Five Essays on Cooperation

with an Application to Climate Change Mitigation

Dissertation

zur Erlangung des akademischen Grades

Doctor Rerum Politicarum

An der

Fakultät für Wirtschafts- und Sozialwissenschaften

der Ruprecht-Karls-Universität Heidelberg

vorgelegt von

Johannes Oliver Lohse

Heidelberg, im Juli 2015

Acknowledgments

This dissertation has profited in various ways from the inputs of a great number of people, who have supported me throughout the past years. To each of them I am deeply grateful.

First and foremost, I would like to thank Prof. Timo Goeschl for supervising this thesis, for being an inspiring co-author, for providing an environment where ideas can grow, and for being an excellent guide to the world of academia.

Furthermore, I am grateful to Prof. Fredrik Carlsson for serving as the second referee for this dissertation.

I thank the members of the CLIMAGE research project for many inspiring discussions and in particular Prof. Christiane Schwieren for being an insightful co-author. Sara Elisa Kettner has not only provided valuable inputs for the papers included in this dissertation, but has also been a fun and helpful colleague.

I enjoyed working with Johannes Diederich on our joint paper and thank him for sharing his thoughts and data. Thanks also go to my other current and former colleagues at the Chair of Environmental Economics, who have made my time here productive and enjoyable. Specifically, I would like to thank Daniel Heyen as well as Tobias Pfrommer for being great colleagues and roommates, and Israel Waichman and Johannes Jarke for their feedback and constructive suggestions regarding my experiments.

I also thank the faculty and colleagues here at the Department of Economics in Heidelberg for their valuable comments regarding my research. Furthermore, I am thankful to Prof. John List and his team for making my research visit at the University of Chicago possible.

During the last five years, some old friends and some new ones have made my time here in Heidelberg special, have provided me with some invaluable memories, and have proven to be resourceful in many ways. To each of them I am thankful for being an entertaining part of my life.

Towards the anonymous German tax payers I have to express my deepest gratitude for providing yet another highly intangible public good by funding my research.

Above all, I am indebted to my parents and my family for their continuous support in all my projects big and small and for giving me countless opportunities to see with my own eyes that the world can be an interesting and formidable place.

Finally, I wish to thank Anna for sharing two truly scarce resources with me in a world of otherwise unprecedented affluence: Her time and her thoughts.

Contents

List of Figures

Chapter 1

Introduction

Introduction

Over three decades ago, the political scientist Robert Axelrod posed a seemingly simple, yet intriguing question: "Under what conditions will cooperation emerge in a world of egoists without central authority?" (Axelrod, 1984, p.3). Thirty years later, variants of the same basic question continue to inspire research at the intersection of psychology (Van Lange et al., 2013), human biology (West et al., 2011; Nowak, 2012), economics (Fehr and Gächter, 2000; Fehr and Fischbacher, 2003; Fischbacher and Gächter, 2010) and the social sciences (Bowles and Gintis, 2011). It is the purpose of this brief introduction to summarize the core results of this rapidly growing literature in economics and highlight the specific contributions of the five articles contained in this dissertation to different aspects of this overarching research question.

Applying (game) theoretic models to cooperation problems provides an answer that is as stark as it is straightforward: in the absence of a disciplining mechanism, typically requiring enforcement powers of a governing authority[1], cooperation among anonymous strangers will *not* emerge in a world of self-interested individuals (Samuelson, 1954; Bergstrom et al., 1986). In contrast to this bleak prediction, there are many real world settings in which cooperation – the act of incurring a personal net cost in order to generate a larger total benefit for others – appears to emerge voluntarily, as expressed, for instance, in effort choices at the workplace (Jones and George, 1998), in the fact that people vote (Downs, 1957), in the observation that some agreements are kept without binding legal contracts (Cheung, 1973), or in the way in which a number of common pool resources are governed by self-regulating institutions (Ostrom, 1990; Vollan and Ostrom, 2010). Since voluntary cooperation is efficient in most instances[2], it is seen as an integral part of a country's "social capital", which can be one important determinant of its economic performance (Knack and Keefer, 1997; Thöni et al., 2012).

The private provision of public goods (PPPG) is probably the most prominent example of voluntary cooperation. By definition, public goods offer benefits that are both non-excludable (i.e., access cannot be limited) and non-rival (i.e., the consumption by one does not influence the consumption possibilities by others). Therefore, theory posits that public goods will not be supplied at an efficient level by decision-makers taking only their private costs and benefits into account, but ignoring the inherent public benefits. This prediction again has to be squared with the outright existence of privately provided public goods: each year, charities in the Untited States receive about 2% of the gross domestic product (GDP) in private donations (List, 2011), open source software or

[1] Without a governing authority, as in cases of decentralized punishment, the costly enforcement of the mechanism constitutes a cooperation problem in itself. Whether and which kinds of self-enforcing mechanisms are effective, is subject to an ongoing discussion of its own (e.g., Ostrom et al., 1992; Nikiforakis, 2008).

[2] Of course, there are also cases in which the exact opposite is true. Collusion in an oligopoly market or bribing officials also require voluntary cooperation among the involved parties, but impose sometimes large costs on outsiders.

the online encyclopedia *Wikipedia*[3] predominantly rely on user contributions of time or money to generate content (Zhang and Zhu, 2011), and customers are sometimes willing to pay a price premium for electricity generated from renewable sources, which indicates private demand for the public good of reducing the risks associated with climate change (Kotchen and Moore, 2007). Naturally, these illustrations cannot serve as evidence that the public goods in question are provided at efficient levels. Furthermore, it would be equally easy to compose a list of examples in which the privately provided quantity is clearly insufficient, resulting, for instance, in the breakup of joint research projects or in the severe degradation of environmental resources. Under which conditions is the former, under which the latter outcome more likely? And in those cases where self-interest cannot explain provision choices, which other motives guide individual behavior?

The quickly progressing integration of laboratory and field experiments into the methodological mainstream of economics (Falk and Heckman, 2009; Croson and Gächter, 2010) has equipped researches with a set of powerful tools allowing them to test and refine hypotheses regarding these recurring questions and uncover causal relationships. In contrast to the purely observational nature of the previous examples, experimental data on cooperative behavior can be collected in controlled decision environments that conform closely to the assumptions made in theoretical models. Early experiments were therefore mainly designed as tests of theory, which would settle the fundamental question of whether individuals would contribute to public goods at all, if the decision environment was stripped down to the theories' core elements (Isaac et al., 1984; Ledyard, 1995). To provide such clean and decontextualized environment the liner public good game (PGG) was devised, which continues to be the standard paradigm for studying cooperative behavior in laboratory experiments. A typical PGG experiment is conducted as follows: Four participants receive a monetary endowment each (e.g., 4 €), which they can keep for themselves or contribute anonymously and simultaneously to a joint public account. Contributions to the public account are doubled by the experimenter and then equally divided among all four group members (i.e., a contribution of 4 € would be doubled to a total of 8 € and each of the four group members, including the initial contributor, would finally receive 2 €). Given this payoff structure, it is the individually dominant strategy not to contribute, while contributing the full endowment to the public account is efficient because it maximizes the payoff of the group as a whole; hence the "social dilemma".

Even when conducted under conditions of full anonymity, no-communication, non-trivial stakes, and no repetition, which altogether should lead to a complete breakdown of cooperation, the prevalence of strict free-riding (i.e., zero contributions by every participant) cannot be confirmed by the data of most PGG experiments (Vesterlund, 2014). These converging findings underline that pure payoff maximization is unlikely the sole motive

[3] At the time of writing this sentence, the English languague version of Wikipedia reports to feature 4,909,260 user generated entries, based on 778,769,995 individual edits (https://en.wikipedia.org/wiki/Special:Statistics [July 5th, 2015]).

within this strategic environment and thus constitute the venture point for an ongoing inquiry into the existence and role of other-regarding preferences in economic decision making.

In the same way Adam Smith anticipated many important concepts of modern economics in his book *"The Wealth of Nations"*, he probably would have been not too surprised by this recent (re-)discovery of behavioral economists (Ashraf et al., 2005), considering that he states in his second book *"The Theory of Moral Sentiments"*:

> "How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it, except the pleasure of seeing it."
>
> Adam Smith, *The Theory of Moral Sentiments* (1790, 6th ed. I.I.1)

In large parts, Smith's early description of human altruism is remarkably similar to the more recent definition of the "warm glow" motive, which has become a central building block of reigning theories of contribution behavior by stating that individuals receive utility from the pure act of giving (Andreoni, 1989, 1990). However, the existence of other-regarding preferences does not rule out selfishness as an important, if not the predominant motive in public good decisions. While the typical experiment refutes strong free-riding, average contributions still remain far from the efficient level. Furthermore, cooperation in PGGs tends to be rather fragile, as quickly declining contributions in repeated settings suggest (Ledyard, 1995). Thus, most decision-makers appear to care both about their own payoff and the payoff of other participants, a trade-off incorporated in current models of other-regarding behavior (e.g. Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). As noted in a second passage from *"The Theory of the Moral Sentiments"*, the concerns for the wellbeing of others indeed need not stretch very far, whereas the concerns for oneself often appear to dominate decision-making (Ashraf et al., 2005):

> "Let us suppose that the great empire of China, with all its myriads of inhabitants, was suddenly swallowed up by an earthquake, and let us consider how a man of humanity in Europe, who had no sort of connexion with that part of the world, would be affected upon receiving intelligence of this dreadful calamity. He would, I imagine, first of all, express very strongly his sorrow for the misfortune of that unhappy people, he would make many melancholy reflections upon the precariousness of human life, and the vanity of all the labours of man, which could thus be annihilated in a moment. [...] And when all this fine philosophy was over, when all these humane sentiments had been once fairly expressed, he would pursue his business or his pleasure [...] The most frivolous disaster which could befal himself would occasion a

> more real disturbance. If he was to lose his little finger to-morrow, he would not sleep to-night"
>
> Adam Smith, *The Theory of Moral Sentiments* (1790, 6th ed. III.I.46)

The current state of experimental evidence is far from a point where a unified theory of other-regarding behavior and cooperation in PGGs could be derived. In order to advance this goal and fill important research gaps, recent efforts have concentrated on two strands of inquiry: The first strand takes the large interpersonal heterogeneity in contribution behavior as a venture point and tries to link these differences to *observable individual attributes*. In other words, by characterizing decision-makers as different cooperative types (Fischbacher et al., 2001), this strand tries to find out more about the question of *who* cooperates. The second strand tries to identify *external factors* such as rules or institutions that affect the average rate of contributions. Thus, it is concerned with the question of *when* voluntary cooperation occurs and *how* it can be promoted (Nikiforakis, 2008).

Variants of these three central themes motivate each of the five articles that constitute this dissertation. Besides their overarching thematic focus, the different articles share a common methodology and a joint application. The common methodology, as the reader might suspect at this point, is the experimental nature of the data on which the five articles are based. While two of the articles (articles 3 and 4) build on a standard PGG experiment conducted in the laboratory with student subjects, two articles (articles 1 and 2) fall under the definition of an artefactual field experiment by drawing on non-student samples (Harrison and List, 2004). Moreover, three articles (articles 1, 2, and 5) embed decisions in a field context by analyzing contributions to a naturally occurring public good: the mitigation of climate change. In sum, each of the five articles imports theories and methods from public economics and behavioral/experimental economics and applies them to environmental economics. Thereby, they jointly contribute to the emerging field of behavioral environmental economics (Croson and Treich, 2014).

The five articles (especially articles 1, 2, and 5) speak to a central literature in environmental economics, namely, that of climate change mitigation. I will, therefore, briefly introduce the core problem of climate change in one paragraph, which deliberately abstracts from most of the more complex issues surrounding this environmental problem. On the most basic level, carbon-based economies inevitably emit the gases carbon dioxide (CO_2) and methane (CH_4) as part of the production processes required to meet their energy demands, to grow food stocks, and to manufacture consumption goods. Unmitigated, the build-up of these stock pollutants in the atmosphere is predicted to lead to large-scale changes in the climate system, with the potential of inflicting non-trivial damages of unknown magnitude to ecosystems and human societies across the centuries to come (IPPC, 2013). Even if average damages should turn out to be smaller than projected, there remains a non-negligible probability of catastrophic damages, in which case mitigation can be seen as a means of insurance (Weitzman, 2009). Mitigation, however, is costly because it requires a restriction of consumption possibilities, either directly or through investments into more energy efficient technologies. Thus, climate change has been coined as the "mother of all externalities"(Tol, 2009, p.29) and its mitigation could equally well be termed as the "archetype of all public goods problems".

Several unique features distinguish mitigation choices from the public good examples given before. First, climate change is by its very nature a global problem. Thus, efficient provision cannot be enforced by a single local or national governing authority. In a world of sovereign countries, each with its own interests, this results in a "Westphalian Dilemma" of international cooperation, which has so far proven hard to resolve (Barrett, 1994; Nordhaus, 2015). Regulation is likely to remain incomplete and voluntary mitigation could, therefore, increase total efficiency. However, in order to realize such efficiency gains, decision-makers would need to take into account the welfare of individuals living on different continents and in different societies. As yet, there exists only scant experimental evidence on the geographical scope of cooperative preferences (Buchan et al., 2009), but the assumption that such preferences decline with an increasing social or spatial distance between actors appears plausible. In the most extreme case, geographically or socially distant individuals might be perceived as out-group members to which no normative obligations or even spiteful preferences apply, as findings on parochial altruism would suggest (Bernhard et al., 2006; Choi and Bowles, 2007). Second, climate change mitigation is an intertemporal problem. Due to the physical inertia of the climate system, mitigation choices today yield hardly any benefits for members of the current generation, but could have significant impacts on the welfare of future generations. The same psychological factors that bias individual decision-making over long time horizons (Laibson, 1997), could play an even larger role when decisions relate to outcomes for temporally distant people partly not even born at the moment of decision making. Moreover, it remains unclear whether intragenerational and intergenerational altruism results from the same set of preferences and thus whether the latter can be easily inferred from existing PGG evidence. Third, the link between mitigation choices and social benefits is subject to several layers of uncertainty. In particular, it is still uncertain how mitigation choices map into climate outcomes and how climate outcomes map into economic or ecological damages. It has been shown that replacing certain with expected benefits reduces contributions in experimental distribution tasks (Brock et al., 2013). Finally, CO_2 is not emitted for its own sake, but is an unintended consequence of individual consumption choices. In carbon-based economies, virtually every decision involves trade-offs between emitting or avoiding different levels of CO_2 or CH_4. However, many of these trade-offs are hidden and complex and thus might not be as salient to decision-makers as in the context of public goods, where contributions are made directly (e.g., when giving money to a charity). This could imply that contribution choices in the context of climate change require higher levels of active deliberation than contribution choices in areas in which the harm or benefit of an action is easy to grasp and therefore follow from a moral intuition (as argued in article 2) (Markowitz and Shariff, 2012).

In sum, voluntary mitigation choices might depend on a number of additional factors, going beyond the typical trade-off between individual cost and public benefits. Considering that most studies which summarize the behavioral dimensions of voluntary climate actions highlight these factors (e.g., Brekke and Johansson-Stenman, 2008; Gowdy, 2008), there is surprisingly little *direct* experimental evidence on voluntary mitigation choices. Instead, most experimental evidence that draws conclusions on climate change policies is *indirect*, in the sense that it has been collected in PGG variants.[4] However, it is as yet an open question whether and under which conditions indirect evidence from generic PGGs readily generalizes to the idiosyncratic context of voluntary climate actions. This is the research question addressed in the first article of this dissertation.[5] Summarizing ahead of a more detailed discussion, our findings suggest that PGGs capture only a fraction of the latent preferences driving voluntary climate actions, but can be implemented in a way that increases their explanatory power.

Experiments have not only established the importance of other-regarding preferences, but have also highlighted the numerous cognitive limits and biases of decision-makers. Cognitive factors could have particular relevance in the context of climate change, where complex relationships and interdependencies are not easily understood. Only recently researchers have turned their attention to the interplay of cognition and other-regarding preferences by systematically studying the role of cognitive abilities and cognitive processes in determining the outcome of cooperation problems. In particular, current findings suggest that decision-makers following a first intuition decide differently from decision-makers reflecting more carefully about the costs and benefits of contributing (Piovesan and Wengström, 2009; Rand et al., 2012, 2014; Nielsen et al., 2014). These findings open up the possibility that voluntary cooperation can be understood and analyzed as a self-control problem (Dreber et al., 2014; Martinsson et al., 2014). Articles 2–4 contribute to this nascent strand of literature by scrutinizing the cognitive basis of cooperative behavior. Article 2 analyzes the cognitive processes underlying choices on voluntary climate actions.[6] Articles 3 and 4 employ a standard PGG design to disentangle how cognitive processes, cognitive abilities, and the understanding of the PGG task interact to shape contribution behavior.[7] Overall, the three articles underline that cognitive processes indeed play an important role in shaping cooperative behavior. In particular, higher levels of deliberation are associated with more cooperative choices.

[4]Direct experimental evidence comes from studies in which subjects can decide about mitigation efforts, for instance, by buying carbon offsets (e.g., Löschel et al., 2013; Diederich and Goeschl, 2014; Kesternich et al., 2014b) or by changing their consumption patterns in the energy sector (Allcott and Mullainathan, 2010; Allcott, 2011). Decisions in these studies, by definition, reflect idiosyncratic factors of climate change such as the temporal dimension or beliefs about the inherent risk. Indirect experimental evidence, on the other hand, comes from studies based on modified PGGs that vary one additional dimension relevant in the context of climate change (e.g., Tavoni et al., 2011; Hauser et al., 2014).

[5]Joint with Timo Goeschl, Sara Elisa Kettner, and Christiane Schwieren.

[6]Joint with Timo Goeschl and Johannes Diederich.

[7]Article 3 is joint work with Timo Goeschl.

Jointly, these findings shed some light on the evolutionary origins of cooperation. Furthermore, they provide some clues regarding the design of decision environments that aim at promoting cooperation. Finally, as in related discussions on biased risk perceptions (Kahneman and Thaler, 2006; Salanié and Treich, 2009), these findings raise the non-trivial methodological question (discussed in article 2) whether intuitive and carefully considered decisions should be given the same weight in regulatory questions.

Finally, despite the "Westphalian Dilemma" of international cooperation, most industrialized countries have by now adopted some regulatory measures to curb carbon emissions. However, some of these policies have been shown to affect the income distribution regressively (Cohen et al., 2013; Grösche and Schröder, 2014). From a perspective of voluntary contributions under incomplete regulation[8], this observation could be central, because it highlights a potential conflict between distributional preferences within a generation and between generations. Will decision makers holding other-regarding preferences continue to contribute to climate change mitigation, even when doing so harms members of the current generation and thus contributing might violate their own fairness norms? This open question motivates article 5 of this dissertation. My experimental results suggest that the outlook of inflicting financial harm on an outsider by contributing does not generally affect behavior, despite the fact that the experimental design renders the level of harm highly salient. However, if the harm results in large (advantageous) inequity between the decision maker and the harmed outsider, contributions are strongly reduced. This finding underlines, how other-regarding preferences, which justify climate policies in the first place, can turn out to be a double-edged sword as soon as those policies ignore distributional aspects.

Synopsis

The first article **"What do we learn from public good games about voluntary climate action? Evidence from an artefactual field experiment"** (with Timo Goeschl, Sara Elisa Kettner, and Christiane Schwieren) examines *whether* and under *which conditions* evidence from PGGs generalizes to voluntary mitigation decisions. This research question follows directly from the rising popularity of experiments, which has gone hand in hand with their promise of providing meaningful insights, which can inform policy-making. Against this background, a high level of generalizability is desirable, since it would allow to obtain causal evidence on mitigation choices based on relatively low-cost laboratory PGG experiments. However, generalizability cannot be readily assumed in all settings, as evidence across a number of studies documents (Galizzi and Navarro-Martinez, 2015). While this concern is not entirely new itself, it has started to receive renewed attention in the context of experiments on other-regarding behavior (Levitt and

[8] For instance, policies like a feed-in tariff for renewable energy subsidize contributions, but still require voluntary actions, as long as private costs outweigh private benefits.

List, 2007). Moreover, as laid out in the introduction, voluntary climate actions – the context to which several articles in this dissertation seek to speak – might be especially prone to be driven by additional factors not captured in standard PGGs.

Our paper constitutes a first step to bring experimental evidence to the – at its heart – empirical question of generalizability in the context of voluntary climate actions. To do so we observe each participant in two contribution situations: a PGG and a real giving task in which contributions are used to reduce CO_2 emissions. This within-subjects design allows to determine whether the same latent preferences drive behavior in both tasks. Once again with the goal of cost-efficiency in mind, the second contribution of this paper is to explore how PGGs could be ideally implemented in order to be more informative on voluntary climate actions. We do so by experimentally varying two central design features, namely, the parameter structure of the PGG and the subject pool under study. In particular we observe ten one-shot PGG decisions elicited for different group-sizes, marginal per-capita returns (MPCR), and degrees of payoff-symmetry. Thereby, we vary the structural resemblance between PGG contribution incentives and voluntary mitigation incentives. We furthermore collect observations both for students and non-students. As the latter are demographically more diverse, their behavior might be more representative for choices made in a climate policy context.

Our results suggest that PGG behavior can generalize to voluntary climate actions, yet not uniformly. Instead, the potential for generalizability crucially depends on the way in which the PGG is designed and conducted. For most parameter constellations we implement in the PGG contributions are uncorrelated with voluntary climate actions. This highlights the importance of idiosyncratic drivers of mitigation behavior that are not captured in standard PGGs. Furthermore, correlations for non-students are overall greatly reduced, suggesting a trade-off between representativeness and generalizability. For student subjects correlations become significant and sizable as soon as the PGG parameters resemble more closely the incentive structure underlying voluntary climate change mitigation. Therefore, analyzing PGGs with a greatly reduced MPCR (implying, in turn, a greatly increased group size) appears to be one promising avenue to study mitigation behavior in cost-efficient abstract lab settings.

The second article **"Giving is a question of time: Response times and contributions to an environmental public good"** (with Timo Goeschl and Johannes Diederich) imports a recent methodological innovation from behavioral economics into environmental economics by addressing the question of whether cooperation can be linked to intuitive or deliberative decision making. In particular, it analyzes response times (i.e., time experimental subjects spend on a decision screen) collected alongside choices in an extra-laboratory experiment on voluntary climate actions (Diederich and Goeschl, 2014) to identify the respective cognitive process at work. Thereby, it contributes to a nascent literature, which attributes cooperation (or a lack thereof) to the resolution of an internal conflict, a self-control problem that can be modeled within a dual-system framework (Dreber et al., 2014; Martinsson et al., 2014; Rand et al., 2014). If true, this could have some important methodological implications both for the interpretation of valuation experiments and for the scope of outcome-based theories of other-regarding behavior. Establishing a link between deliberation and cooperation could provide crucial clues as to how aspects of the decision environment influence contribution behavior, depending on whether a certain aspect or framing of the choice architecture induces more intuition or more deliberation. The predictions of outcome-based theories, on the other hand, should be independent from such aspects of the decision environment.

Dual-system theories posit that each decision-maker either relies on a fast, effortless intuition or on slow, deliberative reasoning (Evans, 2008). Thus, differences in response speed can be used as a proxy for the level of deliberation employed at the moment of decision making (Rubinstein, 2007). The response time data analyzed in the second article of this dissertation come from an experiment in which subjects from the general population faced a dichotomous choice between receiving a monetary payment or reducing CO_2 emissions by one metric ton (Diederich and Goeschl, 2014). Thus, each subject had to resolve a potential conflict between private costs (forgoing a payment between 2 € and 100 €) or public benefits (reducing the impacts of climate change). While the extant literature largely converges on the notion that the cognitive system at work can affect how decision-makers resolve this conflict, the direction of the link remains disputed. The dataset used offers four distinct benefits that facilitate the identification of this link in the context of climate change mitigation: First, participants directly contribute to voluntary climate actions. This could be relevant in light of the results of article 1. Second, with 3483 subjects, the number of independent observations is sufficiently large to approximate different cognitive processes through noisy response time data. Third, subjects come from the general population, which increases the representativeness of our results. Fourth, the dataset contains two observations of response times and associated choices collected on two consecutive decision screens for each subject. Thus, we can employ empirical methods (i.e., first difference regressions) that control for confounds due to the omission of unobserved individual attributes (e.g. preferences for climate change mitigation or general swiftness).

Our results highlight the importance of different cognitive processes for contribution choices. There is a strong and positive association between response times and the propensity to contribute, suggesting that, in the case of climate change, cooperation is a deliberate choice rather than a first impulse. The average response time of contributors – controlling for a set of individual attributes – is approximately 40% longer than that of non-contributors. This positive relationship survives a number of different robustness checks. Importantly, the relationship between response times and the propensity to contribute carries over to the individual level. Switching from defecting in the first to contributing in the second decision significantly increases response times for the same subject and vice versa. This largely excludes alternative explanations based on omitted variables. Our findings could have important implications for studies using experiments or survey methods to assess the preferences for environmental public goods and climate change mitigation in particular. Furthermore, they raise the methodologically non-trivial question whether intuitive and deliberative choices should be given the same weight in preference assessments.

In the third article **"Cooperation in public good games. Calculated or confused?"** (with Timo Goeschl) we test the hypothesis that individuals are more prone to cooperate, if they (can) base their decision on deliberation. Thus it takes the research question and results of the second article as a venture point and expands them along two central dimensions. First, our article contains a tests for the causal effect of constraining deliberation on cooperation. Second, it uses a more general decision task (i.e., two variants of a PGG task) in order to distinguish between different channels by which deliberation could affect contribution behavior: confusion, unconditional cooperation, and conditional cooperation. Employing a context-free decision task also illuminates whether the results of article 2 are specific to voluntary climate actions or apply to cooperation more generally.

To devise a test for causality, we build on the core assumption of dual-system theories that deliberation requires more time and cognitive effort than intuition (Evans, 2008). If true, the use of deliberation can be taxed by forcing participants to decide quickly. In a between-subjects design, we compare contributions under time pressure to contributions in an unconstrained baseline, facilitating the identification of an average treatment effect of constraining deliberation.

Other recent studies (Rand et al., 2012, 2014), which have used similar strategies to constrain deliberation, report that forcing subjects to decide quickly leads to higher average contributions in a one-shot PGG. The authors interpret this as evidence for a causal link between intuition and cooperation. Such a link would contradict our findings in article 2, an observation which more generally mirrors the mixed results in the extant literature, some of which support a positive link between deliberation and cooperation, while the remainder point toward the opposite conclusion (as discussed more throughly in articles 2–4). In light of these contradictory results, we devise a new experimental design, which enables us to isolate different channels through which time pressure could theoretically affect behavior in PGGs. Based on this design, we explore the role of one potentially confounding factor that could result from the use of time pressure. In particular, we hypothesize that giving subjects less time to understand the incentive structure, could have the unwanted side-effect of increasing confusion. In fact, even in the absence of time pressure, there is ample evidence for confused contributions in PGGs (e.g., Houser and Kurzban, 2002; Ferraro and Vossler, 2010). If time pressure increased the fraction of confused subjects (i.e., subjects who fail to understand that free-riding is their dominant strategy), the resulting change in contributions would cease to be equivalent with an increase in cooperation. The true treatment effect of taxing deliberation would then remain unidentified.

As a test for this confound, we employ a behavioral measure of confusion that enables us to observe the effect of time pressure on confused contributions at the moment of decision making. In additional treatment conditions (between-subjects and within-subjects) we apply time pressure in a task that retains the same basic payoff structure of the standard PGG, but removes all gains from or opportunities for cooperation. Human interaction partners are replaced with pre-programmed computer agents, which receive no payment from the public account (Houser and Kurzban, 2002). Given these incentives, confusion is the remaining explanation for contributions.

Contrary to our initial hypothesis we find no evidence that time pressure increases the level of confusion. However, in line with previous findings (Houser and Kurzban, 2002; Ferraro and Vossler, 2010) confusion is high (with and without time pressure), indicating that only approximately 50% PGG contributions can be confidently attributed to cooperation. When retesting for heterogeneous treatment effects, we find that time pressure selectively affects unconfused subjects. Specifically, and in contrast to the results of Rand et al. (2012, 2014), we find that time pressure reduces average contributions and significantly increases free-riding. These effects are strongly increased when controlling for confusion and compliance with the time restriction. Therefore, the causal evidence of the third article conforms with the notion of article 2 that deliberation promotes cooperation.

In the fourth article **"Smart or selfish? When smart guys finish nice"** I analyze parts of the same dataset as in article 3. Moreover, article 4 contributes to the overarching research question of the previous two articles whether deliberation can be linked to cooperative behavior by adding two further aspects. First, it employs a complementary method to assess individual capacities for deliberation. Second, it explores how deliberative capacities and features of the decision environment (i.e., external constraints on using these capacities) interact to shape cooperative behavior. Specifically, it uses a standardized test that identifies subjects' ability to base decisions on reflective rather than intuitive thinking. This ability, which forms a subcategory of a broader set of cognitive abilities relevant for economic decision-making, is assessed by the cognitive reflection test (CRT) (Frederick, 2005).

If deliberation was an important determinant of cooperative behavior (as the results of article 2 and 3 would suggest), contributions in a PGG should be higher among subjects who are more prone to base their thinking on reflection, as measured by a high CRT-score. I test this hypothesis in a standard one-shot PGG without a time constraint. In a second step, I explore how a predisposition towards reflection plays out under time pressure, which could constrain the use of higher reasoning abilities. This contributes to a better understanding how individual traits (i.e., reasoning style) and the decision environment interact to shape heterogeneous contribution behavior. In particular, I explore whether time pressure weighs more heavily on subjects with a high CRT-score, since they would be more prone to rely on deliberation in an unconstrained decision environment or on subjects with a low CRT-score, since they might be less able to cope with the time limit. Finally, I conduct a test whether subjects with a higher CRT-score are more able to identify their dominant strategy of contributing zero. If true, this would complicate the interpretation of the previous results.

In line with the results of the second and third article, I find that subjects with a higher CRT-score contribute more in a one-shot PGG, when there is no time limit. This result is surprising, since most existing studies on cognitive abilities have shown that subjects with a high CRT-score are typically more likely to chose a dominant strategy, if confronted with a dominance solvable decision task (e.g., a beauty contest game) (Brañas–Garza et al., 2012; Grimm and Mengel, 2012). While these extant results could be associated with subjects' understanding of the strategic environment, I find only weak evidence that participants with a high CRT-score display a better understanding of the PGG task. This affirms the results of article 3 that, after controlling for confusion, deliberation drives cooperation. When subjects have to decide under a time limit, there ceases to be a difference in average contributions between those subjects with a high and those with a low CRT-score. This result is mainly driven by a strong effect of time pressure on subjects with a high CRT-score, decreasing their contributions relative to baseline subjects in the same CRT-score category. Taken together, this underlines the higher propensity to cooperate among subjects with higher reasoning abilities, as long as the decision environment allows for the use of these abilities.

The fifth article **"Take from one to give to another"** examines whether individuals will contribute to a public good, even if doing so imposes harm on an outsider. While most PGG experiments have analyzed settings with uniform benefits, the provision of real public goods (including measures to mitigate climate change) can reduce the welfare of some outsiders. Such negative provision externalities could arise in two ways: First, preferences for the public good could be polar, so that some individuals would actually prefer lower or even negative provision levels. Second, in some cases third parties can be forced, implicitly or explicitly, to contribute to a public good, even though they are excluded from its benefits.

From a behavioral economics perspective, two stylized features render this set-up an interesting object of study. On the one hand, decision-makers holding social preferences are now confronted with a conflict between taking from one, yet doing good for another. This form of moral ambiguity, which is not uncommon if individuals have heterogeneous preferences, is rarely included in models of other-regarding behavior and the associated experiments. On the other hand, the share of contribution costs accruing to the decision maker himself can be greatly reduced by the presence of outsiders (for instance, because outsiders have to finance a provision subsidy). An extensive literature finds that reducing the price of giving increases contributions. However, this literature has ignored the question of whether the origin of the price reduction (voluntary vs. involuntary) affects the strength of this effect. When both features are present, the contribution choice could depend on a non-trivial interaction between price effects and the degree to which other-regarding preferences extend to affected third parties.

This setting motivates a simplified contribution experiment, which constitutes one of the first empirical inquiries into the question how third-party externalities affect public good provision. The experiment proceeded in two parts. In the first part, subjects could earn an endowment in a real-effort task. The second part was a contribution task in which this endowment could be spent on a public good. Specifically, like in articles 1 and 2, contributions were used to reduce CO_2 emissions. There were two main treatment variables: the price of giving for the decision-maker and contribution costs (if present) carried by another participant. In total, each participant went through fifteen allocation decisions in which both the price of giving and the intensity of harm inflicted on the outsider varied. Based on the predictions of reigning theories of other-regarding behavior, these allocations tasks were designed to discriminate between different motives that could apply to an outsider.

My results confirm earlier findings, namely that reducing the price of giving leads to higher contributions – if no outsider is present. When an outsider is harmed, my results are twofold. Welfare effects on the outsider are mostly ignored, rejecting linear altruism as a potential motive extending to the outsider. Only when contributing strongly affects (advantageous) inequity, decision-makers refrain from contributing. However, this reduction is large enough to almost fully moderate the price effect.

These results have several implications. First, for standard models of giving (Andreoni, 1989, 1990) they pose the intriguing question how to incorporate trade-offs between decision-makers and outsiders. For instance, does a warm glow of giving still materialize, even if giving implicitly means taking from another person? This could also have implications for the efficacy of policies which subsidize the PPPG, as long as the cost of the subsidy is a factor salient in the calculus of the decision-maker. For the context of climate change these findings highlight the potential conflict between distributional preferences within a generation and across different generations. While climate policy is often motivated by the latter, ignoring fairness considerations captured by distributional preferences within a generation (such as inequity aversion) might undermine its effectiveness.

Outlook

Since each of the five articles will draw separate conclusions regarding its specific research questions and results, there will be no additional concluding section at the end of this dissertation. Instead, the following paragraphs will briefly summarize the joint contribution of the five articles with respect to the common themes of this dissertation. Thereby, I will also provide a short outlook on some open questions that might continue to inspire research on cooperation as an economic, biological, and social phenomenon.

The first question to which several articles of this dissertation speak is the question of *who cooperates*. One way to think about this question is whether there exist observable individual attributes, which are systematically related to unobservable and heterogeneous preferences for cooperation. For instance, it has been shown that differences in cooperative preferences can be linked to a person's gender (Croson and Gneezy, 2009), age (List, 2004) or personality (Volk et al., 2011). Articles 3 and 4 add a person's cognitive abilities to the previous list, as a factor that has, as yet, received only limited attention as a determinant of cooperative behavior. If generalizable to the population level, such a list of factors could be of particular interest to policy makers who wish to tailor their regulatory instruments to different subpopulations. It is conceivable that there are specific subpopulations in which regulation is likely to crowd-out an intrinsic motivation to cooperate and conversely other subpopulations in which the same regulation could be an important prerequisite for cooperative outcomes. Similarly, managers could draw on such insights in order to build more efficient teams by selecting members based on observable characteristics. To the researcher, in turn, it might be of a somewhat lower interest *that* older, more agreeable (according to a personality test), and more deliberative (according to a test of cognitive abilities) females tend to cooperate more in economic experiments, but instead of higher interest *why* such relationships could have emerged. Existing answers to this more fundamental question are far from comprehensive, and thus are likely to motivate further inquiries into the biological (Buser, 2012) and cultural (Henrich et al., 2004) roots of cooperative behavior.

While these findings as well as the way in which social-preference models are typically formulated invite to categorize individuals as belonging to different "cooperative types", the within-subjects analyses contained in most articles of this dissertation paint a slightly more nuanced picture. Mirroring a long-standing controversy in psychology (Ross and Nisbett, 2011), article 1 shows that the same participants who cooperate in one situation (PGG), decide to free-ride in another situation (voluntary climate actions). This highlights the importance of context, often deliberately excluded in abstract laboratory experiments, in shaping cooperative behavior. Furthermore, articles 2 – 4 raise the possibility that cooperation is not only a question of "type", but could rather vary at the individual level depending on the cognitive demands imposed by the decision environment. Finally, article 5 demonstrates that decision-makers do not readily extend the same other-regarding preferences they apply to one group of individuals to another individual. Taken together, these observations tentatively suggest that cooperative preferences could be less stable at the individual level and less generalizable over different domains than theory would typically assume. Therefore, a more thorough investigation of the stability of cooperative preferences across situations (Galizzi and Navarro-Martinez, 2015), across abstract laboratory tasks (Blanco et al., 2011; Peysakhovich et al., 2014) or across time (Volk et al., 2012; Carlsson et al., 2014) could provide important insights into the scope of "cooperative types"at the individual level.

The second question to which several articles in this dissertation speak is the question *when* cooperation occurs and *how* it could be promoted. In particular, articles 2–4 suggest that the average level of cooperation might depend, among other things, on the cognitive demands imposed by the decision environment. Before these findings can provide clues regarding the optimal design of "cooperative decision environments" several further steps of research are certainly warranted. First, while the link between cognitive processes and cooperation appears to be robust across most of the existing studies (including articles 2–4), evidence on the direction of such link is still mixed. This calls for a more thorough investigation of potential moderators that would clarify under which circumstances deliberation can promote and under which circumstances it is more likely to impede cooperation. For instance, one candidate moderator deserving more attention could be the social and psychological distance between those paying for and those benefiting from cooperation. It might well turn out that cooperation between in-group members is intuitive, while cooperation with out-group members requires more deliberation. Another moderator that seems plausible is whether the benefits of cooperation are perceived as rather abstract or as more concrete. Second, article 4 shows that cognitive abilities are linked to cooperative behavior. In this respect, it seems important to clarify to what degree such abilities can be seen as a malleable trait or as a fixed and heritable state (Cesarini et al., 2008; Heckman and Kautz, 2013). Third, as soon as there is robust evidence on the direction of a link between cognitive processes and cooperative behavior one can think about how decision environments might affect the use of specific cognitive processes. In particular it would be of interest to isolate external factors that limit the use of cognitive abilities in economic decision environments. Time pressure, multitasking or other sources of stress are obvious factors. However, recent evidence demonstrates how also less obvious factors can affect the use of cognitive resources (Mullainathan and Shafir, 2013). In particular, this research shows how a (perceived) scarcity of resources, financial or otherwise, can tax the cognitive bandwidth of decision-makers. Extending such findings to cooperative behavior would seem to be a logical next step.

Chapter 2

What do we learn from public good games about voluntary climate action? Evidence from an artefactual field experiment*

*Co-authored by Timo Goeschl, Sara Elisa Kettner, and Christiane Schwieren. The authors gratefully acknowledge financial support by the German Ministry for Education and Research under grant 01UV1012. Furthermore we would like to thank the audiences at ESA New York, ESA Cologne, AURÖ Bern, ZEW Mannheim, WCERE Istanbul, and the IfW Kiel for their valuable comments.

