How much can we learn about voluntary climate action from behavior in public goods games?

Goeschl, Timo; Kettner, Sara Elisa; Lohse, Johannes; Schwieren, Christiane

DOI:
10.1016/j.ecolecon.2020.106591

Document Version
Peer reviewed version

Citation for published version (Harvard):
Goeschl, T, Kettner, SE, Lohse, J & Schwieren, C 2020, 'How much can we learn about voluntary climate action from behavior in public goods games?', *Ecological Economics*, vol. 171, 106591.
https://doi.org/10.1016/j.ecolecon.2020.106591

How much can we learn about voluntary climate action from behavior in public goods games?

Timo Goeschl[a] Sara Elisa Kettner[b] Johannes Lohse[c]
Christiane Schwieren[d]

February 5, 2020

Abstract

Evidence from public goods game experiments holds the promise of informing climate change policies. To fulfill this promise, such evidence needs to demonstrate generalizability to this specific policy context. This paper examines *whether* and *under which conditions* behavior in public goods games generalizes to decisions about voluntary climate actions. We observe each participant in two different decision tasks: a real giving task in which contributions are used to directly reduce CO_2 emissions and an abstract public goods game. Through treatment variations in this within-subjects design, we explore two factors that are candidates for affecting generalizability: the structural resemblance of contribution incentives between the tasks and the role of the subject pool, students and non-students. Our findings suggest that cooperation in public goods games is only weakly linked to voluntary climate actions and not in a uniform way. For a standard set of parameters, behavior in both tasks is uncorrelated. Greater structural resemblance of the public goods game with the context of climate change mitigation produces more sizable correlations, especially for student subjects.

Keywords: Climate Change Mitigation; Generalizability; Lab Experiments; Public Goods Game; Voluntary Cooperation

[a]Email: goeschl@uni-heidelberg.de. Postal address: Department of Economics, Heidelberg University, Bergheimer Str. 20, 69115 Heidelberg, Germany.

[b]Email: kettner@conpolicy.de. Postal address: ConPolicy - Institute for Consumer Policy, Friedrichsstr. 224, , 10969 Berlin, Germany

[c]Corresponding Author: Email: j.lohse@bham.ac.uk. Postal address: Department of Economics, JG Smith Building, Birmingham B15 2SB, UK

[d]Email: christiane.schwieren@awi.uni-heidelberg.de. Postal address: Department of Economics, Heidelberg University, Bergheimer Str. 58, 69115 Heidelberg, Germany

1 Introduction

It has been argued that findings from public goods game (PGG) experiments with student subjects can provide insights into the climate mitigation behavior of the general public and therefore lessons for climate policies (Shogren and Taylor, 2008; Venkatachalam, 2008; Brekke and Johansson-Stenman, 2008; Gowdy, 2008; Gsottbauer and van den Bergh, 2011; Carlsson and Johansson-Stenman, 2012). Along this line of reasoning, researchers have framed experimental studies on public good provision with reference to mitigation decisions or interpreted their outcomes explicitly in a climate policy context (e.g., Milinski et al., 2006, 2008; Tavoni et al., 2011; Brick and Visser, 2015; Vicens et al., 2018). Results of PGG experiments have been used to argue, for instance, that climate policy should showcase individuals that mitigate a lot (Milinski et al., 2006), should involve strong redistributive elements (Vicens et al., 2018), and ensure fairness in international climate negotiations (Brick and Visser, 2015).

The argument that PGG experimental evidence delivers actionable guidance for climate policy is appealing for several reasons, not least because PGG experiments – especially those with student subjects – make it possible to obtain causal evidence quickly and at a low cost. Whether such PGG experiments can truly provide the desired valuable policy insights, however, crucially depends on their *generalizability* (Levitt and List, 2007b). For informing climate policy, generalizability demands that generic behavior, based on observing student subjects in an abstract lab task, broadly transfers to the specific context of voluntary mitigation decisions by a general population. Whether and under which conditions the behavior of student subjects in a PGG experiment generalizes, not only at the aggregate but also at the individual level, to voluntary climate actions (VCA) by the general public is, at heart, an empirical question that can itself be answered through careful experimentation.[1]

In the present paper, we take two steps towards providing an answer to the question of how informative behavior in PGG experiments is for VCA, in particular by the general public. One step consists of testing for generalizability across experimental tasks: When we can observe subjects' choices in abstract PGG experiments and their choices in a concrete and consequential VCA experiment, how well do choices in the former predict choices in the latter? The other step consists of testing for generalizability across subject types: By observing the behavior of student and non-student samples in the different PGG and VCA experiments, we compare the aggregate choices in each task, as well as how well the generalizability of students choices across tasks extends to that of non-students. Jointly, these two steps not only inform about the degree to which behavior in PGG experiments of student subjects generalizes to the behavior of general population subjects in a VCA experiment. They also highlight whether generalizability is threatened more by the choice of experimental subjects, by experimental design features, or by a combination of the two.

Our paper relates to two strands in the literature. The first strand is motivated by a concern that subjects' behavior in abstract game forms under controlled conditions in the laboratory may

[1] Closely related to the concept of generalizability (Levitt and List, 2007b) is the concept of external validity (e.g., Torres-Guevara and Schlüter, 2016; Snowberg and Yariv, 2018). In fact, there is considerable overlap in terms of research questions and methods. For this paper, we adopt the concept of generalizability as a less stringent requirement of transferability of results beyond their immediate experimental context.

not generalize to individual behavior in context-rich situations outside the lab. As our review in the following section shows, such concerns have been voiced particularly in the context of whether social preferences elicited using standard experimental designs are predictive beyond the lab (Levitt and List, 2007a). This question is of obvious relevance for issues of public goods provision such as voluntary mitigation choices.[2] Evidence on generalizability in this context is mixed: The extent to which cooperation in PGG correlates with a broader set of pro-social preferences (Blanco et al., 2011; Peysakhovich et al., 2014) and, more importantly, the extent to which it generalizes to cooperative behavior beyond the laboratory (Benz and Meier, 2008; Laury and Taylor, 2008; de Oliveira et al., 2011; Voors et al., 2012; Torres-Guevara and Schlüter, 2016) varies substantially across studies. The other strand in the literature to which our paper relates has been comparing the behavior of student and non-student samples in experiments. Such comparisons have generated mixed findings on whether student behavior generalizes to non-student behavior (Belot et al., 2015; Snowberg and Yariv, 2018). Based on both strands, the generalizability of behavior of student subjects in the PGG towards voluntary climate actions by the general public can neither be ruled in nor out.

While ecological economists have uncovered evidence for limited generalizability in common-pool resource settings (Torres-Guevara and Schlüter, 2016), the question of generalizability in the climate context has not been addressed so far, despite reasons that generalizability may also run into problems in this context. The first reason is that the deliberately abstract format of the PGG does not capture the richer context of preferences (e.g., risk- or time-preferences), beliefs (e.g., regarding the expected damages from climate change), or attitudes (e.g., regarding the importance of pro-environmental behavior) that are likely to shape voluntary mitigation decisions. This problem may be remediable, however: The experimental paradigm of the PGG can accommodate considerable variation in design features. For instance, a greater resemblance to VCA decisions could be engineered by simple changes to design parameters such as group size or the productivity of the experimental public good. If such variations can capture most of the relevant drivers of mitigation decisions, then generalizability may be accomplishable even by maintaining an abstract setting. The second reason is that it is well known that student samples, which account for the majority of PGG evidence, share only a limited range of individual attributes such as age, cohort, and education with the general population - and differences in these attributes could matter significantly more in contexts such as climate change. For instance, willingness to pay (WTP) studies find a positive association between the WTP for climate actions and education (Diederich and Goeschl, 2014). As a result, the extent to which the behavior of students allows conclusions about the behaviour of non-students is a matter of ongoing discussion (Gächter et al., 2004; List, 2004; Carpenter et al., 2008; Thöni et al., 2012; Anderson et al., 2013; Falk et al., 2013; Belot et al., 2015). If present in the climate context, this problem is unlikely to be remediable without turning to a general population as the target sample.

[2] Levitt and List (2007a) describe several situational factors, present in a typical lab experiment, that might reduce its predictive power for field behavior. For instance, they discuss the extent of scrutiny, the activation of specific norms, or the context in which the decision is embedded as important shift parameters. Their concerns, arguably, carry more weight for experiments conducted in order to inform policymakers than for experiments that try to falsify a theory (Schram, 2005; Sturm and Weimann, 2006; Kessler and Vesterlund, 2015).

To examine whether these reasons threaten the generalizability of student behavior in PGG experiments, we undertake three distinct ventures in one experimental set-up. First, we examine whether estimates of generic cooperative preferences derived from behavior in a PGG experiment can explain a significant portion of individual mitigation behavior in a VCA experiment. Having explanatory power of sufficient size is an important prerequisite for a high level of generalizability (Al-Ubaydli and List, 2015). For this purpose, the researcher would ideally like to be able to observe the totality of an individual's mitigation behavior.[3] Following other examples in the literature (Benz and Meier, 2008; Laury and Taylor, 2008; de Oliveira et al., 2011; Voors et al., 2012; Galizzi and Navarro-Martínez, 2018), we approximate the ideal test by conducting a laboratory experiment in which we instead observe each participant in two contribution situations: A standard public goods game and a real giving task of VCA in which individual contributions are used to reduce CO_2 emissions. Importantly, as contributions in the real giving task are used to reduce real CO_2 emissions, this task captures more closely context-specific preferences that might motivate individual mitigation behavior. Such context-specific preferences might reflect strategic incentives, perceived individual returns, time and risk preferences, and pro-environmental preferences.

The second venture that we undertake is to examine within the same experimental set-up whether the design of abstract PGG experiments can be altered such that generalizability towards VCA increases. We do so by experimentally varying the structural parameters of the PGG, in particular group size, marginal per-capita return (MPCR), and payoff symmetry. The structural parameters of VCA are that the entirety of humanity is involved in the climate problem (large group size), that individual returns to VCA are very low (low MPCR), and that gains from emissions reductions are heterogeneous across the population (high payoff asymmetry). Systematically varying the structural parameters of the PGG task in our experiment allows us to test whether greater structural resemblance between incentives in the abstract PGG task and VCA enhances generalizability.

The third venture embedded in our set-up is a comparison of behavior across two distinct samples of subjects. One is the standard convenience sample of students. The other is a sample recruited from the general population. This third dimension allows us to test whether generalizability in the climate context hinges on subject pool effects.

Our results highlight that informing climate policies based on observing student subject behavior in abstract PGG experiments carries significant risks. The reason is that we find that behavior in PGG can, but in many configurations does not generalize to voluntary mitigation decisions. Research users, among them policy-makers, need to be aware that the potential for generalizability crucially depends on the way the PGG is designed and conducted. For a PGG using the standard parameters, the correlation between contributions in the PGG and the VCA task is small and insignificant. This result holds irrespective of the subject pool. A low correlation indicates that there exist (potentially several) idiosyncratic drivers of mitigation behavior

[3] Under ideal conditions, the researcher would observe two separate decisions by the same individual: Contribution choices in a standard PGG and revealed preferences for voluntary CO_2 mitigation in a field context. The latter would require observing the totality of economic decisions that potentially involve direct or indirect mitigation efforts of CO_2 emissions. In a fossil-fuel economy, this is true for almost all economic decisions. Accurate measurement of the aggregate pure mitigation effort at the level of the individual is therefore empirically daunting, particularly if this measurement should also be obtained in an unintrusive fashion.

that remain unobserved in standard PGG. On the other hand, when PGG parameters resemble more closely the incentive structure characteristics of voluntary climate change mitigation, correlations become more sizable, particularly among student subjects. By implementing simple design changes, some of the apparent differences in individual behavior therefore disappear. On these grounds, future laboratory experiments may be able to contribute to informing discussions about climate policy, provided they show awareness of these design issues.

