,5 Eurocent (0,015 Euro)**. Zusätzlich zum variablen Verdienst erhalten Sie eine feste Teilnahmevergütung von 3,00 Euro.

Experiment Teil 1

Während des ersten Teils des Experiments lesen Sie mehrere Texte auf Papier. Jeder Text enthält **Rechtschreib- und Grammatikfehler**. Finden und zählen Sie die Fehler, und **tragen Sie die Anzahl in das entsprechende Feld auf dem Computerbildschirm ein**. Für jeden richtig erkannten Fehler erhalten Sie **5 Punkte**. übersteigt die Anzahl der Fehler, die Sie eintragen, die wahre Anzahl der Fehler im Text, werden Ihnen für jeden falsch erkannten Fehler wieder 5 Punkte abgezogen. Ihr Verdienst kann jedoch nicht negativ werden, sondern beträgt immer mindestens 0 Punkte.

Beispiel 1: Ein Text enthält 8 Fehler. Sie tragen 5 Fehler ein. Sie erhalten 25 Punkte.
Beispiel 2: Ein Text enthält 8 Fehler. Sie tragen 10 Fehler ein. Sie erhalten 30 Punkte.
Beispiel 3: Ein Text enthält 8 Fehler. Sie tragen 30 Fehler ein. Sie erhalten 0 Punkte.

[*Only the treatment group saw the following paragraph:*] Während dieses Teils des Experiments wird der Raum mit **Geräuschen** beschallt. Der Pegel der Geräusche ist dabei stets gesundheitlich unbedenklich und entspricht Pegeln, denen Sie auch im Alltag begegnen.

Experiment Teil 2

Der zweite Teil des Experiments besteht aus **10 Runden**. Zu Beginn des zweiten Teils werden Sie über Computer per Zufallsauswahl **mit drei anderen Teilnehmern/-innen** im Raum zu einer Gruppe zusammengeschaltet. **Keiner der Gruppenmitglieder kennt die Identität** der anderen Gruppenmitglieder; sie wird auch nach dem Experiment nicht offengelegt. **Ihr Verdienst** in diesem Teil wird sowohl durch Ihre **eigenen** Entscheidungen als auch durch die **Entscheidungen der anderen** Teilnehmer in Ihrer Gruppe beeinflusst.

In jeder Runde sind Sie und die anderen Mitglieder Ihrer Gruppe mit der gleichen Entscheidungssituation konfrontiert: Jede(r) bekommt **20 Punkte zur Verfügung** gestellt und entscheidet, wie sie oder er diese **auf zwei Alternativen verteilen** möchte. Die beiden Alternativen sind ein **privates Punktekonto** und ein **gemeinsames Konto Ihrer Gruppe**.

- **Privates Konto**: Jeder Punkt, den Sie Ihrem privaten Konto zuteilen, erhöht (ausschließlich) Ihren **eigenen** Punkteverdienst, und zwar **um 1 Punkt**.
- **Gruppenkonto**: Jeder Punkt, den Sie dem Gruppenkonto zuordnen, erhöht den Punkteverdienst **bei allen Mitgliedern der Gruppe**, inklusive Ihnen, und zwar **um 0,4 Punkte**. Gleiches gilt,

wenn ein anderes Mitglied Ihrer Gruppe Punkte dem Gruppenkonto zuteilt: Jeder Punkt erhöht den Verdienst bei allen in der Gruppe um 0,4 Punkte.

Ihr persönlicher Verdienst pro Runde lässt sich nach diesen Regeln wie folgt zusammenfassen:

Ihr Verdienst pro Runde = Ihre Zuteilung zum privaten Konto + 0,4 x Summe der Beiträge aller Gruppenmitglieder zum Gruppenkonto

Bitte beachten Sie, dass Ihnen der Einfachheit halber auf dem Computerbildschirm nur **ein** Feld zum Eintragen von Punkten gezeigt wird. Dort tragen Sie ein, wie viele der 20 Punkte Sie dem **Gruppenkonto** zuweisen möchten. Alle übrigen Punkte, die Sie nicht dem Gruppenkonto zuteilen, werden dann automatisch dem privaten Konto zugeteilt.