2.1 Introduction

Economists typically treat climate change mitigation as a public goods problem (Nordhaus, 1991). Consequently, most theoretical models of voluntary mitigation efforts predict that free-riding is the dominant individual behavior. Empirically, however, public good game (PGG) experiments and other social preference tasks have amassed convergent evidence that free-riding may be less prevalent in social dilemmas than predicted (Ledyard, 1995; Zelmer, 2003; Chaudhuri, 2011; Vesterlund, 2014). Does this experimental evidence give reason to rethink the premises of climate policies that are designed with large-scale free-riding in mind? And more generally, can PGG and variants thereof serve as a reliable testbed for predicting behavioral responses to climate change policies?

A number of recent papers tentatively argue that findings from PGG experiments could provide important insights into mitigation behavior and policies in the real world (Shogren and Taylor, 2008; Venkatachalam, 2008; Brekke and Johansson-Stenman, 2008; Gowdy, 2008; Gsottbauer and van den Bergh, 2011; Carlsson and Johansson-Stenman, 2012). In the same spirit, some experimental studies on public good provision have been framed or interpreted with an explicit reference to institutional mitigation decisions (e.g., Milinski et al., 2006, 2008; Tavoni et al., 2011; Brick and Visser, 2015). Such experiments present a theoretically appealing method for obtaining causal evidence at low cost. However, whether or not such PGG experiments can truly provide the desired valuable insights crucially depends on their *generalizability* (Levitt and List, 2007), i.e., the degree to which generic behavior, based on observing subjects in an abstract lab task, transfers to the specific context of mitigation decisions. Whether and under which conditions behavior in a PGG experiment generalizes to voluntary mitigation choices is, at heart, an empirical question. In the present paper, we take a first step towards providing an answer.

Concerns that subjects' behavior in abstract game forms under controlled conditions in the laboratory may not generalize to individual behavior in context-rich situations outside the lab are not new. But their recent recurrence in the context of whether social preferences elicited using standard experimental designs are predictive beyond the lab (Levitt and List, 2007), is particularly relevant for issues of public goods provision such as voluntary mitigation choices.[1] 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)

[1]Levitt and List(2007) describe a number of 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 policy makers than for experiments that try to falsify a theory (Schram, 2005; Sturm and Weimann, 2006; Kessler and Vesterlund, 2015).

varies substantially across studies. On the basis of available evidence, generalizability of behavior in the PGG to voluntary mitigation choices can therefore neither be ruled in nor out.

The climate context, to which one hopes to generalize PGG evidence, provides additional reasons for concern. It could be argued that the deliberately abstract format of the PGG does not capture context-specific 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, at least in theory, should also shape voluntary mitigation decisions. On the other hand, the experimental paradigm of the PGG can accommodate considerable variation in design features. For instance, a greater resemblance to voluntary mitigation decisions might result from simple changes to design parameters such as the group size or the productivity of the experimental public good. If such variations are able to capture most of the relevant drivers of mitigation decisions, then generalizability may be accomplishable at acceptable cost.

This paper brings new experimental evidence to two of the issues raised above. 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, as opposed to unobserved idiosyncratic motives. Such explanatory power of sufficient size is an important prerequisite for a high level of generalizability (Al-Ubaydli and List, 2015). The empirical problem is that the totality of individual mitigation behavior, just like the totality of an individual's charitable behavior towards others, is not observable for the researcher.[2] Following other examples in the literature (Benz and Meier, 2008; Laury and Taylor, 2008; de Oliveira et al., 2011; Voors et al., 2012), we approximate the ideal test by conducting a laboratory experiment in which we observe each participant in two contribution situations: A public goods game and a real giving task in which contributions are used to reduce CO_2 emissions.

The second issue that we examine within this framework is whether abstract PGG experiments can be implemented in a way that increases the generalizability of its output in the direction of voluntary mitigation choices. We do so by experimentally varying two central design features of how PGG evidence is generated, namely its parameter structure and the subject pool. The systematic variation of PGG parameters, in particular group size, marginal per-capita return (MPCR), and payoff symmetry, allows us to test whether generalizability varies with different degrees of structural resemblance between PGG contribution incentives and voluntary mitigation incentives. The comparison of behavior across two samples, one a sample of students and one recruited from

[2]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 a direct or indirect mitigation 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

the general population, allows us to test whether generalizability in a climate context perhaps hinges on subject pool. It is well know that student samples, which account for the majority of PGG evidence, share only a limited range of individual attributes with the general population. As a result, the extent to which the behavior of the former allows conclusions about the latter 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) and at the same time the source of uncertainty over its generalizability to mitigation actions.

Our results suggest that PGG behavior can be indicative of voluntary mitigation decisions, but not in a uniform fashion. Instead, the potential for generalizability crucially depends on the way the PGG is designed and conducted. In a benchmark case employing common PGG parameters, the correlation between contributions in both task is small and insignificant. This result holds irrespective of the subject pool. A low correlation indicates that there exist idiosyncratic drivers of mitigation behavior that remain unobserved in standard PGG. Yet, when PGG parameters resemble more closely the incentive structure underlying voluntary climate change mitigation, correlations - especially those for student subjects - become significant and sometimes sizable. Thus, by implementing simple design changes, some of the apparent differences in individual behavior disappear. This points towards a cost-effective and feasible way of improving current insights into the institutional mechanisms affecting voluntary mitigation behavior that can be gained via laboratory experiments. On the other hand, switching to a subject pool of non-students has more ambiguous effects. In line with previous results, we find that on average, non-students contribute more in both tasks. However, as indicated by strongly reduced correlations, the degree of generalizability is much lower within this more heterogeneous sample. This underlines the existence of a trade-off between representativeness and generalizability unless the apparatus of the experimental design or the sample size are significantly enlarged - at a cost.

The remainder of the paper is organized in the following way: Section 2.2 discusses our research question in relation to the existing literature. In Section 2.3, we describe the experimental set-up and the characteristics of our subject pool. Section 2.4 contains the analyses and core results. Section 2.5 concludes with a discussion of our findings.

2.2 Related literature

There are several studies that examine issues of generalizability, both regarding the relationship of social preferences measured in different abstract lab tasks (e.g., Public Good Game; Dictator Game; Trust Game) and regarding the predictive power of cooperative behavior observed in PGG towards contributions made to a variety of naturally occurring public goods. We follow these studies in their common methodology of employing a within-subjects design.

So far only few studies have analyzed how cooperation in public good games corresponds to social preferences elicited in other abstract tasks. Overall, these studies arrive 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 good game spills over to other abstract game formats. More cooperative subjects are shown to be significantly more likely to give higher amounts in a dictator game and to reciprocate trusting behavior more strongly in a trust game. They furthermore find that more cooperative subjects are also more prone to help the experimenters after the actual experiment, by voluntarily completing an additional questionnaire. Finally, in Galizzi and Navarro-Martinez (2015) public good game behavior is moderately, but significantly correlated with behavior in trust and dictator games.[3] This first strand of 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.

A second strand of literature addresses the same basic question as our paper by investigating the relationship between contributions observed in a laboratory public goods game and contributions to a naturally occurring public good. As in our experiment, these studies largely lack a direct and unintrusive measure of cooperation in the field.[4] 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 an 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 public goods game to the amount they contribute to a real community public good. Closest to our own question, Laury and Taylor (2008) investigate student behavior in a variety of the linear public good game 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) relationship. In a comprehensive literature review, Galizzi and Navarro-Martinez (2015) 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).

[3]They, however, detect no significant relationship with helping or donation behavior in five different field situations which are randomly administered subsequent to the actual experimental sessions.

[4]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.

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 - with the exception of 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 an unified design in which we observe 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.[5]

2.3 Experimental design and implementation

Questions of generalizability from one experimental task to another are typically addressed by a within-subjects design (Benz and Meier, 2008; Laury and Taylor, 2008; 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 good decisions in which we vary experimental parameters along three dimensions (Goeree et al., 2002).[6] In the following, we describe each of the decision tasks in more detail.

[5]Note, however, that the design is explicitly not intended to resolve the broader controversy (Levitt and List, 2007, 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.

[6]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.

2.3.1 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).[7] Prior to 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 at least the same 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. In order 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 personal computer) and the average CO_2 emissions of a German citizen. The instructions also confirmed the public good 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.

2.3.2 Task II: The laboratory public goods game

The average rate of cooperation in PGG has been found to be responsive to changes in experimental parameters such as the group size, the marginal per capita return (MPCR), or the symmetry of payoffs (Isaac and Walker, 1988; Goeree et al., 2002; Nosenzo et al., 2015). We hypothesize on this basis that the choice of these parameters affects the degree of generalizability. To test this proposition, we employed a variant of the standard public goods game (Goeree et al., 2002): Subjects were anonymously and randomly

[7]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.

matched into groups of varying size and completed ten independent one-shot contribution decisions without feedback, displayed on one common decision screen.[8] In each of these decisions participants had to choose how many tokens from their initial endowment they wanted to invest into a public account. Depending on the total number of tokens invested, every public account produced payoffs determined by a distinct combination of MPCR, group size, and payoff symmetry. Table 2.1 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. In the nine other decisions, the parameter constellation systematically shifted the contribution incentives such that they structurally resembled, to greater or lesser degree, those present in voluntary mitigation decisions. In contrast to the benchmark case, the contribution incentives there are characterized by the fact that the 'group of players' is large, the MPCR is small, and payoffs are asymmetric.

The general payoff structure for individual i is summarized by the following expression:

$$\pi_{it} = v(\omega - x_{it}) + m_t^{int} x_{it} + m_t^{ext} \sum_{j}^{N_t - 1} x_{jt}; \forall i = 1, ..., 12/15; \forall t = 1, ..., 10 \qquad (2.1)$$

where v is the value of a token kept and ω is the initial endowment of tokens. t is a subscript denoting each decision and x_{it} is individuals i's contribution to the public account. m_t^{int} and m_t^{ext} are the internal and external value of a token invested in the public account, respectively. For each token subjects invest in the public account they receive m_t^{int} and transfer m_t^{ext} to every other group member. Cases where $m_t^{int} = m_t^{ext}$ are therefore equivalent to a linear PGG with symmetric payoffs. N_t denotes the number of subjects within a group.

In each decision, tokens remaining in the private account yielded a payoff of $v = 20$ Eurocent and subjects were initially endowed with 20 tokens. As the internal returns are always smaller than v, free-riding is a dominant individual strategy. From the group's perspective, it is efficient to contribute the full endowment. Decisions a-d feature parameters that structurally resemble those for voluntary mitigation decisions (small MPCR, larger group size, and asymmetric payoffs) more than those of the benchmark decision f and decisions g-j.[9]

[8]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.

[9]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.

TABLE 2.1: 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 good experiments. The MPCR for each decision is calculated by the following formula: $\frac{1}{Nv}(m_t^{int} + (N-1)m_t^{ext})$

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, prior to entering the decision screen. In these calculations, there was no pre-specified contribution level to avoid setting a standard. At the end of the experiment, one decision was picked randomly with equal probabilities and payed out to the participants. This randomization of payoffs (Starmer and Sugden, 1991) has the advantage that subjects cannot condition their behavior in a given decision on their other choices.

2.3.3 Recruitment and sample characteristics

Participant were recruited from two distinct pools. We compare students to non-students in order 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.[10] 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.[11] The student sample was recruited from the standard subject pool using ORSEE (Greiner, 2015). 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. Naturally, both subject pools consist of self-selected subjects. While this is standard

[10] The *"Rhein-Neckar-Zeitung"* is sold at a price of 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.

[11] 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.

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, 2007). 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 2.2 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 2.2: Demographic attributes of different subpopulations

Demographics	**Total** N=135	**Student** N=43	**Non-Student** N=92
Age (Years)	40.91 (18.76)	22.83 (3.01)	49.36 (16.96)
Gender (1=male)	0.37	0.41	0.35
Individual Net Income (Euro)	1050.83 (902.74)	613.15 (228.59)	1253.65 (1020.73)
Assets (1=Yes)	0.25	0.02	0.36
Education (Years)	14.22 (2.67)	13.86 (1.95)	14.40 (2.94)
Household Size (#)	2.02 (1.44)	1.85 (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)	4.27 (2.72)	4.32 (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")

2.3.4 Experimental procedures

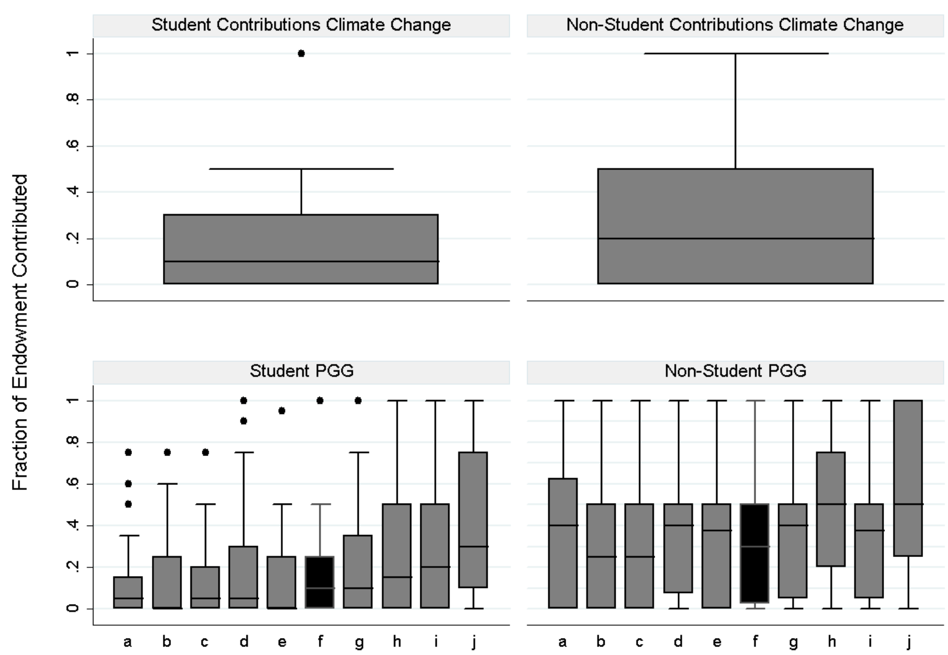
All ten sessions took place at the University of Heidelberg "AWI–LAB" using z-Tree (Fischbacher, 2007). There were either 12 or 15 participants per session. At the beginning of a session, participants were seated at one of the available computer terminals, separated by a divider. A printed version of 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. Average payment was 17.65 € and ranged from 2.68 € to 26.00 €.[12]

2.4 Results

2.4.1 Observed behavior

FIGURE 2.1: 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.

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

Figure 2.1 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.[13] 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; $p < 0.001$).

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[14] confirm that the fraction of endowment contributed increases significantly with group size ($\beta_1 = 0.021$; $p < 0.001$) and internal ($\beta_2 = 0.030$; $p < 0.001$) 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.

2.4.2 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

[13] 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.

[14] We estimate a random effects tobit model controlling for the student status and the set of demographic attributes listed in table 2.2. Full results are shown in the Appendix table 2.8.

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 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.

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

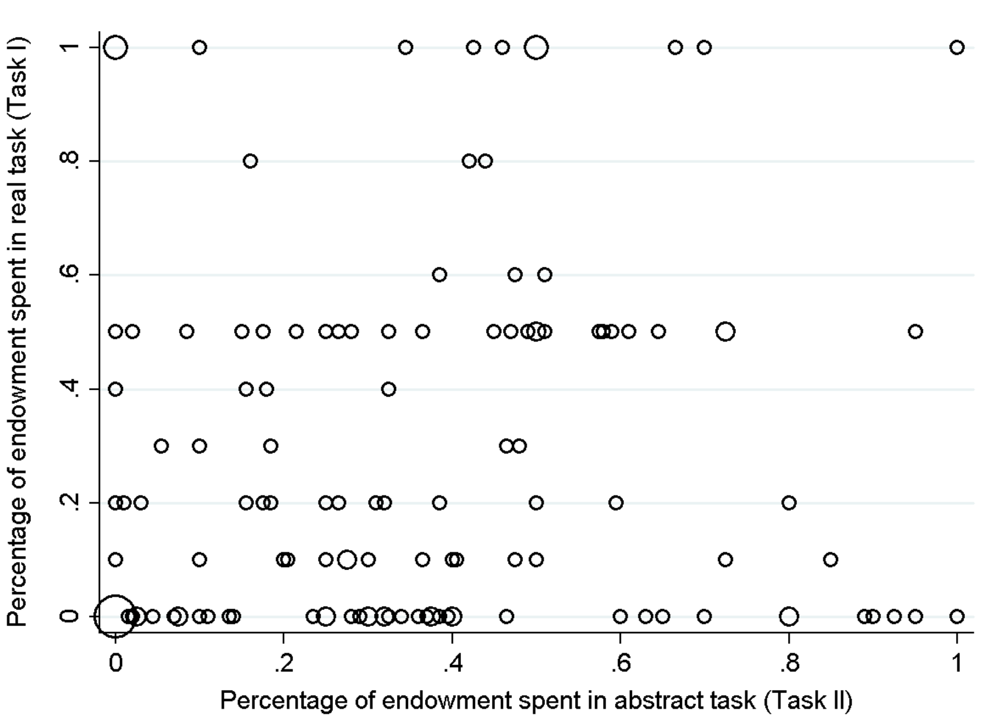
For a simplified first analysis of the relationship between the two tasks, we follow Laury and Taylor (2008) and initially ignore the variation of parameters between the different decisions of Task II. To broadly summarize contribution behavior, we calculate the mean over the ten distinct public good 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 2.2. 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. 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, group-size 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 2.3 displays the corresponding correlation coefficients for the pooled sample of students and non-students.

FIGURE 2.2: Bubble plot 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.

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 public good games. Therefore, 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 2.3 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.[15] 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

[15]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)

to structurally resemble the specific choice context can make an important contribution towards generalizability.

TABLE 2.3: 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.
Significance Levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To further evaluate the potential for generalizability, we now turn to the size of the significant correlation coefficients in table 2.3. 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 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 2.2, we find that the degree of generalizability from Task II to Task I is not smaller than that of PGG contributions to behavior in a number of other context-free social preference tasks. The significant correlations in table 2.3 squarely fall into the range [$r = 0.07$;$r = 0.41$] reported in Blanco et al. (2011) and Peysakhovich et al. (2014).[16]

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

[16]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.

$\alpha = 0.94$) and correlations between single pairs of decisions range from $r = 0.43$ to $r = 0.85$.[17] 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 2.3. 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 2.3. 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 (average) 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

[17] A full correlation table can be found in the Appendix table 2.6.

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 $r_s = 0.2573$. The smallest insignificant correlation $r_s = 0.1153$ is again found in decision j which is characterized by the highest MPCR.[18]

2.4.3 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 2.1 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 2.2 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.

[18] A full table containing decision-wise correlations for free-riding can be found in the Appendix table 2.7.

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 2.4, this disparity is not driven by a single PGG decision. Instead, irrespective of the parametrization, for non-students all correlations are very low.

TABLE 2.4: 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 2.1. Including this outlier reduces correlation in size.
Significance Levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 2.2, we still find significant correlations only for student subjects.[19]

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

[19] 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 2.4 for student subjects. For non-students this hypothesis cannot be rejected for any decision.

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.[20]

2.4.4 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 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.

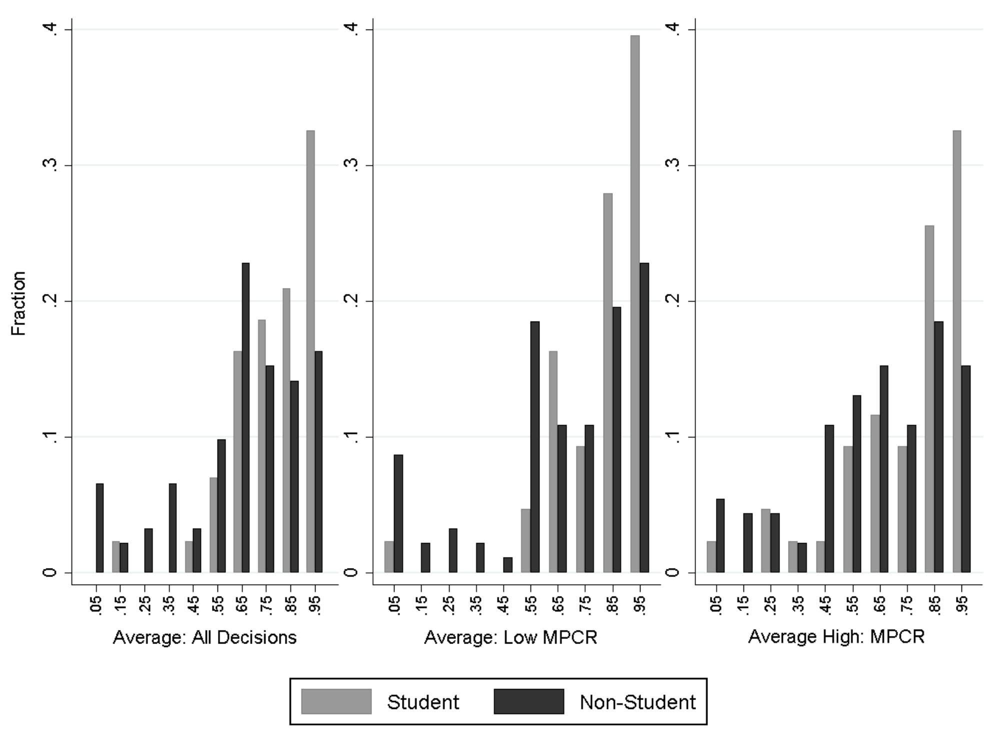
Result 5: Quantitatively, subject pool effects outweigh the effect of game parameters in explaining individual consistency. These differences cannot be attributed to observable characteristics.

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 exactly 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

[20] A full table containing decision-wise correlations for free-riding can be found in the Appendix table 2.7.

two tasks these factors worked in opposite directions, the measure would tend towards zero.

FIGURE 2.3: Distribution of average consistency



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

Figure 2.3 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.[21] The figure reveals similar patterns as the previous sections, but also highlights the extent of individual heterogeneity. A considerable share of participants conform 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

[21] Each of these average measures is calculated according the following formula using the notation introduced in Section 2.3, with g_i denoting the fraction of endowment contributed by individual i in Task I:

$$c_i^z = 1 - \frac{1}{T}\sum_t^T \left|\frac{x_{it}}{\omega} - g_i\right| \tag{2.2}$$

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 2.5.

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 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, with the exception of 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

Table 2.5: 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
Prob > Chi^2	0.000	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. For each specification the dependent variable is one minus the absolute difference between behavior in Task I and Task II in percentage terms.

specific differences have a larger impact on the overall consistency than differences in the parameterization for the range of values we observe.

2.5 Discussion and conclusion

In the past decades, experiments have started to play an increasingly important role in economic research. In line with this development, there is also a growing interest in drawing on experimental methods and evidence to illuminate concrete policy debates, such as those surrounding climate change mitigation (Bohm, 2003). We agree that in this regard much could be learned from experiments, as they offer a cheap and feasible way to gain insights into the behavioral responses to novel policies within a controlled environment. But for experiments motivated by specific policy issues, generalizability becomes a central issue (Schram, 2005; Sturm and Weimann, 2006).

Our analysis highlights that heterogeneity in mitigation decisions is indeed partly attributable to generic cooperative preferences, but also depends on idiosyncratic factors. Of course, for policy advice the main advantage of experiments lies in their ability to isolate the effects of a particular treatment variation on behavior. Given that in our experiment a considerable fraction of individual mitigation decisions are driven by latent variables not observed in the PGG (especially when using standard parameters), it is not obvious whether treatment effects would be highly transferable between these two settings, in a quantitative and maybe even in a qualitative sense.[22] Clearly, this does not mean that such concerns materialize necessarily for all treatment effects of interest. 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).

Importantly, we do not see our finding as discarding the application of (abstract or context-specific) experiments to questions of climate change policy. Rather to the contrary, they call for more experimentation, in the spirit of the arguments raised in Falk and Heckman (2009). Only by obtaining further experimental evidence one can shed additional light on the conditions under which one can safely assume a high level of generalizability. We make a first step into this direction within our own framework and explore two potential shifters of generalizability. Our first treatment variation suggests that the link between PGG behavior and mitigation decisions can be strengthened by bringing the experimental parameters closer to the context of interest. For PGG with a low MPCR the correlation between behavior in both tasks increases, sometimes even substantially. Consequently, in the limit, the best laboratory equivalent to individual mitigation behavior might well turn out to be the standard dictator game in which

[22] As highlighted by Kessler and Vesterlund (2015), a discussion about qualitative transferability might be more fruitful. However, even for qualitative treatment effects with an unknown underlying causal mechanism (Heckman and Smith, 1995; Imai et al., 2011) the potential for transferability is hard to assess, because it is not clear which latent factors (common or idiosyncratic) link the treatment variable to the outcome.

the dictator's private return of contributing is zero. 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. 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.

From our second treatment variation we derive more ambivalent conclusions, regarding questions of generalizability. If it was a central aim to make statements about the level of cooperation, the use of a convenience student sample could be somewhat misleading. We replicate earlier findings that student behavior is only a lower bound for the cooperative behavior that can be expected in a population with broader demographic heterogeneity. On the other hand, we show that students are more responsive to changes in experimental parameters (or conversely less responsive to differences between the tasks) and consequently display a higher consistency between the different decision tasks. Thus, sampling from the general population, with the aim to draw from a more representative subject pool might 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.

Clearly, our experiment is only a first step towards understanding generalizability in the narrow context of climate change mitigation. The larger question, namely whether social preference tasks are generally external valid, cannot be resolved on its basis as our results are, by design, context-dependent. A relevant extension of our design would replace Task I with an actual measurement of voluntary mitigation behavior in a field environment. Such a measure would differ from Task I along several dimensions. Mitigation decisions outside the lab context require individuals to use their own money instead of an experimental endowment, are not scrutinized by an experimenter but instead (in some cases) by the social environment and are often bundled with other attributes of a consumption decision. Each of these shift parameters reduces the artificiality of Task I relative to Task II. It is left for further research to assess how this would affect conclusions about generalizability.

2.6 Appendix

Correlation table task II: Decision a.-j.

Table 2.6 contains the correlation coefficients for each pair of decisions made in task Task II.

TABLE 2.6: 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

Correlations free-riding

Table 2.7 contains Spearman correlation coefficients between free-riding in Task I and Task II. 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 2.7: Spearman correlations between free-riding in the real and in the abstract context for all 10 decisions

(0) Decision	(1) Group Size	(2) Symmetry	(3) MPCR	(4) Correlation	(5) Correlation Non-Students	(6) 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.
Significance Level:* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Regression results task II

Table 2.8 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 2.8: Contributions abstract PGG 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.

2.6.1 Instructions

<table>
<tr><td>Check-in</td><td>Check-in room:

Sign-in and generation of personal code</td></tr>
<tr><td>Experiment</td><td>Laboratory:

Random seat assignment

General instructions read out loud (page 45)

Tasks implemented in z-Tree

• Contribution to climate change mitigation (page 46)
• Laboratory public goods game (page 49)

Payment receipt distributed according to personal code</td></tr>
<tr><td>Payment</td><td>Check-in room:

Subjects exchange payment receipt for cash</td></tr>
</table>

General instructions

General instructions were handed to participants as a print-out and were read aloud by the experimenter.

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 within 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 will receive some general information regarding the procedures. You will decide in several tasks. Please follow the instructions on the screen.
At the end of today's experiment you will receive a 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 <u>anonymous</u>. 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 personalized information on your decisions and payments.
– – –

General instructions
[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.
– – –

General instructions
[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 instance, 70 kg roughly correspond to the amount of CO_2 emitted when driving 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).
– – –

Start real contribution task
[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.
– – –
Decision Screen
[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 any integer between 0 and 10 Euro, i.e., you may fill in whole numbers without decimal place (period or comma).
- You will receive each Euro, not used to purchase CO_2 permits, in cash at the end of the experiment.

<insert decision>
<summary screen displayed>
– – –

Laboratory public good game

Instructions I
[SCREEN 6]

Explanation:
In this task you have the possibility to earn further payments, in addition to the ten Euro you already received at the beginning. Furthermore, during this task you interact with 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.
– – –
Instructions II
[SCREEN 7]

Possible decisions:
In the following 12 decisions you can distribute 20 balls between two bowls which are labeled as 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.
– – –

Example
[SCREEN 8]

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
[SCREEN 9]
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.
Control Questions (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 examples. The actual task will be presented in a table which we explain to you on the following screen.
– – –

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 table above is an example and illustrates how the subsequent decision task will be displayed.
The above table shows only 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.

Chapter 3

Giving is a question of time: Response times and contributions to an environmental public good*

*Co-authored by Timo Goeschl and Johannes Diederich. The authors gratefully acknowledge financial support by the German Science Foundation (DFG) under grant GO1604/1 and the German Ministry for Education and Research under grant OIUV1012. Furthermore, we would like to thank the audiences at IMEBESS Oxford, RGS Bochum, HSC New York, RES Manchester and ZEW Behavioral Environmental Economics Workshop Mannheim for their valuable comments.

3.1 Introduction

A number of articles suggest that much could be learned from advances in behavioral economics when it comes to assessing individual preferences for environmental public goods (Carlsson, 2010; Croson and Treich, 2014). Important contributions in this context have highlighted the possibility that choices in stated or revealed preference studies might not fully reflect the decision makers' true preferences but instead partially express their various cognitive biases—even in the presence of non-trivial financial stakes (Kahneman et al., 1990; Hanley and Shogren, 2005; Beshears et al., 2008). Related to such concerns are recent insights into the different cognitive processes that individuals employ when deciding about public good contributions. In particular, a series of converging experimental findings suggests that some decision makers follow a first intuition (i.e. a simple heuristic) while others make more deliberate choices when resolving a trade-off between selfish and other-regarding motives (e.g. Piovesan and Wengström, 2009; Kocher et al., 2012; Rand et al., 2012, 2014; Nielsen et al., 2014; Kessler and Meier, 2014; Ubeda, 2014). These two principal types of cognitive processes are distinguished by dual-system theories and their applications to economics (Fudenberg and Levine, 2006; Loewenstein and O'Donoghue, 2007; Kahneman, 2011; Dreber et al., 2014): System I, which arrives at decisions through fast and intuitive processes, and System II, which generates decisions based on cognitively more demanding and hence slower deliberative reasoning (Kahneman, 2003; Evans, 2003, 2008; Loewenstein et al., 2008). These theories, and the experiments associated with them, raise the possibility that, everything else equal, individuals might come to a different contribution choice on the sole reason of their decision relying on intuition (System I) or deliberation (System II). If confirmed, this would provide important clues as to how aspects of the decision environment might influence the outcome of a valuation exercise, depending on whether a certain design choice, framing, or other aspect of the study induces more intuition or more deliberation. For example, evidence from stated preference surveys has shown that giving respondents additional "time to think" significantly reduces WTP estimates (Cook et al., 2012) and increases choice precision (Börger, 2015).

The possibility that individual contribution choices are the result of either intuitive or deliberative processes has led to a number of empirical investigations of this hypothesis, in particular, with respect to the question which of the two cognitive systems predisposes towards more cooperation. Most experimental research regarding these two questions involved abstract lab tasks. While their results jointly support the notion that an empirical link between the cognitive system and the choice to cooperate exists, the direction of the link remains disputed. Rand et al. (2012, 2014), for example, find that higher contributions in standard public good games are driven by intuitive decision making. Tinghög et al. (2013) and Verkoeijen and Bouwmeester (2014) fail to replicate this result, as do Duffy and Smith (2014) and Martinsson et al. (2014) using cognitive load or priming designs. In the closely related context of general fairness preferences, Piovesan and Wengström (2009) and Ubeda (2014) conclude that more generous allocations in dictator games are associated with deliberation. Again, Schulz et al. (2014) find the opposite when analyzing the effects of cognitive load, as do Cappelen et al. (2014) in a correlational study. In sum, therefore, the question of how intuition and deliberation relate to public good provision is far from settled and might in-itself depend on the context of the specific public good at hand.

The present paper adds empirical evidence to this discussion that—to our knowledge for the first time—originates from outside the laboratory. Specifically, it assesses the link between cognitive systems and contribution behavior in the area of environmental decision making by exploring contributions to voluntary climate change mitigation—the archetypal public good to environmental economists (e.g. Nordhaus, 1993). Methodologically, we follow the existing experimental literature and identify the unobservable cognitive processes that govern the decision through response time data (Piovesan and Wengström, 2009; Rand et al., 2012; Tinghög et al., 2013; Nielsen et al., 2014; Ubeda, 2014). This strategy relies on the fact that on average, response times (RT) differ between intuition and deliberation. When considering the consequences of a given choice or resolving a moral dilemma, faster decisions are more likely to be the result of intuitive processes while slower decisions are more likely to have involved deliberative reasoning (Rubinstein, 2007, 2013).

Our analysis is based on data from an extra-laboratory experiment[1] on voluntary individual climate action (Diederich and Goeschl, 2014) which allow us to analyze individual differences in RT between contributors and non-contributors. In this experiment, subjects from the general population faced a dichotomous choice between receiving a monetary payment and contributing to voluntary mitigation efforts. Voluntary mitigation efforts took the form of a guaranteed and verifiable reduction of CO_2 emissions by one metric ton (Löschel et al., 2013; Diederich and Goeschl, 2014). This unique dataset of choices and associated RT offers four distinct benefits: First, it is to our knowledge the first set of observations that allows a test of the link between cognitive system and contributions based on a real public good. The specific public good in question (individual voluntary climate action), renders it especially suited for answering the question of whether recent insights from behavioral economics bear implications for assessing individual preferences for environmental public goods. Second, with 3483 subjects, the number of independent observations is large compared to most datasets that examine this link. This is important in light of Rubinstein's (2007, 2013) dictum that the noisy approximation of mental processes through RT data requires large sample evidence. Third, observing a sample of subjects from the general population with a broad range of demographic backgrounds increases the representativeness of our results. Fourth, the dataset contains two RT observations per subject as each subject took two choices between different monetary rewards and mitigating one ton of CO_2. Hence, as in Piovesan and Wengström (2009), it is possible to analyze the within-subject relationship between

[1] Based on the categorization introduced in Charness et al. (2013).

RT and contributions while holding constant unobserved individual attributes or preferences. Thus, our analysis is able to go beyond simple correlation. Our results are threefold: First, we find a clear difference in response times between contributors and non-contributors in an extra-laboratory setting. This is evidence that the link between cognitive systems and contribution decisions detected in previous studies survives in a real choice situation outside the specific setting of a laboratory and when using subjects drawn from the general population. Secondly, we find that intuitive decisions are statistically associated with a choice not to contribute to climate change mitigation while a choice to contribute is more likely when the decision is deliberative. This result also lends support to earlier findings (Piovesan and Wengström, 2009; Ubeda, 2014) that deliberative processes favor pro-social choices. In our extra-laboratory setting with a large and diverse sample of subjects, this effect stands out clearly: The average response time of contributors, controlling for other factors, is approximately 40% longer than that of non-contributors. Thirdly, our finding carries over to the individual level: Subjects who switch from defecting in their first decision to contributing in their second decision need significantly more time for their second decision and vice versa. We interpret this as evidence in support of the hypothesis that voluntary contributions to climate change mitigation are driven by a deliberative weighting of personal costs and social benefits rather than by affect and intuition. Overall, our results suggest that the role of cognitive processes deserve closer scrutiny when assessing preferences for environmental public goods. With this goal in mind, the collection and analysis of supplementary response times data could turn out to be a cheap, unintrusive, and feasible method.

This paper is organized as follows: We summarize the experimental procedure in section 3.2 and present the results in section 3.3 before concluding with a discussion in section 3.4.

3.2 Experimental design

Our identification strategy for different cognitive processes rests on analyzing RT data (Piovesan and Wengström, 2009; Rand et al., 2012; Tinghög et al., 2013; Nielsen et al., 2014; Ubeda, 2014). These data have been collected alongside a previous experiment and exhibit several unique features that render them particularly suited for our purposes here. Among those are a large sample size, non-trivial financial stakes, and two consecutive observations per subjects that allow for a tight econometric control over unobserved individual factors that could otherwise bias the RT-contribution link. In the following paragraphs, we shall focus on the most important design features of the experiment. A more detailed description can be found in Diederich and Goeschl (2014).

3.2.1 Decision task

In the experiment reported on in Diederich and Goeschl (2014), subjects made two consecutive dichotomous choices, deciding each time between receiving a monetary reward or providing a real public good. The real public good took the form of a guaranteed and verifiable reduction of 1 metric ton of CO_2 emissions. Across all observations, there were slight variations in the specific terms of the emissions reduction while retaining the basic design of receiving a personal monetary reward versus contributing to a public good.[2]

The treatment condition in the online experiment consisted of randomly assigning subjects to different monetary rewards. For each subject and each of the two choice situations, the reward was independently drawn from a uniform distribution of even integers between 2 € and 100 €.[3] As a result, the data set contains significant between-subjects and within-subjects variation with respect to the trade-off between own interest (size of the monetary reward) and providing the public good. This variation forms the basis of robustness checks in our analysis, among them a check for the hypothesis that RT is determined by the degree of cognitive difficulty of the decision situation, rather than the cognitive system used (Krajbich et al., 2014, 2015; Evans et al., 2014). One potential issue with employing varying monetary rewards is that this could give rise to field price censoring (Harrison and List, 2004): Participants who would otherwise have chosen to cooperate might choose the monetary reward instead, as they believe that they are able to provide an equivalent reduction of CO_2 emissions at a lower total cost. Based on different robustness checks discussed in Section 3.3.3 we conclude that our results are not affected by this potential confounding factor.

3.2.2 General procedures

The experiment was conducted between May and July 2010 in collaboration with the large online polling organization "YouGov"Ṡubjects were recruited from their existing panel of 65,000 members via e-mail. After following the invitation link participants reached an introduction screen. This screen explained, as common with the pollster's regular surveys, the thematic focus of the poll and the expected duration (ten minutes).[4] Participants then faced a sequence of 10 to 13 computer screens, two of which were "decision screens" that required a choice between either a personal monetary payoff or a public good contribution. Each decision screen presented, through radio buttons, the

[2]There were four variations in total. For example, in some conditions, a contribution decision was made public after the session. Session effects are therefore explicitly included when analyzing pooled data in section 3.3. The main relationship between response times and contribution behavior is unaffected by the different variations.

[3]For each of the 50 reward categories, there are between 56 and 83 observations.

[4]The polling company usually incentivizes panel members participating in a poll through either a piece-rate reward of approximately 1 € for 20 minutes expected survey time or random (lottery) prizes, e.g. in the form of shopping vouchers.

binary choice between the public goods contribution ("reduction in CO_2 emissions of one metric ton") and the specific monetary reward (e.g."46 €") that had been drawn for the subject in this decision, with the order of the cash and contribution button randomly assigned. There was no default and subjects clicked on the desired radio button and on a "proceed" button directly underneath. Before the first decision screen an information screen introduced the specifics of the choice situation.