The choice of the subject pool has more ambiguous implications. In line with previous results, we find that on average, non-students contribute more in both tasks. Climate policy proposals that are informed by PGG experiments with student subjects are therefore likely to overestimate the prevalence of selfish behavior in the general population and be too pessimistic about individuals' willingness to engage in VCA. However, the degree of generalizability is much lower within the more heterogeneous sample of non-students, as indicated by strongly reduced correlations. Under a fixed research budget, this means that the researcher has to trade-off representativeness against generalizability.

The remainder of the paper is organized in the following way: Section 2 discusses our research question in relation to the existing literature. In Section 3, we describe the experimental set-up and the characteristics of our subject pool. Section 4 contains the analyses and core results. Section 5 concludes with a discussion of our findings.

2 Related Literature

There is a growing literature that addresses the same basic question as our paper by investigating the relationship between contributions observed in a laboratory PGG and contributions to a naturally occurring public good.[4] As in our experiment, these studies largely lack a direct and unintrusive measure of cooperation in the field.[5] Instead, they observe contributions to a naturally occurring public good through eliciting choices in a modified dictator game (Eckel and Grossman, 1996). Benz and Meier (2008) investigate the correlation between students' charitable giving in a laboratory setting and their charitable giving in a university fund-raiser. Within a low-income neighborhood, de Oliveira et al. (2011) explore whether subjects who display other-regarding preferences in a linear public goods game also give to local charities. Voors et al. (2012) compare the behavior of subsistence farmers in a linear PGG to the amount they contribute to a real community public good. In the context of environmental goods, we

[4] A different, but related literature highlights that the same individual can behave quite differently even in related abstract social preference tasks, in which idiosyncratic motives should be largely absent and similar preferences should motivate individual behavior. These studies have analyzed how cooperation in public goods games corresponds to social preferences elicited in other abstract tasks and arrive, overall, at mixed results. Blanco et al. (2011) find that contributions made in a standard PGG are significantly correlated with responders' behavior in a sequential prisoners dilemma, but not to other-regarding choices made in ultimatum or dictator games. In an online experiment, Peysakhovich et al. (2014) find stronger evidence that an individual's propensity to contribute in a one-shot public goods game spills over to other abstract game formats. In a similar setting Capraro et al. (2014) find a sizable correlation between dictator game giving and cooperative choices in a one-shot, but not in an infinitely repeated prisoner's dilemma (Dreber et al., 2014). In Galizzi and Navarro-Martínez (2018) public goods game behavior is moderately, but significantly correlated with behavior in trust and dictator games, but not with helping or donation behavior in five different field situations which are randomly administered after the actual experimental sessions.

[5] A notable exception is Fehr and Leibbrandt (2011), in which the overexploitation of a fishery resource is related to behavior in a public good experiment.

are aware of two papers of close relevance to the question posed in this paper. Torres-Guevara and Schlüter (2016) investigate the predictive power of cooperation rates assessed in an abstract setting for the sustainable usage of an existing common pool resource, drawing on a sample of artisanal fishermen in Colombia. Closest to our research question and experimental set-up, Laury and Taylor (2008) investigate student behavior in a variety of the linear PGG and their contributions to a local environmental public good. These studies have brought forth mixed results: some of them find a significant correlation between contributions in the abstract and specific context (Benz and Meier, 2008), whereas others suggest a more moderate (Laury and Taylor, 2008; de Oliveira et al., 2011) or even insignificant (Voors et al., 2012; Torres-Guevara and Schlüter, 2016) relationship. In a comprehensive literature review, Galizzi and Navarro-Martínez (2018) similarly conclude that results vary greatly across studies according to their context (e.g., the real public good offered) and design (e.g., the subject pool under study or the experimental procedures used to assess generic cooperation rates).

In light of the literature reviewed above, the extent to which existing findings are transferable to the specific context of voluntary climate change mitigation is not clear. Several design differences plausibly limit transferability: First, all of the studies above use a particular local public good, while climate change mitigation is a global and intergenerational public good. Second, each of these four studies was conducted with a specific subject pool of either students or aid recipients. This puts into question whether they are sufficiently representative for reaching conclusions about the behavior of broader segments of the population relevant in a climate policy context. Third, each of these studies - except for Laury and Taylor (2008) - uses one specific set of parameters when assessing generic preferences for cooperation within a PGG.

These plausible limitations to transferability inform important design choices in our experiment, with a view to answering the questions raised in the introduction. Our design employs a task directly linked to the reduction of CO_2 emissions. Furthermore, we use a unified design in which we observe the behavior of two different subject pools: One convenience sample of students and a group of subjects that more closely covers demographic attributes of everyday decision-makers. Finally, our design identifies to what degree the correlation between the two tasks depends on the parameter choice in the PGG. These design elements are well suited to provide answers to our research questions with their focus on generalizability to voluntary mitigation.[6]

3 Experimental design and implementation

Questions of generalizability from one experimental task to another are typically addressed in a within-subjects design in which the same participants are asked to complete two or more tasks within the same experimental session.[7] (Benz and Meier, 2008; Laury and Taylor, 2008;

[6]Note, however, that the design is explicitly not intended to resolve the broader controversy (Levitt and List, 2007a, 2009; Falk and Heckman, 2009; Kessler and Vesterlund, 2015; Camerer, 2015) on whether social-preferences assessed in abstract lab tasks are generally externally valid, in any chosen context.

[7]Including several tasks with a moral component within the same session opens up the possibility that some individual choices reflect concerns for 'moral licensing' or 'moral consistency'. The existing psychological and economic literature finds mixed evidence for the existence of both phenomena (Blanken et al., 2014; Gallier et al., 2017; Urban et al., 2019). A design alternative in which both tasks are separated by a longer gap of e.g. several weeks may avoid that individual choices are guided by such considerations. Separating tasks by a longer temporal gap, however, makes it harder to keep other important drivers of behavior (such as beliefs, incomes,

de Oliveira et al., 2011; Blanco et al., 2011; Voors et al., 2012; Peysakhovich et al., 2014). Therefore, we observe for each subject choices in a context-free decision task and in a task related to climate change mitigation. Participants are informed in the initial instructions that there would be several consecutive tasks in which they could earn real money. In *Task I*, we assess individual contributions to the real public good of climate change mitigation. In the subsequent *Task II*, subjects take ten one-shot public goods decisions in which we vary experimental parameters along three dimensions (Goeree et al., 2002).[8] In the following, we describe each of the decision tasks in more detail.

3.1 Task I: The real contribution task

To observe contributions to climate change mitigation in a lab setting, we employ a real giving task (Eckel and Grossman, 1996) in which individual contributions are used to reduce global CO_2 emissions. The transparent and verifiable reduction is executed by retiring emission permits from the EU ETS (Löschel et al., 2013; Diederich and Goeschl, 2014).[9] Before reaching the first decision screen, subjects were informed that they had received 10€ as a reward for taking part in the experiment. Subsequently, they were given the choice to contribute any share of these 10€ (in steps of 1€) towards a common account that would be used by the experimenters to reduce global CO_2 emissions.

Before subjects could select their preferred contribution level on the decision screen, they received a short and neutral description of the public good on an information screen. Thereby we ensured that each subject would have the same basic level of information about greenhouse gas emissions and the procedure by which the emission reductions would be executed by the experimenters. They were also informed about the amount of CO_2 that could be reduced for each 1€-contribution. To render the choice tangible, the instructions related this amount to every-day consumption decisions, expressed in terms of two common activities (car travel; use of a personal computer) and the average CO_2 emissions of a German citizen. The instructions also confirmed the public goods character of CO_2 mitigation by explaining that the particular location of CO_2 reductions would not affect the mitigation of global climate change and by pointing out the temporal delay between the reduction of CO_2 emissions in the atmosphere and the resulting beneficial impacts on climate change.

To avoid potential anchoring effects we made sure that no examples of provision levels were given to subjects before they could select their own contribution. Lastly, participants were informed that documentation from the German Emission Trading Registry would be publicly posted immediately following the last experimental session that would certify that their contributions had been used for the verified emissions reductions.

and information) constant between both points of measurement and may additionally result in selective attrition. This may pose an equally strong or even more serious threat to a valid test of generalizability.

[8] All subjects in the experiment completed the two tasks in this order. We do not explicitly account for order effects, as Laury and Taylor (2008) find no evidence for such effects in a setting comparable to ours. Furthermore, in a small scale pilot of our study (N=30) we find no evidence for order effects.

[9] Obviously, outcomes from Task I are only a proxy for actual field behavior. But they seem to capture, at least to some degree, environmental preferences, since they are significantly correlated with stated donations to environmental organizations.

3.2 Task II: The laboratory public goods game

It is well known that structural parameters of the PGG such as the group size, the marginal per capita return (MPCR), or the symmetry of payoffs affect the average rate of cooperation (e.g., Isaac and Walker, 1988; Goeree et al., 2002; Diederich et al., 2016). We hypothesize on this basis that the choice of these parameters may also affect the degree of generalizability. To test this proposition, we employed a popular variant of the standard public goods game (Goeree et al., 2002) in which the researcher observes each subject in ten separate PGGs that differ structurally. Specifically, subjects were anonymously and randomly matched into groups (large or small). They then completed ten separate one-shot contribution decisions without feedback, displayed on one single decision screen in the same order.[10] The task in each of these decisions is the same: Participants choose how many tokens from an initial endowment they want to invest in a public account. The public account in every decision produced payoffs determined by a distinct combination of MPCR, group size, and payoff symmetry. Table 1 summarizes the ten decisions. In the 'benchmark' or 'reference' case (Decision f), we set the parameters to those used in most existing public good experiments: The group of participants is small, with three members, the payoff structure for investments in the experimental public good is symmetric across participants, and the MPCR is 0.4. The remaining nine decisions capture situations with larger or smaller structural resemblance to contribution incentives present in voluntary mitigation decisions. This is achieved by varying the group size, the MPCR, and the symmetry of payoffs. Decisions a-d feature parameters that structurally resemble those for voluntary mitigation decisions (small MPCR, larger group size) more closely than those of the benchmark decision f and decisions g-j.[11]

All ten decisions have in common that they place experimental participants in a classic public goods dilemma, in which there are no individual monetary incentives to contribute to the public account, while the group as a whole benefits from contributions: per decision, subjects are endowed (ω) with 20 tokens and each token they keep in their private account is worth $v = 20$ cents to themselves. Their contributions from this endowment to the public account (x) earn subjects an internal return of m_t^{int}, which varies across decisions. Tokens invested in the public account additionally yield an external return (m_t^{ext}) that each of the remaining members in their group receives.[12] For all ten decisions, the internal return is lower than 20 cents, such that contributing to the public account is not profitable from the perspective of an individual participant. If all subjects contribute their full endowment to the public account, the sum of earnings available to all group members is maximized in each decision, since $(N-1)*m_t^{ext}+m_t^{int}$ is always larger than 20. Refraining from free-riding and contributing to the public good is thus

[10] This screen also contained two additional decisions, not analyzed in this paper. These decisions only served as a robustness check, as they used parameters for which there was no conflict between individual and group interest, and hence, did not resemble a standard public goods problem.