Beispiel: Angenommen, Sie behalten 10 Punkte für das private Konto und tragen 10 Punkte in das Feld für das Gruppenkonto ein. Ferner sei angenommen, die anderen Gruppenmitglieder tragen ebenfalls zum Gruppenkonto bei. Nehmen wir an, insgesamt kommen auf dem Gruppenkonto 40 Punkte zusammen. Diese 40 Punkte bewirken eine Auszahlung von 0,4 x 40 = 16 Punkte für jeden in der Gruppe. Ihr eigener Verdienst beträgt also die 16 Punkte vom Gruppenkonto und die 10 Punkte auf Ihrem privaten Konto = 26 Punkte.

Die **Tabelle** auf Seite 3 gibt eine übersicht über Ihren Verdienst für einige beispielhafte Kombinationen wieder. Sie lesen die Tabelle wie folgt: **Zeilen** stehen für Beispiele, wie viele Punkte **Sie** dem Gruppenkonto zuteilen (also zwischen 0 und 20). **Spalten** stehen für Beispiele, wie viele Punkte die **anderen drei Gruppenmitglieder insgesamt** auf das Gruppenkonto einzahlen (also zwischen 0 und 60). Jede **Zelle** gibt Ihren persönlichen **Rundenverdienst** wieder, der sich wie oben beschrieben berechnet. Bitte vergewissern Sie sich jetzt, dass Sie die Tabelle vollständig verstehen. Zögern Sie nicht, die Hand zu heben, wenn Sie eine Frage haben.

Vor jeder neuen Runde werden Ihnen Informationen über die vergangene Runde angezeigt, und zwar über:

- Ihren **eigenen Beitrag** zum Gruppenkonto, den Sie in das Feld auf dem Bildschirm eingegeben haben,
- die **Summe der Punkte aller vier Gruppenmitglieder auf dem Gruppenkonto**,
- Ihren **Verdienst** in der vergangenen Runde, der sich aus diesen Entscheidungen ergibt.

Fragebogen

Das Experiment endet mit einem Fragebogen, währenddessen Ihre Auszahlung vorbereitet wird. Die Antworten auf Fragen im Fragebogen geben Sie am besten spontan, ohne lange und wiederholt darüber nachzudenken.

Danke für Ihre Teilnahme!

Tabelle: Ihr Verdienst einer Runde in Teil 2, für beispielhafte Kombinationen

Ihr Beitrag zum Gruppenkonto	Gesamter Beitrag der anderen drei Gruppenmitglieder zum Gruppenkonto								
	0	**3**	**6**	**15**	**30**	**45**	**54**	**57**	**60**
	Ihr Verdienst:								
0	20	21,2	22,4	26	32	38	41,6	42,8	44
1	19,4	20,6	21,8	25,4	31,4	37,4	41	42,2	43,4
2	18,8	20	21,2	24,8	30,8	36,8	40,4	41,6	42,8
5	17	18,2	19,4	23	29	35	38,6	39,8	41
10	14	15,2	16,4	20	26	32	35,6	36,8	38
15	11	12,2	13,4	17	23	29	32,6	33,8	35
18	9,2	10,4	11,6	15,2	21,2	27,2	30,8	32	33,2
19	8,6	9,8	11	14,6	20,6	26,6	30,2	31,4	32,6
20	8	9,2	10,4	14	20	26	29,6	30,8	32

Conclusion

As each article of this dissertation has already drawn conclusions with respect to its particular research question(s) and results, the purpose of this section is to make some concluding remarks regarding the common theme of this dissertation. Shared by all five analyses of the three experiments reported in this dissertation is their relation to the question of the determinants of private provision of public goods (PPPG). For the sake of summarizing, the determinants investigated in the five articles can be grouped into four categories: (A) sociodemographic characteristics, mostly elicited from experimental subjects by a questionnaire, (B) exogenous environmental conditions, either matched to subjects by experimental time and location in the field or administered as a treatment, (C) changes in the exogenous parameters or "rules" of the public good "game", all of them administered as between-subjects (BS) treatment conditions, and (D) determinants related to subjects' (endogenous) information status or expectations, again elicited through questionnaires. Table 1 summarizes the main results regarding these groups of determinants from the five articles. The table reports both the direction and significance levels of the effects but does not distinguish whether the effect on PPPG was at the extensive or intensive margin or both, and whether the public good in the experiment was real (experiment 1) or an experimental "group account" (experiments 2 and 3).