The RT data for the present analysis contain, for each subject, a measure of the time they spent on each of the two "decision screens" that form the core of the experiment. For each decision screen, a subject's RT is defined as the time between entering that screen and clicking on the "proceed" button.

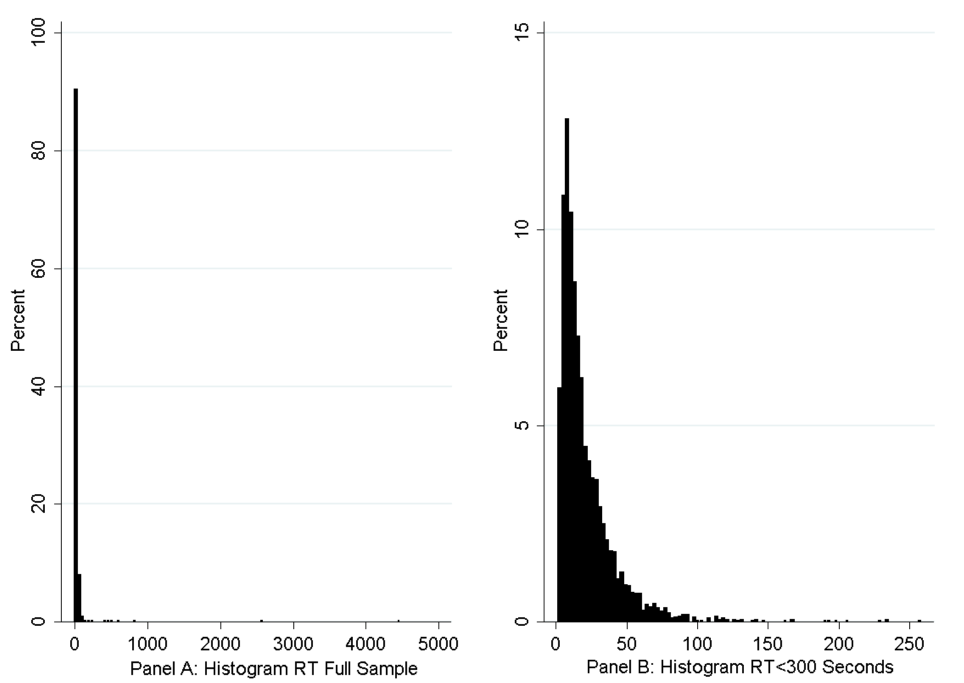
Participants' payoff at the end of the experiment was determined through a random incentive system (Grether and Plott, 1979; Starmer and Sugden, 1991; Lee, 2008) with odds of one in fifty that the actual choice on each of the two decision screens was implemented. This payment procedure was explained to subjects on the introduction screen as taking part in a lottery. This between-subjects random incentive system (Tversky and Kahneman, 1981; Abdellaoui et al., 2011; Baltussen et al., 2012) decreases expected earnings, yet it is incentive compatible by ensuring that the conditional choice between the two options remains at face value.

3.2.3 Subject pool

The sample of subjects was drawn from the representative subject pool of the online pollster "YouGov". This recruitment strategy provides two distinct advantages with regard to the research question, in particular vis-à-vis online labor markets (e.g. Amazon Mechanical Turk) that have become popular tools for running internet experiments (Paolacci et al., 2010; Buhrmester et al., 2011). First, as "YouGov" is predominantly used for conducting surveys rather than experiments, the subject pool has little experience with standard experimental paradigms. This is important in light of evidence documenting the moderating effect of participants' experience on the relationship between RTs and cooperation (Rand et al., 2014). Second, there may be a concern that participants in online labor markets rush through a large number of consecutive surveys and tasks in order to increase their hourly compensation. Arguably, this could result in lower data quality by uninformed decision making. In case of "YouGov" subjects can only take part in one survey at a time and only upon receiving an invitation by email message. The 3483 subjects analyzed here are a representative sample of the internet-using voting-age population of Germany.[5]

[5] We tested for difference to the general population of German voters: Using two-sided t-tests, we reject the hypothesis that the means of socio-demographic characteristics coincide at the 1% level. Our subjects are slightly more likely to be male, younger, and educated than the average German of voting age. Income is self-reported, and therefore the lower average income in the sample is unsurprising. Compared to the full set of subjects who finished the experiment, we exclude observations with missing values in one or more of the variables used in Section 3.3.

FIGURE 3.1: Histogram of response times.



Notes: This figure shows the distribution of the RT variable for Decision 1. RT from Decision 2 follow a comparable pattern. Panel A on the left shows the distribution for all participants. Panel B zooms in on a restricted sample of participants deciding within a $]0; 300]$ sec. interval. Clearly data follow an exponential distribution and are hence log transformed for further analysis.

3.3 Results

3.3.1 Included subjects

Panel A of figure 3.1 shows that RTs collected on the first decision screen follow a highly skewed distribution: 99 % of the study population decide within 115 seconds, yet among the remaining 1 % there are RT outliers of up to 75 minutes. These outliers likely result from subjects leaving the screen or the computer during the experiment to return to the decision screen later. Such RT outliers are not informative about the length of the decision process and potentially bias statistical results. Hence, for the results shown below, we exclude all participants that spent more than 300 seconds on the decision screen. The resulting RT distribution is displayed in panel B of figure 3.1. All core results we present below do not depend on the cutoff criterion, as tests for alternative cutoffs at 30, 60, 120, 180, 240, and 500 sec. and no cutoff demonstrate.

3.3.2 Response times and behavior

The recent experimental literature hypothesizes that a link exists between an observed contribution decision and the time it took to reach that decision, which is seen as an indicator for the decision system responsible. For a first look at the data, we follow Rubinstein (2007, 2013) and Piovesan and Wengström (2009) and classify each decision into one of four categories according to its percentile in the RT distributions: very fast (fastest 10%), fast (10%-50%), slow (50% - 90%) and very slow (slowest 10%). Table 3.1 summarizes, for each of the two subsequent decision screens, the descriptive statistics of the four RT categories and the associated contribution behavior. The average subject spends 21.84 (Median = 15.09) seconds on the first and 21.05 (Median = 15.16) seconds on the second decision screen.[6] Evidently, these RTs vary substantially between the

TABLE 3.1: Categorization of response times

Decision 1

Category	N	Reaction Time (Sec.) Mean (S.D.)	Fraction of Contributors Mean (S.D.)
Very Fast	349	4.17 (0.91)	0.088 (0.28)
Fast	1,393	10.01 (2.70)	0.128 (0.33)
Slow	1,393	26.06 (8.16)	0.203 (0.40)
Very Slow	349	70.02 (31.95)	0.347 (0.48)

Decision 2

Category	N	Reaction Time (Sec.) Mean (S.D.)	Fraction of Contributors Mean (S.D.)
Very Fast	349	4.13 (1.07)	0.140 (0.35)
Fast	1,393	10.36 (2.68)	0.234 (0.42)
Slow	1,393	24.81 (7.52)	0.234 (0.42)
Very Slow	349	65.67 (34.57)	0.300 (0.46)

Notes: This table shows mean RT and associated choices collected on both decision screens. Decision makers are categorized according to their relative decision speed.

four categories. At the first decision screen (Decision 1), subjects in the fastest category responded on average within 4 seconds (Median = 4.28 sec.), while subjects in the slowest category took more than 1 minute (Median = 59.59 sec.). At the second decision screen (Decision 2), average RTs are similar, but hint at a slight acceleration of decision-making

[6] Given these relatively small average RTs it seems unlikely that our observed effects are driven by subjects who leave the decision screen in order to search the internet for additional information on the public good.

relative to Decision 1. Table 3.1 also reports the share of contributors for each RT category. Overall, 17 % of subjects chose to contribute in Decision 1, 23 % in Decision 2. Comparing, for each decision situation, the share of contributors across the four RT categories, we find a positive relationship between response time and contributions that is confirmed by statistical tests. In Decision 1, there are significant differences in contribution behavior between all RT categories (Chi^2-Test: Pairwise comparisons, $p < 0.05$). The difference is most pronounced between the fastest and the slowest group of subjects (Chi^2-Test: $p < 0.001$, $\chi^2 = 68.48$): Among the fastest, only 8.8 % choose to contribute to the public good while among the slowest, a little more than 34 % of subjects do so. The relationship between RTs and contributions gets weaker in the second decision, as Table 3.1 shows. The difference between the fastest and the slowest group of subjects (14.0% of contributors vs. 30.0% of contributors) remains highly significant (Chi^2-Test: $p < 0.001$, $\chi^2 = 26.12$). A pairwise comparison of the groups 'Fast' and 'Slow', however, does not yield a significant difference in contribution behavior (Chi^2-Test: $p = 0.973$, $\chi^2 = 0.0012$).[7]

To sum up, basic tests of correlation between RT categories and average contribution shares within each category are supportive of the hypothesis that faster, more intuitive decisions are associated with a lower probability of contributing while slower, more deliberative decisions are associated with a higher probability. The correlation is strong when subjects encounter the choice for the first time and somewhat attenuated when the contribution choice is presented a second time. Before this result can stand, in the remainder of this section, we test and exclude several confounds and various alternative explanations for these substantial RT differences.

3.3.3 Robustness checks

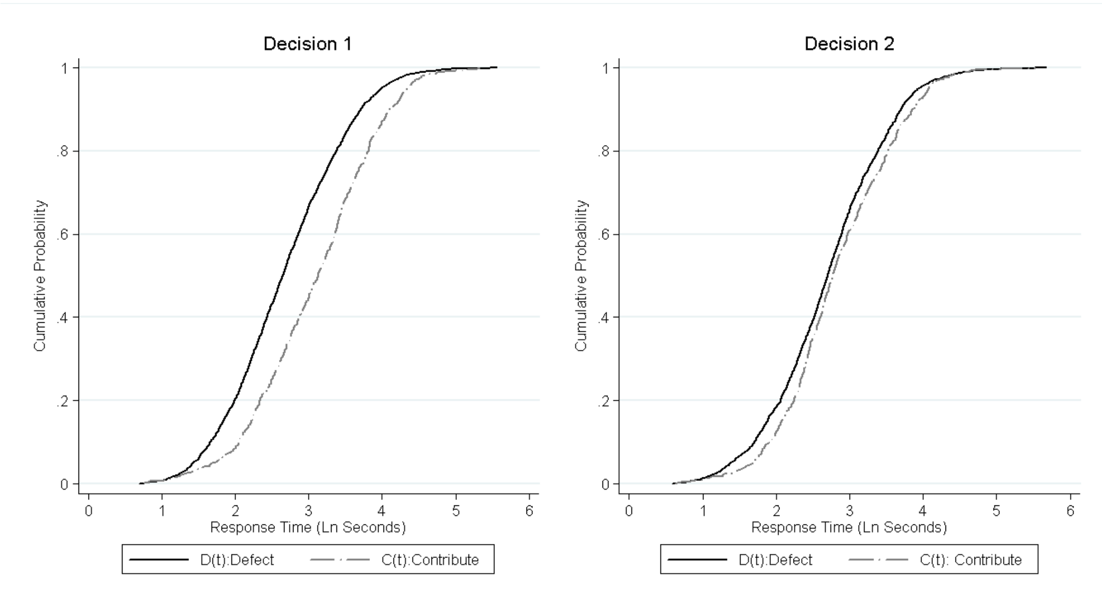
3.3.3.1 Categorization

Correlation tests that compare average contribution shares across categories can be sensitive to the method of categorization. The categorization in section 3.3.2 relies on threshold values for the 10th, the 50th, and the 90th RT deciles as introduced by Rubinstein (2007, 2013) and Piovesan and Wengström (2009). Conceivably, a different choice of thresholds between categories could find different results.

To check for robustness to categorization choice, we examine the entire cumulative distribution function (CDF) of RTs for contributors and non-contributors (Rubinstein, 2013) and find that the results do not depend on the specific categorization. For both decision screens, figure 3.2 shows two CDFs of deciding within t seconds, one for contributors $C(t)$ (grey dashed line) and one for those choosing the monetary reward $D(t)$ (black

[7] We show below that part of this moderation can be attributed to those subjects who contribute in the first decision and do not change their behavior in the second decision.

FIGURE 3.2: CDF of response times separate for contributors and non-contributors



Notes: This figure shows the cumulative distribution function of response times separately for non-contributors (black solid line) and contributors (gray dashed line). The left panel shows data collected on the first decision screen and the right panel data collected on the second decision screen.

solid line). Inspecting the CDFs for Decision 1, $C(t)$ is consistently to the right of $D(t)$ over the full range of observed response times (t). This first-order stochastic dominance represents clear evidence that it takes longer for subjects to contribute than to behave selfishly (Rubinstein, 2013).[8] Inspecting the CDFs for Decision 2, defecting again stochastically dominates contributing, but the difference between the CDFs is smaller. This indicates that the relationship between RTs and contributions, while still present, is weaker in the second decision and thus, supports the evidence in section 3.3.2.

3.3.3.2 Individual heterogeneity

RTs are a noisy proxy for identifying intuition or deliberation (Rubinstein, 2007, 2013). The present experiment responds to the resultant sample size requirement with observations from almost 3500 subjects. However, this large subject pool is highly diverse in terms of its demographic background and is exposed to variations in price and contribution characteristics within the experiment. Furthermore, RT collected on the decision screen might not only capture whether the decision was made intuitively but could also depend on participants' understanding of the task and their general swiftness in handling the survey software (Cappelen et al., 2014). This requires refining the simple analysis above in order to check whether differences in RTs are driven by differences in certain subject characteristics (such as age), by subjects' understanding and swiftness, or by

[8] A CDF $C(t)$ of the action c is said to stochastically dominate a CDF $D(t)$ of the action d if $D(t) \geq C(t) \forall t$

treatment conditions (such as a high price) rather than differences in the use of decision systems. Table 3.2 reports the results of an OLS regression analysis of RT data in which we control for the presence of these potential confounding factors. Specifications (1)–(3) contain estimates for Decision 1 while specifications (4)–(6) report the corresponding estimates for Decision 2. Summarizing ahead of a more detailed discussion, the positive relationship between RT and contribution behavior turns out to be robust to the potential confounds examined here.

TABLE 3.2: Regression of reaction times

	Decision 1			Decision 2		
	(1)	(2)	(3)	(4)	(5)	(6)
	Ln(RT1)	Ln(RT1)	Ln(RT1)	Ln(RT2)	Ln(RT2)	Ln(RT2)
Contributor (1=Yes; 0=No)	0.393****	0.371****	0.306****	0.131****	0.124****	0.082***
	(11.02)	(10.34)	(8.96)	(4.38)	(4.00)	(2.71)
Price (Euro)	0.00002	-0.0001	-0.0004	0.001**	0.001**	0.001**
	(0.04)	(-0.27)	(-0.80)	(2.32)	(2.33)	(2.23)
Age (Years)		0.0063****	0.005****		0.009****	0.008****
		(6.57)	(5.46)		(9.90)	(9.09)
Education (Cat. 1-11)		-0.0146**	-0.0123*		0.006	0.009
		(-2.15)	(-1.92)		(0.98)	(1.42)
Income (Cat. 1-11)		-0.0204****	-0.017***		-0.022****	-0.019****
		(-3.75)	(-3.26)		(-4.02)	(-3.68)
Female (1=Yes; 0=No)		-0.028	-0.027		-0.022	-0.019
		(-1.03)	(-1.05)		(-0.80)	(-0.76)
Time Introduction Screen (Sec.)		0.00001	-0.0003**		-0.00005*	-0.0004***
		(0.35)	(-1.96)		(-1.66)	(-2.76)
Time Information Screen (Sec.)		0.0004****	0.007****		0.0004***	0.006****
		(4.88)	(21.53)		(2.71)	(7.03)
Personal Benefit (Cat. 1- 4)		-0.0483***	-0.0327*		-0.125****	-0.112****
		(-2.67)	(-1.92)		(-7.38)	(-6.88)
Next Generation Benefit (Cat. 1- 4)		0.0768****	0.0550***		0.125****	0.110****
		(4.20)	(3.18)		(7.09)	(6.37)
Constant	2.617****	2.440****	2.294****	2.605****	2.136****	1.999****
	(83.37)	(29.00)	(28.58)	(135.64)	(26.98)	(25.56)
Observations	3483	3483	3456	3483	3483	3458
R^2	0.053	0.081	0.185	0.04	0.09	0.1765
Prob> F	0.000	0.000	0.000	0.000	0.000	0.0000

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

Notes: OLS regression. t-statistics in parentheses. Robust standard errors. Session dummies included to control for the four conditions under which the contribution decision was taken. The dependent variable is RT on decision screen 1 (specifications (1)–(3)) or decision screen 2 (specifications (4)–(6)) in ln seconds. As the distribution of reaction times is close to an exponential distribution (see panel B of figure 3.1), a logarithmic transformation is applied to normalize the dependent variable. The results presented here hold also with untransformed RT. The last specification (3 or 6 respectively) serves as a robustness check for potential outliers, by excluding observations with a high leverage ($L > (2 * k + 2)/N$). All main results continue to hold when we jointly estimate Decision 1 and Decision 2 regressions in a SURE framework that accounts for potential correlation of the error terms.

For the first decision, table 3.2 reports the regression results for three different specifications. Specification (1) regresses the logarithm of RT as a function of the binary contribution decision and the monetary reward (price). An inspection of the coefficient for the contribution variable shows that a contributor took on average approximately 40 % more time to reach a decision than a non-contributor.

Specification (2) adds demographic controls to the analysis, but finds little change in the fundamental relationship between RTs and contribution behavior. As expected, RT increases with age and decreases with education status and income. To proxy for individual variations in general reading and computer handling speed, we use the time spent on the relatively text-intensive introduction screen of the experiment, but find no evidence of a significant relationship with the RT at the decision screen. As a further control variable, we also use the time spent on the information screen. Since this screen contained some details that would become actionable on the decision screen, subjects could conceivably start the decision process before reaching the decision screen, resulting in a negative correlation with our RT measure. Testing this possibility, we indeed find a significant relationship, but it is both quantitatively small[9] and works in the opposite direction: Subjects who spent more time on the information screen tend to also spend more time on the actual decision screen.[10] Specification (2) also includes two variables from the post–experimental survey measuring subjects' attitudes regarding the benefits of CO_2 reductions for themselves and for future generations. We find that these variables relate to RT in a significant way. The relationship mirrors the observed relationship between RTs and the decision to contribute: RT decreases with the strength with which subjects believe that a contribution generates personal benefits, but increases with the strength with which subjects believe that a contribution generates benefits to the next generation. In other words, the more a subject believes that the consequences of the decision affect others, the more likely it is that deliberative processes are involved in the decision.

Specification (3) checks how sensitive the coefficient estimates are to outliers. It excludes observations that display a high leverage when running regression diagnostics after specification (2). The high leverage is mainly driven by a few observations that stand out for the long time spent on the introduction or information screen. Overall 27 observations are discarded. The main relationship between RT and contributions is robust to this change, with an increase in contributors' response time of 31 %. The coefficients for time spent on introduction or information screen gain both in magnitude and significance.

Specifications (4)–(6) in table 3.2 contain the corresponding analysis of RT data from Decision 2. Specification (4) reaffirms a highly significant and positive relationship

[9] One additional second spent on the information screen increases RT by an average of 0.04 %.

[10] As a further robustness check instead of controlling for the time spent on the information screen within the regression, we use the total time spent on both information and decision screen as a dependent variable. We still find a significant difference between contributors and defectors. One interpretation is that subjects who are more oriented towards pro-social goals spent more time acquiring information on how their decision could affect others. Fiedler et al. (2013) provide evidence along these lines.

between the contribution decision and RT. In contrast to the first decision, the RT differences are now smaller, but contributing still increases the RT by 13 %. Decisions are now also significantly slower when a higher monetary reward is at stake, even though this effect remains quantitatively small. Specification (5) demonstrates that the link between RT and the second contribution decision is robust to the inclusion of the same controls considered in the first decision. Similarly, excluding potential outliers in specification (6) does not affect the significance of the main effect, albeit its size is slightly reduced.

The evidence on a positive relationship between RT and contributing in the second decision, as reported in table 3.2, contains one obvious complication: Since Decision 2 is taken subsequent to Decision 1, subjects have already taken a decision once, have therefore greater familiarity with the public good offered, and have seen a specific monetary reward. Table 3.3 takes these concerns into account and contains an additional robustness check for the link between RT and contributions in Decision 2 by explicitly including control variables for behavior and prices from Decision 1. Again, the main result is that the link between RT and contributions remains positive and significant, as we now explain in more detail.

Specification (1) in table 3.3 contains the same demographic controls of specification (5) in table 3.2. Furthermore, it includes variables capturing subjects' prior experience in Decision 1. One pair of variables captures the effect of having contributed in Decision 1 and of contributing in both decisions (through the interaction term), relative to a baseline of contributing in neither. A second pair measures by how much the cost of contributing has increased or decreased relative to the first decision, allowing for a possible asymmetry in the magnitude of the response. On the contribution decision, we find that average RT is 26% higher for those who contribute in Decision 2 for the first time and 4.7% higher for those who contribute in both decisions[11]. Subject who do not change behavior between both decisions spend less time on the second decision screen and even more so, if their first choice was to defect. Among subjects who change their choice from Decision 1 to Decision 2, the average increase in RT is higher for those subjects that change from defecting to contributing than for those that change from contributing to defecting. This provides additional support for the general finding that the decision to defect requires less deliberation. We will follow up on this result in section 3.3.4. On the cost of contributing, we find that exogenously changing the contribution cost has an asymmetric effect on RT. While increases in contribution costs from Decision 1 to Decision 2 are associated with significantly higher RTs, decreasing costs are not significantly associated with lower RTs.

Specification (2) in table 3.3 examines the possibility of an interaction between the change in price and contribution behavior in the second decision. The insignificant interaction terms show that the effects of a price increase or decrease affect contributors

[11] This estimate is the sum of coefficients from contribution decisions 1 and 2 minus the coefficient of the interaction term.

Table 3.3: Regression of response times Decision 2 accounting for Decision 1 behavior

	(1) (LN RT 2)	(2) (LN RT 2)
Contributor Decision 2 (1 = Yes; 0 = No)	0.265****	0.191***
	(5.74)	(3.02)
Contributor Decision 1 (1 = Yes; 0 = No)	0.189**	-0.079*
	(2.57)	(-1.70)
Interaction(Contrib1*Contrib2)	-0.407****	
	(-4.52)	
Negative Difference Price (p2-p1<0)	-0.0006	-0.001*
	(-1.07)	(-1.74)
Positive Difference Price (p2-p1>0)	0.003****	0.003****
	(3.57)	(3.38)
Interaction(Neg Price Diff*Contrib2)		0.001
		(1.03)
Interaction(Positive Price Diff*Contrib2)		0.0004
		(0.21)
Demographic Controls	YES	YES
Constant	2.112****	2.111****
	(26.46)	(26.21)
Observations	3483	3483
R^2	0.10	0.09
Prob> F	0.000	0.000

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

Notes: OLS regression. t-statistics in parentheses. Robust standard errors. Session dummies included to control for the four conditions under which the contribution decision was taken. The dependent variable is RT on decision screen 2 in ln seconds. The same results emerge from an alternative specification in which we estimate separate regressions for Decision 1 contributors and non-contributors.

and non contributors in a uniform way. While the main effect of an increase in the announced price of the contribution does not change in size or significance, a falling price now results in a (weakly) significant negative effect on RT.

3.3.3.3 Field price censoring and response time outliers

Two further factors could conceivably compromise the link between observed RTs and choices as correlates of cognitive processes and cooperation. One is the possible presence of field-price censoring (FPC): Subjects with cooperative intentions may choose the cash option because they know that the public good can be provided more cheaply in the field than in the experiment. Their choices would be erroneously classified as non-cooperative. The other is the possible presence of subjects with a strong earning motive. If present, some of the very fast decisions might not be due to intuitive decision making, but due to subjects' objective to maximize their hourly compensation by rushing through the questionnaire without paying attention to any details of the task. In this section we show via different robustness checks that our results are not driven by these potentially biasing factors.

Potential candidates for FPC are subjects for whom the randomly drawn monetary reward was higher than the prevailing market price for contributing to the public good in the field. FPC requires that a significant share of subjects is familiar with the field price. Yet the follow-up questionnaire reveals that only 17 % of the subjects in our experiment state a price that is close (±10 €) to actual field prices. Using these price estimates and other items from the follow-up questionnaire, we employ three different methods for identifying potentially field price censored subjects[12] and define subsamples that exclude potential instances of FPC. Repeating the analysis conducted in 3.3.3.2 for the three subsamples, Table 3.4 shows that irrespective of the specific method of identification, excluding these subjects does not affect our previous results. As an additional check, we repeat the same analysis with the subsample of subjects that faced a monetary reward that was equal or lower than the field price, excluding FPC as a possibility. Again, the results remain unchanged. By excluding subjects that are well informed about specific characteristics of the public good, one of the tests for FPC (specifically Method 3) can be used to explore a different interpretation of the evidence. This is that differences in RT might be driven by differences in familiarity with the public good rather than differences in altruistic tendencies. A separate analysis of well informed subjects, however, shows that they do not differ quantitatively in the relationship between RT and contributing ($\beta = 0.345$) from other subjects.

Table 3.4: Regression coefficients for the contribution dummy accounting for FPC.

	Decision 1 Coefficient (S.D.)	Decision 2 Coefficient (S.D.)
Method 1	0.390 (0.038)	0.129 (0.033)
	N = 2735	N = 2874
Method 2	0.350 (0.044)	0.119 (0.037)
	N = 2051	N = 2263
Method 3	0.354 (0.040)	0.104 (0.034)
	N = 2866	N = 2866

Notes: This table shows regression coefficients of the contribution dummy when controlling for the full set of demographics and when excluding field-price censored subjects based on three different methods. Method 1 excludes all subjects stating in the questionnaire that they did not contribute because they belief that there are cheaper ways of mitigating CO2. Method 2 excludes all subjects whose field-price estimate is lower than the monetary reward offered. Method 3 excludes all subjects whose field price estimate is in close vicinity (±10 €) of the actual field price.

[12]For a more detailed description of each method refer to Diederich and Goeschl (2013) and the notes of table 3.4.

We already excluded subjects with very long RTs from the analysis because their RT is likely not indicative for the underlying cognitive process. Similarly, very fast RTs could not only be the result of intuitive decision making but also be an expression of a subject's goal to finish the survey as fast as possible out of boredom or to maximize the hourly compensation. If true, such subjects would consistently try to spend as little time as possible on every screen of the experiment. Using additional RT data from the introduction and information screen we find that only 31 subjects out of 3483 are consistently among the fastest 10 % and 169 subjects are among the fastest 25 % on each screen. These numbers indicate that only a negligible fraction of the total study population was actually trying to speed through the questionnaire. In further analyses we find that our results are robust to excluding these subjects.

3.3.3.4 Indifference, indecision, and response times

The result that RT and contribution behavior is linked lends support to the conjecture that underlying decision processes matter for determining outcomes in social dilemmas. However, there is also an alternative interpretation of our results. Rather than reflecting the underlying decision process, RTs could simply reflect the cognitive difficulty of coming to a binary decision when the two options on offer are of similar value to the subject. In this interpretation, those that have a strong preference for one of the options should on average be able to make a faster decision for the preferred option than those subjects who are close to indifference between the two options (Krajbich et al., 2010, 2014; Evans et al., 2014).

In the context of the present experiment, the conjecture of "indecision by indifference" would imply that subjects that are quoted a monetary reward that is sufficiently close to their maximum willingness to pay would find the decision more difficult and therefore require more time for a decision. Those, on the other hand, for whom the reward and the reservation price of contributing are far apart would find it easier and be able to make a fast decision. Which decision is easy depends, under this conjecture, on the reward: If the reward offered is low, the decision to contribute is easy and vice versa. By implication, the RT for contributors is predicted to be low when offered a low price and high when offered a high price while the RT of defectors is predicted to be high for low prices and low for high prices.

To test this prediction, we exploit the fact that in Decision 1, each subject faced a randomly drawn contribution cost in the range of 2 € to 100. Given this random assignment, the testable hypothesis is that all other things equal, contributors should be faster than non-contributors at the lower end of the range while non-contributors should be faster than contributors at the upper end of the range. We implement this test by running specification (2) of the OLS regression model (table 3.2) separately for five equally spaced subsets of the reward range between 2 € and 100. This provides, for each of the five reward bands, a coefficient estimate of how the decision to contribute influences RT. As we discuss below, the prediction that the RT coefficients of a positive contribution decision are negative at low prices and positive at high prices, is not fulfilled. Even a weaker prediction, namely that RT coefficients increase monotonically for higher monetary rewards, is not fulfilled. We therefore find no support for the conjecture of "indecision by indifference".

TABLE 3.5: Contribution dummy coefficient for five ascending price ranges

Price Range	N	Coefficient (S.D.)
EUR 2-20	607	0.224 (0.073)
EUR 20-40	713	0.409 (0.084)
EUR 40-60	668	0.350 (0.086)
EUR 60-80	739	0.351 (0.083)
EUR 80-100	690	0.404 (0.078)

In each of the five reward bands, there is a significantly positive relationship between being a contributor and longer RTs. Strikingly, the effect is also quantitatively comparable across reward ranges. To add robustness, we re-run specification (2) (table 3.2), including an interaction term between the contribution dummy and the reward variable. We find that the main effect of the contribution dummy remains highly significant (Coeff. = 0.347; p = 0.000). The interaction term, predicted to be negative, is not significantly different from zero ($\beta = 0.0004833; p = 0.673$). The alternative interpretation that the difficulty of the decision situation rather than the underlying cognitive processes generates the evidence therefore has little support in the data.

3.3.4 Within-subject differences

The cross-sectional evidence in sections 3.3.2 and 3.3.3 points to a strong and robust relationship between the decision system employed and contribution behavior. On average, subjects that are more likely to be relying on intuitive processes choose the monetary reward while those that are more likely to be relying on deliberative processes choose to contribute to the public good. This finding holds irrespective of RT categorization and controlling for a variety of confounds. The finding can also not be explained by variations in RT resulting from 'indecision by indifference'. Cross-sectional evidence, however, cannot rule out the possibility that the identified correlation is driven by unobserved individual characteristics (e.g. preferences, beliefs, and knowledge regarding climate change).

We address the possible role of unobserved individual characteristics in the link between decision system and contribution choice by exploiting the fact that the online experiment elicited two consecutive contribution decisions and the corresponding RTs for each subject. As all characteristics (observed and unobserved) can confidently be assumed constant for the same individual, a within-subject change in RT that is related to a within-subject change in contribution behavior would provide strong evidence for the existence of a true relationship.

As a first step, table 3.6 compares the changes in decision times for those 426 subjects (12%) who change their contribution decision from Decision 1 to Decision 2 with those subjects who do not.[13] The results shows that we can recover the cross-sectional correlation between contribution decision and RTs also at the individual level: Subjects who switch from contributing to free-riding require on average 4.20 seconds less time for their second choice. In contrast, subjects who switch from free-riding to contributing require on average 1.52 seconds more to come to that decision. The difference is (weakly) significant at the ten percent level (Mann-–Whitney–Test: $p = 0.072$).

Decision 1	Decision 2	RT2 - RT1	Observations
Contributor	Defector	-4.20	117
Defector	Contributor	1.52	309
	No switch	-0.89	3,057

TABLE 3.6: Switching behavior and reaction times

Table 3.7 presents the results of an analysis with full controls for changes in the incentive structure at the subject level. There, we estimate the effect of changing contribution behavior on RT in a first-difference estimation framework. This within-subject framework captures the potential effects of all observable time varying factors during the experiment while differencing eliminates potential biases due to observed and unobserved individual time-constant characteristics. Specification (1) reports the coefficient estimates for regressing a change in RT on a change in contribution behavior and a change in price. Table 3.7 shows that on average, the same subject takes 8.2% more time for a contribution decision than for a free-riding decision, compared to a baseline of subjects that do not change their contribution behavior. Changing the monetary reward does not affect a change in reaction times. Under the premise that this analysis includes all time-varying factors between the two decisions,[14] this evidence can be restated to say that on average, more deliberative decision-making leads to more cooperative behavior.

[13]Note, that roughly three quarters of those who switch, change their behavior from being a non-contributor to being a contributor. This would be expected if defection truly followed from a (potentially error prone) first impulse.

[14]Potential candidates for unobserved time-varying factors could be boredom or fatigue by the subjects. Their role can be considered minor in light of the fact that the median subject completed the experiment within 6 minutes.

Table 3.7: Decision times first difference equation

	(1) OLS		(2) IV	
Δ contributions (contrib2 - contrib1)	0.0822**	(1.99)	0.423**	(2.11)
Δ price (p2 - p1)	0.0001	(0.23)		
Constant			-0.037**	(-1.98)
Observations	3483		3483	
R^2	0.10		-	
Prob> F	0.000		0.0350	
First Stage F	-		28.60	

t statistics in parentheses; session dummies included.

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

As an additional robustness check, specification (2) accounts for the possible omission of unobserved time varying factors that conceivably bias the results of specification (1). The strategy is to employ an IV estimation framework in which the exogenous variation in the monetary rewards through random assignment is used as a instrument for changes in the contribution behavior. The randomly drawn rewards are uncorrelated with any unobserved time-varying factors and, under the validity of the exclusion restriction, valid instruments by design (Smith, 2013).[15] In this framework, the coefficient estimate reports a positive within-subjects relationship between switching to cooperation and RT that confirms the previous results at the individual level.

3.4 Conclusion

In this paper we analyze RT data from an online experiment on the choice between a monetary reward and the provision of a carbon emissions reduction, in order to answer the questions whether and which cognitive processes drive the decision to contribute to a real public good. We detect a positive relationship between RT and contributions, which survives various robustness checks and attempts to find alternative explanations consistent with our data. We therefore conclude that our analysis, which benefits from a large sample and two consecutive choices in the data that allow to exclude confounds due to omitted variables, provides strong evidence that deliberate reasoning rather than intuition drives individual contributions towards voluntary climate action.

There are several good arguments for a closer integration of behavioral economics and environmental economics (Carlsson, 2010; Croson and Treich, 2014). In this spirit, our

[15] The first stage regression F statistic returns $F = 28.60$. This indicates that the instruments are not weak.

paper seeks to import a recent methodological innovation from behavioral economics. Specifically, when we apply RT analysis to open a window into the otherwise unobservable decision process (Spiliopoulos and Ortmann, 2014), we find that the extent to which subjects deliberate about their choice is positively associated with their contributions towards climate change mitigation. To the environmental economist, who has traditionally focused on stated or revealed preference methods rather than taking the underlying decision processes into account, our findings raise the question whether intuitive and deliberative choices should be given the same weight in preference assessments. A radical step would be to base estimates on deliberative choices only and consequently discard all observations from subjects who fail to cross a certain RT threshold. However, from a positive perspective, assigning different weights to decision makers would only be warranted if the elicitation method induces more or less deliberation than the actual "natural" decision environment of interest. One might, for instance, be worried that any form of study (survey or experiment), being mostly artificial and novel to participants, could induce more deliberation than a more familiar decision environment. To the degree that differences in deliberation matter for outcomes, resultant estimates will be biased in one direction or the other. From a normative perspective, the distinction is less clear cut and largely rests on whether intuitive choices are more prone to decision errors, as suggested in Rubinstein (2013) and Recalde et al. (2014). Objective and subjective definitions of decision errors need not coincide, especially in the given context. Similar discussions have addressed the question of whether the regulation of environmental risks should be based on subjective (and wrong) probabilities or objective (and right) probabilities (Pollak, 1998; Salanié and Treich, 2009), and there are valid arguments on both sides. Attaching a lower weight on intuitive choices could be justified from a normative perspective if decision makers, e.g., out of regret, would be willing to revise their intuitive choice upon giving them an opportunity to deliberate (Kahneman and Thaler, 2006; Gilboa, 2009). Whether this would be indeed the case for decisions such as the present one is left for further research.

At the same time, our results speak directly to an ongoing discussion about the cognitive underpinnings of cooperation. One strand in this discussion holds that in social dilemma situations in which the decision to cooperate is costly, individuals employing intuitive processes are more likely to cooperate while those employing deliberative processes are more likely to act selfishly (e.g. Rand et al., 2012, 2014; Zaki and Mitchell, 2013; Nielsen et al., 2014).[16] The evidence used to support this claim mostly comes from public good game experiments, some of them involving the exogenous application of time pressure or cognitive load. Another strand in the behavioral literature arrives at the opposite conclusion: In dictator games, giving is often associated with deliberative reasoning, and selfish behavior with intuitive processes (Piovesan and Wengström, 2009; Fiedler

[16]Note that several follow-up studies have failed to replicate these findings reporting a null result or even pointing in the opposite direction (Tinghög et al., 2013; Duffy and Smith, 2014; Verkoeijen and Bouwmeester, 2014; Martinsson et al., 2014)

et al., 2013; Ubeda, 2014; Corgnet et al., 2015).[17] The context of the online experiment analyzed here is ostensibly one of public goods. Yet, our results fall squarely into the second strand. There are at least three explanations for this. One potential reason are design differences: Suter and Hertwig (2011) highlight that the decision context is essential for triggering different mental processes. Voluntary mitigation choices, as the real public good used in the present experiment, differ from public good contributions in standard lab experiments along several important dimensions. The marginal per capita return (MPCR) for climate protection is low on account of the large number of potential beneficiaries and the temporal structure of climate change. This means that the appropriate experimental paradigm among laboratory experiments to compare our results to may well turn out to be the standard dictator game experiment where the dictator's private return of contributing a token is zero. There is a second candidate explanation. Subjects in a standard public goods game face strategic uncertainty (e.g. Gangadharan and Nemes, 2009) as their own payoff depends on the strategic behavior of others. Subjects in our experiment have complete control over their own monetary payoff. Lastly, differences between our finding and other findings in the literature could be explained by design elements affecting the psychological distance between participants and thus the degree of empathy towards the potential beneficiaries of a contribution. Different cognitive processes are found to favor giving behavior depending on the psychological distance between the recipient and the potential contributor (Small and Loewenstein, 2003; Loewenstein and Small, 2007; Small et al., 2007; Ein-Gar and Levontin, 2013). Compared to lab experiments in which cooperation can emerge among a close group of fellow students, potential beneficiaries of climate change mitigation are a highly dispersed group that is temporally and spatially more distant from the participants in our experiment.

[17]This result is more pronounced if no fifty-fifty split is possible. For deviating results, see Schulz et al. (2014) and Cappelen et al. (2014). In particular, the former find that cognitive load can increase giving in a binary dictator game, while the latter find that choosing a fifty-fifty split is faster than other allocations.