[11] The emphasis here is on structural resemblance. Numerically, of course, the largest feasible group size in a typical lab experiment is still much smaller than the number of beneficiaries of climate change mitigation. The largest group we observe consists of all participants present in a given session, which were either 12 or 15. As a consequence, the lowest MPCR feasible under this constraint is, arguably, still far higher than the potential MPCR from avoiding 1 Ton of CO_2.

[12] Jointly, the internal and external return determine the MPCR, which can be calculated by the following formula: $\frac{1}{Nv}(m_t^{int} + (N-1)m_t^{ext})$. Separating the two returns allows for having decisions with asymmetric payoffs. In cases where $m_t^{int} = m_t^{ext}$, our setup is fully equivalent to the linear PGG.

commonly seen as an expression of cooperative behavior. The general payoff structure for individual i can be summarized as:

$$\pi_{it} = v(\omega - x_{it}) + m_t^{int} x_{it} + m_t^{ext} \sum_{j}^{N_t - 1} x_{jt}; \forall i = 1, ..., 12/15; \forall t = 1, ..., 10 \quad (1)$$

where t is a subscript denoting each decision and x_{it} is individuals i's contribution to the public account. N_t denotes the number of subjects within a group.

Table 1: Parameterization of the 10 PGG Decisions

Decision	Group Size (N)	Internal Return (m_t^{int})	External Return (m_t^{ext})	MPCR	Symmetry
a	12/15	2	2	0.10	Symmetric
b	12/15	3	2	0.10	Advantageous Asymmetric
c	12/15	2	3	0.15	Disadvantageous Asymmetric
d	12/15	4	4	0.20	Symmetric
e	3	8	6	0.33	Advantageous Asymmetric
f	3	8	8	0.4	Symmetric
g	12/15	2	9	0.42	Disadvantageous Asymmetric
h	3	12	8	0.46	Advantageous Asymmetric
i	3	8	12	0.53	Disadvantageous Asymmetric
j	3	16	16	0.80	Symmetric

Notes: This table shows the parameters used in decisions a-j. Internal and external returns are displayed as Eurocent per token contributed to the public account. Decision f is used and marked as reference case, as it is characterized by a combination of parameters that is common in most public goods experiments.

Our elicitation procedures follow the standards in the PGG literature to account for several known concerns: First, to minimize potential bias due to confusion (Houser and Kurzban, 2002; Ferraro and Vossler, 2010), subjects had to go through hypothetical payoff calculations for themselves and other group members, before entering the decision screen. In these calculations, there was no pre-specified contribution level to avoid anchoring effects. Second, at the end of the experiment, one decision was picked randomly with equal probabilities and paid out to the participants. This randomization of payoffs (Starmer and Sugden, 1991) has the advantage that subjects have no incentive to condition their behavior in a given decision on their other choices.

3.3 Recruitment and sample characteristics

Participants were recruited from two distinct pools. We compare students to non-students to analyze, whether the prior focus on student subjects influences the conclusions that can be drawn from existing experiments. To recruit from the general population, we used advertisements in two different local newspapers.[13] As a further recruitment tool, notices about the experiment were posted in all neighborhoods and public places of the city of Heidelberg. Prospective participants contacted a research assistant for further information and were invited to a session.[14] The student sample was recruited from the standard subject pool using ORSEE (Greiner, 2015).

[13]The *"Rhein-Neckar-Zeitung"* is sold for 1,40 € and has a daily readership of 88.649 within the Heidelberg region. The "Wochen-Kurier" is distributed for free to all households in the Heidelberg region with a run of 74.000 copies.

[14]The research assistant assured that subjects would be able to use a computer. The response rate to the different recruitment methods was comparable and no significant differences can be found with respect to demographic attributes or behavior.

To keep the two distinct subject pools comparable in terms of their experience with economic experiments, only subjects who had not taken part in previous studies were included in the experiment. For the same reason, both pools were incentivized for their participation in the same way. Naturally, both subject pools consist of self-selected subjects. While this is standard practice in almost all economic experiments, there are some concerns that the use of self-selected subjects could overestimate the prevalence of other-regarding preferences (Levitt and List, 2007a). Empirically, these concerns have not been confirmed, so far (Anderson et al., 2013; Exadaktylos et al., 2013).

Overall, we recruit 135 subjects for the experiment: 92 from the general population and 43 from the student population. Table 2 gives an overview over the demographic attributes used in the analyses below. The two samples differ significantly with respect to socio-demographics directly related to the student status such as age, income, assets, or number of children. Apart from that, the two pools do not differ significantly regarding their education, stated risk aversion, or stated concern about the consequences of climate change. Obviously, despite being more diverse, the non-student participants in our study are also a convenience sample, but one with a somewhat higher resemblance to the average population.

Table 2: Socio-demographic attributes of different subpopulations

Demographics	**Total**	**Student**	**Non-Student**
	N=135	N=43	N=92
Age (Years)	40.92 (18.76)	22.84 (3.01)	49.37 (16.96)
Gender (1=male)	0.38	0.42	0.36
Individual Net Income (Euro)	1050.83 (902.74)	613.16 (228.59)	1253.66 (1020.73)
Assets (1=Yes)	0.25	0.02	0.36
Education (Years)	14.23 (2.67)	13.86 (1.95)	14.40 (2.94)
Household Size (#)	2.02 (1.44)	1.86 (1.22)	2.10 (1.54)
Has Children (1 = yes)	0.39	0.09	0.53
Stated Risk Aversion (Scale 1 - 11)	4.31 (2.72)	4.28 (2.72)	4.33 (2.73)
Concern Climate Change (Scale 1-7)	5.13 (1.77)	5.04 (1.57)	5.17 (1.87)

Notes: Income is self reported. Assets are coded as a dummy variable that takes the value of one if subjects state that they own either a flat or a house. Risk aversion is self-reported based on a question adapted from the German social survey (G-SOEP) ("How do you see yourself: are you, in general, a person fully prepared to take risks or do you try to avoid taking risks?"). Concerns about climate change are assessed by a questionnaire item ("On a scale of 1-7: How concerned are you about the consequences of climate change")

3.4 Experimental procedures

All ten sessions took place at the Heidelberg University Economics Computer Lab using z-Tree (Fischbacher, 2007). Most experimental sessions were run with a mix of student and non-student subjects participating in the same sessions. Two sessions were run with only student subjects. There were either 12 or 15 participants per session. At the beginning of a session, participants were seated at one of the available computer terminals, separated by a divider. A printed version of instructions explaining general procedures was handed out and read to subjects before they could begin with the actual decision tasks. All other instructions were

fully computerized. Communication between participants was not allowed at any point of the experiment, while questions addressed at the experimenter were answered quietly. All sessions were conducted under full anonymity. Furthermore, communication before the experiment was held at a minimum due to a separate check-in room that reduced common waiting times. In the check-in room subjects also generated a personal code. They were informed up-front that this personal code had the purpose to guarantee their anonymity during the experiment and anonymous payment at the end of a session: Experimenters provided sealed envelopes with earning receipts, only distinguishable by the subjects' personal code. The payment itself was conducted in a different room by a research assistant who was not present at any time of the experimental sessions. With this payment procedure, subjects could be assured that their overall earnings and identity would not be revealed to the experimenter at the end of the session. Sessions lasted around 75 minutes. The earnings of the average subject were 17.65€ and ranged from 2.68€ to 26.00€ [15].

4 Results

4.1 Observed behavior

Figure 1 gives a first overview over the distribution of contributions in Task I and Task II. The box-plots in the top panel show the fraction of the initial endowment contributed to climate change mitigation during Task I separately for the two different subject pools. The two diagrams in the bottom panel contain information on contribution behavior in Task II. Each box summarizes data for one of the ten distinct public good decisions. In the left diagram, we show data for student subjects and in the right one data for non-students. The benchmark case (Decision f) is depicted in a different color.

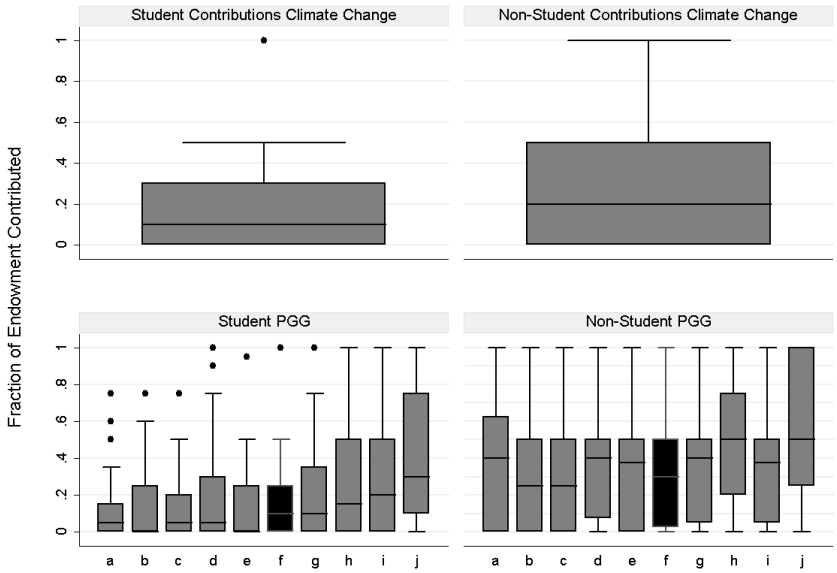
Median and mean contributions are positive in both tasks and for most parameters values in Task II, contributions in Task I and Task II fall into a similar range.[16] Overall, average contributions in Task I are slightly lower than in Task II, especially for high MPCR decisions.

In line with previous findings (Gächter et al., 2004; List, 2004; Carpenter et al., 2008; Thöni et al., 2012; Anderson et al., 2013; Falk et al., 2013; Belot et al., 2015), student subjects contribute a lower fraction of their initial endowment. Both for the abstract public good decisions in Task II (Mann-Whitney Rank-Sum Test: $p < 0.05$ for each decision) and contributions to climate change mitigation in Task I (Mann-Whitney Rank-Sum Test: $p < 0.05$) this difference is statistically significant. Furthermore, in both tasks, a more compressed interquartile range suggest that students' contributions are less dispersed. This observation is also supported by significance tests, which reject the hypothesis of equal variances both for average contributions in Task II (Levene's Robust Test; $p < 0.05$) and contributions in Task I (Levene's Robust Test; $p < 0.001$). On an aggregate level, the observation of significant subject pool differences implies that existing evidence, based on student samples, underestimates the willingness to contribute to public goods in a larger population as well as the degree of heterogeneity in contribution

[15] This value includes earnings from incentivized follow-up questions that are not part of the analysis.

[16] This observation is also supported by non-parametric significance tests (Sign Rank Test: $p < 0.05$) that find significant differences between the tasks for only two out of ten decisions.

Figure 1: Box-plots of contributions across tasks and subject pools



Notes: The top row shows the fraction of endowment contributed to climate change mitigation in the real giving task. The bottom row displays for each decision in the PGG the fraction of endowment contributed to the public account. The black line indicates median contributions. The lower and upper quartiles are marked by the gray box and whiskers are used to display values within 1.5 times the interquartile range. Outliers from this range are displayed as a dot.

behaviour. This applies both in the general context of abstract PGG and in the specific context of VCA.