Our results on the effect of various observable determinants on contributions can help to target the "right" motives when extending the existing research in the area of *why* people voluntarily contribute. In the literature on PPPG, this question has been focal, and a lot of research has been undertaken in this area (see, e.g., Kolm and Ythier 2006). Nevertheless the discipline has not yet arrived at a unifying behavioral theory reconciling the neoclassical prediction and the empirical reality. Experimentally disentangling causal motivations that mediate observable changes in giving behavior induced by changes in

Table 1: Determinants of PPPG collected and analyzed in this dissertation

Determinant	Effect on PPPG[a]	Experiment	Remark
A. Sociodemographic characteristics[b]			
Female	++ / − − −	1,2	
Age	+ / ∘	1,2	
Education	+++	1,2	
Income	∘	1	
Children in HH	∘	1	
Rural residential environment[c]	+	2	
Eastern Germany	− − / ?	1,2	Significant only if younger (age < 27) subjects are excluded
B. Exogenous environmental conditions			
Ambient outdoor temperature	++	1	
Media coverage of real public good	∘	1	
Acute ambient noise	−	3	BS treatment effect
C. Exogenous "rules of the game"			
Price of providing a unit of PG	− − −	1	BS treatment effect
Visible, observed contribution	+	1	BS treatment effect
Group size	+++	2	BS treatment effect, pure public good, MPCR=0.3
D. Information / expectations			
Endogenous knowledge about real PG	+++	1	
Past/present negative contributions[b]	− − −	1	
Egoistic benefit expectations[b]	+++	1	
Altruistic benefit expectations[b]	+++	1	Estimates larger in size than for egoistic benefits[a]

Notes: Experiment 1 is the framed field experiment of Papers 1 to 3, experiment 2 is the artefactual field experiment of Paper 4, and experiment 3 is the lab experiment of Paper 5. +++ / − − − Significance up to 1% level. ++ / − − Significance up to 5% level. + / − Significance up to 10% level. ∘ Not significant. [a] Results jointly report the effect on giving at the extensive and intensive margin as well as to real (experiment 1) and "laboratory" (experiments 2 and 3) public goods. [b] Self-reported variable(s). [c] Variable constructed from self-reported ZIP (PLZ) codes.

other observable factors can be complex (Imai et al. 2011). Ideally, a successful research for mediators would identify a couple of main causal drivers of providing public goods within humans which in fact are the ones mediating the observed effects of all the identified determinants. Questions for mediators in the context of the determinants identified in our results would be, for example: What causes the observed differences between males and females in giving? Are "biological" differences in altruism over time responsible for an age effect, or are differences in the socialization of different age cohorts driving the effect, pointing to social or moral norms driving the effect? Do people who are more inclined to contribute move to rural areas, or does a rural environment nurture cooperation? If the latter dominates, is this based on social norms, potentially based on the dependency on reciprocity in rural areas, or is this a physiological effect on some intrinsic motivation, e.g. from a quieter environment?

To the policy maker (or fundraiser), who is less likely to be interested in causal mechanisms, our results suggest that targeting certain subgroups within the population is likely to have significant effects on the extent of voluntary giving to public goods. The educated Western German of the countryside may be such a stereotype, for example. However, some of our results caution against over-extrapolating the observed effects as some determinants are not consistent in their effect across experiments. This observation is most pronounced in the case of the difference between males and females. An obvious explanation is that the effect is contingent on the specific context or type of public good: While the public good in experiment 1 was real (climate change mitigation), the public good in experiments 2 and 3 was experimental "group accounts". This raises the question to which extent people perceive social dilemmas as social dilemmas, i.e. to which extent they consider the decision on whether or how much to contribute to some particular public good as a strategic environment of a multi-person prisoners' dilemma. Further research in this area is clearly warranted as it speaks to the external validity of many experimental results from public good games.

The same question extends to our results that changes in environmental conditions and the "rules of the game" may affect contributions. While there is evidence that some of these determinants matter and some not, the question arises which of these effects depend on the specific context (e.g. the observed temperature effect) and which do not.

Lastly, some of the most pronounced effects are correlations of contribution choices with variables related to subjects information status and their expectations and beliefs about the public good. Apart from pointing to possible mediators such as altruism or moral norms, this finding strengthens the role informational campaigns can play in determining contributions. However, all the pronounced effects of informational variables are based on endogenous information status. In contrast, media coverage as an exogenous informational variable does not affect contributing in the particular case of climate change mitigation. This points to important subtleties in the effects of providing, receiving, and acquiring information regarding public goods.

References

Imai, K., Keele, L., Tingley, D. and Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies, *American Political Science Review* **105**(4): 765–789.

Kolm, S.-C. and Ythier, J. M. (2006). *Handbook of the Economics of Giving, Altruism and Reciprocity*, Elsevier.