3.5 Appendix

3.5.1 Instructions

Original in German available from the authors on request. Translated instructions for the relevant screens (screen-shots below) and treatments. Further information regarding the procedures available in (Diederich and Goeschl, 2013, 2014).

Introduction Screen

Dear participants,
we would like to invite you to participate in two lotteries and to answer some questions about CO_2-emissions and climate change. Your participation will take approximately ten minutes. In the lotteries, you have the chance to win points worth up to a three-digit amount in Euros. As usual, all your information will be treated confidentially.

Information Screen

In the lotteries, you may choose between the following two prizes:

• A cash prize in points
or
• the reduction of carbon (CO_2) emissions by 1 ton

How will the reduction of the CO_2 emissions take place? We will make use of a reliable opportunity provided by the EU emissions trading system: We will purchase and delete an EU emissions allowance for you. Emissions allowances are needed by power plants and other large installations within the EU in order to be allowed to emit CO_2. Since there is only a fixed overall amount of allowances in place, deleted ones are no longer available to facilitate emissions. Emissions in Germany and other EU countries decrease by exactly one ton through one deleted allowance. Because of the way in which CO_2 mixes in the air, it does not matter for the effect on the climate where CO_2 emissions are reduced. What counts is only total emissions worldwide.In the lotteries, 100 winners will be randomly selected out of about 5.000 participants. The following two lotteries may differ in the prizes offered as well as in the payoff procedures.

Decision Screen

Order of prizes randomized.

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

If you choose the cash amount and win, then the corresponding amount of points will be transferred to your points account within the next few days. All winners will receive a short notification email.

For every winner who chooses the emissions reduction one additional allowance will be deleted. Winners will receive a short notification email containing a hyperlink to Heidelberg University web pages where they can reliably verify the deletion.

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

• (*Radiobutton*) The reduction of CO_2 emissions by one ton through the deletion of one EU emissions allowance

• (*Radiobutton*) *"Random Cash Price"* Euro in bonus points

Follow Up Questions Used

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

• Do you think that future generations in Germany (for instance your children and grand-children) will benefit if climate change mitigating CO2 emissions reductions are undertaken in the present time?

Screenshots of German Version

FIGURE 3.3: Information Screen

FIGURE 3.4: Decision Screen

Chapter 4

Cooperation in Public Good Games. Calculated or Confused?*

*Co-authored by Timo Goeschl. The authors gratefully acknowledge financial support by the German Ministry for Education and Research under grant OIUV1012. Furthermore, we would like to thank the audiences at ESA Zürich, Aurö Frankfurt, RGS Bochum, University of Heidelberg, University of Chicago, and ZEW Mannheim for their valuable comments. This chapter has also profited from several comments by David Rand.

4.1 Introduction

In many areas of economics, experiments have contributed to a better understanding of the motives behind individual behavior. Yet, even sophisticated experimental designs can sometimes not distinguish sufficiently between competing theories, solely by their behavioral predictions. In these cases, analyzing the cognitive processes underlying decisions could provide additional insights (Schotter, 2008; Alós-Ferrer and Strack, 2014; Krajbich et al., 2014; Schotter and Trevino, 2014; Agranov et al., 2015). In order to gain a window into the otherwise unobservable decision process, researchers have turned to collecting supplementary non-choice data. Especially the analysis of response times (i.e., the time experimental subjects spend on selecting their preferred alternative) and the application of time pressure have become popular methods among those interested in the role that different levels of deliberation could play in driving experimental outcomes (Rubinstein, 2007; Spiliopoulos and Ortmann, 2014).

When analyzing the joint distribution of response times and choices, Rubinstein (2007) detects a systematic relationship between the two variables in a number of decision tasks. From this observation he concludes that in each task different strategies vary with respect to their cognitive demands: Some strategies are chosen instinctively and hence require little time for deliberation. Other strategies are selected more frequently by participants who expend some cognitive effort on deliberation, which, in turn, slows down their decision making. The present paper contributes to a rapidly expanding literature that exploits such differences in decision speed to address the question of whether cooperation and defection in social dilemmas can be attributed to different levels of deliberation. In a series of one-shot public good experiments, Rand et al. (2012, 2014) have found that participants under time pressure and hence constraint deliberation contribute more, on average, than participants deciding in a time delay condition. They interpret this as evidence for a causal link between intuition and cooperation and, more generally, as support for applying a dual-self model to cooperation problems.[1] Such models could shed new light on previously conflicting experimental findings that are not easily explained by existing models of other-regarding preferences (Dreber et al., 2014).[2] Considering these potentially far-reaching implications (Gächter, 2012), it is not surprising that the initial findings of Rand et al. (2012) have sparked a rapidly expanding experimental

[1]Deliberation and its counterpart intuition form central building blocks of dual-self models (Evans, 2008; Kahneman, 2011) and their application to economics (Fudenberg and Levine, 2006; Loewenstein and O'Donoghue, 2007; Dreber et al., 2014). These models and the associated experiments have previously been applied to other areas of decision making such as risk (e.g., Kocher et al., 2013) or intertemporal choice (e.g., Benjamin et al., 2013).

[2]In their paper Dreber et al. (2014) describe a dual-self model of pro-social behavior that could reconcile some of these conflicting findings. For instance, standard social preference models cannot explain why a significant number of individuals chooses to avoid opportunities for altruistic giving, when there is a possibility to do so (Dana et al., 2007; Andreoni et al., 2011; DellaVigna et al., 2012). This behavior, however, would be consistent with a decision maker's deliberate choice to avoid the temptation of impulsive giving (Vesterlund, 2014; Dreber et al., 2014). Similarly, the difficulty to generalize from some lab results to other regarding behavior in the field (Levitt and List, 2007) could be due to the higher level of deliberation induced by the unfamiliar experimental environment.

literature on the cognitive basis of pro-social behavior. Several further studies support the existence of a positive relationship between intuition and cooperation, drawing on either correlational (Lotito et al., 2012; Nielsen et al., 2014) or causal evidence (Cone and Rand, 2014; Rand et al., 2015). However, there is also a rising number of studies that find evidence contradicting a general altruistic predisposition. Tinghög et al. (2013) and Verkoeijen and Bouwmeester (2014) fail to replicate the original findings. Based on alternative methods, Duffy and Smith (2014) and Martinsson et al. (2014) also find no evidence for intuitive cooperation in repeated public good games. Regarding altruistic behavior in non-strategic distribution tasks, like modified dictator games, there is also correlation (e.g., with response times) and causal (e.g., using cognitive load designs) evidence that paints a more nuanced picture of intuitive pro-sociality (Piovesan and Wengström, 2009; Hauge et al., 2009; Fiedler et al., 2013; Ubeda, 2014; Grossman et al., 2014; Lohse et al., 2014; Corgnet et al., 2015).[3]

Given these mixed findings, some argue that using response times to identify intuitive cooperation or defection in correlational studies is more complex and might require paying closer attention to the role of different moderating factors (Recalde et al., 2014; Evans et al., 2015; Krajbich et al., 2015). In this paper, we provide an experimental test for the related concern that applying time pressure to gather causal evidence in this area might have the unwanted side-effect to increase confusion among those subjects, who would need more time to fully understand the incentive structure. In fact, even in the absence of time pressure, converging experimental evidence documents confused contributions in public good games (Andreoni, 1995; Houser and Kurzban, 2002; Ferraro and Vossler, 2010; Burton-Chellew and West, 2013; Bayer et al., 2013). In these studies, up to 50 percent of participants display signs of confusion in the first rounds of a public good game (PGG). With repetition and hence the possibility to gain experience with the task format this fraction tends to decline. The presence of confused subjects could be problematic regarding the conclusions that can be drawn from Rand et al. (2012) for two reasons: First, time pressure could increase the fraction of subjects who fail to understand that free-riding is their dominant strategy. If true, higher contributions under time pressure cease to be equivalent to more cooperation. In fact, there is evidence suggesting that faster decisions could be positively correlated with errors. In Rubinstein (2013) subjects who decide quickly are significantly more likely to choose dominated actions. Recalde et al. (2014) find similar evidence in a non-linear PGG. In their setting, which is slightly more complex than a standard PGG, because subjects have to locate an interior optimum (Keser, 1996), faster participants are more likely to contribute at suboptimal levels. Second, even if time pressure would not affect the level of confusion,

[3]Then again, several other studies provide evidence for intuitive giving in the dictator game. Schulz et al. (2014) find increased giving in dictator games under cognitive load. Kessler and Meier (2014) report higher, lower, or unchanged levels of charitable giving, depending on subtle changes in the their cognitive load design. Finally, Cappelen et al. (2014) find that a fifty-fifty split correlates with faster response times in a standard dictator game.

there could still be heterogeneous treatment effects: those participants who fail to understand the incentive structure could react to time pressure in a different way than subjects who fully understand how choices map into payoffs. Since only for the second group contributions can confidently be interpreted as cooperation, treatment effects for this group are of particular interest.

These two related concerns call for an experimental design that can identify confusion at the individual level and can pick up changes in confusion due to time pressure. Therefore, we apply time pressure in two different variants of a PGG, one of which servers as a behavioral measure of confusion. The first variant (henceforth *human condition (HC)*) is a standard PGG in which contributions can reflect both cooperation and confusion. The second variant (henceforth *computer condition (CC)*) retains all features of the standard PGG except for one difference: human interaction partners are replaced with computer agents, which automatically contribute a predetermined amount (Houser and Kurzban, 2002; Ferraro and Vossler, 2010). Since the CC does not offer an opportunity for or gains from cooperation, confusion is the remaining explanation for positive contributions. Furthermore, we observe each participant both in the HC and the CC, controlling for order effects. Behavior in the CC can, thus, be used in a test for the presence of heterogeneous treatment effects.

Contrary to the correlational evidence in Recalde et al. (2014), we find no support for our initial hypothesis, namely that time pressure increases confusion per se. However, classifying each subject as “confused” or “unconfused” according to their choices in the CC reveals that time pressure in the HC selectively affects unconfused subjects. This heterogeneous treatment effect could be one candidate explanation why we - in line with Tinghög et al. (2013) and Verkoeijen and Bouwmeester (2014) - fail to replicate the original results of Rand et al. (2012, 2014) and even find results that tentatively point towards the opposite conclusion. In our setting, time pressure reduces average contributions and significantly increases free-riding. The reduction of average contributions is weakly significant in the full sample and becomes highly significant, when we restrict the sample to control for confusion and compliance with the time limit. These main findings continue to hold when we move from a one-shot to a repeated setting, in which subjects can gain more experience with the task format.

In the next section we describe the experimental design in more detail. Section 4.3 spells out our main results. We conclude in section 4.4 with a discussion and implications.

4.2 Experimental design

Our design explores two different channels through which confusion could theoretically confound the inference from existing time pressure experiments (Recalde et al., 2014). The main distinction between both channels is whether or not time spent on the decision screen influences how well subjects understand the incentive structure.[4] If confusion depends on the cognitive effort a participant is willing or able to invest, taxing deliberative capacities by applying time pressure will lead to higher levels of confusion. The resulting increase of contributions in a standard PGG would no longer be equivalent to an increase in cooperation and the true treatment effect would remain unidentified.

Second, confusion could also exist independently from deliberation. For instance, some subjects might simply have misread the instructions before they reach the decision screen and thus additional time to deliberate on the actual decision screen would not influence their confusion status. But even if confusion was exogenous in this sense, heterogeneous treatment effects between confused and unconfused subjects could still affect the sign of the average treatment effect. As a consequence, the effect of time pressure across different experiments would depend on the (unobserved) level of confusion in each study population, which might, in turn, be driven by its demographic composition or the format of the instructions (Ferraro and Vossler, 2010). We test for heterogeneous treatment effects by splitting the sample into confused and unconfused subjects according to their behavior in the CC.

4.2.1 Basic setup

To disentangle confusion from social preferences we compare two different public good conditions, closely following the design of Houser and Kurzban (2002). In the *human condition (HC)* participants were randomly and anonymously matched into groups of four to participate in a standard PGG. Each participant could decide how to divide an initial endowment (v) of 20 tokens between a private and a public account. A token was worth 0.20 € in the private account and contributions (x) to the public account lead to a payoff of 0.10 € for all subjects in the group. In other words, each token contributed to the public account was doubled in value (0.40 €) and was then split evenly among four group member so that the marginal per capita return (MPCR) equals 0.5. Hence, free-riding is a dominant strategy while full contributions maximize the payoff of the group as a whole. Equation (1) summarizes the linear payoff function for subject i.

$$\pi_i = 0.2(v - x_i) + 0.1(\sum_j^3 x_j + x_i) \qquad \forall i \neq j \tag{4.1}$$

Positive deviations from this dominant strategy have been attributed to social preferences (e.g., Fischbacher and Gächter, 2010), but also to confusion (Andreoni, 1995; Houser and Kurzban, 2002; Ferraro and Vossler, 2010).

[4]From here onwards, we will refer to confusion as a subject's failure to understand that free-riding is his pay-off maximizing strategy. More generally, Chou et al. (2009, p.160) use the term *game form recognition* to specify different sources of confusion. Subjects can be said to display a perfect game form recognition (i.e. they are unconfused), if they understand "(1) the sets of strategies available [...], (2) the information conditions, (3) the relationship between strategy choices and outcomes, and (4) the relationships between outcomes and payoffs."

The second condition, henceforth *computer condition (CC)*, retains all features of the HC with the only difference that the gains from cooperating are removed: Subjects shared their public account with a computer program instead of human interaction partners. Retaining the same basic payoff structure, subjects lost 0.10 € for each token contributed without generating additional gains for a group of other participants, as the computer agents did not receive any payoff. Thus, contributions in this condition cannot be attributed to cooperative preferences.[5] Employing this behavioral measure of confusion is central to our design, as it enables us to observe the direct effect of time pressure on the level of confusion, at the moment of decision making. In contrast, an ex-post survey measure would not capture the full effect, as subjects by then would have had additional time to understand the incentives. To analyze the effect of constraining deliberation, the HC and CC were each conducted both in a *baseline setting* (BL) with unconstrained decision time and under *time pressure* (TP). Time pressure was randomly assigned between-subjects. In total we compare four different combinations of treatments: HC-BL, HC-TP, CC-BL or CC-TP.

To assess the individual confusion status needed in a test for heterogeneous treatment effects, we add a within-subjects dimension to the design of Houser and Kurzban (2002). Each subject was observed in one of the HC and in the corresponding CC, controlling for order effects. Under *normal order* (NO) the first task was in the HC, while under *reverse order* (RO) subjects began in the CC. In both order conditions subjects were informed that there would be a second task, but were uninformed about the specifics of this second task. In order to compare our results to other studies in the literature, we are primarily interested in the outcomes of the one-shot PGGs. However, to assess the role of confusion over time we also conduct a repeated public good game in which subjects are given a possibility to gain additional experience. In each treatment condition subjects also interacted in nine rounds of a repeated public good game with feedback. While subjects knew that they would take additional decisions, the specifics of the repeated protocol were only revealed after the one-shot game. Table 4.1 summarizes the succession of the different tasks and the corresponding sample sizes. Each condition is described in more detail below.

4.2.2 Computer condition (CC-BL & CC-TP)

Our behavioral measure of confusion replicates all central elements of the Houser and Kurzban (2002) design. We slightly deviate from their design in the following two aspects: We provide no payoff table in the instructions or on the decision screen in order to rule out that differences in information seeking interact with the effects of

[5] As discussed thoroughly in Houser and Kurzban (2002) and Ferraro and Vossler (2010), subjects theoretically could contribute to benefit the experimenter. While this cannot be ruled out by design, a questionnaire administered at the end of our experiment finds little empirical support for this concern. Furthermore, if altruism towards the experimenter was indeed present, then there is no obvious reason to assume that it is only selectively present in the CC.

Table 4.1: Treatment conditions and order

	Normal Order		**Reverse Order**	
	Baseline	*Time Pressure*	*Baseline*	*Time Pressure*
First Task	**(HC-BL)**	**(HC-TP)**	**(CC-BL)**	**(CC-TP)**
	One-Shot	One-Shot	One-Shot	One-Shot
	Repeated	Repeated	Repeated	Repeated
Second Task	**(CC-BL)**	**(CC-TP)**	**(HC-BL)**	**(HC-TP)**
	One-Shot	One-Shot	One-Shot	One-Shot
	Repeated	Repeated	Repeated	Repeated
Sample Size	N=108	N=112	N=64	N=64

time pressure. Furthermore, in each round of the repeated CC subjects did not receive feedback about the actions of the computer agents prior to, but after stating their own decision in this round. We altered this feature to make the CC more comparable to the HC. Subjects in the CC received the same set of instructions explaining the payoff structure of the standard public good game as subjects in the HC. The only difference was that CC subjects were explicitly informed that their group would consist of three computer agents who (naturally, as they are a computer program) would not receive any payoffs generated through contributions to the public account. On each decision screen we reminded participants of this fact.

To exclude other reasons for contributing in the CC, it is essential that subjects understand the difference between human and computerized interaction partners. Particularly, they should not wrongly assume that the computer was programmed to react to their contribution choices. Therefore, we instructed subjects that the computer agents would contribute predetermined amounts. In order to make this information credible, contributions were written on a concealed poster in the room prior to the experiment and were revealed to subjects at the end of the session. This procedure was described in the instructions before subjects could make any decision. A manipulation check based on two questionnaire items confirms that 92 percent of the subjects understood that they had interacted with a computer program and 93 percent believed that they were not able to influence the computer's contribution.

4.2.3 Time pressure (HC-TP & CC-TP)

In the one-shot decision of the time pressure treatments, subjects had to decide within seven seconds. This is a slightly stricter limit than in Rand et al. (2012, 2014) and Tinghög et al. (2013). In the later rounds of the repeated tasks (5-9) the limit was tightened to four seconds, to account for the possibility that subjects adapt to the time constraint. These limits were constructed by subtracting one standard deviation from mean decision time in the first two sessions of the baseline condition. In accordance with the existing literature, subjects were informed about the time constraint only after going through all instructions, right before reaching the decision screen. This procedure prevents subjects in the time pressure condition from changing their behavior on the instruction screen in anticipation of the time constraint. On the decision screen a counter displayed the remaining decision time. There are different approaches, how to deal with subjects violating the time limit. One alternative would be a binding constraint, which shuts down the decision screen after reaching the time limit and automatically chooses a default contribution. We decided against a binding limit as this would complicate the game structure, by adding the option of strategic inaction for subjects in the TP condition. Instead, subjects could violate the time constraint. However, to reduce statistical problems associated with non-compliance (Tinghög et al., 2013), we introduced an incentive. For each violation of the time constraint subjects lost 0.20 € of their show-up fee.

4.2.4 One-shot and repeated decisions

The majority of studies analyzing the effects of time pressure in public good games were conducted in a one-shot environment. To allow for a comparison with these studies, the first decision in our experiment is one-shot as well. After taking their one-shot decision, subjects received no feedback regarding the choices of other group members.

Experience could play an important role in reducing initial confusion. Therefore, we conduct a repeated version of the same PGG subsequent to the one-shot task. While subjects knew that they would take further decisions in the experiment, they only learned about the specifics of the repeated decisions after stating their choice in the one-shot game. Specifically, they were instructed that there would be nine consecutive rounds within a fixed group of subjects and that they would receive feedback after each round. Between each decision screen there was a feedback screen displaying the total contributions of their group members. To keep BL and TP comparable, in both conditions the feedback screen was only available for ten seconds after which the next decision screen appeared automatically.

4.2.5 Experimental procedures

The experiment was conducted at the University of Heidelberg "AWI–LAB" between December 2012 and November 2013. We ran twenty-six sessions with sixteen or twelve subjects per session for a total of 348 participants. The participants were recruited from a standard subject pool of undergraduate and graduate students and randomly assigned to the different treatment conditions. The subjects were from mixed disciplines, including economics (34%). There was a nearly balanced ratio of female (53%) to male (47%) participants.[6] Using ORSEE (Greiner, 2015), subjects who had previously taken part in a public goods experiment at the "AWI–LAB" were excluded from recruitment to the experiment. No participant took part in more than one session of the experiment and all sessions were run by the same experimenter. Upon arrival, participants were seated at their computer terminal, generated a random password to ensure their anonymity and received a set of general instructions that were read aloud by the experimenter. All other instructions were fully computerized. The decision tasks were implemented using z-Tree (Fischbacher, 2007). During the experiment subjects were only allowed to ask questions in private. Participants were not allowed to communicate with one another. After the decision task, subjects had to complete a set of demographic survey questions and two standardized psychological tests to measure their predisposition for cognitive reflection (Frederick, 2005) and their working memory span (Wechsler, 1955). Furthermore, they were asked to answer an incentivised comprehension question in which they had to state their payoff maximizing strategy and a set of control questions.[7] At the end of the experiment, participants were paid their earnings from one randomly drawn round and task in private. All sessions lasted approximately 75 minutes and participants earned an average of 9.51 € (Min.:4.80 €;Max.:15.00 €), including a show-up fee of 3 €.

4.3 Results

We first discuss results from the one-shot PGGs, subjects encountered first in each condition (i.e., NO-HC and RO-CC). These outcomes are directly comparable to the evidence in Rand et al. (2012, 2014), Tinghög et al. (2013), and Verkoeijen and Bouwmeester (2014). We then proceed with the evidence from the reverse order condition and the repeated games to explore the role of experience and strategic interaction.

4.3.1 One-shot decisions

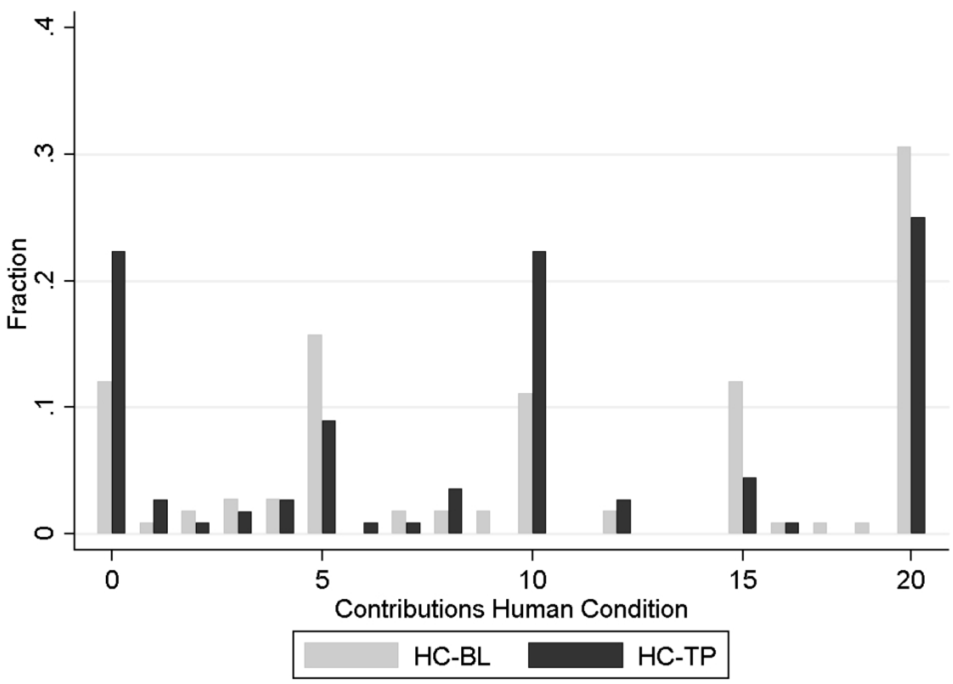
In Figure 4.1 we display the effect of time pressure on the distribution of contributions for those subjects making their first choice in the HC (gray bars: *HC-BL* vs. black bars: *HC-TP*).[8] Despite essentially reproducing the design of Rand et al. (2012), Figure 4.1 does not support their conclusion regarding a tendency to cooperate instinctively in one-shot public good games. Instead, we find a higher incidence of free-riding (BL: 12%; TP: 22%) and a slightly reduced fraction of full contributions (BL: 30%; TP: 25%)

[6] Further summary statistics are contained in Table 4.9 of the Appendix.

[7] Some of the subjects (N=96) answered these control questions as part of their demographic survey, while others answered them as part of the instructions (N=252). As we find no differences in one-shot contributions, both with (M.W. Rank Sum Test: z=0.213, p=0.831) and without applying time pressure (M.W. Rank Sum Test: z=-0.082, p=0.935), we pool observations for the analyses below.

[8] Remember that these subjects were assigned to the normal order condition, so that behavior cannot be influenced by the subsequent tasks.

Figure 4.1: Distribution of contributions HC



Notes: This graph shows the distribution of contributions to the public account separately for subjects in the baseline and under time-pressure. Gray bars are used for BL subjects and black bars for TP-subjects. Data from the normal order condition only.

in the treatment group. Furthermore, the higher fraction of subjects who split their endowment equally between both accounts under time pressure (BL: 11%; TP: 22%) could either indicate increased confusion (Ledyard, 1995; Ferraro and Vossler, 2010) or point towards a fairness heuristic (Roch et al., 2000; Cappelen et al., 2014; Capraro et al., 2014). A comparison of mean behavior corroborates these first observations. In the baseline, subjects on average contribute 56 percent of their endowment, which falls into the range typically observed for public good games (Ledyard, 1995). Contributions in the treatment condition are lower, at an average of 47 percent. This difference is weakly significant (M.W. Rank Sum Test: z=1.66, p=0.097) at the ten percent level. We reach an even stronger conclusion, when restricting our analysis to the most extreme forms of defection or cooperation. Time pressure significantly increases free-riding (Chi^2: χ^2=4.07, p=0.044) while it does not affect the fraction of subjects who contribute their full endowment (Chi^2: χ^2=0.85, p=0.357). These results are robust to controlling for additional demographic (age, sex, risk aversion, correct answer to control question) and psychometric (time spent reading the instructions, test scores from cognitive reflection test, and working memory test) variables, as shown by multiple regressions in Table 4.8 of the Appendix.

Table 4.2: Instrumental variable estimates of the effects of fast decision making

	(1) Contributions	(2) Contributions	(3) Free-Riding	(4) Free-Riding
Panel A: Second Stage (DV: Contributions)				
Response Time (Log10 Sec.)	3.654**	4.164***	-0.441***	-0.489***
	(2.36)	(2.71)	(-2.67)	(-2.81)
Age (Years)		-0.224		-0.009
		(-0.81)		(-0.26)
Sex (1=Male)		3.510**		-0.139
		(2.45)		(-0.85)
Risk Aversion (1-11)		0.214		-0.003
		(0.70)		(-0.09)
Unconfused (1=Yes)		-2.192		0.240
		(-1.48)		(1.45)
Cognitive Reflection Score (0-3)		1.965***		0.045
		(2.90)		(0.60)
Readingtime Instructions (Log10 Sec.)		3.104*		-0.048
		(1.84)		(-0.24)
Working Memory Score (0-12)		0.230		-0.042
		(0.59)		(-1.04)
Constant	2.209	-10.36	0.102	0.632
	(0.63)	(-1.05)	(0.28)	(0.53)
Observations	348	335	348	335
Panel B: First Stage (DV: Response Time)				
Treatment(I): Normal Order + Time Pressure (1=Yes)	-0.855****	-0.814****	-0.855****	-0.814****
	(-12.49)	(-11.78)	(-12.49)	(-11.78)
Treatment(II): Reverse Order (1=Yes)	-0.583****	-0.610****	-0.583****	-0.610****
	(-6.67)	(-7.36)	(-6.67)	(-7.36)
Treatment(III): Reverse Order + Time Pressure (1=Yes)	-1.324****	-1.327****	-1.324****	-1.327****
	(-16.93)	(-15.77)	(-16.93)	(-15.77)
Age (Years)		-0.001		-0.001
		(-0.12)		(-0.12)
Sex (1=Male)		-0.019		-0.019
		(-0.36)		(-0.36)
Risk Aversion (1-11)		0.005		0.004
		(0.37)		(0.37)
Unconfused (1=Yes)		0.068		0.068
		(1.25)		(1.25)
Cognitive Reflection Score (0-3)		-0.029		-0.029
		(-1.18)		(-1.18)
Readingtime Instructions (Log10 Sec.)		0.243****		0.243****
		(3.44)		(3.43)
Working Memory Score (0-12)		-0.027**		-0.027**
		(-2.16)		(-2.16)
Constant	2.830****	2.167****	2.829****	2.167****
	(48.84)	(6.18)	(48.71)	(6.18)
F-Statistic First Stage	108.66	34.84	108.66	34.84

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

Notes: Specifications (1) and (2): Tobit-IV maximum likelihood estimation to account for censoring from below (0) and above (20). Specifications (3)-(4): Probit-IV maximum likelihood estimation. t-statistics in parentheses. Robust standard errors. Estimates for the pooled sample. Treatment effects are robust to using the following alternative specifications: using only observations from the normal order condition, using OLS instead of Tobit, using a dummy variable for fast decisions (response times either <= 5 or <= 7 seconds) instead of a continuous response time variable, using time pressure as the only instrument. The natural logarithm of response times is used to give less weight to outliers. Alternatively, excluding these outliers leads to equivalent results.

Taken together, we therefore reject the hypothesis of intuitive cooperation and state the following result:

Result 1: ***There is no evidence for an intuitive tendency to contribute to the public account. Instead, time pressure significantly increases the incidence of zero contributions and weakly decreases average contributions.***

As expected, time pressure induces subjects in the treatment condition to spend significantly less time on the first decision screen (M.W. Rank Sum Test: z=10.48, p<0.001). Median response times are 16 seconds in the BL and 7 seconds in the TP condition.[9] However, only 57.2 percent of subjects under time pressure make their decision within the set time limit, whereas 7.4 percent of subjects in the baseline decide within seven seconds. Decisions from subjects who chose to spend more time on the decision screen are less informative for identifying the effects of intuition (Myrseth and Wollbrant, 2015). Consequently, the more conservative intention-to-treat effect of forced intuition (i.e., the effect of treatment assignment) corresponds to the weighted average of the zero (or reduced) effect on non-compliers and the true treatment effect on compliers (Bloom, 1984). In other words, it most likely understates the true impact of constraining deliberation. Therefore, we now adjust previous results for compliance. When we simply compare fast subjects (response times $\leqslant$ 7 seconds) to slow subjects (response times $>$ 7 seconds) the negative effect of constrained deliberation on the size of contributions as well as on the probability to contribute at all, increases in size and significance. Fast subjects contribute 40 percent of their endowment and slow subjects 57 percent (M.W. Rank Sum Test: z=3.11, p<0.01). Similarly, 31 percent of fast subjects free-ride as compared to 10 percent of slow subjects (Chi^2: χ^2=16.12, p<0.001). Before this result can stand, the analysis needs to account for potential selection effects, since fast and slow subjects might differ in observable or unobservable ways (Tinghög et al., 2013). The data generated by our experiment do not point towards the presence of selection bias on the basis of observed characteristics.[10] The only exception could be working memory capacity which is significantly higher for fast subjects (M.W. Rank Sum Test: z=-4.63, p<0.001), but which is already at increased levels for subjects randomized to the TP condition (M.W. Rank Sum Test: z=-2.88, p<0.01). There is, however, still the possibility that some unobservable subject characteristic is correlated with both response times and contribution behavior. To account for potential problems resulting from self-selection, we follow Imbens and Angrist (1994) and Angrist et al. (1996) and use assignment to one of the treatment conditions as an instrument for potentially endogenous response times. By design, treatment assignment is random and hence truly exogenous, while still being highly correlated with faster decisions. Table 4.2 displays estimates from four

[9] These values are computed for subjects in the normal order condition. An overview over the full distribution of response times across all treatment conditions is given in Table 4.7 of the Appendix.

[10] An overview over subject characteristics by treatment and compliance status is given in Table 4.9 of the Appendix.

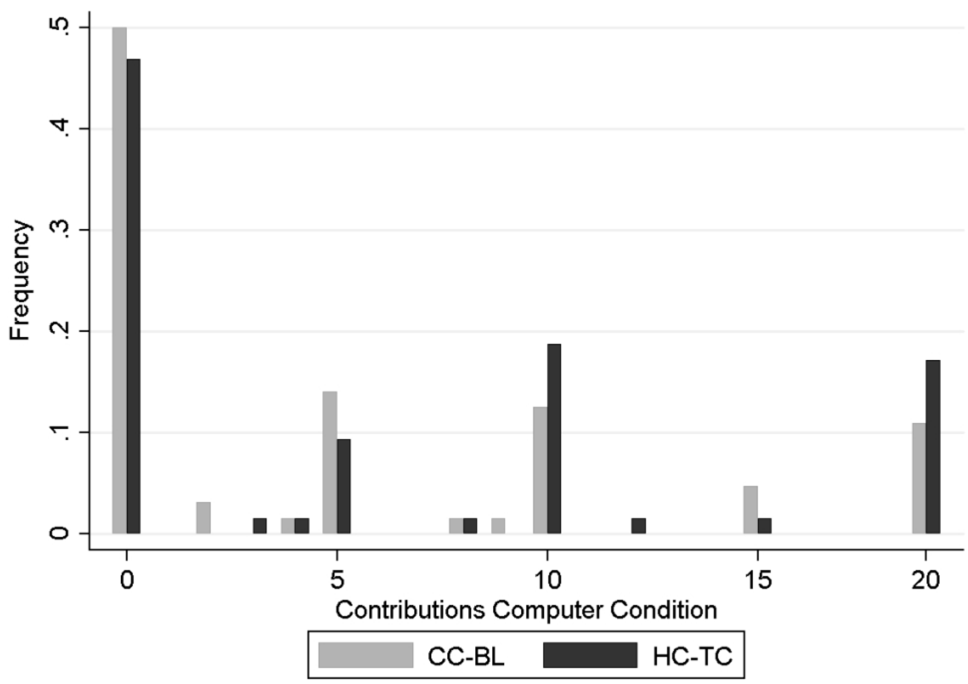
instrumental variable regressions. Tobit regressions (1) and (2) take contributions as the dependent variable, probit regressions (3) and (4) the decision to free-ride or not. First stage regressions (Panel B) show that random assignment to treatment significantly decreases response times in each case relative to the unconstrained baseline. Furthermore, two psychometric variables are correlated with fast decision making. Subjects with a higher working memory capacity and subjects who spend more time reading the instructions make faster choices on the actual decision screen. Plausibly, investing more time to understand the task and a better ability to remember the details of the task speed up decision making. Second stage regressions in Panel A show the main effect of interest. Across specifications (1)–(4) faster decisions lead to significantly lower contributions and significantly more free-riding. Thus, adjusting for potential selection-bias, IV results confirm a positive effect of more deliberation on contributions: a ten percent increase of time spent on the decision screen increases contributions by 0.35 tokens. This positive link is robust to controlling for additional demographic and psychometric variables in regressions (2) and (4). Male subjects contribute significantly more. Confusion status, assessed by a simple survey question is not correlated with contribution behavior. Yet, two of the included psychometric variables are related to the experimental outcome. The amount of time subjects spend on the instructions screen is included as a general measure of their engagement and reading speed. Furthermore, it could be related to the amount and kind of information subjects acquire when reading the instructions. It has been shown (Fiedler et al., 2013) that subjects who care more about the payoffs of other participants acquire more information about the payoff structure and consequently might spend more time on reading the instructions. This interpretation would be in line with the weakly positive relationship shown in regression (2). Subjects could also differ in their propensity to rely on their intuition. We control for these differences by scores from a cognitive reflection test (CRT) (Frederick, 2005). In line with the main treatment effect we find that subjects who are more prone to rely on deliberative thinking, as measured by the CRT, contribute more to the public good.[11] Both psychometric measures are not associated with the rate of free-riding (4).

Result 2: ***Faster subjects contribute significantly less to the public account than slower subjects. After controlling for potential selection effects, we still find support for a causal link between more deliberation and higher contributions.***

We continue by analyzing choices of those participants who took their first one-shot decision (i.e., reverse order) in the computer condition (CC). Figure 4.2 illustrates how time pressure affects behavior in a situation of comparable complexity to the human condition, but in which gains from cooperation cannot motivate behavior. Time pressure only slightly increases the occurrence of confused contributions: fewer participants stick

[11] In a companion paper (Lohse, 2014) we explore and interpret this relationship more thoroughly using parts of the same dataset. This paper is included in this dissertation as Chapter 5.

Figure 4.2: Contribution Frequencies CC



Notes: This graph shows the distribution of contributions to the public account separately for subjects in the baseline and under time-pressure. Gray bars are used for BL subjects and black bars for TP-subjects.

to their dominant strategy of contributing zero tokens (BL: 50%; TP: 47%), whereas there are more participants who give up half of (BL: 12%; TP: 19%) or even their full (BL: 11%; TP: 17%) endowment. None of these differences reaches statistic significance at conventional levels. This continues to hold when we adjust results for compliance with the time constraint. Fast subjects are neither significantly less likely to contribute zero (Chi^2: χ^2=0.04 $p = 0.851$), nor do they contribute more on average (M.W. Rank Sum Test: z=-0.55, $p = 0.584$).[12] Therefore, in contrast to the concerns raised by correlational evidence in Racalde et al. (2014), we conclude that taxing participants' deliberative capacities by applying time pressure does not increase confusion levels in our setting.

Result 3: ***In the one-shot CC, we observe no effect of time-pressure on contributions.***

As in Houser and Kurzban (2002) and Ferraro and Vossler (2010), approximately half of the participants in the CC contribute positive amounts, despite the fact that this reduces their own payoffs without benefiting any other group member. This substantial presence of confusion could complicate the interpretation of the link between contributions and

[12] Results from the corresponding IV regressions confirm this finding and are available on request.

cooperation. Only for subjects who show no sign of confusion in the computer condition a treatment effect in the HC can confidently be attributed to a change in cooperative behavior. Furthermore, time pressure could affect subjects selectively, according to their confusion status. We exploit the within-subjects structure of our data to devise two different tests for these potential concerns.

The first test is based on contribution data from the reverse order condition. Subjects who decide in the HC after making choices in the (one-shot and repeated) CC could accustom themselves with the task format. In line with a reduction of confusion, 52 percent of the subjects contribute in the initial one-shot game of the CC while this fraction drops significantly to 38 percent of the subjects for the last round of the repeated game (Wilcoxon Signed-Rank Test: $z=3.71$, $p<0.001$). Despite improvements in understanding, we continue to find no evidence for a tendency towards intuitive cooperation: compared to baseline subjects from the reverse order condition (HC-BL), time pressure does neither affect average contributions nor free-riding significantly. Thus, in line with Rand et al. (2014), we find that exposing participants to additional experience prior to making their first decision, moderates the effects of time pressure. One plausible further explanation could be that participants become not only more accustomed to the game itself, but also to decision making under a time limit. This could, arguably, dampen the taxing effects of time pressure on deliberation.[13] Finally, compared to baseline participants from the normal order condition (NO:HC-BL), who have neither been subjected to time pressure nor additional experience, time pressure in the reverse order condition still reduces contributions (M.W. Rank Sum Test: $z=2.03$, $p=0.043$) and increases free-riding (Chi^2:$\chi^2=7.04$, $p<0.01$) significantly.