In Task II, the contribution average varies substantially across decisions a-j. In line with previous findings, contributions increase with rising returns from the public good (Goeree et al., 2002). This positive relationship is more pronounced for students than for non-students. Regression results[17] confirm that the fraction of endowment contributed increases significantly with group size ($\beta_1 = 0.021$; $p < 0.001$) and internal ($\beta_2 = 0.029$; $p < 0.001$) or external returns ($\beta_3 = 0.013$; $p < 0.001$). The observation that behavior in Task II depends on the choice of parameters provides a first indication that this design choice could also influence the degree of generalizability from one task to another.

4.2 Individual Behavior: The role of experimental parameters

In this section, we study behavior at the individual level to analyze whether and under which conditions PGG experiments capture the main motivational drivers underlying voluntarily carbon emissions reductions, as observed in the real giving task. We answer these two related questions by successively exploring the within-subjects relationship between behavior in Task I and Task II at different levels of aggregation across individuals and Task II decisions. At each

[17]We estimate a random effects tobit model controlling for the student status and the set of demographic attributes listed in table 2. Full results are shown in the Appendix table 8.

of these levels, a high correlation would suggest that contextual factors play a negligible role and behavior in both tasks is driven by generic preferences that favor cooperation.[18]

Result 1: ***There is no significant correlation between average contributions in the abstract public goods game and contributions to the real public good of climate change mitigation.***

For a simplified first analysis of the relationship between the two tasks, we follow Laury and Taylor (2008) and initially ignore the variation of parameters between the different decisions of Task II. To broadly summarize contribution behavior, we calculate the mean over the ten distinct public good decisions ($\frac{1}{T}\sum_{t}^{T=10} x_{it}$). Across all decisions, the average participant contributed 33.85 percent (Median: 32 percent) of his initial endowment to the public account. This average value is close to the cooperation rate (29 percent) reported in Laury and Taylor (2008), who use a similar PGG design. In comparison, average contributions to climate change mitigation in Task I are only slightly lower at 27.48 percent (Median: 10 percent).

Similar average behavior across tasks need not reflect similar individual behavior. This is, in fact, the main message of figure 2. It shows a bubble plot of realized choices, with the percentage of endowment spent by each individual across all decisions in Task II on the x-axis and that spent in Task I on the y-axis. Visual inspection of the bubble plot does not hint at an association between the size of contributions in the two tasks. The same conclusion arises when employing a relative instead of the absolute scale of contributions: For no more than a quarter of participants do contributions fall into the same quintile in both tasks.

The largest overlap can be found within the bottom quintile, a result mostly driven by consistent free-riders that are the focus of result 3. As we will discuss in more detail below, there is some evidence for consistent free-riding at the individual level, which hints at a higher level of generalizability for extensive margin decisions.

The descriptive results are corroborated by the small and insignificant correlation between contributions in Task I and average contributions in Task II ($r = 0.1303$; $p = 0.132$). In contrast to Laury and Taylor (2008), therefore, behavior in the two distinct tasks in our experiment is only loosely related when the analysis relies on the average decision in Task II.

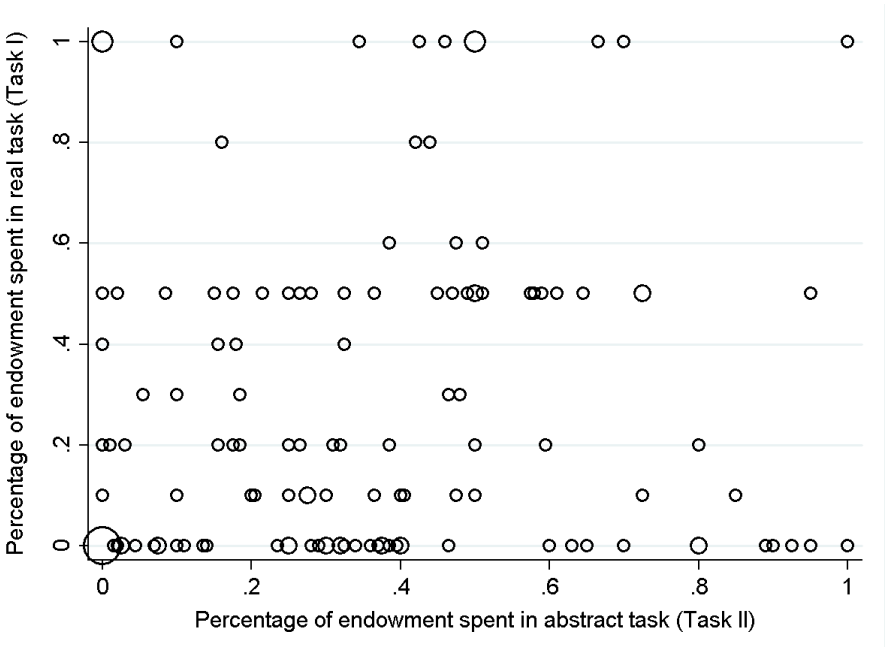
Result 2: ***Correlations are higher when the MPCR in Task II is low, groupsize is large, or payoffs are asymmetric.***

We now move on to explore the correlation structure at a lower level of aggregation of Task II decisions. Thereby we aim to assess how changes in the incentive structure across the ten PGG decisions affect the correlation between contributions made in Task II and Task I. For each decision, table 3 displays the corresponding correlation coefficients for the pooled sample of students and non-students.

Our analysis proceeds in two steps. We first examine the results for decision f. By the choice of parameters (Columns 1-3), this benchmark case is representative for standard PGGs. Therefore,

[18] All results presented in this and later sections hold both when analyzing the full sample and when discarding of eight subjects who have stated that their confidence in the existence of climate change is low.

Figure 2: Scatterplot of average contributions in the PGG and real giving task.



*Notes:*Bubble plot with frequency weights. The size of the bubbles is proportional to the frequency of a pair of contribution choices.

decision f is most informative regarding the question to what degree findings from the existing PGG literature readily transfer to the context of climate change. Comparing Task I and decision f of Task II, we find that behavior in the two tasks is not significantly correlated ($r = 0.1404; p = 0.1043$). This cautions against immediate transferability from PGG results to the climate policy context.

As a second step, we turn to the nine other decisions of Task II. Table 3 reports on the correlations. We now see that the relationship between contributions in Task I and Task II strengthens slightly for those Task II decisions that structurally resemble voluntary mitigation decisions: When the MPCR is lower and groups larger than in the benchmark case, we find contribution behavior that is significantly correlated across tasks. The highest significant correlation is reported for decision c, in which there was a low MPCR, a high group size, and an asymmetry of payoffs.[19] Conversely, for those decisions in which the MPCR increases relative to the benchmark case, correlation coefficients drop to a highly insignificant size. Taken together, this decision-wise analysis raises the possibility that simple adjustments in experimental parameters of the PGG to structurally resemble the specific choice context can make an important contribution towards generalizability.

To further evaluate the potential for generalizability, we now turn to the size of the significant correlation coefficients in table 3. Interpreting their strength requires some point of reference. We propose two reference categories: Correlations between PGG contributions and other abstract tasks that elicit social preferences and pairwise correlations across Task II decisions. The

[19]These findings continue to hold when we adjust p-values to address concerns regarding multiple testing. We employ the method of Dubey, which accounts for the fact that behavior in Task II is highly correlated across decisions. A detailed description of this method can be found in Sankoh et al. (1997)

Table 3: Decision-wise correlations between Task I and Task II

(0) Decision	(1) Group Size	(2) Symmetic	(3) MPCR	(4) Correlation Pooled Sample
a	Large	Yes	0.1	0.0985
b	Large	No	0.1	0.1822**
c	Large	No	0.15	0.2003**
d	Large	Yes	0.2	0.0737
e	Small	No	0.33	0.1713**
f	Small	Yes	0.4	0.1404
g	Large	No	0.42	0.0446
h	Small	No	0.46	0.0956
i	Small	No	0.53	0.0042
j	Small	Yes	0.8	0.0491

Notes: Decision f constitutes the benchmark case.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

first is a plausible upper limit for the size of correlations between Task I and Task II contributions since behavior in structurally similar games (e.g., a public goods game and a prisoner's dilemma) should be more highly correlated than that across structurally less similar decisions. Based on the results of the literature reviewed in Section 2, we find that the degree of generalizability from Task II to Task I is not smaller than that of PGG contributions to behavior in several other context-free social preference tasks. The significant correlations in table 3 squarely fall into the range $[r = 0.07; r = 0.41]$ reported in Blanco et al. (2011) and Peysakhovich et al. (2014).[20]

The second reference category, pairwise correlations across single decisions of Task II, relies on data generated by our own experiment and is a more restrictive measure. With the general task structure constant within that task, all variance in individual behavior across single decisions should only reflect changes in experimental parameters. Comparing correlations, we find that the relationship between Task II and Task I is much weaker than that between decisions under changing contribution incentives within Task II. Overall, subjects behave highly consistently across all ten PGG decisions (Cronbach's $\alpha = 0.94$) and correlations between single pairs of decisions range from $r = 0.43$ to $r = 0.85$.[21] Even when contribution incentives strongly differ as, e.g., between decisions b and j, the respective correlation coefficient is larger than any correlation shown in table 3. This apparent difference in size is further corroborated by formal statistical testing: a test for correlated correlation coefficients, as described in Steiger (1980) and Meng et al. (1992), shows that even the highest observed correlation between Task I and Task II (Decision c) is significantly smaller than any correlation observed across different decisions of Task II.

There are at least two potential explanations for the moderate size of correlations in table 3.

[20] The fact, that even for these more comparable contribution tasks some correlations are weak to negligible mirrors findings from social psychology (Ross and Nisbett, 2011) which underline that individual behavior is often strongly influenced by situational factors and only to a limited degree attributable to stable traits.

[21] A full correlation table can be found in the Appendix table 6.

One is that even the MPCRs in decisions a-d are not sufficiently low to reflect the actual incentives underlying voluntary climate change mitigation efforts in Task I. If so, participants would see Task I and Task II as generally equivalent and the differences in individual behavior between tasks would solely reflect differences in the experimental parameters. In light of the high behavioral consistency throughout Task II, despite substantial parameters changes, such reasoning can only provide a partial explanation of the moderate correlations between tasks. Another potential explanation is that context-specific factors influence individual behavior beyond a generic preference for cooperation. This reasoning is supported by the observation that even when the same participant faces very similar contribution conditions (i.e., sharing money with fellow students in a PGG and a sequential prisoners dilemma), there is only limited evidence for identical behavior at the individual level (Blanco et al., 2011).

Result 3: ***Extensive-margin behavior generalizes better than average behavior. A variation of experimental parameters has little impact on the correlation between free-riding in Tasks I and II.***

So far, we have analyzed behavioral consistency based on comparisons between the amounts contributed to the respective public goods. There is reason to believe, however, that extensive-margin decisions (whether or not to contribute at all) could be determined by different factors than the subsequent decision about the size of the contribution (Bergstrom et al., 1986; Smith et al., 1995; Kotchen and Moore, 2007). If so, the previous analysis could have overlooked an aspect of Task II that indeed generalizes to Task I. We, therefore, repeat the main steps of the previous analysis, now examining extensive-margin behavior.

A first, rough summary measure of the extensive margin is the percentage of decisions in which subjects contribute zero tokens in Task II. Based on this measure, 12.6 percent of subjects are categorized as strict free-riders because they never contribute to the public account. By comparison, 39.3 percent of subjects do not contribute to the public good of climate change mitigation in Task I. While these mean rates of free-riding differ substantially, we now find evidence for similar behavior at the individual level: Free-riding in the two tasks is correlated in a weakly significant way ($r_s = 0.1521$; $p = 0.0783$) when looking at all Task II decisions. There, 59 percent of strict free-riders also do not contribute in the mitigation task. The evidence becomes stronger when we look at distinct decisions within Task II. For the benchmark case, we find a significant correlation ($r_s = 0.1992$; $p < 0.05$) between individual free-riding behavior in decision f and in the mitigation task. For eight out of ten decisions there is a significant ($p < 0.05$) positive correlation in the narrow range from $r_s = 0.1905$ to $r_s = 0.2573$. Extensive margin behavior therefore generalized better, particularly for student subjects. The smallest insignificant correlation $r_s = 0.1153$ is again found in decision j which is characterized by the highest MPCR. Beyond this, however, there are no clear patterns connecting structural parameters and extensive-margin behavior.[22]

[22] A full table containing decision-wise correlations for free-riding can be found in the Appendix table 7.

4.3 The role of subject pool

A considerable number of studies have examined whether conducting experiments with a convenience sample of students affects the conclusions that can be drawn from economic experiments on social preferences (Gächter et al., 2004; List, 2004; Carpenter et al., 2008; Thöni et al., 2012; Anderson et al., 2013; Falk et al., 2013; Belot et al., 2015). The main concern is that students share only a limited range of individual attributes with the general population and, hence, could lack an important determinant of population behavior. It is subject to an ongoing discussion whether this concern mainly applies to level effects (e.g., in our case the size of contributions) or also to treatment effects (Harrison and List, 2004). Figure 1 clearly shows that the average student contributes significantly less in both tasks than the average non-student. Thus, our results conform to prior evidence that the behavior of students can be seen as a lower bound for the extent of pro-sociality one can expect among a more heterogeneous population. But does this significant level effect also imply that more could be learned about voluntary mitigation decisions from conducting a conventional PGG experiment with participants from a more diverse, and therefore more policy-relevant, study population? This would only be the case if behavior from PGGs transferred equally well to the mitigation context for students and non-students. The mixed results of the studies reviewed in Section 2 raise the possibility that this is not necessarily the case. For instance, some of the studies - especially those drawing on student subjects (Laury and Taylor, 2008; Benz and Meier, 2008) - have found significant correlations while studies conducted among a more diverse population (Voors et al., 2012) have not detected a significant relationship. Yet, as each of these studies observes contributions to a specific real public good, it is unclear whether their opposing results indeed arise from systematic differences between their respective subject pools. By contrast, we observe participants drawn from two distinct subject pools interacting with the same public good. Hence, we can analyze if correlations differ between those two subject pools.

Result 4: ***For student subjects, behavior in the PGG is more strongly correlated with behavior in the real giving task than for non-student subjects.***

When breaking down our prior analysis by student status, we find that the results reported above are mainly driven by the consistent choices of students. The correlation between average contributions in the PGG and contributions in Task I is slightly larger, yet still insignificant, for students ($r = 0.1531$; $p = 0.3288$). For non-students this correlation is negligible ($r = 0.0312$; $p = 0.7196$). As shown in table 4, this disparity is not driven by a single PGG decision. Instead, irrespective of the parametrization, for non-students all correlations are very low.

For students, however, there are significant correlations for some of the decisions in Task II. The choice of experimental parameters again influences the strength of these correlations. Only when the MPCR is smaller or the group size is larger than in the benchmark case of decision f, correlations are sizable. This difference between subject pools is robust to accounting for the higher demographic heterogeneity among non-student subjects. By calculating partial correlation coefficients, which hold constant the set of observed characteristics contained in table 2,

Table 4: Decision-wise correlations between Task I and Task II

(0) Decision	(1) Group Size	(2) Symmetric	(3) MPCR	(4) Correlation Non-Students	(5) Correlation Students
a	Large	Yes	0.1	0.0027	0.1689
b	Large	No	0.1	0.1081	0.3723**
c	Large	No	0.15	0.1319	0.3516**
d	Large	Yes	0.2	-0.0184	0.2939*
e	Small	No	0.33	0.0906	0.2964*
f	Small	Yes	0.4	0.0827	0.1455
g	Large	No	0.42	-0.0074	0.0570
h	Small	No	0.46	0.0242	0.1880
i	Small	No	0.53	-0.0452	0.1308
j	Small	Yes	0.8	-0.0719	0.1376

Notes: Decision f is the benchmark case. For student subjects we exclude one apparent outlier shown in figure 1. Including this outlier reduces correlation in size.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

we still find significant correlations only for student subjects.[23] Importantly, this difference is not driven by a higher level of confusion about the PGG task among non-student subjects. In a comprehension check administered before the PGG task, students (9.3 percent) display a similar frequency of incorrect responses as non-students (15.5 percent).

An additional analysis of free-riding behavior mirrors these findings. Only students display a (borderline) significant correlation when averaging over all ten decisions ($r_s = 0.2967; p = 0.0534$) of the PGG. Students who free-ride in Task I, on average contribute a significantly smaller fraction of their endowment in Task II (13.35 percent vs. 27.05 percent; Mann-Whitney Rank-Sum Test: $p = 0.01$). These results do not carry over to non-students. For them, the correlation between average free-riding in the abstract task and contributing zero in the real contribution task is negligible ($r_s = 0.0511$; $p = 0.6287$). Similarly, free-riding in the real contribution task is unrelated to average contributions in Task II. A decision-wise analysis of free-riding retains the previous result that the correlation structure is largely unaffected by the choice of parameters. For students there is a significant correlation for almost every decision ($r_s = 0.28$ to $r_s = 0.39$), while non-students reveal no significant correlation for any single decision.[24] In sum, mirroring result 3, free-riding appears to be more generalizable than average behavior; especially among student subjects. This also implies that the more sizable correlation coefficients for students are partly driven by consistent free-riders. Removing them from the analysis shown in Table 4 reduces correlation coefficients for decisions (b) - (e) in size and significance ($r_b = 0.2324$; $r_e = 0.1480$).

4.4 The joint role of task format and individual characteristics

The sections above have highlighted how both the experimental parameters in the PGG and the choice of the subject pool can influence the degree to which results on contribution behavior

[23] Alternative robustness checks yield equivalent results. In a SURE framework, using the same demographic controls, Breusch-Pagan tests reject the hypothesis that residuals are independent for three out of four decisions shown to be significantly correlated in table 4 for student subjects. For non-students this hypothesis cannot be rejected for any decision.

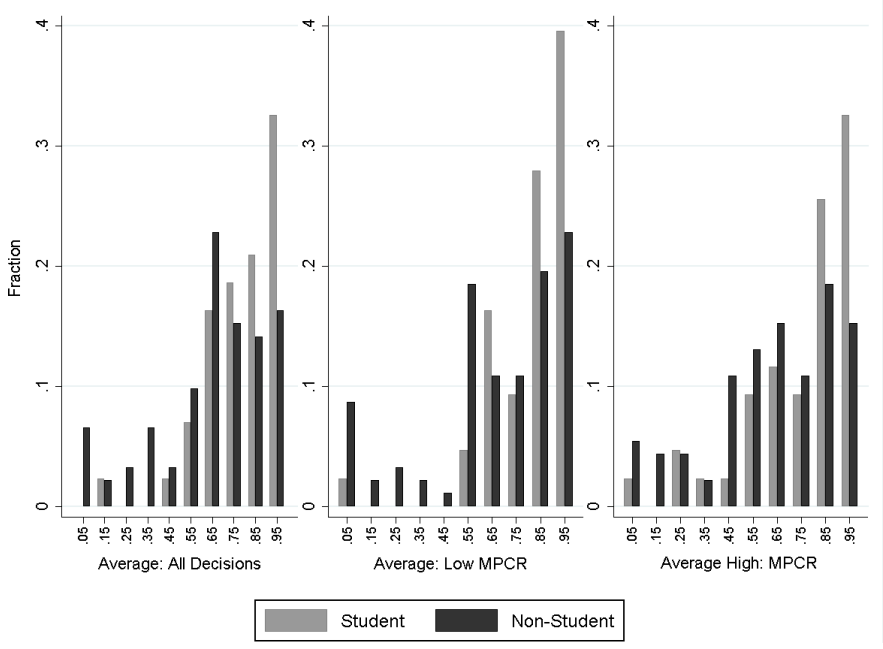
[24] A full table containing decision-wise correlations for free-riding can be found in the Appendix table 7.

are readily transferable to the context of voluntary climate change mitigation. In this section, we expand these previous results along two dimensions. First, we explore the joint role of subject-pool effects and task format. Second, we look at key attributes beyond student status that could account for subject pool effects. This second step might help to identify specific segments of the population for which PGG behavior is particularly generalizable. If possible, this characterization could provide some guidance when targeting specific study populations, for which one can expect results to be meaningfully interpretable in the mitigation context.

Result 5: Quantitatively, subject pool effects outweigh the effect of game parameters in explaining individual consistency. These differences cannot be attributed to observable characteristics.

As a first step, we define a measure of individual behavioral consistency. By our stylized definition, a pair of choices would count as perfectly consistent if a decision-maker selected identical actions in an identical setting. As a simple measure that conforms with this definition, we calculate the absolute difference between the fractions of endowment contributed in Task I and Task II and subtract it from one. Clearly, whether or not a given decision-maker indeed perceives choices in Task I and Task II as equivalent could depend on context-specific factors (e.g., game parameters and framing), individual characteristics determining his preferences in each task, and the interaction of these factors (Furr and Funder, 2004). Applied to our experiment, if behavior in both tasks was driven by the same set of individual characteristics and contextual factors did not matter, our measure would be one for the same individual. In contrast, if for the two tasks these factors worked in opposite directions, the measure would tend towards zero.

Figure 3: Distribution of average consistency



Notes: Histogram displaying the distribution of different average consistency measures by subject pool.

Figure 3 displays the distribution of this consistency measure for the two distinct subject pools. From left to right, we show three different averages: One average across all ten decisions of Task I, another only for low MPCR (< 0.4) decisions, and the third only for high MPCR (≥ 0.4) decisions.[25] The figure reveals similar patterns as the previous sections but also highlights the extent of individual heterogeneity. A considerable share of participants conforms to our definition of "perfect consistency". Across the three panels, between 15 and 40 percent of subjects select almost identical contributions in both tasks. Comparing the middle panel to those to its left and right shows that identical choices are most common among students taking the low MPCR decisions. Consistent free-riding accounts for more than half of this fraction. However, especially among non-students, there is also a large group of subjects who reach only a low to medium level of consistency.

In a more refined analysis, we now check whether this heterogeneity can be linked to the variation of individual attributes and contextual factors. The resulting regression model makes use of the full panel structure of our data. For each individual we observe ten decision-wise consistency measures, which is our dependent variable ($1 - |\frac{x_{it}}{\omega} - g_i|$). Across all 1,350 observed realizations of this variable, we find 118 instances of perfect inconsistency and 335 instances of perfect consistency. The largest part (63.5 percent) of consistent decisions are by subjects who free-ride in both tasks, followed by subjects who contribute half of their endowment (23.9 percent) and full contributors (5.3 percent). This conforms with the findings of others, stating that free-riding is the most stable individual behavior within the same task, across different cooperation tasks and across time (Brosig et al., 2007; Ubeda, 2014). To quantify to what degree behavioral differences in the two tasks are driven by parameter choices and to what degree they are linked to individual characteristics, we estimate different specifications of a random effects tobit model shown in table 5.