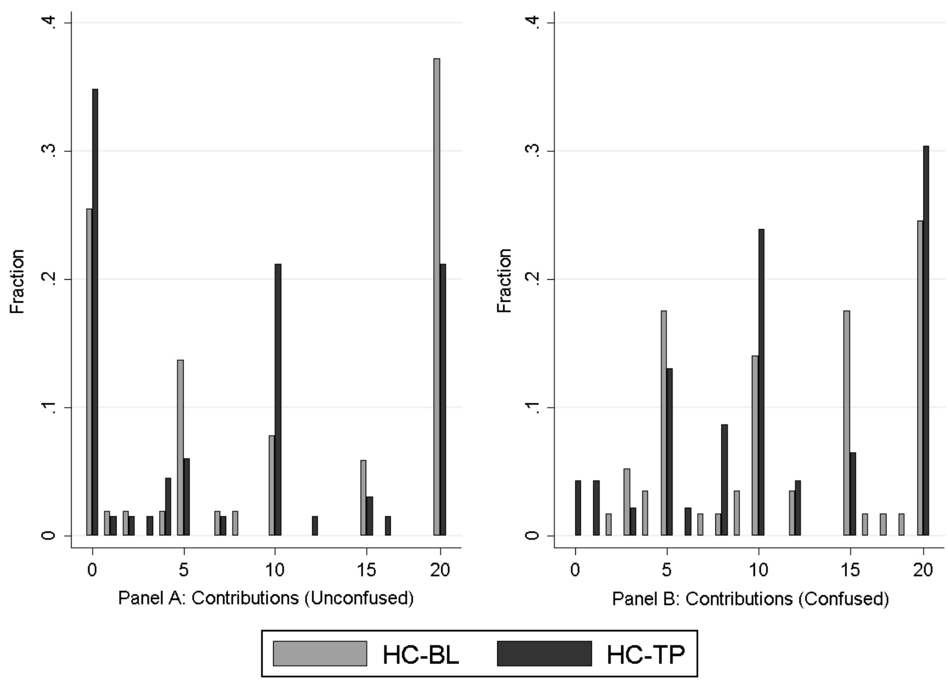
For the second test, we exploit the within-subjects design to split the sample into "confused" and "unconfused" subjects. We do so by sorting a participant into the "confused" bin if we observe positive contributions in the one-shot game of the CC and into the "unconfused" bin if we observe zero contributions. Panels A and B of Figure 4.3 compare the effect of time pressure on contributions between subjects in the "confused" and "unconfused" bins.[14] Two observations stand out clearly: First, for baseline subjects the distribution of contributions differs by their confusion status. None of the confused subjects contribute zero tokens. However, confused subjects are not more cooperative in general, as they are also less prone to contribute their full endowment. Instead, they more frequently choose a contribution from within the contribution range.[15] Overall, the average contributions of confused subjects are significantly higher than for unconfused

[13]Response time data suggest that decision making under a time limit becomes less difficult. In the reverse order condition only 6.25 percent of subjects violate the time limit, as compared to 42.8 percent of subjects in the normal order condition.

[14]We display results for the normal order condition to allow for a clean comparison to Figure 4.1. Results, however, do not differ, when using pooled data.

[15]Remember, intermediary contributions are not consistent with the predictions of many standard social preference models, which posit that decision makers either contribute nothing or their full endowment, depending on the strength of their other-regarding concerns. Therefore, it would not be surprising if intermediary contributions were more common among confused participants, who might mistakenly think that contributing half of the endowment equalizes payoffs.

FIGURE 4.3: Distribution of contributions by confusion status



Notes: This graph shows the distribution of contributions to the public account separately for subjects in the baseline and under time-pressure using observations from the normal order condition. Gray bars are used for BL subjects and black bars for TP subjects.

ones (M.W. Rank Sum Test: z=-3.06, p<0.01) Second, the effect of time pressure appears to work in opposite directions, by confusion status. For unconfused subjects time pressure increases free-riding and decreases full contributions. For confused subjects time pressure slightly increases full contributions.

When retesting for a selective treatment effect, we find time pressure to reduce average contributions only among unconfused subjects. This effect gets stronger when we adjust results for compliance to time pressure. On average, fast subjects contribute significantly less (fast: 34.6%; slow: 52.8%) if unconfused (M.W. Rank Sum Test: z=2.99, p<0.01), but approximately the same average amount if confused (M.W. Rank Sum Test: z=0.13, p=0.89). Instrumental variable regressions in Table 4.3 confirm that these results are not driven by selection effects. Specifications (1) and (2) contain estimates for unconfused subjects. Potentially endogenous response times (1) or a dummy indicating fast decisions below seven seconds (2) are again instrumented by exogenous treatment assignment. In both specifications faster decisions lead to significantly lower contributions. In contrast, specifications (3) and (4) show that subjects classified as confused are largely unaffected by their decision speed. Those who decide within seven seconds (4) do not differ from slower decision makers in their contribution behavior. The effect for a continuous response time variable (3) remains weakly significant, but is quantitatively much smaller that the corresponding effect (1) for unconfused subjects. These findings

TABLE 4.3: Instrumental variable estimates of the effects of fast decision making separated by confusion status

	Unconfused		Confused	
	(1)	(2)	(3)	(4)
	Contributions	Contributions	Contributions	Contributions
Second Stage (DV: Contributions)				
Response Time (Log10 Sec.)	7.864**		2.495*	
	(2.23)		(1.74)	
Response Time ≦ 7 Sec. (1=Yes)		-12.190**		-4.030
		(-2.29)		(-1.56)
Age (Years)	-0.455	-0.278	-0.053	-0.043
	(-0.74)	(-0.45)	(-0.21)	(-0.17)
Sex (1=Male)	4.789	5.170*	3.421**	3.401**
	(1.53)	(1.65)	(2.54)	(2.50)
Risk Aversion (1-11)	-0.240	-0.311	0.464*	0.506*
	(-0.35)	(-0.47)	(1.76)	(1.86)
Cognitive Reflection Score (0-3)	2.945*	2.578*	1.956***	1.882***
	(1.80)	(1.67)	(3.20)	(3.04)
Readingtime Instructions (Log10 Sec.)	4.846	5.761	1.210	1.645
	(1.31)	(1.60)	(0.82)	(1.10)
Working Memory Score (0-12)	0.704	0.954	0.085	0.197
	(0.80)	(1.09)	(0.25)	(0.50)
Constant	-26.10	-10.96	-3.942	0.982
	(-1.21)	(-0.54)	(-0.45)	(0.12)
Observations	170	170	165	165
First Stage F-Statistic	19.40	12.12	19.03	14.33

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

Notes: Tobit-IV maximum likelihood estimation to account for censoring from below (0) and above (20). Specifications (1)-(2): subjects classified as unconfused. Specifications (3)-(4): subjects classified as confused. t-statistics in parentheses. Robust standard errors. Estimates for the pooled sample. First stage available on request.

are robust to switching to a more regressive criterion by which we sort participants into the "confused" bin. When only sorting subjects into the "confused" bin, because they were unable to identify the strategy that would have maximized their own payoff in a control question and additionally made a positive contribution in the CC, we again find that time pressure selectively affects unconfused subjects. From these observations we state the following result:

Result 4: ***Those subjects who both understand the incentive structure and decide quickly under time pressure can be said to cooperate less. Those giving reason to doubt whether they understand the incentive structure of the PGG are largely unaffected by time pressure***

Overall, our results from the one-shot games show that constraining deliberation by applying time pressure reduces contributions to the public account. Contrary to our initial expectations, we do not find time pressure to directly increase confusion in the CC. However, there is evidence that time pressure selectively affects participants who display no signs of confusion in the CC. This points towards one potential explanation why we, in line with Tinghög et al. (2013) and Verkoeijen and Bouwmeester (2014), fail to replicate evidence on an intuitive predisposition towards cooperation (Rand et al., 2012, 2014).

4.3.2 Repeated decisions

Subsequent to each one-shot decision, participants remained in their assigned treatment conditions (HC-BL, HC-TP, CC-BL, CC-TP) and took decisions in nine rounds of a finitely repeated PGG. Therefore, in total, every participant completed two distinct repeated PGG, one in the human and one in the corresponding computer condition (compare Table 4.1). Prior to taking their first decision, participants were instructed that they would receive feedback regarding the total contributions of the other three group members (predetermined total contributions of the three computer agents) at the end of each round. Based on the additional observations from the repeated games we explore two issues which have not been addressed in the previous literature on time pressure in the PGG: First, by comparing aggregate behavior across the different conditions, we assess the persistence of treatment effects to repetition. Second, analyzing the evolution of individual decisions across different rounds, we evaluate how confusion, experience, and time pressure interact to shape strategic behavior.

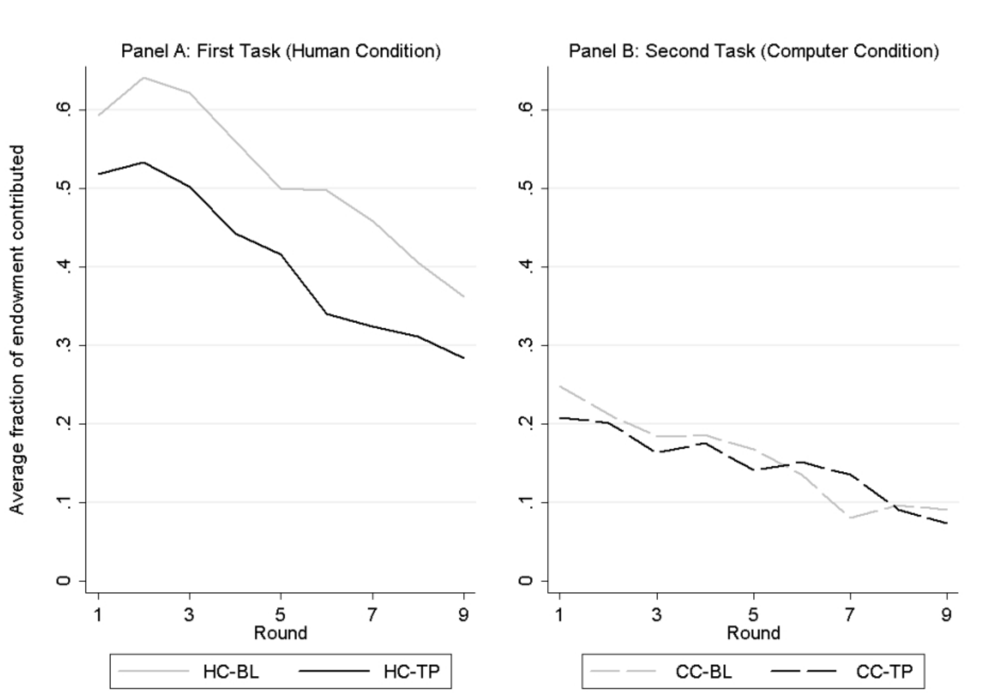
Table 4.4: Contributions averaged over nine rounds across treatment conditions

	(I) Normal Order		**(II) Reverse Order**	
	Baseline	*Time Pressure*	*Baseline*	*Time Pressure*
First Repeated Public Good Game	HC-BL (N=108)	HC-TP (N=112)	CC-BL (N=64)	CC-TP (N=64)
Contribution Average (% of endowment) (s.d.)	0.52 (0.28)	0.41 (0.27)	0.18 (0.21)	0.20 (0.22)
Second Repeated Public Good Game	CC-BL (N=108)	CC-TP (N=112)	HC-BL (N=64)	HC-TP(N=64)
Contribution Average (%of endowment) (s.d.)	0.16 (0.21)	0.15 (0.21)	0.35 (0.31)	0.42 (0.31)

Table 4.4 contains summary statistics on contribution rates averaged over all rounds. Between-subjects comparisons suggest that the treatment effects in the one-shot games are robust to repetition. The top row summarizes decisions from those repeated games which subjects encountered first under each condition. Consequently, subjects in these games have only been exposed to a limited amount of experience by deciding in the preceding one-shot game. We continue to find a significantly negative effect of time pressure in the HC (Group-level M.W. Rank Sum Test: z=2.20, p = 0.028) and no significant effect in the CC (Group-level M.W. Rank Sum Test: z=-0.62, p = 0.534). Moving to the second repeated games (i.e., more experienced subjects) displayed in the

bottom row, time pressure does neither affect average contributions in the HC (Group-level M.W. Rank Sum Test: z= -0.68, $p = 0.498$), nor in the CC (Group-level M.W. Rank Sum Test: z= 0.24, $p = 0.814$). Irrespective of task order, average contributions in the CC are significantly smaller than average contributions in the HC and overall confusion accounts for up to 40 percent of all tokens contributed in the human condition. This is slightly below the rates of confusion reported in Houser and Kurzban (2002) and Ferraro and Vossler (2010).

FIGURE 4.4: Round-wise average contributions (Normal Order)



Notes: Contributions averaged over 9 rounds across the different treatments of the normal order condition.

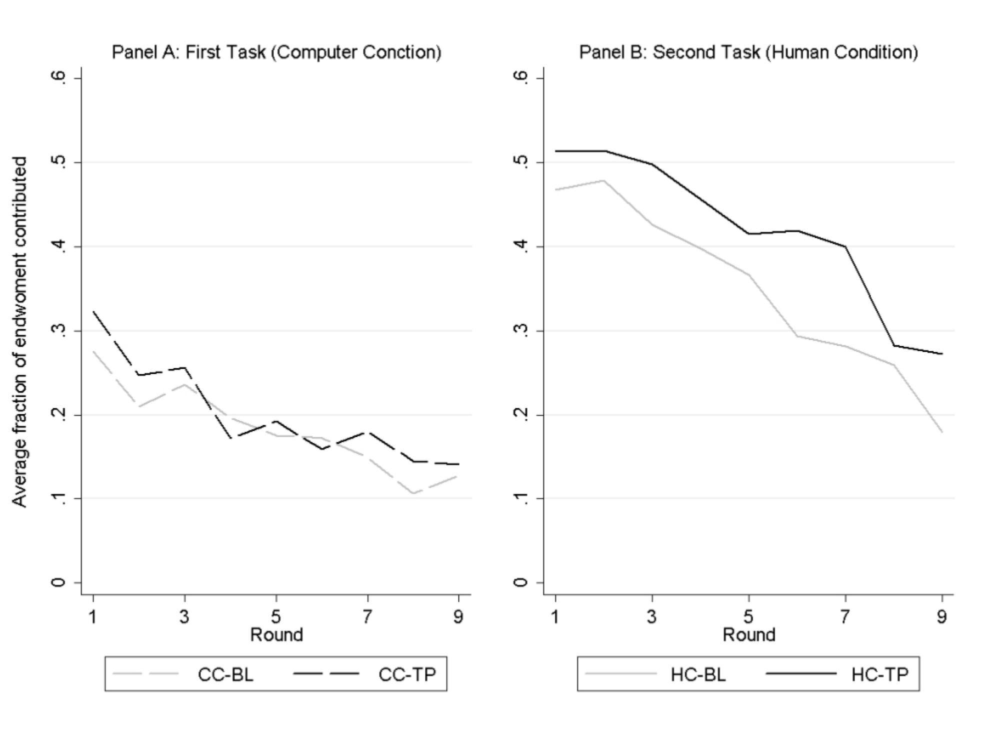
Figure 4.4 displays the evolution of average contributions (as a fraction of endowment) over time for each of the four conditions conducted under normal task order. Panel A shows contributions in the human condition (First Task: HC-BL and HC-TP), Panel B contributions in the subsequent computer condition (Second Task: CC-BL and CC-TP). Across all four conditions, contributions exhibit the typical convergence towards the equilibrium. In the HC the share of zero contributions nearly doubles from 23 percent in the first round to 41 percent in the final round. At this lower level of aggregation we continue to find no evidence for intuitive cooperation: participants under time pressure contribute less in each of the nine rounds and converge towards equilibrium at a comparable speed. Irrespective of time pressure, we observe no pronounced end-game effects in the last round. Moving to the CC, there is again no evidence, that time pressure affects the level of confusion. In the first round 46 percent of subject contribute a positive amount compared to 25 percent in the last round. The decline of contributions is steeper in the HC than in the CC. This lends support to the interpretation that declining contributions in the CC mostly represent a reduction of confusion, while declining contributions in the HC could additionally be due to "frustrated attempts" (Andreoni, 1995, p.892) at unreciprocated cooperation.

Figure 4.5 displays the contribution patterns for the four conditions conducted in reverse task order. In Panel A we show average contributions from the computer condition (First Task: CC-BL and CC-TP) and in Panel B contributions from the subsequent human condition (Second Task: HC-BL and HC-TP). For subjects taking their first repeated decisions in the CC we continue to find no evidence for increased confusion under time pressure or a slower convergence towards zero contributions. However, consistent with learning, the initial level of confused contributions is higher and the subsequent decline steeper than in the corresponding rounds from the CC conducted under normal order. Similarly, subjects deciding in the HC after completing the CC start at a lower level of contributions when there is no time limit. A significant restart effect between the CC and the HC suggests, that learning accounts only partially for the decline of contributions (Andreoni, 1988). Finally, in the HC there is no significant difference in average contributions between subjects in the baseline and subjects under timer pressure. Thus, subjects who are more familiar with the task and the time pressure manipulation display neither an intuitive tendency to cooperate or to defect.

Random effects regressions in Table 4.5 confirm the observations from Figures 4.4 and 4.5, when pooling data across both task orders. Regressions (1) and (2) display results from the HC, using individual contributions in round t as the dependent variable. Regressions (3) and (4) similarly model contribution behavior in the CC. In each specification subjects from the normal order condition serve as the left-out baseline category, against which we compare behavior in the other three randomly assigned treatment conditions. Regressions (1) and (2) show that applying time pressure significantly reduces contributions in the normal order condition of the HC. Furthermore, exogenously increasing subjects experience by assigning them to the reverse order condition reduces contributions: the coefficients of both the *Treatment(II)* and *Treatment(III)* dummies are negative and significant. However, applying time pressure under reverse order does not further reduce contributions.[16] In the CC there is little evidence that the different treatment conditions affect contribution behavior. Only inexperienced subjects deciding under time pressure (*Treatment(III): Reverse Order + Time Pressure*) display marginally increased contribution levels. The decline in contributions is captured by the *Round* variable, which is negative and significant across all specifications. The decline is steeper in the HC than in the CC. To test for potential end-game effects we include an additional dummy variable indicating the last round, which is insignificant, both in

[16]This can be verified by comparing the size of the coefficients of *Treatment(II)* and *Treatment(III)*. A Wald test fails to reject the hypothesis that they are the same (Chi^2: χ^2=0.22 p=0.6383). The same conclusion can be drawn from an alternative specification including interaction terms between a time pressure and an order dummy or by estimating separate regressions for observations from the normal and reverse order condition.

FIGURE 4.5: Round-wise average contributions (Reverse Order)



Notes: Contributions averaged over 9 rounds across the different treatments of the reverse order condition.

the HC and CC. Finally, by including interaction terms in regressions (2) and (4) we analyze, if deciding under a time limit or deciding in the reverse order condition affects the decay of contributions. One plausible hypothesis would be that constraining deliberation via time pressure negatively affects the rate of learning, because subjects can invest lower cognitive efforts to understand the game form or the behavior of their group members. We find no support for this hypothesis in the HC. Despite constraining deliberation via time pressure or giving subjects additional experience in the reverse order conditions, contributions decline at comparable speeds. In the CC there is weak evidence that subjects under time pressure converge slower in the early rounds, but display faster convergence in the last round. Taken together we state the following results:

Result 5: ***In the repeated games there is no evidence for intuitive cooperation: time pressure significantly reduces contributions in the normal order condition and does not affect contributions in the reverse order condition.***

Result 6: ***In the repeated games there is only weak evidence that time pressure increases confusion in the CC. Furthermore, time pressure marginally affects the rate at which confusion is reduced.***

Finally, our data can shed some light on the question how deliberation and confusion interact to shape strategic behavior in a repeated setting. This question has not been

Table 4.5: Repeated decisions: Decay of contributions HC and CC

	Human Condition (HC)		Computer Condition (CC)	
	(1)	(2)	(3)	(4)
	Contributions	Contributions	Contributions	Contributions
Treatment(I): Normal Order + Time Pressure (1=Yes)	-2.071**	-1.969*	0.035	-0.806
	(-2.54)	(-1.84)	(0.06)	(-0.85)
Treatment(II): Reverse Order (1=Yes)	-3.003**	-2.840*	0.494	0.186
	(-2.52)	(-1.94)	(0.74)	(0.16)
Treatment(III): Reverse Order + Time Pressure (1=Yes)	-2.300*	-3.017**	1.195*	0.958
	(-1.85)	(-2.20)	(1.77)	(0.86)
Round (1-9)	-0.658****	-0.664****	-0.402****	-0.484****
	(-9.68)	(-5.22)	(-8.26)	(-5.81)
Last Round (1=Yes)	-0.374	-0.674	0.294	0.643*
	(-1.00)	(-0.90)	(1.34)	(1.69)
Treatment(I)*Round		-0.043		0.191
		(-0.25)		(1.60)
Treatment(II)*Round		-0.024		0.062
		(-0.12)		(0.44)
Treatment(III)*Round		0.140		0.048
		(0.70)		(0.33)
Treatment(I)*Last Round		1.023		-1.036**
		(1.09)		(-1.99)
Treatment(II)*Last Round		-0.369		-0.031
		(-0.35)		(-0.05)
Treatment(III)*Last Round		0.176		-0.016
		(0.15)		(-0.02)
Constant	7.900**	7.963**	3.137	3.511
	(2.05)	(2.05)	(1.04)	(1.16)
Demographic Controls	YES	YES	YES	YES
Observations	3015	3015	3015	3015
Individuals	335	335	335	335
Groups (Clusters)	87	87	87	87
R^2	0.15	0.15	0.07	0.07
Prob > Chi^2	<0.001	<0.001	<0.001	<0.001

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

Notes: OLS random effects estimation. z-statistics in parentheses. Robust standard errors, clustered at group level (HC) or individual level (CC). Estimates for the pooled sample. Included demographic controls: age, sex, risk aversion, correct answer to control question, reading-time, CRT-score, and working memory score. Estimates of treatment effects are robust to the following alternative specifications: Analyzing data at the group level, using Tobit models, clustering all standard errors at the individual level, estimating specifications from the HC and CC as seemingly unrelated regressions.

addressed by previous time pressure experiments which have almost exclusively analyzed one-shot games. In a typical PGG about fifty percent of participants can be classified as conditional cooperators (Fischbacher et al., 2001; Chaudhuri, 2011): they increase (decrease) their own contributions to the public account from one round to another if (they believe that) the other group members also contribute more (less). The one-shot evidence in Rand et al. (2012, 2014) cannot be used to disentangle, whether time pressure affects conditional or unconditional cooperation in their setting (Gächter, 2012).[17] Furthermore, related evidence from the prisoner's dilemma (Milinski and Wedekind, 1998; Duffy and Smith, 2014) demonstrates that constraining deliberation via cognitive load can have an impact on strategic behavior. Specifically, these studies find that subjects under cognitive load are less able to condition their own decisions on their partner's past decisions. If time pressure has similar effects, we would expect to observe less conditional cooperation (defection) among treated subjects. To test this hypothesis and identify conditional cooperation in our data, we follow the empirical strategies described in Croson et al. (2005), Croson (2007), and (Ashley et al., 2010). We estimate a set of panel regressions in which we model how individual contributions change from round t-1 to round t. This first difference in contributions can depend on the behavior of the other group members in round t-1. In theory, a conditional cooperator will contribute more in round t, if his contributions are below the group average in round t-1. Similarly, he will reduce his contributions in round t, if his contributions exceed the group average in t-1. We define two dummy variables to capture this relationship in our regression framework.[18] Subjects contributing the same amount as the group average serve as the reference category.

Table 4.6 contains results from six different regressions. Specifications (1) and (2) use pooled data from the HC. We find no evidence that changes in behavior depend on treatment assignment. The significant negative coefficient of the *Round* variable captures a general decline in contributions. The effects for the two main variables of interest (*Above Group Average in t-1* and *Below Group Average in t-1*) point towards the presence of conditional cooperation. Subjects contributing more than the group average, decrease their contributions significantly in the subsequent round. Similarly, subjects who contribute less than the group average increase their contributions significantly in the following round. In line with Ashley et al. (2010), the coefficients of both variables differ in their strength. The fact that subjects react more strongly to lower contributions of their group members could be one important factor shaping the decline in average contributions across rounds. In specification (2) we test whether time pressure affects subjects in their ability to condition their behavior on the choices of the other group members. We capture these effects by interacting the treatment dummies with the main variables of interest. The interaction terms provide weak evidence contradicting our hypothesis that time pressure would decrease subjects' responsiveness to the

[17]Based on the strategy method, Nielsen et al. (2014) provide correlational evidence that conditional cooperation is faster than defection. However, as for other correlational studies based on endogenous response times, this relationship cannot be interpreted as causal evidence that intuition favors conditional cooperation.

[18]While common in the literature, one problem with this empirical strategy could be that the behavior of other subjects cannot be seen as truly exogenous, as it might, in turn, depend on the past choices of the decision maker. Therefore, we devise a robustness check in which we only use data from the first two rounds. In these rounds the behavior of other subjects can be treated as exogenous, given that group composition is random. This robustness check arrives at similar conclusions. Table 4.10 of the Appendix furthermore shows that replacing the dummy variables with continuous regressors indicating the absolute deviation leads to comparable results.

Table 4.6: Repeated decisions: conditional cooperation

	HC: Full Sample		HC: Unconfused		HC: Confused	
	(1)	(2)	(3)	(4)	(5)	(6)
	Δ Contributions	Δ Contributions	Δ Contributions	Δ Contributions	Δ Contributions	Δ Contributions
Treatment(I): Normal Order + Time Pressure (1=Yes)	-0.093	1.444	-0.127	2.843*	0.066	-2.046
	(-0.50)	(1.09)	(-0.54)	(1.70)	(0.17)	(-1.03)
Treatment(II): Reverse Order (1=Yes)	-0.215	2.091*	-0.084	3.466	-0.543	-1.626
	(-1.03)	(1.84)	(-0.32)	(2.75)	(-1.27)	(-0.85)
Treatment(III): Reverse Order + Time Pressure (1=Yes)	-0.202	0.389	-0.175	1.720	-0.341	-2.679
	(-0.87)	(0.36)	(-0.56)	(1.50)	(-0.79)	(-1.38)
Above Group Average in t-1 (1=Yes)	-2.912****	-1.441	-2.716****	-0.0119	-3.330****	-5.140***
	(-7.20)	(-1.33)	(-4.63)	(-0.01)	(-6.98)	(-2.91)
Below Group Average in t-1 (1=Yes)	1.654****	3.302***	1.569***	3.887****	1.861****	1.447
	(4.34)	(3.19)	(2.90)	(4.35)	(4.10)	(0.88)
Round (1-9)	-0.144***	-0.202**	-0.147***	-0.217**	-0.131**	-0.195
	(-3.28)	(-2.10)	(-2.74)	(-1.99)	(-2.14)	(-1.53)
Treatment(I)*Above		-2.229*		-4.229**		2.109
		(-1.70)		(-2.57)		(1.09)
Treatment(II)*Above		-2.087*		-4.001***		2.395
		(-1.70)		(-3.28)		(1.21)
Treatment(III)*Above		-0.832		-2.463*		2.738
		(-0.67)		(-1.76)		(1.47)
Treatment(I)*Below		-2.099		-3.547**		1.379
		(-1.58)		(-2.17)		(0.75)
Treatment(II)*Below		-2.502**		-3.327***		-0.168
		(-2.23)		(-3.23)		(-0.09)
Treatment(III)*Below		-1.700		-2.226*		-0.132
		(-1.46)		(-1.71)		(-0.07)
Treatment(I)*Round		0.101		0.142		0.0678
		(0.88)		(1.05)		(0.38)
Treatment(II)*Round		-0.0257		-0.0256		-0.0125
		(-0.20)		(-0.17)		(-0.07)
Treatment(III)*Round		0.126		0.0704		0.182
		(0.87)		(0.40)		(1.00)
Constant	0.719	-0.437	0.179	-1.878	2.322	2.728
	(0.65)	(-0.32)	(0.14)	(-1.34)	(1.12)	(0.94)
Demographic Controls	YES	YES	YES	YES	YES	YES
Observations	2680	2680	1672	1672	1008	1008
Individuals	335	335	209	209	126	126
Groups(Clusters)	87	87	83	83	79	79
R^2	0.10	0.09	0.11	0.11	0.08	0.07
Prob > Chi^2	0.000	0.000	0.000	0.000	0.000	0.000

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

Notes: OLS random effects estimation. z-statistics in parentheses. Robust standard errors, clustered at group level. Estimates for the pooled sample. Included demographic controls: age, sex, risk aversion, correct answer to control question, reading-time, CRT-score, and working memory score. Main results are robust to the following alternative specifications: estimating a Tobit model, using continuous instead of dummy variables to capture the behavior of group members.

choices of other group members. Instead, time pressure causes subjects to reduce their own contributions more strongly when observing lower average contributions by their group members. On the other hand, subjects under time pressure who contribute less than their group members do not increase their contributions in the following round as much as unconstrained subjects. This second interaction effect is, however, insignificant for the pooled sample. To confidently interpret changes in contribution behavior as conditional cooperation, subjects should display low levels of confusion regarding the underlying incentive structure.[19] Therefore, in specifications (3) - (6) we provide separate estimates based on subjects' confusion status by once more splitting the sample

[19] One plausible alternative explanation why confused subjects might condition their behavior on the choices of others could be that they see these choices as containing an informative signal about the game form (Burton-Chellew and West, 2013).

into two 'bins' according to behavior in the CC.[20] Specifications (3) and (5) show that both confused and unconfused subjects condition their own behavior on the past choices of their group members. A comparison of specifications (4) and (6) reveals that time pressure selectively affects the strategic behavior of unconfused subjects. As for the pooled sample, subjects under time pressure react more strongly to negative experiences with their group members. Time pressure, however, does not lead to an overall increase in conditional cooperation. Treated subjects are also more prone to exploit the higher cooperation levels of their group members by not increasing their own contributions. Especially this second observation contradicts an intuitive predisposition towards cooperative behavior.

Result 7: ***We find no evidence that subjects under time pressure are less able to condition their behavior on that of other subjects. They are more likely to reduce their contributions upon a negative experience with their group members, while they are less likely to increase their contributions after a positive experience. These interaction effects are only present among unconfused subjects.***

4.4 Discussion and conclusion

In this paper we investigate whether confusion is an important confound in a recent literature employing time pressure to investigate the effects of intuition on cooperation in PGG. Before we turn to the role of a confusion confound, we note that we fail to replicate the initial finding of Rand et al. (2012, 2014) namely that the application of time pressure leads to significantly higher contributions in PGG. We therefore discuss at first, whether this discrepancy could result from differences between our and their experimental design and procedures. The majority of observations in Rand et al. (2012, 2014) have been sampled from the online labor market on *Amazon Mechanical Turk (AMT)*. While this methodological innovation allows for an unusually large sample size, there are also several idiosyncratic features of AMT that might influence how well resultant data compare to those of typical laboratory experiments, like the present one. First, as common on AMT (Amir and Rand, 2012), stake sizes are very low: the maximum loss a full cooperator will incur, given that all of his other group members choose to defect, is 20 cents. Therefore, a large part of the evidence in Rand et al. (2012, 2014) actually shows that cooperation might be intuitive, when it is, arguably, not very costly and thus risky to cooperate (Myrseth and Wollbrant, 2015). While some evidence suggests that differences in stake size do not influence average cooperation rates (List, 2006; Kocher et al., 2008), there are also examples from the social preference literature in which

[20] To account for learning, we classify a subject as confused if his contributions across the nine rounds of the CC are above those of the average subject. Results are similar if we classify subjects according to their behavior in the final round.

stake size matters (Andersen et al., 2011; Raihani et al., 2013). Furthermore, Amir and Rand (2012) find, that hypothetical choices and low stake choices do not differ in a PGG experiment conducted via AMT. In both conditions average cooperation rates are somewhat higher than in typical PGG under higher stakes. Even if average cooperation rates are unaffected by differences in stakes, it remains unclear if this observation readily transfers to treatment effects, such as those of time pressure manipulations. Different stakes are clearly not the only plausible explanation why we fail to replicate a positive treatment effect, considering that the only study in Rand et al. (2012) employing stakes comparable to those in our experiment, still finds a (weakly) significant positive effect of time pressure on contributions.[21]

One further explanation could be related to differences in sample composition. Subjects on AMT typically have more diverse demographic backgrounds than the students taking part in our experiment. One important moderator that could differ between both populations is subjects' familiarity with the PGG or economic experiments more generally (Rand et al., 2014). When sampling from AMT it is, in fact, hard to exercise control over subjects' experience (Chandler et al., 2014). In contrast, using a typical lab sample, we explicitly excluded subjects with prior experience with PGG from our experiment. In line with Rand et al. (2014), we find that exogenously increasing subjects' experience in the reverse order condition moderates the time pressure effect. This observation still does not account for the fact that treatment effects reverse in the normal order condition containing subjects with little experience. Sample composition could also play a role, in light of the observation that cooperation norms and their effects on intuitive decision making could vary across different cultures and nationalities (Capraro and Cococcioni, 2015).

There is also a number of other more subtle design differences. With seven instead of ten seconds, the time limit for the treatment group in our experiment is stricter. Subjects in the baseline group, on the other hand, have unlimited decision time instead of a minimum decision time. Tightening the time limit in comparison to most of the original experiments in Rand et al. (2012, 2014) should increase rather than reverse the observed effect on contributions. Likewise, the absence of a minimum decision time actually permits intuitive decision making in the baseline. This makes the baseline group more similar to the treatment group, making it harder to detect a statistically significant difference. Finally, subjects in Rand et al. (2012) entered their contribution choice through a slider that was initialized at 50% of the contribution space. In our experiment subjects had to enter a number to select their preferred contribution level. Thus, it is at least possible that subjects in Rand et al. (2012) were unintentionally given the additional option to stick with a (fast) default, when having to decide under a time limit. Obviously, based on our present experiment, we can only speculate which (if

[21] Contrary to what is reported in Rand et al. (2012), the effect of time pressure on contributions is only weakly significant in this study (p=0.089). Their paper reports the significant effect for compliers (p=0.032), which might be affected by selection bias (Tinghög et al., 2013). If IV methods are used to control for selection bias, the effect remains only borderline significant (p=0.105).

any) of the mentioned differences shifts our results in the opposite direction. Identifying moderators (other than confusion) would require a more elaborate design in which these candidate explanations are randomly varied. This is left for further research.

Regarding the role of confusion as a moderator our results are twofold. Contrary to previous concerns voiced in the context of response time studies (Recalde et al., 2014), we find no evidence that forcing subjects to decide quickly increases confusion. Thus, the results in Rand et al. (2012, 2014) are not an artifact of inducing more confusion through the application of time pressure. Of course, this does not exclude time pressure from increasing confusion in tasks in which it is even more difficult for subjects to recognize their dominant strategy. Even within the comparably simple setting of the linear PGG, behavior in the CC points towards substantial levels of confusion. Based on this behavioral measure we, furthermore, show that confusion status affects the level of contributions, the distribution of contributions, and the strength of observed treatment effects in the HC. In particular, time pressure selectively affects unconfused subjects by reducing their cooperative behavior. This moderating factor could complicate the comparison of time pressure effects across experiments if the level of confusion differs between experimental populations. Furthermore, the presence of confusion complicates the interpretation of results as contributions can only be confidently equated with cooperation for unconfused subjects.

More generally, our results caution against treating confusion in PGG experiments simply as additional statistical noise. While non-trivial levels of confusion have been documented in the PGG literature, our within-subjects design reveals that confusion can affect the identification of treatment effects and, thus, pose a more serious threat to the internal validity of experimental results than previous evidence would have suggested.

4.5 Appendix

Response Time Distribution

Table 4.7 summarizes the distribution of response times in the one-shot public good game across the four HC.

TABLE 4.7: Response times percentiles across different treatments

	Normal Order		**Reverse Order**	
Percentile	(HC-BL)	(HC-TP)	(HC-BL)	(HC-TP)
1%	6	4	4	2
5%	7	4	4	3
10%	8	5	5	3
25%	11.5	6	6.5	3
50%	16	7	10	4
75%	23	8	13	6
90%	33	10	18	7
95%	56	17	20	8
99%	109	25	40	26

Notes: Response time percentiles for the one-shot public good game across the different order and treatment conditions.

Regression results: treatment effects one-shot public good game

In tobit regression models (1) and (2) of Table 4.8 we analyze the effect of treatment assignment on contributions. Relative to observations from the normal order condition without time pressure, subjects under time pressure contribute less under normal task order (*Treatment(I): Normal Order + Time Pressure*). This effect is weakly significant at the 10 percent level. Being assigned to the reverse order condition (*Treatment(II): Reverse Order*) reduces contributions, but not significantly. Applying time pressure in the reverse order condition (*Treatment(III): Reverse Order + Time Pressure*) significantly reduces contributions relative to subjects in the normal order condition without time pressure, but not relative to subjects in the reverse order condition (Wald Test: p=0.5545). These results continue to hold when controlling for the same demographic variables as in Rand et al. (2012) (age, sex, ability to answer comprehension question correctly), a survey measure of risk aversion, and several psychometric variables (time spent on the instruction screen, CRT-score, and working memory test score). Probit regression models (3) and (4) estimate the effect of treatment assignment on the propensity to contribute zero tokens. Again time pressure significantly increases free-riding without further control variables and when using the same covariates as for contribution behavior.

Table 4.8: Effects of treatment assignment

	(1) Contributions	(2) Contributions	(3) Free-Riding	(4) Free-Riding
Treatment(I): Normal Order + Time Pressure (1=Yes)	-3.256*	-3.270*	0.412**	0.430**
	(-1.80)	(-1.83)	(2.01)	(2.04)
Treatment(II): Reverse Order (1=Yes)	-3.384	-3.298	0.343	0.406*
	(-1.61)	(-1.62)	(1.44)	(1.67)
Treatment(III): Reverse Order + Time Pressure (1=Yes)	-4.820**	-5.571***	0.594***	0.653***
	(-2.22)	(-2.64)	(2.60)	(2.73)
Age (Years)		-0.228		-0.009
		(-0.86)		(-0.27)
Sex (1=Male)		3.390**		-0.132
		(2.33)		(-0.80)
Risk Aversion (1-11)		0.244		-0.008
		(0.82)		(-0.24)
Unconfused (1=Yes)		-1.867		0.209
		(-1.26)		(1.25)
Cognitive Reflection Score (0-3)		1.840***		0.055
		(2.76)		(0.73)
Readingtime Instructions (Log10 Sec.)		4.171**		-0.165
		(2.49)		(-0.81)
Working Memory Score (0-12)		0.102		-0.028
		(0.27)		(-0.71)
Constant	12.82****	-1.407	-1.173****	-0.443
	(10.01)	(-0.16)	(-7.50)	(-0.41)
Observations	348	335	348	335

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

Notes: Specifications (1)-(2): Tobit estimation to account for censoring from below (0) and above (20). Specifications (3)-(4): Probit estimation. t-statistics in parentheses. Robust standard errors. Estimates for the full sample. Results are robust to using the following alternative specifications: using only observations from the normal order condition.