In the first specification, we jointly estimate the effect of an exogenous variation of the MPCR and moving from a student to a non-student sample. Increasing the MPCR inflates contribution differences between Task I and Task II significantly. Furthermore, for a given MPCR, students display more behavioral consistency than non-students. Quantitatively, the increase in consistency caused by reducing the MPCR from the highest (0.8) to the lowest (0.1) parameterization amounts to approximately two-thirds of the effect observed when switching from a non-student to a student subject pool. In specification 2 we show that changes in the MPCR affect students and non-students differently. The weakly significant interaction term indicates that a *ceteris paribus* reduction of the MPCR increases the consistency of students more strongly than that of non-students. In other words, students react more strongly to changes in contextual factors. In practice, this would mean that a PGG would have to be adapted more strongly when administered to non-students compared to students in order to achieve a similar effect on generalizability. Using only the student status to differentiate between the two subject pools

[25] Each of these average measures is calculated according to the following formula using the notation introduced in Section 3, with g_i denoting the fraction of endowment contributed by individual i in Task I:

$$c_i^z = 1 - \frac{1}{T}\sum_{t}^{T} |\frac{x_{it}}{\omega} - g_i| \tag{2}$$

Table 5: Differences in behavior, Task Format and Individual Characteristics

	(1) Consistency	(2) Consistency	(3) Consistency
MPCR	-0.218****	-0.310****	-0.219****
	(-6.17)	(-4.91)	(-6.16)
Non-Student (1=Yes)	-0.233***	-0.282****	-0.242**
	(-3.07)	(-3.49)	(-2.31)
Non-Student*MPCR		0.134*	
		(1.77)	
Age (Years)			0.003
			(0.93)
Male (1=Yes)			-0.101
			(-1.27)
Assets (1=Yes)			0.035
			(0.34)
Years of Education			0.011
			(0.86)
Household Size			-0.019
			(-0.69)
Parent (1=Yes)			-0.230**
			(-2.07)
Stated Risk Aversion (1-11)			-0.004
			(-0.32)
Fear Climate Change (1-7)			-0.009
			(-0.44)
Constant	0.982****	1.016****	0.915****
	(15.26)	(15.10)	(3.70)
Observations	1350	1350	1320
Individuals	135	135	132
Chi^2	47.23	50.35	56.06

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Notes: Random effects tobit maximum likelihood estimation to account for censoring from below (0) and above (1). z statistics in parentheses. For each specification the dependent variable is one minus the absolute difference between behavior in Task I and Task II in percentage terms.

masks a number of individual characteristics that could drive behavioral differences in the two tasks. Thus, specification 3 contains additional controls for individual characteristics. Some of these characteristics, such as gender (Croson and Gneezy, 2009) or age (List, 2004) have been included because they have been shown to influence contribution behavior in standard PGG. Other characteristics such as risk preferences, parenthood, or the fear of climate change could be especially relevant for the decision to contribute to climate change mitigation (Löschel et al., 2013; Diederich and Goeschl, 2014). Thus, these two groups of variables are plausible correlates of context-specific preferences in either Task II or Task I. However, except for being a parent, the included characteristics provide no additional information for individual consistency. As the student dummy remains significant and nearly unchanged in size, despite the further control

variables, there are likely unobserved individual characteristics that underlie subject-pool differences. Overall, the regression results point out that moving to a more diverse subject pool but retaining the standard task format of a PGG does not necessarily increase the generalizability of results in our context. Subject-pool specific differences have a larger impact on the overall consistency than differences in the parameterization for the range of values we observe.

5 Discussion and conclusion

As experimental economics matures, there is a growing interest in drawing on experimental methods and evidence to inform concrete policy debates (Bohm, 2003). We fundamentally agree that much can be learned from experiments, especially since they offer a sometimes unique opportunity for identifying causal relationships where field data are inadequate or scarce. But when experiments are motivated by specific policy issues, their generalizability becomes a central issue (Schram, 2005; Sturm and Weimann, 2006; Alm et al., 2015). In this paper, we have investigated whether and under which conditions behavior in generic PGGs generalizes to the context of voluntary climate action (VCA).

Our analysis highlights that individuals' willingness to contribute to climate change mitigation is at least partly attributable to generic cooperative preferences assessed in PGG experiments. At the same time, it highlights the failure of standard PGG experiments to capture many idiosyncratic factors that are candidate drivers of individual VCA decisions. This implies that informing climate policies based on generic behavior observed in standard PGG experiments risks lacking generalizability. Our results show that these risks are partially remediable: Researchers can increase generalizability by decreasing the MPCR and increasing the group size, thereby creating greater structure resemblance between the abstract game and the VCA context. This may imply that, in the limit, the best laboratory equivalent to VCA decisions may be the standard dictator game in which the dictator's private return of contributing is zero.[26]

Our analysis also speaks to the second threat to the generalizability of PGG experiments towards VCA, namely when the representativeness of the experimental sample for the target population is low. Most evidence from previous laboratory PGG experiments draws on student subjects. Students share only a limited range of individual attributes with the general population so that their behavior may not generalize. Here, our comparison of contribution behavior across subject pools reveals that student behavior in both experimental tasks is only a lower bound for the extent of cooperative behavior predicted for a population with broader demographic heterogeneity. Experiments with student subjects appear to overestimate the prevalence of selfish behavior in the general population and to be too pessimistic about individuals willingness to engage in VCA. At the same time, we find that students are more responsive to changes in experimental parameters (or conversely less responsive to differences between the tasks) and

[26]So far, there is only limited experimental evidence on contribution behavior from PGG under conditions of very low MPCR (Weimann et al., 2012). While some general patterns persist, there is also some emerging evidence that well-known mechanisms for fostering cooperation such as peer-punishment (Xu et al., 2013) are much less effective given a reduced MPCR. Similarly, strongly reducing the MPCR and delaying the payoffs have been shown to strongly reduce successful cooperation in a threshold public goods game (Jacquet et al., 2013; Hauser et al., 2014). Further research in this direction could be of great interest for those who wish to study the behavioral mechanisms of cooperation in the context of climate change.

consequently display a higher consistency between the different decision tasks. This is especially true for the domain of selfish behavior. Thus, sampling from the general population can serve the aim of drawing from a more representative subject pool, but appears to impose stronger demands on the experimental design: The higher diversity of the subject pool might not only call for a larger sample size but also for additional treatment variations.

In sum, when users of laboratory experiments go beyond falsifying theories and instead aim at deriving concrete recommendations for policymakers in areas such as climate change mitigation, questions of generalizability (Levitt and List, 2007a; Alm et al., 2015; Al-Ubaydli and List, 2015), parallelism (Plott, 1987) and representativeness (Anderson et al., 2013) require more attention. In our study, generic cooperative preferences of students in the PGG are not sufficient to explain individual-level behavior in the richer context of VCA, even though PGG contributions and VCA are both observed within the same lab setting that holds many other contextual features constant. Generalizability towards VCA outside the lab, i.e. in policy-relevant field settings, is then likely to be affected by additional shift factors (Levitt and List, 2007a). For instance, VCA decisions outside the lab context require individuals to use their own money instead of an experimental endowment; are not scrutinized by an experimenter, but possibly by a social environment; and are often bundled with other attributes of a consumption decision. Each of these contextual shift parameters further affects generalizability beyond the scope of our own study. These observations make it likely that future experimental research that aims at informing climate policies will benefit from moving beyond simple PGG tasks in the lab and creating experimental paradigms that can better capture the complex motivational structure that may underlie individual VCA decisions. To investigate decisions about climate change in their full complexity, a variety of methods including lab experiments, field experiments, surveys, and observational data are probably best suited to be complemented by each other to derive reliable policy conclusions (Czibor et al., 2019).

Of course, the main benefit from and motivation for conducting abstract or context-rich experiments is their ability to isolate the causal effect of a particular treatment variation on behavior. Given that our findings show that only a small fraction of VCA behavior is driven by generic cooperative preferences observed in the PGG, it is not obvious whether treatment effects would be highly transferable between both settings. The potential for transferability, both in a quantitative and qualitative sense,[27] is hard to assess for a treatment effect when the underlying causal mechanism is unknown (Heckman and Smith, 1995; Imai et al., 2011). In short, our findings do not rule out that some of the treatment effects identified in standard PGG would transfer to the specific context of climate change mitigation.[28] However, they highlight that such transferability across different contexts cannot be simply assumed without understanding the contextual features and causal mechanism that underlie a specific treatment effect of interest.

Acknowledgement: *The authors gratefully acknowledge financial support by the German Min-*

[27] As highlighted by Kessler and Vesterlund (2015), a discussion about qualitative transferability might be more fruitful.

[28] For instance, the qualitative predictions regarding the effects of providing social information have been largely unaffected by the setting under which they were obtained, be it for contributions in abstract laboratory PGG tasks (Bardsley, 2000), in different field settings (Alpizar et al., 2008; Shang and Croson, 2009), or in the specific context of mitigation decisions (Allcott, 2011; Goeschl et al., 2018).

istry for Education and Research under grant 01UV1012.

References

O. Al-Ubaydli and J. A. List. On the generalizability of experimental results in economics. In G. R. Fréchette and A. Schotter, editors, *Handbook of Experimental Economic Methodology*, pages 420–463. Oxford University Press, 2015.

H. Allcott. Social norms and energy conservation. *Journal of Public Economics*, 95(9):1082–1095, 2011.

J. Alm, K. M. Bloomquist, and M. McKee. On the external validity of laboratory tax compliance experiments. *Economic Inquiry*, 53(2):1170–1186, 2015.

F. Alpizar, F. Carlsson, and O. Johansson-Stenman. Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica. *Journal of Public Economics*, 92(56):1047–1060, 2008.

J. Anderson, S. V. Burks, J. Carpenter, L. Götte, K. Maurer, D. Nosenzo, R. Potter, K. Rocha, and A. Rustichini. Self-selection and variations in the laboratory measurement of other-regarding preferences across subject pools: evidence from one college student and two adult samples. *Experimental Economics*, 16(2):170–189, 2013.

N. Bardsley. Control without deception: Individual behaviour in free-riding experiments revisited. *Experimental Economics*, 3(3):215–240, 2000.

M. Belot, R. Duch, and L. Miller. A comprehensive comparison of students and non-students in classic experimental games. *Journal of Economic Behavior & Organization*, 113:26–33, 2015.

M. Benz and S. Meier. Do people behave in experiments as in the field? evidence from donations. *Experimental Economics*, 11(3):268–281, 2008.

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

M. Blanco, D. Engelmann, and H. T. Normann. A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, 72(2):321–338, 2011.

I. Blanken, N. van de Ven, M. Zeelenberg, and M. H. Meijers. Three attempts to replicate the moral licensing effect. *Social Psychology*, 45(3):232, 2014.

P. Bohm. Experimental evaluations of policy instruments. In Karl-Göran Mäler and Jeffrey R. Vincent, editor, *Handbook of Environmental Economics*, pages 437–460. Elsevier, 2003.

K. A. Brekke and O. Johansson-Stenman. The behavioural economics of climate change. *Oxford Review of Economic Policy*, 24(2):280–297, 2008.

K. Brick and M. Visser. What is fair? An experimental guide to climate negotiations. *European Economic Review*, 74(0):79–95, 2015.

J. Brosig, T. Riechmann, and J. Weimann. Selfish in the end?: An investigation of consistency and stability of individual behavior. *FEMM Working Paper No. 05*, 2007.

C. Camerer. The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List. In G. R. Fréchette and A. Schotter, editors, *Handbook of Experimental Economic Methodology*, pages 249–296. Oxford University Press, 2015.

V. Capraro, J. J. Jordan, and D. G. Rand. Heuristics guide the implementation of social preferences in one-shot Prisoner's Dilemma experiments. *Scientific reports*, 4, 2014.

F. Carlsson and O. Johansson-Stenman. Behavioral economics and environmental policy. *Annu. Rev. Resour. Econ.*, 4(1):75–99, 2012.

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

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

E. Czibor, D. Jimenez-Gomez, and J. A. List. The dozen things experimental economists should do (more of). Technical report, National Bureau of Economic Research, 2019.

A. C. de Oliveira, R. T. Croson, and C. Eckel. The giving type: Identifying donors. *Journal of Public Economics*, 95(56):428–435, 2011.

J. Diederich and T. Goeschl. Willingness to pay for voluntary climate action and its determinants: Field-experimental evidence. *Environmental and Resource Economics*, 57(3):405–429, 2014.

J. Diederich, T. Goeschl, and I. Waichman. Group size and the (in) efficiency of pure public good provision. *European Economic Review*, 85:272–287, 2016.

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

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

F. Exadaktylos, A. M. Espín, and P. Brañas Garza. Experimental subjects are not different. *Nature: Scientific reports*, 3, 2013.

A. Falk and J. J. Heckman. Lab experiments are a major source of knowledge in the social sciences. *Science*, 326(5952):535–538, 2009.

A. Falk, S. Meier, and C. Zehnder. Do lab experiments misrepresent social preferences? The case of self-selected student samples. *Journal of the European Economic Association*, 11(4): 839–852, 2013.

E. Fehr and A. Leibbrandt. A field study on cooperativeness and impatience in the tragedy of the commons. *Journal of Public Economics*, 95(9-10):1144–1155, 2011.

P. J. Ferraro and C. A. Vossler. The source and significance of confusion in public goods experiments. *The BE Journal of Economic Analysis & Policy*, 10(1):1935–1682.2006, 2010.

U. Fischbacher. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178, 2007.

R. Furr and D. C. Funder. Situational similarity and behavioral consistency: Subjective, objective, variable-centered, and person-centered approaches. *Journal of Research in Personality*, 38(5):421–447, 2004.

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

M. M. Galizzi and D. Navarro-Martínez. On the external validity of social preference games: a systematic lab-field study. *Management Science*, 2018.

C. Gallier, C. Reif, and D. Römer. Repeated pro-social behavior in the presence of economic interventions. *Journal of behavioral and experimental economics*, 69:18–28, 2017.

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

T. Goeschl, S. Kettner, J. Lohse, and C. Schwieren. From social information to social norms: Evidence from two experiments on donation behaviour. *Games*, 9(4):91, 2018.

J. M. Gowdy. Behavioral economics and climate change policy. *Journal of Economic Behavior & Organization*, 68(3):632–644, 2008.

B. Greiner. An online recruitment system for economic experiments. *Journal of the Economic Science Association*, (1), 2015.

E. Gsottbauer and J. van den Bergh. Environmental Policy Theory Given Bounded Rationality and Other-regarding Preferences. *Environmental and Resource Economics*, 49(2):263–304, 2011.

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

O. P. Hauser, D. G. Rand, A. Peysakhovich, and M. A. Nowak. Cooperating with the future. *Nature*, 511(7508):220, 2014.

J. J. Heckman and J. A. Smith. Assessing the case for social experiments. *The Journal of Economic Perspectives*, 9(2):85–110, 1995.

D. Houser and R. Kurzban. Revisiting kindness and confusion in public goods experiments. *The American Economic Review*, 92(4):1062–1069, 2002.

K. Imai, L. Keele, D. Tingley, and T. Yamamoto. Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, 105(04):765–789, 2011.

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

J. Jacquet, K. Hagel, C. Hauert, J. Marotzke, T. Rhl, and M. Milinski. Intra-and intergenerational discounting in the climate game. *Nature Climate Change*, 2013.

J. Kessler and L. Vesterlund. The external validity of laboratory experiments: The missleading emphasis on quantitative effects. In G. R. Fréchette and A. Schotter, editors, *Handbook of Experimental Economic Methodology*, pages 391–407. Oxford University Press, 2015.

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

S. K. Laury and L. O. Taylor. Altruism spillovers: Are behaviors in context-free experiments predictive of altruism toward a naturally occurring public good? *Journal of Economic Behavior & Organization*, 65(1):9–29, 2008.

S. D. Levitt and J. A. List. What do laboratory experiments measuring social preferences reveal about the real world? *The Journal of Economic Perspectives*, 21(2):153–174, 2007a.

S. D. Levitt and J. A. List. Field experiments in economics: The past, the present, and the future. *European Economic Review*, 53(1):1–18, 2009.

S. D. Levitt and J. A. List. On the generalizability of lab behaviour to the field. *Canadian Journal of Economics*, 40(2):347–370, 2007b.

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

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

X.-L. Meng, R. Rosenthal, and D. B. Rubin. Comparing correlated correlation coefficients. *Psychological Bulletin*, 111(1):172, 1992.

M. Milinski, D. Semmann, H.-J. Krambeck, and J. Marotzke. Stabilizing the earths climate is not a losing game: Supporting evidence from public goods experiments. *PNAS*, 103(11): 3994–3998, 2006.

M. Milinski, R. D. Sommerfeld, H.-J. Krambeck, F. A. Reed, and J. Marotzke. The collective-risk social dilemma and the prevention of simulated dangerous climate change. *PNAS*, 105 (7):2291–2294, 2008.

A. Peysakhovich, M. A. Nowak, and D. G. Rand. Humans display a cooperative phenotype that is domain general and temporally stable. *Nat Commun*, 5, 2014.

C. R. Plott. Dimensions of parallelism: Some policy applications of experimental methods. *Laboratory experimentation in economics: Six points of view*, pages 193–219, 1987.

L. Ross and R. E. Nisbett. *The person and the situation: Perspectives of social psychology*. Pinter & Martin Publishers, 2011.

A. J. Sankoh, M. F. Huque, and S. D. Dubey. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Statistics in Medicine*, 16(22):2529–2542, 1997.

A. Schram. Artificiality: The tension between internal and external validity in economic experiments. *Journal of Economic Methodology*, 12(2):225–237, 2005.

J. Shang and R. Croson. A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *The Economic Journal*, 119(540): 1422–1439, 2009.

J. F. Shogren and L. O. Taylor. On behavioral-environmental economics. *Review of Environmental Economics and Policy*, 2(1):26–44, 2008.

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

E. Snowberg and L. Yariv. Testing the waters: Behavior across participant pools. Technical report, National Bureau of Economic Research, 2018.

C. Starmer and R. Sugden. Does the random-lottery incentive system elicit true preferences? An experimental investigation. *American Economic Review*, 81(4):971–78, 1991.

J. H. Steiger. Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87 (2):245, 1980.

B. Sturm and J. Weimann. Experiments in environmental economics and some close relatives. *Journal of Economic Surveys*, 20(3):419–457, 2006.

A. Tavoni, A. Dannenberg, G. Kallis, and A. Löschel. Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *PNAS*, 108(29):11825–11829, 2011.

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

L. E. Torres-Guevara and A. Schlüter. External validity of artefactual field experiments: A study on cooperation, impatience and sustainability in an artisanal fishery in colombia. *Ecological Economics*, 128:187–201, 2016.

P. Ubeda. The consistency of fairness rules: An experimental study. *Journal of Economic Psychology*, 41(0):88–100, 2014.

J. Urban, Š. Bahník, and M. B. Kohlová. Green consumption does not make people cheat: Three attempts to replicate moral licensing effect due to pro-environmental behavior. *Journal of Environmental Psychology*, 2019.

L. Venkatachalam. Behavioral economics for environmental policy. *Ecological Economics*, 67 (4):640–645, 2008.

J. Vicens, N. Bueno-Guerra, M. Gutiérrez-Roig, C. Gracia-Lázaro, J. Gómez-Gardeñes, J. Perelló, A. Sánchez, Y. Moreno, and J. Duch. Resource heterogeneity leads to unjust effort distribution in climate change mitigation. *PloS one*, 13(10):e0204369, 2018.

M. Voors, T. Turley, A. Kontoleon, E. Bulte, and J. A. List. Exploring whether behavior in context-free experiments is predictive of behavior in the field: Evidence from lab and field experiments in rural Sierra Leone. *Economics Letters*, 114(3):308–311, 2012.

J. Weimann, J. Brosig, H. Hennig-Schmidt, C. Keser, and C. Stahr. Public-good experiments with large groups. *Magdeburg University Working Paper 9/2012*, 2012.

B. Xu, C. B. Cadsby, L. Fan, and F. Song. Group size, coordination, and the effectiveness of punishment in the voluntary contributions mechanism: An experimental investigation. *Games*, 4(1):89–105, 2013.

6 Appendix: Supplementary tables and regressions

6.1 Correlation table task II: Decision a.-j.

Table 6 contains the correlation coefficients for each pair of decisions made in Task II.

Table 6: Correlation Matrix of Task II decisions

Decisions	a	b	c	d	e	f	g	h	i	j
a	1.000									
b	0.681	1.000								
c	0.697	0.849	1.000							
d	0.706	0.731	0.696	1.000						
e	0.674	0.716	0.701	0.642	1.000					
f	0.617	0.691	0.598	0.696	0.758	1.000				
g	0.658	0.626	0.572	0.749	0.516	0.611	1.000			
h	0.587	0.564	0.480	0.613	0.597	0.655	0.691	1.000		
i	0.555	0.494	0.528	0.583	0.579	0.559	0.613	0.721	1.000	
j	0.467	0.431	0.436	0.544	0.469	0.504	0.588	0.762	0.625	1.000

6.2 Correlations free-riding

Table 7 contains Spearman correlation coefficients between being a free-rider in Task I (coded as 1) and being a free-rider in Task II (coded as 1). For the pooled sample (4) there are significant correlations for eight out of ten Task II decisions. These mainly reflect consistent free-riding among student subjects (6).

Table 7: Spearman correlations between free-riding in the real and in the abstract context for all 10 decisions

(0)	(1)	(2)	(3)	(4)	(5)	(6)
Decision	Group Size	Symmetry	MPCR	Correlation	Correlation Non-Students	Correlation Students
a	Large	Sym	0.1	0.2085**	0.1196	0.3486**
b	Large	Asym	0.1	0.1924**	0.0919	0.3603**
c	Large	Asym	0.15	0.2221***	0.1196	0.3908***
d	Large	Sym	0.2	0.2573***	0.1738	0.3841**
e	Small	Asym	0.33	0.1261	0.0067	0.3072**
f	Small	Sym	0.4	0.1992**	0.13	0.2969*
g	Large	Asym	0.42	0.2051**	0.1201	0.3341**
h	Small	Asym	0.46	0.1905**	0.0378	0.3841**
i	Small	Asym	0.53	0.2133**	0.11	0.3812**
j	Small	Sym	0.8	0.1153	-0.0045	0.2861*

Notes: Decision f constitutes the benchmark case.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6.3 Regression results task II

Table 8 displays results from a random effects tobit regression with the fraction of endowment contributed as the dependent variable. The most basic specification (1) corroborates a positive and significant relationship between contributions and the internal return, external return, group size in each decision of Task II. Furthermore, non-students contribute higher amounts.

These relationships are robust to controlling for further demographic variables and attitudes in specification (2).