Summary Statistics

Table 4.9 contains summary statistics for the control variables used in all regressions above. As expected under random assignment, there are no significant differences between the BL and TP apart from working memory scores. This does not change when comparing slow and fast subjects in columns (4) - (6).

TABLE 4.9: Summary statistics by time pressure and compliance

	(1)	(2)	(3)	(4)	(5)	(6)
	BL	**TP**	BL vs. TP	**Slow**	**Fast**	Slow vs. Fast
	Mean (s.d.)	Mean (s.d.)	p-Value	Mean (s.d.)	Mean (s.d.)	p-Value
	N=172	N=176		N=196	N=152	
Age (Years)	22.71 (2.89)	22.83 (2.52)	0.38	22.65 (2.86)	22.93 (2.49)	0.13
Sex (1=Male)	0.46 (0.50)	0.49 (0.50)	0.58	0.46 (0.50)	0.49 (0.50)	0.53
Risk Aversion (1-11)	4.91 (2.43)	4.79 (2.37)	0.63	4.83 (2.44)	4.88 (2.34)	0.84
Unconfused (1=Yes)	0.52 (0.50)	0.58 (0.49)	0.24	0.53 (0.50)	0.58 (0.49)	0.44
Cognitive Reflection Score (0-3)	1.78 (1.10)	1.85 (1.12)	0.52	1.76 (1.12)	1.89 (1.10)	0.26
Readingtime Instructions (Log10 Sec.)	3.38 (0.38)	3.33 (0.42)	0.13	3.38 (0.41)	3.33 (0.39)	0.30
Working Memory Score (0-12)	4.64 (2.08)	5.45 (2.33)	< 0.01	4.56 (2.07)	5.68 (2.30)	< 0.001

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

Notes: Individual characteristics by treatment assignment and treatment compliance. Pooled sample across order conditions. Fast subjects (Response times <= 7 seconds) and slow subjects (Response times > 7 seconds). P-Values in (3) and (6) are from M.W. ranks sum test for ordinal variables and from Chi^2 tests for binary variables.

Results conditional cooperation with continuous regressors

Table 4.10 shows that results of table 4.6 continue to hold, when replacing dummy variables with continuous regressors (*Below Group Average in t-1; Above Group Average in t-1*) indicating absolute lagged deviations and controlling for contributions in the one-shot PGG (*Contributions One Shot PGG*). Deviations are measured on a log scaled to account for non-linear responses towards large deviations.

TABLE 4.10: Repeated decisions: conditional cooperation

	HC: Full Sample		HC: Unconfused		HC: Confused	
	(1)	(2)	(3)	(4)	(5)	(6)
	Δ Contributions	Δ Contributions	Δ Contributions	Δ Contributions	Δ Contributions	Δ Contributions
Treatment(I): Normal Order + Time Pressure (1=Yes)	0.242	0.367	0.181	0.798	0.410	-0.746
	(1.03)	(0.55)	(0.70)	(0.86)	(0.79)	(-0.53)
Treatment(II): Reverse Order (1=Yes)	0.199	2.070***	0.448*	2.513**	-0.506	1.196
	(0.75)	(2.72)	(1.66)	(2.49)	(-0.99)	(0.84)
Treatment(III): Reverse Order + Time Pressure (1=Yes)	0.347	-0.287	0.527	-0.412	-0.138	-0.649
	(1.31)	(-0.47)	(1.57)	(-0.47)	(-0.26)	(-0.53)
Above Group Average in t-1 (0–20)	-1.824****	-1.381****	-1.630****	-0.870**	-2.165****	-2.208****
	(-9.57)	(-3.83)	(-6.54)	(-2.18)	(-8.51)	(-3.68)
Below Group Average in t-1 (0–20)	1.438****	1.893****	1.291****	1.819****	1.807****	2.288****
	(8.27)	(5.93)	(6.36)	(5.26)	(6.18)	(3.99)
Round (1-9)	-0.119***	-0.191**	-0.123**	-0.217**	-0.0971	-0.158
	(-2.82)	(-2.11)	(-2.26)	(-1.98)	(-1.57)	(-1.22)
Contributions One Shot PGG (0–20)	0.125****	0.131****	0.100****	0.107****	0.165****	0.171****
	(6.43)	(6.70)	(4.30)	(4.54)	(4.95)	(4.96)
Treatment(I)*Above		-0.588		-1.138**		0.293
		(-1.28)		(-2.05)		(0.40)
Treatment(II)*Above		-1.170**		-1.532***		-0.701
		(-2.34)		(-2.80)		(-0.92)
Treatment(III)*Above		-0.058		-0.364		0.547
		(-0.11)		(-0.51)		(0.77)
Treatment(I)*Below		-0.515		-0.757*		0.215
		(-1.33)		(-1.73)		(0.28)
Treatment(II)*Below		-1.209***		-1.204***		-1.574**
		(-2.89)		(-2.96)		(-2.27)
Treatment(III)*Below		-0.058		0.149		-0.661
		(-0.10)		(0.21)		(-0.69)
Treatment(I)*Round		0.132		0.145		0.150
		(1.22)		(1.07)		(0.82)
Treatment(II)*Round		-0.040		-0.0150		-0.072
		(-0.29)		(-0.09)		(-0.38)
Treatment(III)*Round		0.164		0.227		0.103
		(1.19)		(1.32)		(0.59)
Constant	-0.704	-0.822	-0.915	-1.477	0.361	0.0408
	(-0.62)	(-0.70)	(-0.65)	(-1.03)	(0.16)	(0.02)
Demographic Controls	YES	YES	YES	YES	YES	YES
Observations	2680	2680	1672	1672	1008	1008
Individuals	335	335	209	209	126	126
Groups(Clusters)	87	87	83	83	79	79
R^2	87	87	83	83	79	79
Prob Chi^2	87	87	83	83	79	79

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

Notes: OLS random effects estimation. z-statistics in parentheses. Robust standard errors, clustered at group level. Estimates for the pooled sample. Included demographic controls: age, sex, risk aversion, correct answer to control question, reading-time, CRT-score, and working memory score. Main results are robust to the following alternative specifications: estimating a Tobit model, using continuous instead of dummy variables to capture the behavior of group members.

Instructions

Experiment	Laboratory: Random seat assignment Personal code for anonymity Tasks implemented in z-Tree • General instructions *(page 110)* • Public Good Game HC *(page 110)* • Public Good Game CC *(page 114)* Payment according to personal code

General instructions
Introduction public good game
[SCREEN 2]

The first task is about to start. From now on please do not communicate with other participants in the room. In case you do so, we unfortunately will have to exclude you from the study. In this case you will not receive any compensation.

On the following screens you will find detailed instructions for the decision task. Please read them carefully. This ensures that you will know how to influence your earnings by your own decisions.

Instructions Public Good Game (HC)
[SCREEN 3]
Decision Task

Your main task in this study is to decide, how to divide 20 balls between two different bowls marked with **A** or **B**. You interact with **3 other participants** in this room. Thus including yourself, there are **4 participants** in a group. It will be impossible for you and all the other participants to observe who got matched with whom. Each of the other participants can also distribute the same number of balls (20) as yourself. You final payoff will depend on how you and the other participants distribute the balls between the two different bowls. The rules are identical for you and the other participants and all participants have received these instructions.

- **Bowl A:** Only you can fill bowl A. For each ball you put in your own bowl A, only you receive 20 Cent.
- **Bowl B:** You and the other 3 participants in your group can fill bowl B. The amount that you and all the other participants receive from bowl B depends on the total number of balls that are in bowl B. For each ball in bowl B you and each of the other 3 participants receive 10 Cent each.
- **The other 3 participants**: Each of the participants also receives 20 balls. For each ball that one of the other participants puts in his own bowl A, only he himself receives 20 Cent. For each ball that one of the other participants puts in bowl B, you, he and the other two participants receive 10 Cent each.

  So the payout rules are the same for all participants.
- **The final payoff:** Your final payoff depends on how you and the other participants fill the bowls. You will receive the payoff from **your bowl A**, as well as the payoff from the **joint bowl B**.

Procedure
[SCREEN 4]
Decision Task
Part I:

Overall, you will carry out the distribution task ten times.

First, you will take a decision only once. After stating your **first decision** you will receive new instructions that are only going to apply for the remaining nine decisions.

You will be matched anonymously with the same three participants in this room.

Part II:

After stating the first 10 decisions there will be a short questionnaire. After the questionnaire you will once again complete the distribution task for another 10 times.

For that purpose you will receive again new instructions. Please read these new instructions again carefully, as this can affect your earnings.

After the **first decision** round you will again receive additional instructions that are going to apply for the remaining 9 decisions.

Your final payoff:

At the end of this study, one of the 20 decisions is going to be selected at random. The probabilities for selecting a certain decision are the same (Like throwing a dice with the numers 1–20). Only this decision will be used to calculate your final earnings. So each decision is equally important for your final earnings.

End Instructions
[SCREEN 5]

You have completed all instructions and examples successfully.

You are now going to begin with the first 10 decisions.

(FOR TIME PRESSURE ONLY)

You have only a limited time budget available to enter your decision.

- Your time budget for the first 5 decisions is 7 seconds.
- For the second 5 decisions your time budget is 4 seconds.

For each round in which you take longer than the time limit, 20 Cent will be deducted from your 3 € show-up fee.

Decision Screen
[SCREEN 6]
(FOR TREATED ONLY: Counter $<< +1 >>$**)**

Please indicate in the blue field how many balls you want to put in bowl B. You have to distribute exactly 20 balls in total. All balls that you do not want to put in bowl B remain automatically in bowl A. You are free to choose any number of balls between 0 and 20.

- Bowl A: You receive 20 cents per ball.
- Bowl B: You receive 10 cents per ball. Each of the other 3 participants also receives 10 cents per ball.

$<< Entry : Contribuion(0-20) >>$

Additional Instructions
[SCREEN 7]
Additional rules for rounds 2-10

Additional information:

From now on you will be informed after each round how many balls the other participants have put into bowl B in total. The other participants that you interact with receive this information as well. The feedback screen will be left after a short time (10 Sec.) and the next round begins automatically.

Additional Decision Screens
Screens for decisions 2-10. Equivalent to screen 6.

Instructions Computer Condition (CC)
[SCREEN 8]

Description of payoffs equivalent.
Change of rules.
The other participants: As in the first 10 rounds, you will interact with three other players. However, these players **are not** other participants in this room. Instead these three players are controlled automatically by a computer program. Thus your interaction partners are no real human beings. Each of the three computer players has (like you) 20 balls that it divides up between bowl A and bowl B. The way the three computer players are going to divide up their balls between bowl A and bowl B has been determined prior to you first decision. Therefore, you cannot influence the computer players by your own choices. The contributions of the three computer players have been written on a poster here in this room that will be uncovered after your last decision at the end of the experiment. Thereby you can verify that the computers indeed act according to a preprogrammed contribution sequence.

While **you can earn actual money** from the balls in bowl A and B, the computer players naturally receive **no earnings** (as they are only a computer program)

Screen 10: Decision Screen Computer Condition

Please indicate in the blue field how many balls you want to put in bowl B. You have to distribute exactly 20 balls in total. All balls that you do not want to put in bowl B remain automatically in bowl A.

- **Bowl A:** You receive 20 cents for each ball.
- **Bowl B:** You receive 10 cents for each ball. Each of the other computer players “receives” 10 cents for each ball.

$<< \mathit{Entry} : \mathit{Contribuion}(0-20) >>$

Chapter 5

Smart or Selfish? When Smart Guys Finish Nice.*

*Financial support by the German Ministry for Education and Research under grant OIUV1012 is gratefully acknowledged. Furthermore, I would like to thank the audiences at ESA Zürich, Aurö Frankfurt, RGS Bochum, University of Heidelberg, University of Chicago, and ZEW Mannheim for their valuable comments. This chapter has also profited from comments by three anonymous referees.

5.1 Introduction

Recently, researchers have started exploring the role of cognitive abilities as one potentially important determinant of economic behavior (Rustichini, 2015). Since individuals differ in their cognitive abilities, a better understanding of the interplay between cognition, preferences, and behavior could shed light on the drivers of behavioral heterogeneity in economic experiments (Frederick, 2005; Benjamin et al., 2013; Deck and Jahedi, 2015) and, more generally, could illuminate the sources of differences in market outcomes (Heckman et al., 2006; Cunha and Heckman, 2009; Heineck and Anger, 2010; Christelis et al., 2010; Mazzonna and Peracchi, 2012).

Compared to non-strategic choices[1], less is known about the relationship between cognitive abilities and strategic choices. A large part of the evidence on this link comes from experiments on participants' levels of strategic sophistication (Nagel, 1995; Costa-Gomes et al., 2001). More recently, cognitive abilities have also started to attract attention as a predictor of strategic behavior in public good, trust, or ultimatum games. The present paper adds to this growing literature by providing experimental evidence on the relationship between contributions in a public good game (PGG) and cognitive abilities assessed by the *Cognitive Reflection Test* (CRT) (Frederick, 2005). This short test of cognitive abilities has been designed to capture the propensity to override a first, intuitive response that quickly comes to mind with a more cognitively reflected and demanding one. Therefore, the contribution of this paper is twofold. First, it provides evidence for the presence and direction of a link between cognitive reflection and strategic choice in a one-shot PGG. Second, by examining two additional treatment conditions, it illuminates the nature of this link. The first condition assesses whether CRT-scores are linked to preferences for cooperation or rather to a better understanding of the incentive structure. The second condition explores the link between contributions and CRT-scores, if the choice setting is cognitively more demanding. In particular, in this more demanding setting, participants have to decide under time pressure, which should limit their ability to base their choices on cognitive reflection.

It is far from obvious whether a more reflective cognitive style should be associated with a higher or lower level of contributions in a one-shot PGG. In a variety of strategic decision tasks, individuals with higher cognitive abilities have been shown to be more likely to select strategies that are in line with game theoretic equilibrium predictions.[2] One-shot

[1]For instance, individuals with higher cognitive abilities have been found to display lower levels of small-stakes risk aversion, (e.g., Burks et al., 2009; Dohmen et al., 2010), to discount future payments at lower rates (e.g., Frederick, 2005; Benjamin et al., 2013), and to be less affected by biases in financial decision making (e.g., Oechssler et al., 2009; Hoppe and Kusterer, 2011; Kiss et al., 2015).

[2]Most of the findings on strategic sophistication have been observed in games of iterated dominance. For instance, participants with higher cognitive abilities have been found to submit lower entries in beauty contest games (e.g., Burnham et al., 2009; Rydval et al., 2009; Brañas–Garza et al., 2012; Carpenter et al., 2013; Gill and Prowse, 2015). Similarly, Grimm and Mengel (2012) find that subjects with higher CRT-scores are more likely to choose according to the Nash prediction in a series of 3x3 normal form games.

PGG have a dominant strategy equilibrium in full free-riding, assuming that decision-makers hold purely selfish preferences. Hence, if the existing evidence on equilibrium selection also applied to one-shot PGGs and preferences for cooperation and CRT-scores were otherwise unrelated, subjects with higher CRT-scores should be observed to free-ride more often. This prediction is in line with previous findings in Kanazawa and Fontaine (2013), who observe more free-riding in a one-shot prisoner's dilemma (PD) among subjects with higher cognitive abilities. Similarly, when comparing the cognitive abilities of different cooperative types, Nielsen et al. (2014) find that strict free-riders have significantly higher CRT-scores than conditional or unconditional cooperators.

Thus, if cognitive abilities and social preferences were uncorrelated, a negative relationship between CRT-scores and contributions would seem plausible. However, there is experimental evidence that points towards a positive relationship between social preference and cognitive abilities. Burks et al. (2009) report that participants with higher scores in the *Raven's IQ Test* cooperate more frequently in a sequential PD as first-movers and retaliate more against defection as second-movers. Similarly, a meta-study by Jones (2008) finds that students from schools with higher SAT and ACT entry scores are significantly more likely to cooperate in repeated PD games.[3] In repeated or sequential settings, a positive link between cognitive abilities and contributions could be due to long-term strategic considerations rather than social preferences (Keser and Winden, 2000).[4] Yet, evidence from simple allocation tasks, in which such strategic considerations are typically not present, also suggest that cognitive abilities and social preferences could be related. Chen et al. (2013) find a positive correlation between SAT scores and dictator game giving. For CRT-scores the evidence is somewhat mixed and depends on the specifics of the decision task. Ponti and Rodriguez-Lara (2015) and Corgnet et al. (2015) both find that more reflective dictators are less generous in standard dictator games, but more generous when the price of giving is low (or zero).

In sum, cognitive abilities could be related to behavior in one-shot PGGs through two different channels: subjects with higher CRT-scores could be less (or more) cooperative and it could be easier for them to identify their dominant strategy. In a standard PGG, as used in the baseline of this study, it is not always possible to tease these two distinct channels apart. For instance, a negative correlation between CRT-scores and contributions (Nielsen et al., 2014) could indicate that reflection is required to find the dominant strategy or that more reflective decision-makers hold less cooperative preferences. Therefore, I employ an additional treatment condition that helps to distinguish between both explanations. This condition (Variant 1: *Computer Condition (CC)*) retains all structural features of a one-shot linear PGG, apart from one difference: Instead

[3]Further evidence on a positive relationship between cognitive abilities and cooperation in repeated or sequential tasks is found in Terhune (1974), Segal and Hershberger (1999), Cappelletti et al. (2011), Jones (2014), and Al-Ubaydli et al. (2014).

[4]In line with this interpretation, Milinski and Wedekind (1998) and Duffy and Smith (2014) find that imposing cognitive load on subjects through a memory task reduces their ability to condition their strategies on previous rounds in repeated PDs.

of interacting with human partners, subjects interact with a computer algorithm that mechanically contributes a predetermined amount to the public account (Houser and Kurzban, 2002; Ferraro and Vossler, 2010). Therefore, contributing zero is a dominant strategy that is independent from cooperative preferences towards other participants.

Finally, several papers have pointed out that choices in PGG might depend on the cognitive resources available to an individual at the moment of decision making (Rand et al., 2012, 2014). In the context of PGG, this literature on the effects of time pressure has ignored interaction effects with individual cognitive abilities. However, such effects plausibly exist, either because subjects with higher CRT-scores are more able to cope with having to decide under a time limit, or conversely because their better reasoning capacities are less useful when having to decide quickly.[5] Employing time pressure as a second between-subjects treatment (Variant 2: *Time Pressure Condition (TP)*), I test for the presences of an interaction effect of this kind.

My results show that subjects with higher CRT-scores tend to contribute significantly more in a one-shot PGG. This result is surprising in light of a large literature, which finds that higher cognitive abilities typically enable decision makers to identify their dominant strategy more easily. To some degree, it can be explained by observations from the CC. Here, subjects across all CRT-score categories display similar contribution levels. This suggests that identifying the dominant strategy in a PGG might depend less on cognitive reflection than in other, more complex decision tasks (Benito-Ostolaza et al., 2016). Finally, behavior in the TP condition demonstrates that specific features of the decision environment can strongly influence the relationship between CRT-scores and contribution behavior. When subjects have to decide under time pressure and it is hence harder to engage in cognitive reflection, there is no significant correlation between PGG contributions and CRT-scores.

The remainder of this paper is structured in the following way: section 5.2 outlines the experimental design and procedures. Section 5.3 contains results and robustness checks. Section 5.4 closes with a short discussion of the main findings.

5.2 Methods and procedures

5.2.1 Measuring cognitive abilities

In order to measure cognitive abilities, I administered the cognitive reflection test (CRT) in its original version, as introduced in Frederick (2005). This simple test assesses subjects' predisposition to base their decisions on reflective rather than intuitive thinking. The test consists of the following three items:

[5] For instance, Jones (2014) finds a positive relationship between ACT-scores and cooperation in a repeated PD. However, this relationship is only observed when the implementation complexity of cooperative strategies is low, but not when the complexity is high.

- A tennis racket and a ball cost 1.10 € in total. The bat costs 1.00 € more than the ball. How much does the ball cost?
- If it takes 5 machines 5 minutes to make 5 widgets, how long will it take 100 machines to make 100 widgets?
- In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Each question has an intuitive but incorrect answer (10 cents, 100 minutes, 24 days). The correct answer (5 cents, 5 minutes, 47 days) can be found with a sufficient amount of reflection and is easy, at least in the sense that the solution is "easily understood when explained"(Frederick, 2005, p.27). Following Frederick (2005), I construct an overall CRT-score ranging from 0 (lowest reflection abilities) to 3 (highest reflection abilities), by counting the number of correct answers. This score is then used to classify subjects in the analyzes of section 5.3.

Clearly, the ability to rely on reflection rather than on a first impulsive thought is not a measure of general intelligence and only represents one specific subcategory of a broad set of cognitive abilities that could be relevant for economic decision making. One could also argue that the test is too short and narrow to reliably capture such abilities. In spite of these objections, the CRT seems particularly suited for the purpose of this paper for the following three reasons: First, the CRT is used in several papers that find a positive relationship between cognitive abilities and the ability to identify a dominant strategy (e.g. Brañas–Garza et al., 2012; Grimm and Mengel, 2012; Carpenter et al., 2013). Second, papers that report a significant effect of time pressure in one-shot PGGs, attribute this result to the fact that subjects under time pressure rely more on intuition and less on reflection (Rand et al., 2012, 2014). These are the exact cognitive styles distinguished by the CRT-score. Third, despite its simple design, CRT-scores have been shown to be significantly correlated with more general and sophisticated measures of cognitive abilities such as scores from the SAT, the *Wonderlic Personnel Test*, the *Wechsler Matrix Test*, and several tests for numerical abilities (Frederick, 2005; Toplak et al., 2011, 2014).

To increase the reliability of measured CRT-scores and avoid the potential issue that the intrinsic motivation to reach a good test result could vary across subjects (Chen et al., 2013), participants could earn a monetary reward of up to 4 € for giving a correct answer to each question[6]. There are also some concerns (Toplak et al., 2014) that, by now, subjects could be familiar with some of the CRT items, due to the test's rising popularity in research. If true, correct answers might not always coincide with higher levels of reflection. To address this second concern, only subjects with limited prior lab

[6]Since a current meta-study finds no significant relationship between incentives and test-scores, these concerns might not have been warranted (Brañas–Garza et al., 2015).

experience[7] were invited to take part in the experiment. Furthermore, after completing all CRT questions, participants were asked to indicate whether they were already familiar with a given item.

5.2.2 Measuring cooperation

In order to analyze the link between cooperation and cognitive abilities, I employed three different variants of a one-shot PGG. Subjects were randomly assigned to one of these three conditions in a between-subjects design. In section 5.3, each of the two treatment conditions is contrasted against the baseline condition, which I will describe first.

Baseline condition (N=108)

The baseline is a standard linear PGG, as reviewed in Ledyard (1995) or Vesterlund (2014). Subjects received an initial endowment of $\omega = 20$ tokens, each worth 0.20 € when kept in their private account, from which they could contribute $x_i \in [0, 20]$ tokens to a public account. This public account was shared with a group of three other subjects, who could also contribute up to twenty tokens each[8]. Tokens in the public account created a payoff of 0.10 € for every subject in the group. Group matching was random and anonymous and the one-shot contribution choices were made simultaneously without feedback. The following equation summarizes the payoff structure:

$$\pi_i = 0.2(\omega - x_i) + 0.1x_i + 0.1\sum_{j \neq i}^{N=3} x_j \tag{5.1}$$

From this equation it is straightforward to verify that contributing zero is a dominant strategy for subjects who aim at maximizing their own payoff: taking the contributions of the other three group members ($\sum_{j \neq i}^{N=3} x_j$) as given, subjects lose 0.10 € for each token they contribute to the public account. However, full cooperation would be the efficient choice: contributing all twenty tokens to the public account will maximize the payoff of the group as a whole, as each token contributed generates a total payoff of 0.40 €. In other words, the public account has a marginal per capita return (MPCR) of 0.5 and a marginal total return of 2. By fully cooperating instead of free-riding all subjects could thus double their payoff from the task to 8 €. Yet, deviating from cooperation unilaterally generates an additional payoff of 2 € and is therefore the dominant equilibrium strategy for payoff-maximizers.

[7] The median subject had participated in no more than two prior studies.

[8] In the instructions tokens were labeled as *balls* and the PGG framing was neutral (instead of framing it as an investment decision), in the sense that the private account was simply labeled as *Bowl A* and the public account as *Bowl B*

Variant 1: Computer condition (N=64)

A number of studies suggest that a significant fraction of subjects typically fails to identify their dominant strategy in standard PGG and contributes out of confusion (Andreoni, 1995; Houser and Kurzban, 2002; Ferraro and Vossler, 2010; Bayer et al., 2013). None of these studies have looked into differences in cognitive abilities as a potential source of subjects-specific heterogeneity in confusion levels. However, as outlined in the introduction, cognitive abilities could serve as a candidate explanation, since it might be easier for subjects with higher CRT-scores to recognize their dominant strategy of contributing zero. If true, there should be a negative correlation between contributions and the CRT-score, as long as preferences for cooperation are unrelated to subjects' cognitive abilities. Yet, in a standard PGG the same outcome would emerge, if cognitive abilities were unrelated to the ability to identify the dominant strategy, but subjects with higher CRT-scores were genuinely less cooperative.

The first treatment condition (henceforth: *computer condition (CC)*) was designed to differentiate between both explanations. To do so, I built on a variant of the standard PGG, initially introduced by Houser and Kurzban (2002). Specifically, I kept all structural elements of the baseline PGG[9], but removed possibilities for or gains from cooperation: subjects shared their public account with three computer agents instead of sharing it with three human interaction partners. Subjects hence lost 0.10 € for each token contributed, without generating additional gains for other participants. Thus, contributions in this variant cannot be attributed to cooperative motives vis-á-vis other subjects and contributing zero is a dominant strategy that is independent from the presence of social-preferences. In the CC, a negative correlation between contributions and CRT-scores would provide unambiguous evidence that being able to identify the dominant strategy in a PGG environment depends on cognitive abilities.

In order to successfully exclude other reasons for contributing in the CC, it is essential that subjects understand the difference between human and computerized interaction partners and acknowledge that the computer agents are not programmed to react to the amounts chosen. Clarifying the latter point is even relevant in a one-shot setting, as subjects could still (wrongly) assume that the computer is following a sequential contribution algorithm. To exclude this sort of bias, subjects were informed that the computer agents would contribute a predetermined amount[10] and (naturally as they are only a computer program) would not receive any payoffs generated from the public account.

[9]Under structural elements I broadly subsume all game parameters that could be relevant to decision makers when reasoning about their optimal strategy in the baseline, such as the endowment, the value of token in the private account, the individual return from the public account and the group size.

[10]To strengthen this point, the predetermined contribution by the computer agents had been written on a concealed poster in the room prior to the experiment and was revealed to subjects at the end of a session. Subjects were informed about this procedure before making any decision in this task. A manipulation check confirmed that 92 percent of subjects understood that they had interacted with a computer program and 93 percent believed that they were not able to influence the computer's contribution.

Variant 2: Time pressure (N=112)

Decision environments which offer an opportunity for cooperation can vary greatly by the level of cognitive demands they impose on a decision maker. For instance, some decisions have to be taken under time constraints or simultaneously with a large number of other decisions. Recent studies have investigated the role of such cognitive constraints in one-shot and repeated cooperation tasks. Rand et al. (2012, 2014) find a significant increase of one-shot contributions under time pressure, while this result is not replicated by Tinghög et al. (2013) and Verkoeijen and Bouwmeester (2014)[11]. None of these studies has investigated, whether increasing the cognitive demands of the decision environment (via time pressure) affects decision makers with different reasoning abilities differently. There are two equally plausible directions in which a potential interaction effect could go. On the one hand, decision makers with higher cognitive abilities could be better equipped to cope with the increased cognitive demands of the decision environment. In this case the biggest effect should materialize for subjects with low CRT-scores. On the other hand, the reverse could be true. Time pressure could particularly affect subjects who would otherwise rely on their reasoning abilities (high CRT-score), while it could have little effect on subjects who are anyhow prone to rely on their intuition (low CRT-score) in their decision making. Variant 2 (henceforth: *time pressure condition (TP)*) provides a test for these two possibilities. It keeps the baseline PGG unchanged, but forces subjects to decide within a short time limit. Subjects under time pressure were asked to decide within 7 seconds, a limit constructed by subtracting one standard deviation from the average decision time in the first two baseline sessions. In order to keep subjects from adapting to the time pressure by already choosing a contribution level while reading the instructions, they were only informed about the time limit right before seeing the decision screen. A counter on the screen reminded subjects of the remaining time.

5.2.3 General procedures

The experiment was conducted at the University of Heidelberg "AWI Lab" between December 2012 and March 2013 during 21 sessions. The 284 participants were recruited from a standard subject pool of undergraduate and graduate students using ORSEE. The subjects were from mixed disciplines, including economics (33%). There was a nearly balanced ratio of female (52%) to male (48%) participants. Subjects who had previously taken part in a public goods experiment were excluded from recruitment to the experiment. Upon arrival, participants were seated at their computer terminal (separated by dividers), generated a random password to ensure their anonymity and received a set of general instructions that were read aloud by the experimenter. All other instructions, tasks, and questionnaires were fully computerized using z-Tree (Fischbacher, 2007). For all subjects the CRT was conducted subsequent to the PGG and

[11] In repeated PDs, Milinski and Wedekind (1998) and Duffy and Smith (2014) find evidence for less strategic cooperation under cognitive load.

several questionnaires. At the end of the experiment, participants were paid their earnings in private. All sessions lasted approximately 75 minutes and participants earned an average of 9.55 € (Min.:5 €; Max.:15.00 €), including a show-up fee of 3 €[12].

5.3 Results

5.3.1 Observed behavior

Table 5.1 summarizes the main results of the CRT. As expected under random assignment, average CRT-scores are balanced across the three treatment conditions (Chi^2-test: χ^2=2.07, p=0.913). The first item (Ball and Bat) has a solution frequency lower than the other two items, as common for this test. The last row contains the overall CRT-score. On average 1.81 of the three questions are answered correctly. This average results from

Table 5.1: CRT-score and single items across treatments

	Baseline (N=108)	Variant 1 (CC) (N=64)	Variant 2 (TP) (N=112)	Overall (N=284)	Oechsler et al. (2009) (N=564)	Brañas-Garza et al. (2015) (N=41,825)
Item 1 (Ball and Bat)	0.46 (0.40)	0.43 (0.37)	0.59 (0.53)	0.51 (0.42)	0.54	0.32
Item 2 (Widget)	0.62 (0.59)	0.67 (0.62)	0.60 (0.55)	0.62 (0.55)	0.70	0.40
Item 3 (Lilly Pad)	0.70 (0.63)	0.66 (0.64)	0.68 (0.61)	0.68 (0.61)	0.78	0.48
CRT-score (0–3)	1.78 (1.54)	1.76 (1.64)	1.87 (1.61)	1.81 (1.58)	2.05	1.19

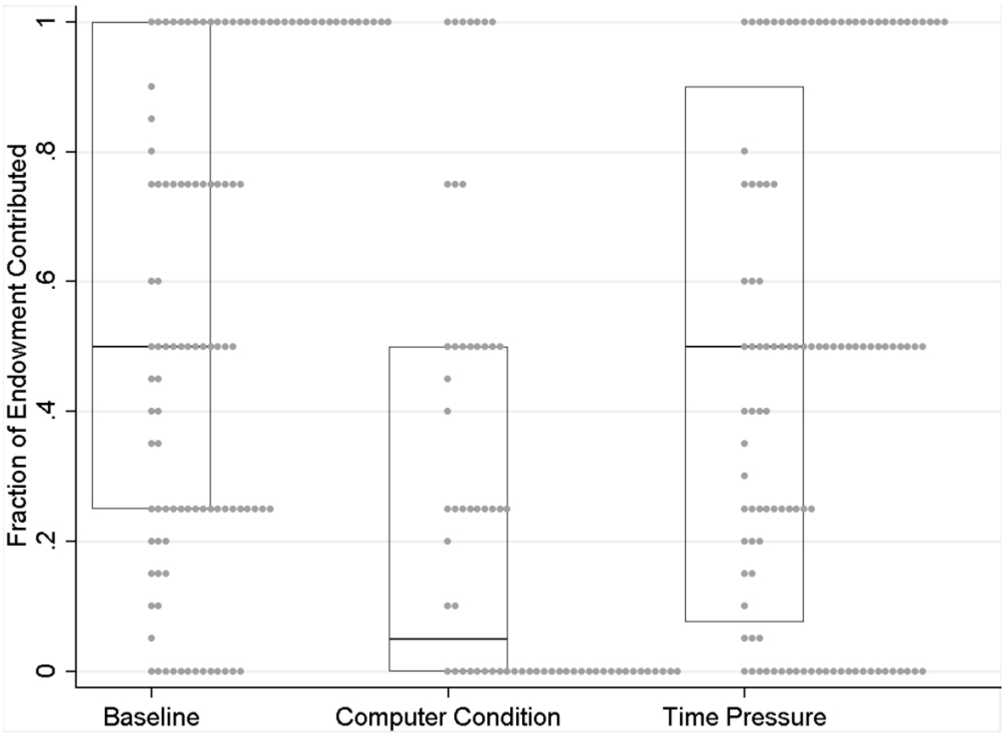
Notes: CRT-scores across the different treatment conditions. Values in brackets display corrected scores when discarding observations from subjects who indicate in the follow-up questionnaire to have known this item in advance. The last two columns contain reference values for CRT scores found in other studies.

the following distribution of individual CRT-scores: 34.9 percent of subjects answer all items correctly, 28.5 percent of subjects answer two items correctly, 20.1 percent of subjects answer one item correctly and the remaining 16.5 percent of subjects answer none correctly. The last two columns of Table 5.1 provide reference values for CRT-scores from two different earlier studies. Oechssler et al. (2009) use a subject pool comparable to that of the present experiment by collecting student samples from three different large German universities. CRT-scores observed in the present experiment are only slightly below their reported averages. The second set of reference values comes from a current comprehensive meta-study (Brañas–Garza et al., 2015) covering a total of 41,824 single CRT observations from students (41%) and non-students (59%), collected across a span of eight years. Subjects in the present experiment reached significantly higher scores than the average participant in Brañas–Garza et al. (2015), which puts them closer to the average student from two top US universities (MIT:2.18, Princeton:1.63) in the original sample of Frederick (2005). The values in brackets show solution rates

[12] There were also further incentivised questionnaires and decision tasks not analyzed in this paper. Since, each of these additional tasks was conducted subsequent to the one-shot PGGs they could not influence choices analyzed here. A companion paper discusses the effects of time pressure and confusion in these tasks.

when discarding observations from those 12-19 percent of subjects who indicate in the follow-up questionnaire that they already knew a given item[13].

FIGURE 5.1: Box plots of contributions across PGG variants



Notes: This graph shows a box plot of the fraction of endowment contributed, separately for the three PGG variants. The black line indicates median contributions. The lower and upper quartiles are marked by the surrounding box. The stacked gray dots overlay a histogram of individual contributions plotted in a horizontal direction.

Figure 5.1 displays the distribution of contributions across the three different PGG variants. The baseline contribution average ($\mu = 0.56$; s.d.=0.369) falls into the typical range of 40 to 60 percent of the initial endowment (Ledyard, 1995). Furthermore, there are three spikes in the distribution indicating that a majority of participants either free-rides, contributes their full endowment, or splits it evenly between the private and public account.

For the time pressure condition, the distribution of contributions is close to the baseline with an identical median, but a significantly higher incidence of free-riding (Chi^2-test: χ^2=4.069, p=0.044). Thus, in contrast to Rand et al. (2012, 2014), but closer to the null results reported in Tinghög et al. (2013) and Verkoeijen and Bouwmeester (2014), the application of time pressure results in a contribution average ($\mu = 0.476$; s.d.=0.374) below baseline at a weakly significant level (M.W. rank sum: z=1.661, p=0.096).

[13]Note that, despite claiming to know the item, not all of these subjects actually state the correct answer. Furthermore, excluding them from the analyzes below does not affect the main conclusions.

In the computer condition behavior deviates substantially from the baseline. When interacting with a computer program, subjects contribute 0.261 percent (s.d.=0.340) of their endowment on average, well in line with values for the first round in Houser and Kurzban (2002). This reduction, relative to the baseline, is mainly driven by the 50 percent of subjects who are able to identify their dominant strategy and contribute zero tokens. The positive contributions of the other half of subjects can be interpreted as a strong sign for the presence of confusion (Houser and Kurzban, 2002).

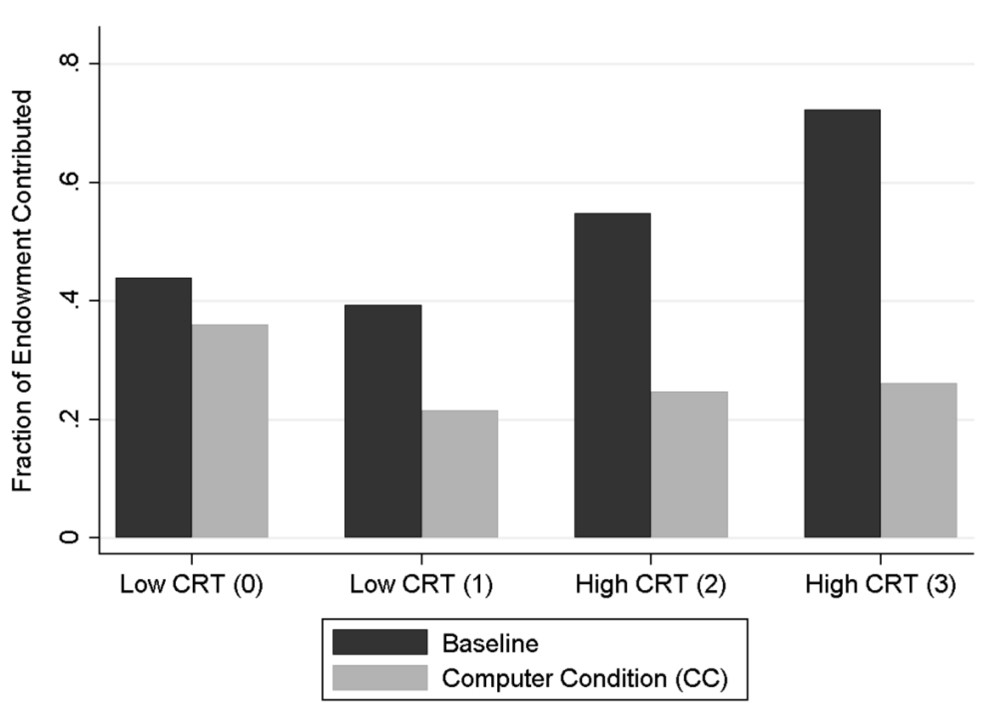
5.3.2 Comparing baseline (BL) and computer condition (CC)

Figure 5.2 displays the relationship between the number of correct answers in the CRT and contributions to the public account, separately for the BL (black bars) and the CC (gray bars). Contributions in the BL can be seen as an expression of cooperative preferences, at least for those subjects who fully understand the underlying incentives. In contrast, in the CC in which subjects' contributions go to a public account shared with a computer program, cooperative preferences towards other group members cannot motivate behavior. The CC, therefore, provides a more direct way to test whether subjects with higher CRT-scores are more able to identify their dominant strategy in a decision environment that is structurally similar to the baseline PGG.

In the BL there is a clear positive relationship between CRT-scores and contributions, which is highly significant (Spearman's Rho=0.338, $p<0.001$). On average, subjects with the lowest CRT-score contribute 44 percent of their initial endowment to the public account, whereas subjects with the highest CRT-score contribute 73 percent. This increase, by two thirds, is mainly attributable to a growing fraction of subjects who contribute their full endowment, while the fraction of subjects contributing zero tokens remains close to constant across all CRT-score categories. Only 15 percent of subjects in the lowest CRT-score category contribute their full endowment, while this is true for 50 percent of subjects in the highest CRT-score category. Comparing contributions across all CRT categories furthermore points at significant differences by category (ANOVA: $F=5.17$, $p=0.002$). Finally, splitting the sample into a low (0–1) and high (2–3) CRT group provides further evidence that contributions of the low group are significantly smaller than contributions of the high group (M.W. rank sum: $z=-3.181$, $p<0.01$)

In the CC there is a negative, but insignificant (Spearman's Rho=-0.129, $p=0.310$) relationship between contributions and CRT-scores. On average, subjects within the lowest CRT category contribute 36 percent of their initial endowment and this average drops to approximately 25 percent for subjects falling into the other three CRT categories. When comparing all CRT-score categories there is no evidence for significant differences in average contributions. (ANOVA: $F=0.38$, $p=0.771$). Splitting the sample into a low (0–1) and high (2–3) CRT group yields similar results (M.W. rank sum: $z=1.122$, $p=0.262$). If

Figure 5.2: Average contributions by CRT-scores: BL vs. CC



Notes: This graph shows the fraction of endowment contributed for subjects grouped by their CRT-score. Black bars are used for the BL and gray bars for the CC.

contributions in the CC are mainly driven by subjects' inability to recognize their dominant strategy, then subjects within the low CRT-score group display stronger signs of confused behavior: within this group only 30 percent of subjects contribute zero tokens as compared to 60 percent of subjects in the high CRT-score group.[14]

Although there are no differences in average contribution behavior, subjects with a low CRT-score are less likely to choose their dominant strategy[15]. Lower confusion levels could be one plausible channel for the finding of Nielsen et al. (2014) that subjects with higher CRT-score tend to free-ride more in their experiment. However, their finding is not in line with the positive relationship found for BL subjects. Of course, both experiments are only imperfectly comparable. While both use a one-shot PGG task, the experiment of Nielsen et al. (2014) is based on the strategy method, which might induce a different kind or level of reflection about the contribution incentives.

[14]Another observation that squarely falls in line with the notion of higher confusion levels among subjects in the lowest CRT category, is that for this group the proportion of choices that do not fall in one of the three distributional spikes (at 0%, 50% or 100%) is substantially larger, than in the highest CRT category. This is true both for the BL and the CC (BL: 70% vs. 26%; CC: 50% vs. 19%). This could be attributed to more random choices among confused subjects.

[15]Note, however, that despite a large quantitative difference (Low-CRT(0): 30%; High-CRT(3)60%) this result is not significant at conventional levels (Chi^2-test: χ^2=1.697, p=0.193).

Given that the link between CRT-scores and contributions goes in opposite directions in the BL and CC, confusion could mask an even stronger relationship between genuine cooperation and CRT-scores. BL contributions for subjects with the lowest CRT-score are statistically indistinguishable (M.W. rank sum: z=0.810, p=0.418) from contributions in the CC. Hence, relative to subjects in higher CRT categories, only a smaller fraction of BL contributions of low CRT subjects can be attributed to cooperative motives and a larger fraction to confusion. In contrast, for the high CRT-score group BL contributions are significantly larger (M.W. rank sum: z=4.690, $p<0.01$). This change of the relative importance of contribution motives across different CRT groups could bias the baseline correlation between contributions and CRT-scores downward, so that it is only a lower bound for the true relationship between cooperation and CRT-scores.

One additional concern could be that high CRT-scores overstate the true cognitive abilities for those subjects already familiar with parts of the test (Toplak et al., 2011, 2014). As a robustness check, I thus repeat the previous steps of analysis but restrict attention to the subset of subjects (87%) who are not familiar with more than one test item. This procedure does not affect the main conclusions. For the baseline there is still a high and significant correlation (Spearman's Rho=0.331, $p<0.001$) and average contributions are significantly different between all CRT categories (ANOVA: F=4.73, p=0.004). For the CC, in contrast, I do not find a significant correlation (Spearman's Rho=-0.1003, p=0.471) and subjects do not show different average contribution levels according to their CRT-score (ANOVA: F=0.80, p=0.499).

Taken together, the comparison of the BL and the CC leads to two central observations for the link between cognitive abilities and cooperation in a one-shot PGG. First, there is a strong and positive relationship between contributions and CRT-scores in the BL. This is a surprising result, to the degree that previous research strongly suggests that higher cognitive abilities typically help individuals to select their dominant strategy. These findings, however, mainly originate from decision tasks like beauty contest games, in which there is no prominent trade-off between individual and group interest. In the CC, in which no such trade-off is present, there is still only weak evidence that subjects with the lowest CRT scores are less able to identify their dominant strategy. Thus, detecting the dominant strategy in a PGG seems to be only marginally affected by a subject's ability to reflect about the problem.

5.3.3 Comparing baseline (BL) and time pressure condition (TP)

Figure 5.3 illustrates the relationship between the number of correct answers in the CRT and contributions to the public account, separately for the BL (black bars) and the TP (gray bars). By comparing these two conditions I analyze whether the link between cognitive abilities and contribution behavior is independent of the cognitive demands imposed by a given decision environment. Previous research has shown that imposing a time limit on participants in a one-shot PGG can increase contributions (Rand et al., 2012, 2014)[16]. Existing studies have, however, not analyzed whether this effect is the same for subjects with low and high cognitive abilities.

A potential interaction effect between time pressure and CRT-scores could, a priori, go in both directions. The CRT assigns a high score to decision makers who easily override a choice resulting from intuition in favor of a choice that is based on reflection. Time pressure, in turn, should make it harder for subjects to rely on reflection, as deliberating requires more time relative to deciding intuitively (Evans, 2008). On the one hand, decision makers with higher CRT-scores could be better equipped to cope with the time limit and still make a reflected decision. If true, the biggest effect of time pressure (relative to the BL) should materialize for subjects with low CRT-scores. On the other hand, if the time constraint was felt equally strong by all subjects, exactly the reverse outcome could emerge. Time pressure could particularly affect subjects who would otherwise rely on their reasoning abilities (high CRT-score), while it would have little effect on subjects who would have based their decision on a first intuition (low CRT-score) anyhow. Summarizing ahead of a more detailed analysis, figure 5.3 clearly points towards the second proposition.

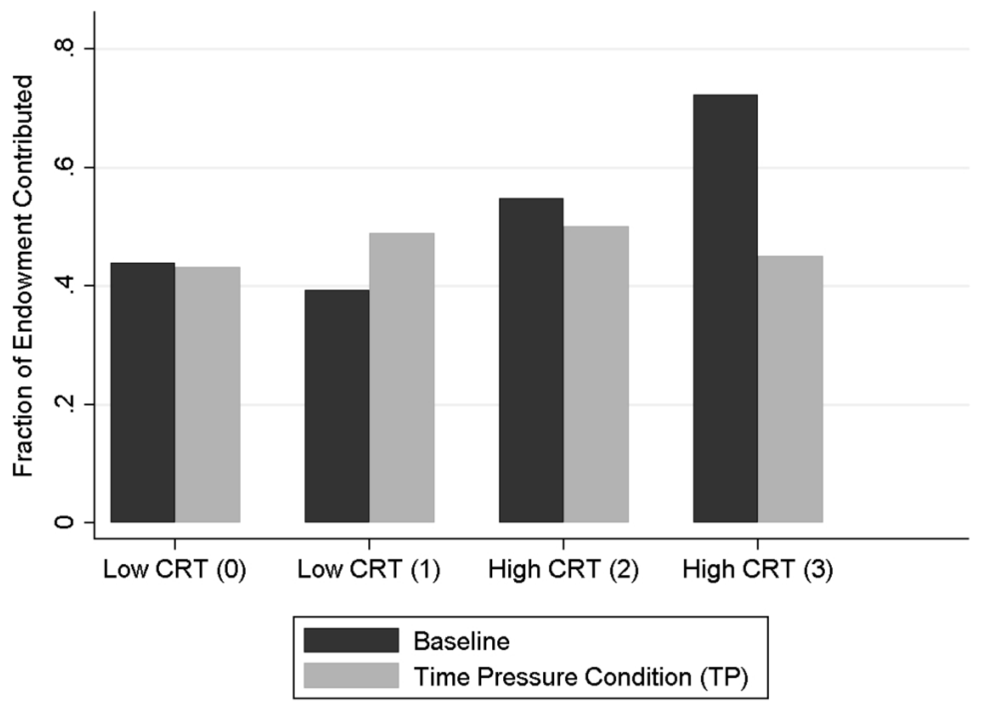
As discussed at length in the previous section, there is a clear positive relationship between CRT-scores and contributions in the BL. On average, subjects with the highest CRT-score contribute 66 percent more than subjects with the lowest CRT-score.

Turning to behavior in the TP, there is no significant correlation (Spearman's Rho=-0.022, p=0.817) between CRT-scores and contributions. Furthermore, the fraction of subjects contributing their full endowment in the TP is comparable for subjects in the highest (27%) and lowest (12%) CRT-score category (Chi^2-test: χ^2=1.697, p=0.193) and average contributions do not differ between all CRT-score groups (ANOVA: F=0.19, p=0.905). The incidence of free-riding is not linked to CRT-scores, and splitting the sample into a low (0–1) and high (2–3) CRT group yields a null result for the size of contributions (M.W. rank sum: z=-0.095, $p = 0.924$).

Comparing behavior between the TP and the BL condition demonstrates that subjects with high CRT-scores cease to contribute more, as soon as the decision environment impedes the use of reflection. In comparison to the BL, contributions under time pressure are at significantly lower levels for subjects within the high CRT group (M.W. rank sum: z=2.571, p<0.001). Subjects within the low CRT group do not contribute different amounts under time pressure, compared to those in the BL deciding without a time constraint (M.W. rank sum: z=-0.743, p=0.457). Excluding subjects with prior knowledge of the CRT affects these results only marginally. For high CRT subjects (2–3) differences between the BL and the TP remain weakly significant (M.W. rank sum:

[16]The authors explain their finding by proposing a *Social Heuristics Hypothesis*, according to which subjects under time pressure employ an intuitive decision heuristic that is often advantageous in their repeated daily interactions (cooperate), but is maladapted to the atypical one-shot decision context of a lab experiment (Rand et al., 2014).

Figure 5.3: Average contributions by CRT-scores: BL vs. TP



Notes: This graph shows the fraction of endowment contributed for subjects grouped by their CRT-score. Black bars are used for the BL and gray bars for the TP.

z=1.704, p= 0.081). For the highest CRT category differences between the BL and the TP remain significant (M.W. rank sum: z= 1.966, p=0.049).

To sum up, this second comparison underlines the higher propensity to cooperate among subjects with higher cognitive abilities, as long as the decision environment allows for the use of these abilities. When this is not the case, contributions across all CRT groups converge to the level of low CRT subjects. Taken together, this strongly suggest that reflection is a core determinant of higher contributions. While interesting in itself, this overlooked interaction effect could also provide one plausible explanation for the conflicting findings across different studies on the effects of time pressure in one-shot PGGs (Rand et al., 2012, 2014; Tinghög et al., 2013; Verkoeijen and Bouwmeester, 2014), if the single studies were based on samples with diverging CRT-score distributions. CRT-scores reported in Frederick (2005) show that mean scores already differ significantly between student subjects from different academic institutions (Min.:0.57; Max.:2.18).

5.3.4 Regression results

CRT-scores are typically related to a number of other individual attributes such as risk preferences, general cognitive abilities or gender (Frederick, 2005; Oechssler et al., 2009). Subjects in this study are no exception. Male participants reach a significantly higher average CRT-score (2.01) than female participants (1.63) (M.W. rank sum: z = -2.981, p=0.003)[17]. Stated risk attitudes were assessed by a standardized question from the German socio economic panel (*"How do you see yourself: are you, in general, a person fully prepared to take risks or do you try to avoid taking risks?"*), which has been shown to have a high behavioral validity (Dohmen et al., 2011). Surprisingly, and in contrast to existing findings based on risk lotteries, high and low score subjects do not differ in their (stated) risk attitudes (M.W. rank sum: z=1.268, p=0.205). Furthermore, CRT-scores are significantly correlated with a proxy of general cognitive abilities, namely better grades[18] in the math and language section of German A-level exams (Spearman's Rho=-0.2709, p<0.001). To the degree that each of these attributes is also related to contribution behavior in PGGs, the previously reported correlations could be biased (Croson and Gneezy, 2009; Teyssier, 2012). By employing a multiple regression framework, I control for these potential confounds. Obviously, this method cannot exclude the possibility that CRT-scores are related to an unobserved subject characteristic driving the results.

TABLE 5.2: Regression results

	(1)		(2)		(3)		(4)		(5)		(6)	
	Baseline		Computer Condition		Time Pressure		Full Sample		Unknown CRT		Unconfused	
CRT-score (0–3)	0.085**	(2.29)	-0.029	(-0.63)	0.019	(0.50)	0.110***	(3.22)	0.121***	(3.16)	0.143***	(3.17)
Male (1=Yes)	0.198**	(2.25)	0.068	(0.62)	0.015	(0.20)	0.076	(1.54)	0.061	(1.14)	0.077	(1.17)
Risk (1–11)	0.016	(1.02)	0.001	(0.07)	-0.011	(-0.72)	0.002	(0.21)	-0.002	(-0.16)	0.013	(1.03)
Average Grade (1–6)	-0.031	(-0.70)	-0.016	(-0.25)	-0.053	(-1.03)	-0.036	(-1.18)	-0.026	(-0.84)	-0.072*	(-1.87)
Age (Years)	0.006	(0.44)	-0.018	(-0.82)	-0.009	(-0.50)	-0.006	(-0.69)	-0.005	(-0.55)	-0.006	(-0.45)
Graduate Student (1=Yes)	-0.073	(-0.63)	-0.193*	(-2.00)	0.044	(0.51)	-0.048	(-0.82)	-0.072	(-1.15)	-0.126*	(-1.70)
Economics Major (1=Yes)	-0.005	(-0.05)	0.075	(0.67)	0.036	(0.40)	0.038	(0.69)	0.051	(0.88)	0.081	(1.04)
Prior Lab Experience (1=Yes)	0.006	(0.08)	0.163	(1.63)	-0.065	(-0.69)	0.015	(0.30)	0.042	(0.73)	0.019	(0.25)
Known CRT-items (0–3)	0.064	(1.52)	0.032	(0.56)	-0.079*	(-1.93)	-0.006	(-0.21)	0.039	(0.67)	0.018	(0.47)
CC(1=Yes)							-0.057	(-0.55)	-0.008	(-0.07)	0.129	(0.98)
CC*CRT-score							-0.139***	(-2.62)	-0.171***	(-3.06)	-0.207***	(-3.22)
TP(1=Yes)							0.122	(1.25)	0.141	(1.35)	0.196	(1.52)
TP*CRT-score							-0.114**	(-2.49)	-0.102**	(-2.01)	-0.161**	(-2.55)
Constant	0.243	(0.84)	0.730*	(1.76)	0.794**	(2.10)	0.592***	(2.82)	0.546**	(2.31)	0.609**	(2.08)
Observations	83		55		102		240		211		136	
R^2	0.2514		0.1903		0.0538		0.1700		0.1934		0.2349	
Prob > F	0.0001		0.2148		0.6712		0.0000		0.0000		0.0000	

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

Notes: This table shows OLS regression coefficients for the CRT-score when controlling for a set of individual attributes. T-statistics based on robust standard errors are shown in parentheses. The lower number of observations results from incomplete observations, especially for the variable *average grade*. Excluding this variable does not change results. Using a Tobit Model as an alternative specification that accounts for potential censoring in the contribution data does not alter the results for CRT-scores either. Including dummies for each CRT-score category, instead of a categorial variable, shows that the largest effect is between high (3) and low (0) score subjects.

[17] Rather than exclusively reflecting differences in upbringing or formal education, this gender effect could also be related to prenatal hormone exposure, as recent evidence suggests (Bosch-Domènech et al., 2014).

[18] According to the German convention the best grade is coded as a low number (1) and the worst grade as a high number (6). Consequently, a negative correlation indicates that higher CRT-scores are related to better grades.

Table 5.2 shows several specification of an OLS regression with contributions as a fraction of the initial endowment being the dependent variable. The main coefficient of interest is that of the *CRT-score* variable. Specifications (1) - (3) estimate separate models for the baseline and each of the treatment conditions. In addition to the potentially confounding factors discussed before, I also include several control variables that could have an impact on how well subjects understand the PGG. Holding constant these variables does not change the conclusions from the previous sections. CRT-scores are significantly correlated with contributions in the BL, but in neither of the two treatment conditions. With few exceptions none of the control variables is related to behavior in the different PGGs variants. Male subjects contribute higher amounts in the BL. Furthermore, graduate students give slightly less in the CC, which could relate to their better understanding of the incentives. Specification (4) pools observations across treatments and tests for heterogeneous effects of cognitive abilities by including dummies for assignment to the CC or TP condition, and their interaction with CRT-scores. Again, previous results persist. The main effect of CRT-scores remains positive and significant, while each of the interaction effects is significantly negative. Thus, compared to the BL, the relationship between CRT-scores and contributions is significantly smaller and even reverses in sign for the CC.

Specifications (5) and (6) comprise additional robustness checks. In specification (5) I restrict the sample to subjects, who have only limited prior knowledge of the CRT-test, by excluding subjects who state to have known more than one item in advance. This procedure leaves estimated coefficients for the remaining "inexperienced" subjects unchanged.

Specification (6) looks at a restricted sample as well, by excluding all subjects who failed to correctly answer an item intended as a confusion check. In the follow-up questionnaire subjects were asked to identify the strategy that would have maximized their own payoff in the one-shot PGG. Despite the fact that the question was incentivised[19], only 55 percent of subjects selected "zero tokens". Subjects stating this correct answer are included in the regression model, while the remaining observations are discarded. This robustness check again does not point towards a different conclusion. Rather the relationship between CRT-scores and contributions is slightly stronger. When excluding confused subjects, the relationship between grades and contributions becomes also weakly significant showing that subjects with better (i.e. lower) grades tend to contribute more. This replicates findings from earlier studies based on SAT scores and less complex sharing tasks, where confusion arguably plays a smaller role (Chen et al., 2013).

[19] In particular, subjects had a chance to win up to 4 € if stating the correct answer ("zero tokens").

5.4 Discussion and Conclusion

This paper highlights the role of cognitive abilities as one factor that could underly the heterogeneity of individual behavior observed in most PGG experiments. Specifically, I find a positive relationship between cognitive abilities and cooperation in a one-shot PGG. As for all correlational evidence, this finding should be interpreted with the necessary care, since it cannot be fully ruled out that unobserved factors drive parts of the reported relationship. The optimal empirical strategy to get around this potential source of bias and establish causality would be to randomly assign cognitive abilities to different individuals, which is infeasible for obvious reasons. Instead, like in the TP condition of this study, cognitive load (Milinski and Wedekind, 1998; Duffy and Smith, 2014) or time pressure (Rand et al., 2012, 2014) have been used as methods that inhibit the use of existing cognitive abilities in cooperation problems. However, as shown here in section 5.3.3 as well as in Carpenter et al. (2013), these methods can have heterogeneous effects on subjects with higher or lower existing cognitive abilities. Furthermore, subtle changes in the experimental design seem to affect the efficacy of cognitive load manipulations (Kessler and Meier, 2014). This should be kept in mind, when comparing treatment effects across studies based on these methods, especially when they collect samples from populations for which it is plausible to assume that the unobserved distribution of cognitive abilities is not the same, e.g., a convenience sample of students and a field sample from the general population.

Even though the interaction is one-shot, I observe higher average cooperation levels for subjects with higher cognitive abilities. Similar findings have emerged in other studies using finitely repeated (Burks et al., 2009; Jones, 2008) cooperation tasks or non-strategic (Chen et al., 2013) distribution tasks. From a theoretical perspective these findings are surprising: one-shot settings lack a *shadow of future interactions*, which could discipline free-riders and hence would make cooperation a profitable and rational long-term strategy (Axelrod and Hamilton, 1981). Consequently, in several of the existing experiments - including this one - "smarter" subjects actually appear to be less able to adapt their behavior to the specific incentives created by a one-shot interaction in the lab. Based on the evidence provided here, I can only speculate on the underlying causes. Positive contributions in one-shot PGGs have been attributed to the possibility that some subjects fail to discriminate between the lab context and their daily interactions and thus behave "as if" the PGG was infinitely repeated. If one is willing to assume that this failure is unrelated to cognitive abilities, this could partly resolve the seeming contradiction. Alternatively, Millet and Dewitte (2007) propose that unconditional cooperation in a one-shot setting could serve as a costly signal of cognitive abilities that is cheaper for individuals with high cognitive abilities to the degree that these abilities allow them to earn higher incomes more easily.

Obviously, it would be of substantial interest to learn more about the question of *why* smarter subjects tend to cooperate more in my experiment and some related studies.

One unexplored behavioral channel through which CRT-scores could be related to cooperation is via time preferences and self-control. Subjects with higher cognitive abilities have been shown to be more patient (Frederick, 2005) and to be able to exert higher levels of self-control (de Wit et al., 2007). Experimental evidence as well as theoretical considerations suggest that both of these attributes could drive behavior in PGGs. Martinsson et al. (2014) demonstrate that subjects with higher self-control capabilities cooperate more. Similarly, Fehr and Leibbrandt (2011) show that patience assessed by a lab measure predicts cooperative behavior in the field. Thus exploring the relationship between self-control, cognitive abilities, individual discount rates and cooperation could be of great interest for further research.

One potential limitation of this study is that both by using the relatively short CRT and through sampling from a student population, the observed variance of cognitive abilities is fairly limited. Yet taking this limitation into account, it is rather more surprising to which extent subjects behave differently across different CRT-score categories.

5.5 Appendix

Instructions

CRT-Questionnaire
[SCREEN 1]
Each of the questions on this screen has only one correct answer. If you answer all questions on this screen correctly, you can win up to 4 €. If several participants should answer the questions correctly, the 4 € will be divided equally among these participants.

- A tennis racket and a ball cost 1.10 € in total. The bat costs 1.00 € more than the ball. How much does the ball cost?
- If it takes 5 machines 5 minutes to make 5 widgets, how long will it take 100 machines to make 100 widgets?
- In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

CRT-Post Questionnaire
[SCREEN 2]
Please indicate for each question of the previous questionnaire whether you knew the question prior to this study.

- Question 1 (Ball and Racket) << Radio Button: Yes/No >>
- Question 2 (Machines) << Radio Button: Yes/No >>
- Question 3 (Lilly Pad) << Radio Button: Yes/No >>

Chapter 6

Take from one to give to another*

*Financial support by the German Ministry for Education and Research under the project "SOKO ENERGIEWENDE" is gratefully acknowledged. Furthermore, I would like to thank the audiences at Universität Heidelberg and ZEW Mannheim for their valuable comments. This chapter has also profited from comments by Tobias Cagala, Timo Goeschl, and Israel Waichman.

6.1 Introduction

Standard economic theory predicts that the private provision of public goods is inefficient because individuals have strong incentives to free ride on the contributions of others (Samuelson, 1954). However, confronting participants in laboratory or field experiments with a provision decision reveals that a significant fraction of them is willing to cooperate by contributing, despite such incentives (e.g. Ledyard, 1995; Chaudhuri, 2011; Vesterlund, 2014). The fraction of contributors is even larger in repeated interactions, which allow for the punishment of free-riders (Fehr and Gächter, 2000; Nikiforakis, 2008) or offer room for reciprocity (Keser and Winden, 2000; Fischbacher et al., 2001; Fischbacher and Gächter, 2010).

Public good games have been established as the standard vehicle for analyzing cooperation problems, and in the typical experiment every participant benefits uniformly from higher provision levels. Yet, for many naturally occurring public goods, the assumption of uniform benefits is, arguably, somewhat oversimplified. Instead, individuals are regularly affected in different ways by the contribution choices of others. Do such asymmetries promote or hinder the private provision of public goods? While there is already a rich experimental literature on the effects of heterogeneous benefits (e.g. Fisher et al., 1995; Reuben and Riedl, 2013; Kesternich et al., 2014a), much less is known about a different kind of asymmetry in which the provision of a public good – although beneficial to one group of individuals – reduces the welfare of another group of individuals (Engel and Zhurakhovska, 2014). Such negative provision externalities could arise in two, sometimes related, ways: First, preferences for the public good could be polar in such a manner that it actually constitutes a public bad for a subset of the population (Isaac et al., 2013). Thus, the provision of the public good per se, even if globally efficient, could affect some individuals negatively. For instance, most parents in a neighborhood will benefit from the construction of a new playground, while the adjoining residents might experience disutility from the increase in noise. Second, in some cases, third parties can be forced to contribute to a public good, while being excluded from its benefits. The current paper will mainly focus on this second case. Two further examples illustrate constellations in which the money of a third party can be spent on public good contributions without its explicit consent: A CEO giving a fraction of a firm's revenue to charities provides a public good at the expense of its shareholders. Similarly, if donations to charities are tax-deductible, the costs of the rebate are (sometimes involuntarily) distributed over all taxpayers.

In this paper, I will describe the results of a simplified contribution experiment that reproduces stylized features of the decision problem portrayed in the last two examples. These stylized features recreate two changes in the incentive structure which follow directly from the presence of an affected third party: First, decision-makers are now confronted with trade-offs between the welfare of several parties. If voluntary contributions to public goods result from social preferences, as generally assumed in the literature (e.g. Andreoni, 1990), similar preferences should prevent decision-makers from ignoring losses due to involuntary contributions by third parties. Second, the presence of a third party reduces the share of contribution costs accruing to the decision-maker – in the most extreme case to zero. If decision-makers treat public good contributions as an ordinary good (Peloza and Steel, 2005; Karlan and List, 2007), this price change should increase provision levels. The total effect on contributions could therefore depend on a non-trivial interaction between price effects and the degree to which social preferences extend to third parties.

Employing a one-shot decision task that deliberately does not allow for cases of strategic interaction over time (Engel and Rockenbach, 2011; Engel and Zhurakhovska, 2014), I explore whether the presence of a negative externality, borne by one other participant, changes contribution behavior. In each condition there is a single decision-maker and, for some decisions, one harmed outsider. The single decision-maker can contribute parts of his endowment – earned in a real-effort task – to a real-world public good at five different prices of giving. As a treatment variation, this decision is either made in isolation or in the presence of an outsider who has to co-finance the different rebates. By comparing these two settings, I analyze if a reduction in the price of giving affects contribution behavior differently when the costs of this reduction have to be borne by an identifiable outsider. By varying both the price of giving and the intensity of harm inflicted on the outsider, I furthermore explore different motives that could underly such changes in behavior.

My findings suggest that the price of giving and other regarding preferences towards the outsider jointly determine average contributions: As expected by arguments of standard price theory (e.g. Bergstrom et al., 1986), reducing the price of giving leads to higher contributions, albeit under one important qualification. When the outsider has to bear the majority and up to the full share of provision costs, a price reduction ceases to increase giving. This result is best explained by (advantageous) inequity aversion. In line with this motive, negative welfare effects are mostly ignored, in those cases where the decision-maker and the outsider face the same prices of giving. Therefore, in contrast to Engel and Zhurakhovska (2014), high absolute levels of harm do not always lead to lower contributions. For some combinations of contribution costs, this leads to substantial losses in total efficiency. Each of these core results continues to hold within a second treatment group in which the decision-maker and the outsider earn heterogeneous incomes in the real-effort task.

The remainder of this paper is structured as follows: Section 6.2 discusses the research question in relation to the existing literature. Section 6.3 introduces the experimental design and procedures in greater detail. Section 6.4 states some predictions, and the corresponding results are reported in section 6.5. Finally, Section 6.6 concludes with a brief discussion.

6.2 Related Literature

The main research question of this paper relates to two distinct strands within the experimental literature. The first strand has looked at the role of prices for contribution behavior, while the second one has started to explore the role of third party externalities.

There is converging evidence, both from the laboratory and the field, that decision-makers react to changes in the price of giving. In laboratory experiments (Eckel and Grossman, 2003, 2006; Davis et al., 2005) a lower price of giving - resulting from either matching contributions or mathematically equivalent rebates[1] - leads to higher contribution levels, with quantitatively larger effects observed under a matching protocol. Findings from field experiments on charitable giving mirror these results. By investigating different matching ratios, Karlan and List (2007) find a 1:1 match to significantly increase generosity, whereas larger ratios lead to no additional increases. Further studies have largely replicated these results for various matching ratios (Huck and Rasul, 2011) and rebate rates (Eckel and Grossman, 2008). Moreover, for the real public good studied here, direct price reductions (Löschel et al., 2013; Diederich and Goeschl, 2014), as well as indirect price reductions resulting from a rebate or a matching subsidy (Kesternich et al., 2014b) have also been found to increase contributions. This first strand of literature highlights the role of prices for giving while there are still open questions regarding the monotonicity of price effects and whether they operate at the intensive or extensive margin (Karlan and List, 2007; Huck and Rasul, 2011). Furthermore, existing studies have largely ignored the question whether the origin of the match (voluntary vs. involuntary) affects the strength of a price effect. This is mainly due to methodological reasons. Most of the relevant field studies are embedded into the context of philanthropic giving in which matches are typically financed by a lead donor. Since matching in this form can be assumed to be voluntary, welfare concerns toward the lead donor should be immaterial. But, as the introductory examples illustrate, not all forms of (implicit) rebates can be said to be voluntary. These cases can give rise to negative contribution externalities if the preferences of the decision-maker and the preferences of those financing the rebate are not perfectly aligned.

The second nascent literature scrutinizes the role of third party externalities in various experimental settings. Several recent papers look at such effects in strategic decision tasks. Engel and Rockenbach (2011) and Engel and Zhurakhovska (2014) come to mixed conclusions when analyzing contribution externalities in either a public good game or a prisoner's dilemma. In the public good game provision levels are unaffected by additional positive or negative effects on three outsiders. Yet, in the prisoner's dilemma conditional

[1]Under a matching protocol a third party contributes a fixed additional amount of money for each 1 € contribution made by the decision-maker. When a 1:1 matching ratio is in place, the third party gives 1 € for each 1 € that was contributed. Similarly, under a rebate scheme, the third party returns a fixed percentage of the initial contribution to the decision-maker. Under a 50 % rebate scheme, decision-makers receive back 1 €, if they contribute 2 €. Therefore, a 1:1 matching ratio and a 50% rebate rate are mathematically equivalent in the sense that they reduce the final price of giving by the same amount.

cooperation decreases when contributing inflicts increasing levels of absolute harm on an outsider. This is also in line with findings in Bartling et al. (2014). They observe the behavior of producers and buyers in an experimental market in which successful trades can create a negative externality for a third party. Their results demonstrate that, contrary to the equilibrium prediction under standard preferences, a significant fraction of buyers is actually willing to accept a price premium in exchange for avoiding the negative externality. Thus, concerns toward externality-bearing third parties seem to persist in a competitive market setting. In the context of non-strategic decision tasks, several studies use modified dictator games to study third party effects. Schumacher et al. (2014) find that dictators give less to a recipient if the costs of giving are dispersed over a group of outsiders. However, as dictators ignore how many outsiders are affected by their decision, this leads to losses in net efficiency as the group of outsiders grows in size. While Engel and Zhurakhovska (2014) also look at third party externalities in a public goods setting, closest to the design used in my experiment are the studies of Carlsson et al. (2011) and Chavanne et al. (2011). In the experiment of Carlsson et al. (2011), a decision-maker could give to an environmental charity either only from his own endowment or by dictating a minimum contribution for a group of other participants. Their findings indicate that average contributions decline as soon as others are affected. Chavanne et al. (2011) study behavior in a group dictator game. In their treatment condition, giving reduces the payoffs of other dictators but is, in turn, cheaper than in a baseline setting without the externality. In this setup, a strong in-group bias prevents the resultant price effect to increase giving in the treatment group.

This paper contributes to both strands of literature as follows: It explores the role of rebates for increasing contribution behavior when these are explicitly financed by an identifiable outsider. This adds a new aspect to the existing experimental literature on rebates that has largely ignored their potential welfare implications for third parties. Furthermore, the present design allows one to differentiate more thoroughly between several other-regarding motives that could underly the hitherto mixed evidence on contribution externalities. By varying two experimental parameters, I discriminate between a set of other-regarding preferences that could become actionable in the presence of an affected outsider. First, the (relative) level of harm inflicted on the outsider by contributing could influence behavior based on the decision-makers' preferences for efficiency, equity, and altruism. Second, each of these three preferences could be enhanced or attenuated as soon as the decision-maker and the outsider earn unequal endowments in the real-effort task.

6.3 Design and procedures

The experiment proceeded in two stages. The first stage was a real-effort task in which subjects earned an endowment that they could use in the subsequent contribution task.

In the real-effort stage the experiment implemented two treatment conditions between-subjects: In the *equal income condition (EI)* subjects received the same payment, independent of their performance, while in the *different income condition (DI)* payment was based on relative performance. The second stage contained the actual contribution task. Treatment effects in this task are analyzed on the basis of a within-subjects variation of the *price of giving* and the effects that contributions have on a *second participant (outsider)*. I will now describe the single stages in greater detail.

The real-effort task

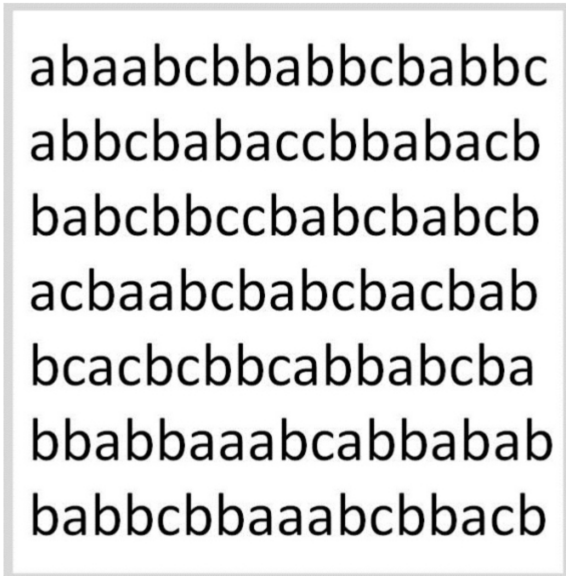
After completing a basic demographic questionnaire, the participants were randomly matched into groups of two and received instructions for the real-effort task. This task served two purposes in the experiment. On the one hand, it aimed at increasing the level of parallelism to real world contribution decisions by generating earned income. In field decisions, incomes are typically earned rather than emerging out of a windfall profit, and this crucial difference has been shown to affect contribution behavior (Cherry et al., 2002; Reinstein and Riener, 2012; Oxoby and Spraggon, 2013; Carlsson et al., 2013)[2]. On the other hand, the task was used to introduce a natural between-subjects variation of initial endowments. In the DI condition (N=74) the higher-scoring participant received an endowment of 12 Euros, whereas the participant with the lower score received 8 Euros.[3] Accordingly, in the subsequent contribution stage, some decision-makers were in an advantageous position while some others in a disadvantageous one, relative to the outsider. In the EI condition (N=76) both participants received the same endowment (8 Euros), independent of their performance. By comparing both treatments, I analyze whether the initial distribution of earned income affects the weight that the decision-maker attaches to the final payoff of the outsider in the second stage.

In the real-effort task, participants faced a succession of several screens on which they had to count how often a certain letter, digit, or geometrical object was shown. Figure 6.1 displays an example screen on which subjects had to count the instances of the letter b. Participants received one point per correct count. This task was designed to be simple enough to prevent subjects with higher cognitive abilities or other attributes related to performance from being overrepresented in the high payment condition. Similarly, while subjects were informed up-front that there would be a subsequent task in which they could use the earned endowment, there was no reference to the content of this task in order to prevent subjects with higher social preferences from expending more effort (Erkal et al., 2011). Also, the real-effort task was designed to be sufficiently tedious to promote a feeling among the participants that they have earned the endowment. Overall, the subjects had three minutes to solve as many screens as possible. One example screen was shown prior to the actual task to familiarize subjects with the

[2]For deviating results see Clark (2002) and Cherry et al. (2005)

[3]In cases in which both participants scored equally, earnings were determined randomly.

Figure 6.1: Example screen of the real-effort task



Notes: This figure shows an example screen of the real-effort task in which participants had to count the instances of the letter b. In other variants of this task, the subjects had to count the instances of the number 1 or instances of a triangle.

setup. The average solution rate was 3.72 (min.:0; max.:7). Surprisingly, average scores do not differ significantly between the equal and unequal payment scheme. The number of correct answers uncorrelated with all the observed individual attributes, including a proxy for cognitive abilities.