Table 8: Contributions Abstract Public Goods Game and Demographic Variables

	(1) Contributions	(2) Contributions
Non-Student (1=Yes)	0.333****	0.225*
	(3.93)	(1.94)
Internal Return	0.029****	0.029****
	(6.98)	(6.88)
External Return	0.012****	0.013****
	(3.89)	(4.04)
Group Size	0.021****	0.021****
	(5.79)	(5.80)
Age (Years)		0.007*
		(1.82)
Male (1=Yes)		-0.031
		(-0.36)
Assets (1=Yes)		-0.245**
		(-2.46)
Years of Education		-0.008
		(-0.57)
Household Size		0.017
		(0.56)
Number of Children		0.049
		(1.00)
Fear Climate Change (1-7)		-0.036
		(-1.51)
Constant	-0.424****	-0.304
	(-4.82)	(-1.12)
Observations	1350	1320
Individuals	1350	1320
Prob > Chi^2	0.000	0.000

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Notes: Random effects tobit maximum likelihood estimation to account for censoring from below (0) and above (1). z statistics in parentheses.

7 Instructions

Check-in	Check-in room: Sign-in and generation of personal code
Experiment	Laboratory: Random seat assignment General instructions read out loud *(page 34)* Tasks implemented in z-Tree • Contribution to climate change mitigation *(page 35)* • Laboratory public goods game *(page 37)* Payment receipt distributed according to personal code
Payment	Check-in room: Subjects exchange payment receipt for cash

General Instructions

General instructions were handed to participants as a print-out.

General Information

Dear participant,

Thank you for participating in this study. In the following we will inform you about the rules and procedures. You have the opportunity to earn real money. Your final payment depends on your decisions in the experiment. Every participant has received the same printed instructions as you did. Please take your time and read the instructions carefully.

No communication with other participants

Please do not communicate with the other participants. Otherwise we are forced to exclude you from the experiment and you will receive no payment. If you have any questions, please raise your hand. The experimenter will answer your question in quiet.

Procedure

Please make sure that you created your personal code. During today's experiment, you will be asked to enter your personal code. Your personal code ensures that your decisions during the study remain anonymous. The experiment is taking place at the computer and each task is explained step-by-step. Please read the instructions on the screen thoroughly. If amounts of money are mentioned in the explanations for a given task, these amounts refer to real payments which we will pay you in cash – according to your decisions– at the end of the experiment.

It is important that you answer all questions; your personal data is treated anonymously.

Thank you!

Real Contribution Task

General Instructions (Screen 1)

Dear participant,

Thank you for supporting our research. On this screen you receive general instructions on the procedures. You will take several tasks. Please follow the instructions on the screen.

At the end of today's experiment you will receive your payment. At several points you can influence this payment by your own decisions. Whenever this is the case you will be informed on the respective screen and you will receive information on the specific rules of each task.

Your decisions are anonymous. Your anonymity is ensured by your personal code. In addition, you receive your payment at the end in room 00.005a (check-in room). Therefore, the experimenters will not receive information on your decisions and payments.

– – –

(Screen 2)

For your participation in this study you will receive **ten Euro**.

These **ten Euro** are paid to you at the end of today's experiment in cash.

Alternatively, we offer you to use any share of these ten Euro to reduce global CO_2 emissions.

In the following we explain how it is possible to reduce global CO_2 emissions.

– – –

(Screen 3)

What is CO_2?

CO_2 is a gas which is emitted by burning oil, coal, or fuel. It accrues from the manufacturing of goods or the production of electricity as well as from travel by car or airplane.

Why reduce CO_2?

The more CO_2 gets into earth's atmosphere, the more likely is the occurrence of the environmental problem climate change. Scientists expect climate change to cause consequences such as the rise of sea levels, the stronger spread of tropical diseases, or smaller yields in agriculture.

How is it possible to reduce CO_2 emissions?

Within the European Union a binding limit has been installed which constitutes how much CO_2 may be emitted by large industrial companies. In order to emit CO_2, these companies need emission permits. These permits can be purchased from the emission-trading-registry of the Federal Environmental Agency. After purchase these permits are not available to companies anymore. In this way, European CO_2 emissions are reduced by the amount of purchased permits. As the climatic system reacts inertly to a change in CO_2 emissions, the reduction action contributes only in approximately 50 years towards noticeable climate change mitigation.

What do we offer to you?

As soon as you have completed reading this information, we offer you to purchase permits from the German emission-trading-registry of the Federal Environmental Agency using your ten Euro. For each Euro you can mitigate emissions of approximately 70 kg CO_2, i.e., with your ten Euro you can reduce CO_2 emissions by a total of 700 kg. For example, 70 kg correspond to CO_2 emissions arising from a drive from Frankfurt am Main to Hamburg by car.

On average a German citizen emits 9 tons of CO_2 per year (one ton equals 1000 kg). Therefore,

700 kg, which may be reduced with your 10 Euro, correspond to a little less than the monthly CO_2 emissions of an average German.

How can you verify that your contribution was used to retire CO_2 permits?

As permits for CO_2 emissions are purchased through the emission-trading-registry of the Federal Environmental Agency, the procedure can be monitored transparently. At the end of this study a certificate of reduction –issued by the emission-trading-registry– will be posted at the notice board of the Chair of Behavioral Economics (Prof. Dr. C. Schwieren).

– – –

(Screen 4)

Purchase of CO_2 permits

On the following screen you may indicate the share of your ten Euro you would like to spend on CO_2 permits.

– – –

(Screen 5)

On this screen you may purchase emission permits using your ten Euro.

- Please insert into the blue field how much money you would like to use to retire CO_2 permits and thus reduce global CO_2 emissions.
- You are free to choose every integer between 0 and 10 Euro, i.e., you may fill in whole numbers without decimal place (period or comma).
- Each Euro you are not using to purchase CO_2 permits, you will receive in cash at the end of the experiment .

<insert decision>

<summary screen displayed>

– – –

Laboratory Public Goods Game

Screen 1: Instructions I

Explanation:

In this task you have the possibility to receive further payments, in addition to the ten Euro you already received at the beginning. Furthermore, during this task you interact with the other participants in this room. They will be matched to you randomly and you will not be informed who is matched to you.

Payment:

Your own decisions determine how much money you receive at the end. In addition, the decisions of the other matched participants influence your payment.

This part of the study contains a total of 12 decisions.

As soon as you took all decisions, a random mechanism will determine which of the 12 decisions will be relevant for payment at the end of the study. For the other decisions which are not selected, you will not receive payment. Each decision will be chosen with the same probability. Therefore, each decision is equally important for your final payment.

– – –

Screen 2: Instructions II

Possible Decisions:

In the following 12 decisions you can distribute 20 balls between two bowls which are labelled A and B.

Bowl A can be filled by you only.

Bowl B can be filled by you and the other participants you interact with.

While you make your decision, it is not possible to observe how many balls are placed into Bowl B by the other matched participants.

Anonymous Matching:

For this task the computer will match participants anonymously. This procedure determines the other participants who can place balls into Bowl B.

In some decisions you will execute the task with two other participants (i.e., in total three); in other decisions with eleven other participants (i.e., in total twelve).

If you are interacting with two other participants, you and the others cannot observe who these participants are. How many participants interact will change between decisions.

– – –

Screen 3: Example

Calculation of Payment:

This numerical example illustrates how payments in the decision task are determined.

The amounts shown here are only valid for the example and will differ in each of the actual 12 decisions.

You and the other participants can distribute 20 balls between Bowl A and Bowl

B:

Each participant fills his own Bowl A.

Bowl B can be filled by you and the other participants you interact with.

Bowl A: For each ball placed in Bowl A you receive 20 cent and the other matched participants receive 0 cent.

Bowl B: For each ball placed in Bowl B you receive 5 cent and the other matched participants receive 15 cent each.

The calculation is the same for all participants: Hence, all other participants can also distribute 20 balls.

Bowl A: For each ball another participant places in his/her own Bowl A, he/she receives 20 cent and you receive 0 cent.

Bowl B: For each ball another participant places in Bowl B, he/she receives 5 cent and all other matched participants **(including yourself)** receive 15 cent each.

– – –

Example:

Please choose how many balls you would like to place in Bowl B. Remember, balls which are not placed in Bowl B are automatically placed in Bowl A.

This is only an example.

Bowl A: This Bowl is only filled by you. You receive 20 cent per ball. The other participants receive 0 cent per ball.

Bowl B: This Bowl is filled by you and the other (two or eleven) matched participants.
You receive 5 cent per ball. The other participants receive 15 cent per ball each.

Your choice:

Please indicate in the blue field how many of the 20 balls you would like to place in **Bowl B**. The remaining balls are automatically placed in **Bowl A**.

<insert choice for example>

– – –

Your decision

You decided to place *<example choice>* of 20 balls in Bowl B. Hence, you placed the remaining *<20 minus example choice>* in Bowl A.

Per ball placed in Bowl A you receive 20 cent.

Per ball placed in Bowl B you receive 5 cent and the other participants receive 15 cent.

Calculation of Payment:

Please indicate how much you would receive for the decision.

In the example you placed *<20 minus example choice>* in Bowl A. Hence, you receive from Bowl A: *<insert calculation for example>*

In the example you placed *<example choice>* in Bowl B: You receive *<insert calculation for example>*

In the example you placed *<example choice>* in Bowl B: Hence, every other participant receives *<insert calculation for example>*

In addition, your own payment may change depending on how much the other participants place in Bowl B. For each ball another participant places in Bowl B, the other matched participants (including yourself) receive 15 cent per ball.

<feedback screen on calculation of example. If correct, continue. If incorrect, repeat example>

– – –

You have now completed the numerical examples. The actual task will be presented in a table.

– – –

Example

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6
Decision	Bowl A per ball you receive	Bowl B per ball you receive	Bowl B per ball the other participants receive	Bowl B number of participants	Your decision
1	20	5	15	3	

Example for table:
The above table is an example and illustrates the display of the subsequent decision task.
The above table displays a single row. The actual decision table will consist of twelve rows. Each row corresponds to one decision.
Explanation of table:
In this explanation you receive information on the (numbered) columns in the table.
Column 2 This column displays the amount of cents which you will receive for each ball remaining in Bowl A.
Column 3 This column displays the amount of cents which **you** will receive for each ball remaining in Bowl B.
Column 4 This his column displays the amount of cents which **each other matched participant** will receive for each ball remaining in Bowl B.
Column 5 This column displays the number of participants who can place balls in in Bowl B. This number includes you.
Column 6 In this column you will indicate how many balls you would like to place in Bowl B.
– – –
You have completed the examples. Now the actual task will begin! All decisions are equally relevant for payment. We will chose one of the 12 decisions randomly (with equal probabilities) at the end of the experiment and determine your payment.
– – –

Decision Task

The table displays the 12 decisions. Each row corresponds to a new decision.

Bowl A per ball you receive	Bowl B per ball you receive	Bowl B per ball the other participants receive	Bowl B number of participants	Your decision
20	2	9	12	<*insert choice*>
20	2	2	12	<*insert choice*>
20	4	4	3	<*insert choice*>
20	4	4	12	<*insert choice*>
20	16	16	3	<*insert choice*>
20	12	8	3	<*insert choice*>
20	8	12	3	<*insert choice*>
20	8	8	3	<*insert choice*>
20	8	6	3	<*insert choice*>
20	3	2	12	<*insert choice*>
20	1	1	12	<*insert choice*>
20	2	3	12	<*insert choice*>

Please indicate in the blue fields how many balls you would like to place in Bowl B. The remaining balls are placed in Bowl A.