The contribution task

After completing the real effort-task and receiving information about their earnings, the participants entered the second stage of the experiment. At the beginning of this stage, subjects were re-matched with a new participant.[4] Within each group of two, subjects could either take the role of the decision-maker or the role of the outsider. Roles were assigned randomly at the end of the experiment after choices were collected for each subject in the role of the decision-maker. No choices could be made in the role of the outsider. The contribution task itself was based on a modified dictator game (Eckel and Grossman, 1996) in which the decision-maker could allocate a part of his endowment to an existing organization providing a public good. The naturally occurring public good in this experiment was chosen from the behaviorally rich sphere of climate change mitigation. The public goods nature of voluntary climate actions is widely agreed on by economists (Nordhaus, 1991, 1993). Each decision-maker could select how much money would go to an offsetting program of CO_2 emissions. This particular choice serves two

[4]There could be concerns that putting participants under competitive pressure in the DI condition of the real-effort task would lead to feelings of antipathy toward the outsider in the subsequent contribution task. In order to minimize such potential carry over effects, I assign participants to a new partner they have not competed against in the real-effort task.

purposes in the experiment. On the one hand, for this global and inter-temporal public good, it is credible that its marginal benefits from consumption are in the neighborhood of zero, both for the decision-maker and the outsider. Thus, in comparison with laboratory public good games used in the previous literature (Engel and Rockenbach, 2011), the decision-maker cannot gain financially by taking from the outsider.[5] On the other hand, benefits from mitigation accrue mainly to future generations due to the physical nature of the climate system. Thus, behavior observed in this experiment can also shed some light on the weight attached by decision- makers to members of a future generation, when trading off their welfare against that of members of the current generation. Before any decision could be made, a screen informed participants about the specific public good.[6]

Similar to comparable studies (e.g. Andreoni and Miller, 2002; Eckel and Grossman, 2003), each decision- maker selected contribution levels for a list of allocation problems, each implementing a distinct combination of incentives. At the end of the experiment, one of the 15 resulting choices was randomly executed with equal probability. The different variants of the allocation problem can be grouped into three distinct conditions, each containing five allocation problems. Under the *baseline condition (BL)*, the allocation problems closely resemble those of the rebate treatments of Eckel and Grossman (2003, 2006). Each decision-maker could pass on up to 8 € (in steps of 1 €)[7] to the receiving organization *without affecting the outsider*. Allocation decisions could be made at five different *prices of giving* ($P_s = 1.0; 0.8; 0.50; 0.2; 0.0$). These prices were explained to subjects as representing their individual costs of contributing 1 € to the organization. It has been shown that this explanation is easier to understand for subjects and thus likely reduces bias from confusion (Davis et al., 2005)[8]. In the BL, decision-makers were explicitly informed that their choices had no influence over the payoffs received by the other participant in their group. This design feature changed in the two outsider conditions. Here, at each price of giving, the decision-maker was informed that there were also *provision costs borne by the outsider*. In the five allocation decisions under the *additive condition (ADD)* the price faced by the outsider (P_o) and the price faced by

[5]This eliminates the scope for "moral biases"(Croson and Konow, 2009) in the trade-off between the decision-maker and the outsider, which would otherwise confound the pure effect of the contribution externality.

[6]The text on this information screen included a reference to the amount of CO_2 that could be reduced for each Euro contributed. To make this amount more tangible, it was also expressed in terms of two common activities (car travel; use of personal computer). As this particular public good was chosen because of its minimal marginal consumption value for both the decision-maker and the outsider, the temporal delay between the reduction of CO_2 emissions in the atmosphere and the impact of this reduction on climate change was highlighted. In order to enable participants to verify that all contributions were passed on to the organization, an anonymous confirmation of the recipient was posted publicly on campus after the last session. The subjects were informed about this procedure before they were able to make any decisions.

[7]While a discrete action space affects the flexibility of allocations, it has been shown in a comparable setting that hardly any participant made use of intermediate values, when given the possibility (Davis et al., 2005).

[8]An implementation that is arguably more prone to error would state the full price (1 €) and convert each price below 1 € into a rebate, paid back to contributors at the end of the experiment as in Eckel and Grossman (2003)

the decision-maker (P_s) add to one ($P_s + P_o = 1$). This represents a case in which the rebate is fully financed by an outsider. For several prices, positive contributions under this condition affect payoff equality. They increase inequality when participants have initially equal endowments (EI) and decrease inequality when decision-makers start out with a relatively lower endowment (DI). Based on this difference, I analyze if potential changes in contribution behavior are driven by inequality aversion. To differentiate between equity and efficiency concerns, there was a second outsider condition. In the allocation problems of the *equal condition (EQU)*, the price for the decision-maker and the outsider was equivalent ($P_s = P_o$). Accordingly, there are no changes in pay-off inequality resulting from positive contributions, but changes in efficiency (as measured by the total cost of contributing (1 €). Table 6.1 gives an overview of the three allocation problems for each price P_s.

TABLE 6.1: Observed combinations of P_s and P_o

		Cost of 1 Euro contributed for the outsider (P_o)		
	P_s	BL	ADD	EQU
Price of giving 1 Euro (P_s) to the organization [Equivalent Rebate]	1.00 [r-0%]	Not Affected	0.00	1.00
	0.80 [r-20%]	Not Affected	0.20	0.80
	0.50 [r-50%]	Not Affected	0.50	0.00
	0.20 [r-80%]	Not Affected	0.80	0.20
	0.00 [r-100%]	Not Affected	1.00	0.00

Notes: This table displays the 15 allocation problems under the different treatment variations. By design, at $P_s = 1$, both under the BL and the ADD condition the task corresponds to a standard modified dictator game (Eckel and Grossman, 1996). Contributions at $P_s = 0$ under BL and EQU can be seen as measuring the value decision-makers attach to the provision of the public good, when no monetary costs have to been borne by either party. Note that in one instance ($P_s = 0.50$), the EQU treatment and the ADD would be equivalent using the standard rule. Instead, here I set the price $P_o = 0.00$ in the EQU condition, which allows for one more point of comparison with the BL.

The different allocation problems were displayed as five blocks[9] in three different orders. Based on pairwise comparisons and regression-based tests, there is little evidence for systematic order effects.[10]

Procedures and sample

The experiment was conducted at the University of Heidelberg AWI-Lab for Experimental Economics between May and December 2014 in 15 sessions. All instructions and tasks were fully computerized using z-Tree (Fischbacher, 2007), and each session was run

[9] On each of five decision screen there was a block of three (BL,ADD,EQU) allocation problems for a given level of P_s

[10] Out of the 45 resulting pairwise comparisons I find significant differences in behavior for only five decisions (M.W. Rank Sum Test, $p < 0.05$). Thus, I pool data from the different order conditions for further analysis and control for order effects in all regressions.

by the same experimenter. The 150 participants were recruited via ORSEE (Greiner, 2015) from the standard subject pool of undergraduate and graduate students, excluding subjects who had prior experience with similar distribution tasks. The participants were from mixed disciplines, including economics (approx. 25 %). At the beginning of a session, they were seated at one of the available computer terminals, a random password was generated to ensure their anonymity, and they received a set of general instructions that was read aloud by the experimenter. The two tasks were followed by a set of demographic questionnaires. At the end of the experiment, the randomly selected choice was executed and subjects were paid in private in accordance with their randomly determined role. All sessions lasted for 45–60 minutes, and the average payment excluding an additional show-up fee of 2 € was 6.50 € (Min: 0 €; Max:8 €) in the EI treatment and 9.16 € (Min:4 €; Max:12 €) in the DI treatment.[11]

6.4 Predictions

In order to derive behavioral predictions, the experimental outcome can be written as one of two related variables. On the decision screen subjects select the final contributions received by the organization. This primary outcome variable indicates the provision level that results from the different treatment conditions. The contributions actually given by the decision-maker (i.e., finally subtracted from his initial endowment after accounting for P_s) and, where relevant, by the outsider (after accounting for P_o) can be used to differentiate between different underlying motives. In the following, let $C^T_{P_s}$ denote the average contribution received by the organization for condition $T \in (BL; ADD; EQU)$ and price of giving $P_s \in (1.0; 0.8; 0.5; 0.2; 0.0)$. Likewise, let $G^T_{P_s}$ denote the average contribution given. These two outcomes are related by the functional relationship $G^T_{P_s} = P_s C^T_{P_s}$. The following testable hypotheses can be derived from theoretical considerations that build on standard models of giving (Andreoni, 1989, 1990; Karlan and List, 2007) and other regarding preferences (Fehr and Schmidt, 1999), as well as from existing experimental findings.

If participants were exclusively motivated by maximizing their own monetary payoff, contributions should be zero for any positive price of giving. Likewise, pure payoff maximizers should weakly prefer the maximum contribution level, when the price of giving is zero. As accumulated experimental findings document, participants often deviate from this strong theoretical prediction (e.g. Eckel and Grossman, 1996; Vesterlund, 2014).

The first two predictions relate to behavior in the BL, in which other regarding preferences only apply to the potential beneficiaries of climate change mitigation. In this context, several motives for positive contributions have been discussed (e.g. Karlan and List, 2007; Huck and Rasul, 2011), and theories commonly distinguish between pure

[11] Total earnings were higher for some subjects as there were additional earnings from a follow-up question not analyzed here, which was run after the contribution task in each session.

altruism and warm glow (Andreoni, 1989, 1990). A participant purely motivated by the "warm glow" felt through the act of giving itself, would keep constant $G^T_{P_s}$ for all (positive) levels of P_s. Then, by the relationship $G^T_{P_s} = P_s C^T_{P_s}$, contributions received increase automatically for lower levels of P_s. A participant motivated by pure altruism, and hence the total provision level, would increase $C^T_{P_s}$ due to the price effect of lowering P_s (Karlan and List, 2007). If participants, however, engage in pure donation targeting (Huck and Rasul, 2011), they would keep $C^T_{P_s}$ constant, thereby ignoring price changes. This, in turn, would automatically decrease $G^T_{P_s}$ by the arithmetic of the rebate mechanism. There is ample experimental evidence (e.g. Crumpler and Grossman, 2008; Huck and Rasul, 2011) that each of these motives affects behavior to different degrees. Thus, a theoretical prediction is not clear cut and depends on the relative strength of each of these motives. Drawing on the previous experimental findings (Eckel and Grossman, 2003, 2006; Karlan and List, 2007; Huck and Rasul, 2011; Kesternich et al., 2014b), I expect that $C^{BL}_{P_{s_t}} \geq C^{BL}_{P_{s_k}} \forall P_{s_k} \geq P_{s_t}$ and $G^{BL}_{P_{s_t}} \geq G^{BL}_{P_{s_k}} \forall P_{s_t} \geq P_{s_k}$. In words:

Hypothesis 1: ***In the BL, lower prices lead to (weakly) higher contributions received but (weakly) lower actual giving.***

Theoretical consideration backed by empirical and experimental evidence point towards a positive relationship between income and giving. From the existing laboratory evidence, it is, however, unclear if this relationship will persist for endowments earned in a real-effort task (Eckel and Grossman, 2003, 2006; Erkal et al., 2011; Tonin and Vlassopoulos, 2013). In the context of this study, I still expect that subjects with the higher endowment in the DI condition would contribute more.

Hypothesis 2: ***In the BL, those subjects who earn a higher endowment in the DI treatment contribute more.***

As soon as contributions affect an outsider in the ADD and EQU conditions, additional motives could become actionable. For subjects who are only motivated by their own monetary payoff, contributions should remain at zero in these conditions. Yet, for subjects whose positive contributions in the BL indicate the presence of other-regarding motives, these motives could extend to the payoff received by the outsider. There are three candidate preferences that could, in theory, be in line with contributions in the BL and could, in turn, keep decision-makers from contributing as soon as it decreases the payoff of an outsider: Linear altruism, inequity aversion, and efficiency concerns.

If linear altruism (e.g. Ledyard, 1995; Ahn et al., 2003) was driving contributions in the BL, the participants' utility would also depend on the absolute payoffs received by the outsider in the ADD and EQU conditions. Whenever contributing reduces theses payoffs, decision-makers should (weakly) reduce their contributions relative to the BL.

This effect is expected to be present both in the ADD and EQU conditions, and should be increasing in the harm inflicted on the outsider. Furthermore, this reaction should also be independent of the initial distribution of endowments. Thus, if linear altruism was the main underlying motive, I would expect the following outcome:

Hypothesis 3: ***For each price level, contributions in the ADD and EQU are weakly lower than in the BL. This holds independent of the initial distribution of incomes.***

Inequity aversion, as discussed in Engel and Zhurakhovska (2014)[12], could be another important motive. Theoretical models (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) and the associated experiments, strongly suggest that individuals experience disutility from both advantageous and disadvantageous inequity. If true, the presence of an outsider should influence contributions only in the ADD condition but not in the EQU condition. By design, contributions in the EQU do not affect relative earnings and there is no outsider affected in the BL. Hence, an inequity-averse decision-maker should contribute less in the ADD for the price pairs $(P_s; P_o) = \{(0.80; 0.20); (0.20; 0.80); (0.00; 1.00)\}$. For each of these three combinations, positive contributions are strictly inequity-increasing when the decision-maker and the outsider have equal endowments (EI). In these cases, I expect that $C^{ADD}_{P_{s_k}} \leq C^{BL}_{P_{s_k}} = C^{EQU}_{P_{s_k}}$. This relationship could change for those decision-makers who earn a relatively lower endowment in the DI condition. Positive contributions are now inequity-decreasing for the price pairs $(P_s; P_o) = \{(0.20; 0.80); (0.00; 1.00)\}$. In words, if decision-makers account for the outsider's welfare and are purely inequity averse, I would expect the following outcomes for the EI and DI treatment, respectively:

Hypothesis 4a (EI): ***For each price level, contributions in the ADD condition are weakly smaller than contributions in the BL or the EQU condition.***

Hypothesis 4b (DI): ***For each price level, contributions in the ADD condition are weakly larger (smaller) than contributions in the BL or the EQU condition, depending on whether the decision-maker is initially in the disadvantageous (advantageous) endowment position.***

If participants perceive the outsider as part of the decision problem, their choices could also be influenced by the total cost borne jointly by both parties for each Euro received by the organization. Such efficiency concerns have been shown to be an important driver of individual decisions in comparable settings (e.g. Charness and Rabin, 2002; Engelmann and Strobel, 2004). In the ADD condition the total cost are in sum always 1. Conversely, in the EQU, the total cost of provision are greater than 1 for prices $(P_s = 1.0; 0.80)$.

[12]In contrast to the strategic setting of their paper, in the present non-strategic context, beliefs about the actions of another decision-maker should not play a role.

Thus, decision-makers only concerned about efficiency should contribute weakly less in the EQU than in the ADD and BL condition at these prices ($C^{EQU}_{P_{s_k}} \leq C^{BL}_{P_{s_k}} = C^{ADD}_{P_{s_k}}$).

Hypothesis 5: ***For each price level, contributions in the EQU condition are weakly smaller than contributions in the BL or the ADD condition.***

6.5 Results

6.5.1 Observed behavior

In absence of an externality and a rebate (BL; $P_s = 1$), subjects gave on average 17.35 % of their initial endowment[13], which is less than the 30 % observed in a comparable setting (Eckel and Grossman, 1996). This lower level of giving could reflect the different sources of income in the two experiments, as contributions are typically smaller when endowments have to be earned (Cherry et al., 2002; Reinstein and Riener, 2012; Oxoby and Spraggon, 2013; Carlsson et al., 2013). Additionally, some subjects might simply attach a lower value to the provision of the specific public good offered in the present study.[14]

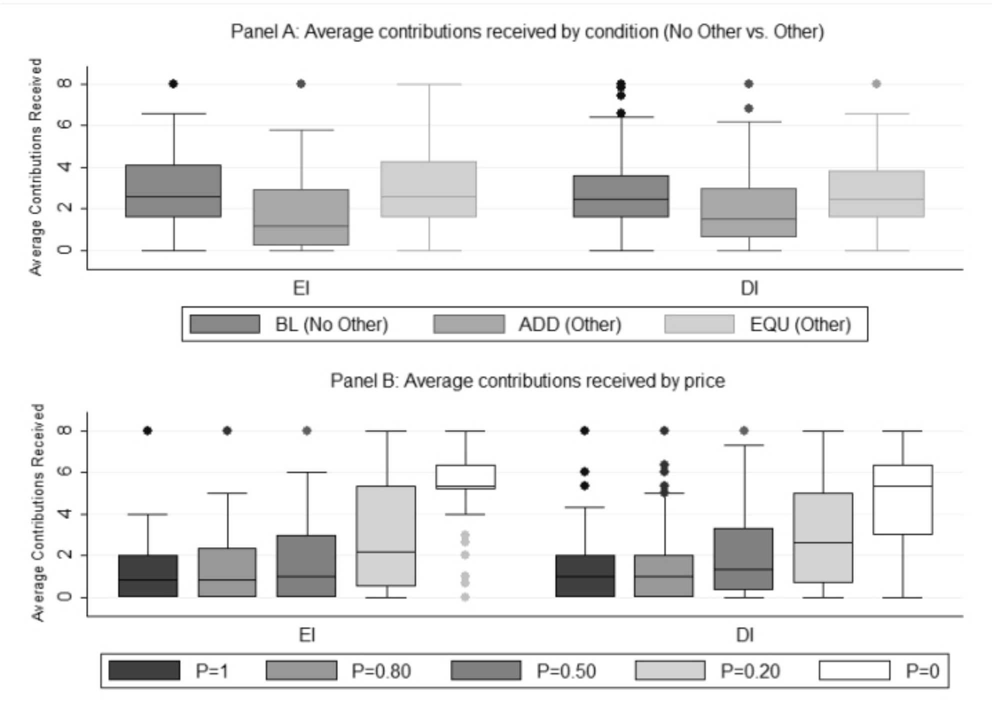
Figure 6.2 provides a first overview of $C^{T}_{P_s}$ along the three dimensions varied in the experiment. The distributional plots in the top panel illustrate that, compared with the BL (Med:2.6; Mean:2.82), the presence of an affected outsider reduces the median and mean contribution in the ADD (Med:1.4; Mean:1.92) condition, yet not in the EQU (Med:2.6; Mean:2.85) condition. The separate box plots for the EI and DI treatments furthermore indicate a remarkably small impact of the initial distribution of incomes. The bottom panel compares contributions received for the different prices of giving, separately for the EI and the DI treatments. Independent of the initial distribution of endowments, the organization received higher contributions when the price of giving was smaller.

Obviously, by pooling over data from different conditions, this first overview could mask important interactions between the treatment variables. Table 6.2, therefore, lists summary statistics of the two main outcomes ($C^{T}_{P_s}$ and $G^{T}_{P_s}$) for all 15 allocation problems. Column (1) and (2) contain the values for the endowment and price of giving in each allocation problem. Average contributions received (including the rebate) are shown in columns (3)–(5) for the three different ways in which the outsider was affected. Similarly, the average contributions given (excluding the rebate) are displayed in columns (6)–(8).

[13]In this baseline allocation problem, 59 % of subjects contributed a positive amount and the conditional mean is 30 % of the initial endowment.

[14]There should be, however, no concerns that a majority of participants does not value the public good at all, given that the organization received high contributions (Conditional Mean = 6.76 €) by nearly 90 % of the subjects in the BL ($P_s = 0$) condition.

FIGURE 6.2: Average contribution behavior



Notes: The top panel displays contributions received across the varying conditions (no-outsider vs. outsider) and different endowment treatments (EI vs. DI). Values are calculated by pooling over the different price levels. The bottom panel displays contributions received for all price levels and endowment treatments (EI vs. DI). Values are calculated by pooling over all conditions (no-outsider vs. outsider). The median contribution is plotted as a black line. The lower and upper quartiles are illustrated by the surrounding gray boxes, and whiskers are used to display values within 1.5 times of the interquartile range. The contributions outside this range are shown as a dot.

TABLE 6.2: Average contributions

		(C) Contributions Received [Mean (s.d.)]			(G) Contributions Given [Mean (s.d.)]		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Endowment	Price of Giving (P_s)	BL	ADD	EQU	BL	ADD	EQU
EI = 8	1.00	1.25 (1.64)	1.25 (1.64)	1.22 (1.69)	1.25 (1.64)	1.25 (1.64)	1.22 (1.69)
	0.80	1.39 (1.77)	1.40 (1.76)	1.44 (1.98)	1.11 (1.41)	1.12 (1.40)	1.15 (1.58)
	0.50	1.84 (2.20)	2.06 (2.43)	1.81 (2.13)	0.92 (1.10)	1.03 (1.21)	0.90 (1.06)
	0.20	3.32 (3.05)	2.15 (2.53)	3.39 (3.18)	0.66 (0.61)	0.43 (0.50)	0.67 (0.63)
	0.00	6.47 (2.78)	2.22 (2.65)	6.60 (2.77)	-	-	-
DI = 8	1.00	1.51 (2.02)	1.40 (2.03)	1.62 (2.05)	1.51 (2.02)	1.40 (2.03)	1.62 (2.05)
	0.80	1.72 (2.24)	1.56 (2.12)	1.70 (2.24)	1.38 (1.79)	1.25 (1.70)	1.36 (1.79)
	0.50	2.00 (2.30)	2.48 (2.61)	1.91 (2.32)	1.00 (1.15)	1.24 (1.30)	0.95 (1.16)
	0.20	3.13 (2.86)	2.94 (2.59)	3.35 (2.98)	0.62 (0.57)	0.58 (0.51)	0.67 (0.59)
	0.00	5.62 (3.22)	2.62 (2.69)	5.89 (3.12)	-	-	-
DI = 12	1.00	1.54 (2.02)	1.59 (2.01)	1.16 (1.40)	1.54 (2.02)	1.59 (2.01)	1.16 (1.40)
	0.80	1.54 (2.16)	1.54 (2.00)	1.18 (1.50)	1.23 (1.73)	1.23 (1.60)	0.95 (1.20)
	0.50	2.02 (2.61)	2.16 (2.38)	2.13 (2.66)	1.01 (1.30)	1.08 (1.19)	1.06 (1.33)
	0.20	3.16 (3.13)	1.78 (1.93)	3.13 (3.06)	0.63 (0.62)	0.35 (0.38)	0.62 (0.61)
	0.00	5.56 (3.34)	2.16 (2.15)	5.91 (3.18)	-	-	-

Notes: G is computed by multiplying C with the price of giving. No values of G are shown for a price of giving of $P_s = 0$, as they cannot be interpreted meaningfully.

Comparing values row-by-row in columns (3)–(5) suggests a negative relationship between $C^T_{P_s}$ and the price of giving. The strength of this relationship differs by condition. Average contributions rise by more than a factor of five when prices drop from one to zero in the BL and EQU conditions, whereas they less than double in the ADD condition. The lower contribution average in the ADD condition – as displayed in Figure 6.2 – is mainly driven by two allocation problems: When the outsider had to bear higher ($P_s = 0.20$; $P_o = 0.80$) or the full ($P_s = 0.00$; $P_o = 1.00$) provision cost, decision-makers increase their contributions only moderately, despite the low price of giving. In contrast, in the BL (3) and the EQU (5) conditions, no such reluctance to contribute is present for the same level of P_s. Rather, in all allocation problems of the EQU condition, the presence of an outsider does not change behavior relative to the BL. This remains true even when the outsider had to pay the same high contribution cost (P_s=P_o=1.0) as the decision-maker.

The fact that some subjects earn a 50% increased income in the DI condition affects behavior only marginally. This can be seen both by comparing the EI and DI treatments and also by comparing high and low earners within the DI treatment. In the ADD condition, better-off subjects were slightly more reluctant to further increase inequality (i.e., lower contributions), while worse-off subjects were slightly more inclined to reduce inequity (i.e., higher contributions). Columns (6)–(8) show that subjects actually give lower net amounts when the price of giving falls. In other words, $C^T_{P_s}$ rises not as strongly as would be expected if participants were keeping constant $G^T_{P_s}$.[15] In the following, I will explore these observations more rigorously along the lines of the hypotheses outlined in Section 6.4.

6.5.2 The role of price and endowment in the BL condition

I begin by analyzing the changes in contributions behavior in reaction to different prices of giving and endowments in the BL. For these five choices, there are clear predictions from the experimental literature on rebates. Column (3) in Table 6.2 shows a strictly increasing pattern of contributions for falling prices. Relative to the allocation problem with no rebate ($P_s = 1.00$), each reduction of prices leads to a significant increase in contributions (Pairwise Sign-Rank-Test; $p < 0.05$). Yet, contributions given (6) decline significantly with each fall in prices (Pairwise Sign-Rank-Test; $p < 0.05$). To allow for a quantitative comparison with existing findings (Eckel and Grossman, 2003, 2006; Karlan and List, 2007; Huck and Rasul, 2011), I estimate a panel regression model in

[15]Some of this decline might reflect the fact that subjects selected contributions received in steps of 1 Euro. Hence, giving cannot be perfectly balanced between the different prices. Furthermore, in order to keep all treatments comparable, the maximum contribution was held constant at 8 €. Therefore, for instance, subjects who already contribute 8 € at a price of 1.0, cannot further increase their giving at lower prices. However, as the average contribution was only 1.25 € at P_s=1.0, this concern is unlikely to be the sole driver of the positive relationship between $G^T_{P_s}$ and price.

the commonly used log linear specification:

$$log(C_{it}) = \beta_0 + \beta_1 log(P_{s_t}) + \beta_2 log(E_i) + \epsilon_{it} \tag{6.1}$$

C_{it} denotes the contribution[16] by individual $i = 1, ..., 150$ in allocation problem $t = 1, ..., 5$. P_{s_t} is the price of giving faced by the decision-maker in allocation problem t and E_i is his earned endowment. In this specification, the coefficients β_1 and β_2 can be interpreted as the elasticities of giving with respect to price and income. Table 6.3 contains estimates of these core parameters using either C_{it} or G_{it} as the dependent variables.[17]

Table 6.3: Price and income elasticities BL condition

	(1)		(2)		(3)	
	C		C if $p > 0$		G if $p > 0$	
Log(Price) (Euro)	-0.814****	(-14.43)	-0.619****	(-7.57)	0.176***	(2.68)
Log(Endowment) (Euro)	-0.216	(-0.32)	-0.108	(-0.14)	-0.037	(-0.06)
Constant	-0.385	(-0.26)	-0.550	(-0.33)	-0.692	(-0.50)
Order-Dummies	Yes		Yes		Yes	
Observations	750		600		600	
Subjects	150		150		150	
R^2	0.17		0.04		0.01	
$Prob > \chi^2$	0.000		0.000		0.099	

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

Notes: This table shows random effects OLS estimates for the elasticities with respect to price and income. t-statistics based on robust standard errors are shown in parentheses. Standard errors are clustered at the subject level. Order effects are controlled for and are insignificant in all specifications. Using a Tobit Model as an alternative estimation method, which accounts for potential censoring in the contribution data, does not change results qualitatively, but leads to higher in higher elasticities. Similarly, the inclusion of further individual controls (age, gender, points in real-effort task, stated risk aversion, number of experiments, beliefs regarding the existence and severity of climate change) does not affect results.

In line with previous laboratory and field studies, column (1) shows a highly significant and negative price elasticity of giving. Column (2) displays results for the same model, restricting observations to strictly positive prices of giving. The marginal effect of a price change is slightly lower but remains highly significant. Both estimates of the price elasticity lie within the range [-.340;-1.488] reported in the two laboratory experiments of (Eckel and Grossman, 2003, 2006) and also in the vicinity of those found in field studies using a matching protocol (Karlan and List, 2007; Huck and Rasul, 2011). As their absolute value is smaller than one, there is evidence for a partial crowding effect.

[16] As common in the literature, a small number (0.10) is added to the contributions to circumvent the problem that log(0) is not defined.

[17] As derived formally in Huck and Rasul (2011), the former specifications (C_{it}) yield the own price elasticity of giving, while the latter (G_{it}) yields the cross price elasticity of consumption with respect to the price of giving.

This can also be seen from the positive elasticity for actual amounts given in column (3). The change of signs of the two elasticities, when moving from column (2) to column (3), indicates that both warm glow and pure altruism could motivate contributions in this experiment (Huck and Rasul, 2011). Based on this evidence, I fail to reject Hypothesis 1 and state the following result:

Result 1: ***When no outsider is affected, a decrease in the price of giving leads to significantly higher contributions received and significantly lower contributions given.***

In contrast to the existing experimental evidence (Eckel and Grossman, 2003, 2006), the elasticity of giving with respect to income is insignificant in all specifications. There are two plausible explanations. First, in comparison with existing studies, there is less variation in the endowments. Second, it could play an important role that higher incomes were awarded for better performance in the real-effort task, instead of being randomly assigned to the participants as a windfall. This would be in line with findings in Tonin and Vlassopoulos (2013). Similarly, non-parametric tests (M.W. Rank Sum Test $p < 0.05$) find no evidence for differences in average contributions between high and low income subjects at each price level. In other words, in the BL, giving does not show characteristics of a normal good. Given this evidence, I reject Hypothesis 2.

Result 2: ***There is no significant increase in contributions or giving for subjects with a higher earned endowment.***

6.5.3 The role of the negative externality

6.5.3.1 Aggregate analysis

I now turn to the main question of the paper by analyzing how the presence of an affected outsider influences decision-making. First evidence on this question comes from a series of pairwise non-parametric significance tests, comparing average contributions received between the different conditions shown in columns (3)–(5) of Table 6.2. Between the BL (3) and the EQU (5) condition, I find no significant differences in contributions received or given for each price level (Pairwise Sign-Rank-Test; $p < 0.05$). This holds both for the pooled sample and for separate tests within the DI and EI samples. In the allocation problems of the EQU condition, the presence of an affected outsider has little influence on behavior, even when the outsider loses as much as 12.5% of his initial endowment for each Euro contributed. This clearly contradicts a general concern for the absolute payoffs of the outsider, as proposed in Hypothesis 3. Accordingly, average contributions are not significantly lower when contributing leads to a total loss of efficiency (due to $P_s + P_o > 1$). This outcome contradicts Hypothesis 5.

Table 6.4: Elasticities and treatment effects all conditions

	(1)		(2)		(3)	
	C		C		C if $p > 0$	
Log(Price) (Euro)	-0.637****	(-13.75)	-0.814****	(-14.45)	-0.619****	(-7.58)
Log(Endowment) (Euro)	-0.107	(-0.17)	-0.107	(-0.17)	-0.093	(-0.13)
Dummy ADD (1 = Yes)	-0.326****	(-5.42)	0.164****	(4.33)	0.042	(1.23)
Dummy EQU (1 = Yes)	-0.032	(-0.85)	-0.095	(-1.60)	-0.120*	(-1.91)
ADD*Log(Price)			0.609****	(9.55)	0.201**	(1.98)
EQU*Log(Price)			-0.078**	(-2.32)	-0.161**	(-2.19)
Constant	-0.424	(-0.31)	-0.567	(-0.41)	-0.521	(-0.34)
Order-Dummies	Yes		Yes		Yes	
Observations	2250		2250		1800	
Subjects	150		150		150	
R^2	0.11		0.14		0.04	
$Prob > \chi^2$	0.000		0.000		0.000	

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

Notes: This table shows random effects OLS estimates for the elasticities with respect to price and income. t-statistics based on robust standard errors are shown in parentheses. Standard errors are clustered at the subject level. Order effects are controlled for and are insignificant in all specifications. Using a Tobit model as an alternative estimation method that accounts for potential censoring in the contribution data does not change results qualitatively but leads to higher elasticities. Similarly, the inclusion of further individual controls (age, gender, points in real-effort task, stated risk aversion, number of experiments, beliefs regarding the existence, and severity of climate change) does not affect results.

A comparison of the BL and the ADD conditions arrives at more nuanced results. For those allocation problems in which the decision-maker had to pay for the larger share of the provision cost, the presence of the outsider does not significantly influence average contributions. This is different for the remaining allocation problems. As soon as the outsider had to carry a larger fraction ($P_s = 0.20; P_o = 0.80$) or the full contribution cost ($P_s = 0.00; P_o = 1.00$), average contributions drop significantly (Pairwise Sign-Rank-Test; $p < 0.05$). When the decision-maker and the outsider share cost equally ($P_s = 0.50; P_o = 0.50$), contributions even rise slightly. This statistically significant increase (Sign-Rank-Test; $p < 0.05$) is most pronounced for low-income subjects in the DI treatment. Taken together, these observations provide limited support for Hypothesis 4a: When contributing has neutral effects on the final distribution of payoffs or increases inequality in the disadvantageous direction (from the point of view of the decision-maker), the effects on the outsider are largely ignored. Implied increases in advantageous inequality, however, lead to a reduction of contributions. This finding is not driven by the complete absence of an outsider in the BL. When comparing contributions in the ADD with contributions in the EQU condition, I also find that decision-makers only significantly decrease their contributions when the outsider carries a larger share of the total contribution cost (Sign-Rank-Test; $p < 0.05$).

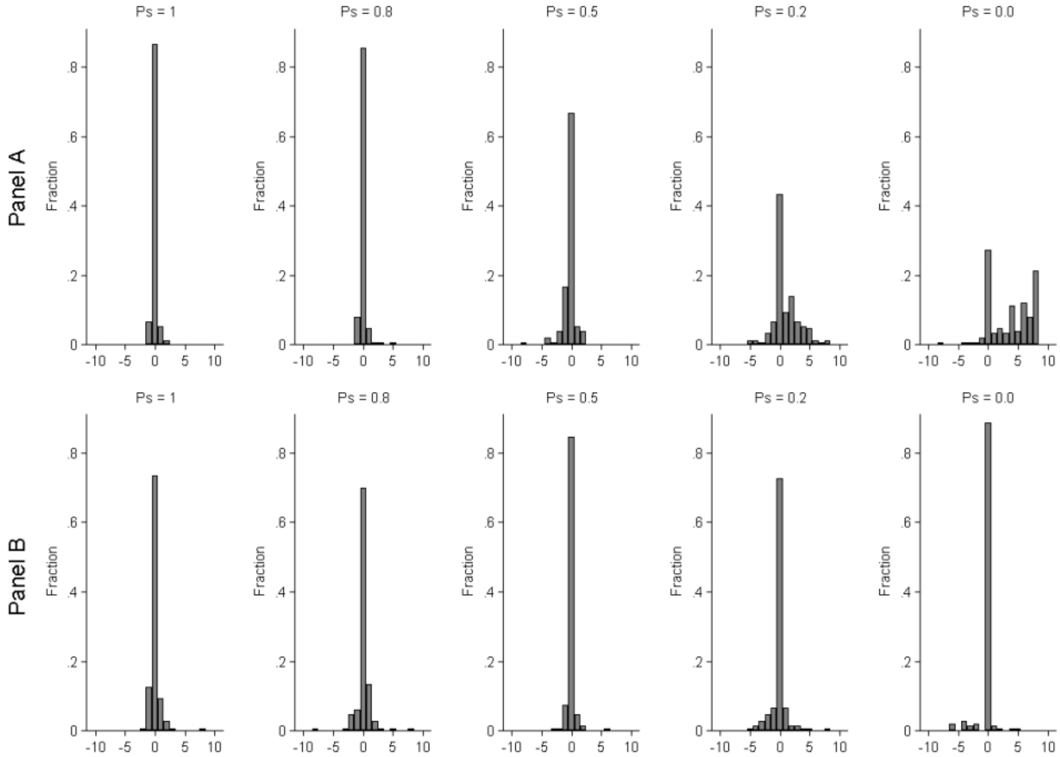
Analyzing the full sample, the regression results in Table 6.4 quantify how imposing harm on an outsider affects average provision levels under different prices of giving. In specification (1) I estimate a log linear model of contributions introduced in equation (1). Additional dummy variables indicate the presence of an affected outsider in the ADD and EQU conditions. Specifications (2) and (3) test, through including interaction terms between these dummies and the price variable, if and how strongly the elasticity changes in the presence of an outsider. Specification (1) replicates the non-parametric result that average contributions are significantly lower in the ADD condition. The contributions decline by 32.6 % relative to the baseline. There is no comparable decline of contributions in the EQU condition.

The significant interaction terms in specifications (2) and (3) demonstrate that the elasticity of giving is lower (in absolute terms) in the ADD condition than in the BL condition and that it is slightly larger in the EQU condition. Accordingly, rebates are a far less effective method for increasing the provision level, when they are fully financed by an identifiable third party. Specification (3) shows that these results are qualitatively robust to excluding the three allocation problems offering a full rebate. Quantitatively, the elasticity of giving is smaller, and its reduction in the ADD condition is less pronounced.

6.5.3.2 Individual analysis

As a final step of analysis, I will now examine changes in behavior at the individual level. This within-subjects perspective sheds light on the additional motives that could become actionable as soon as an outsider is harmed. Figure 6.3 displays how decision-makers change their behavior when they move from the baseline to one of the outsider conditions. The top panel shows the distribution of $D1 = C_{P_s}^{BL} - C_{P_s}^{ADD}$, and the bottom panel the distribution of $D2 = C_{P_s}^{BL} - C_{P_s}^{EQU}$ for the five different prices of giving. In eight out of 10 allocation problems, a large fraction of decision-makers (70–85%) does not adapt their contribution behavior when an outsider is affected by their choices. Only in two allocations problems of the ADD condition a majority of decision-makers (60% and 75%) reduce their contributions. When $P_s = 0.0$, and consequently contributions are fully financed by an outsider ($P_o = 1.0$), the fraction of decision-makers who reduce contributions from eight to zero is nearly as large as the fraction who do not behave differently. As in the aggregate data, this strong reduction indicates that many decision-makers refrain from contributing only if this increases advantageous inequity. Table 6.5 explores this observation more thoroughly via regression analysis. The dependent variable ($C_{P_{s_{it}}}^{BL} - C_{P_{s_{it}}}^{T}$ for $T \in (ADD, EQU)$) in each OLS regression takes a positive value if the decision-maker contributes more in the absence of an outsider, and a negative

FIGURE 6.3: Histograms of D1 and D2



Notes: Panel A shows the fraction of subjects who change their behavior between the BL and the ADD conditions for the five different price levels. Panel B shows this change between the BL and the EQU conditions.

value if he contributes less. The two price variables capture reactions to differences in the price of giving, allowing for asymmetric effects. Allocation problems in which $P_s = P_o$ serve as the left out baseline category. The coefficients of the two price variables in specification (1) confirm the role of inequity aversion as an important motive. When $P_s \neq P_o$, contributions are inequality-increasing. Thus, the significant and positive coefficients support the proposition that decision-makers reduce their contributions when these are inequity increasing. Fehr and Schmidt (1999) posit that decision-makers might perceive inequality differently, depending on whether it is advantageous or disadvantageous to them. In line with this proposition, the two coefficients strongly differ in magnitude – by a factor of 15. Yet, different from the assumptions in Fehr and Schmidt (1999), decision-makers are more averse to advantageous than to disadvantageous inequity in the current setting.

The two coefficients of the income variables furthermore indicate whether decision-makers change their behavior when their initial endowment is lower (higher) than that of the outsider. Decision-makers in the EI treatment serve as the baseline category in which both the outsider and the decision-maker had the same income. Against this baseline, richer decision-makers in the DI treatment are not more prone to changing their behavior if contributing affects an outsider. There is (weakly) significant evidence that

Table 6.5: Changes in contribution behavior

	(1)		(2)	
	Δ Contributions		Δ Contributions	
Negative Price Difference ($\lvert P_s - P_o \rvert$ if $P_s < P_o$)	3.349****	(13.76)	3.827****	(11.85)
Positive Price Difference ($\lvert P_s - P_o \rvert$ if $P_s > P_o$)	0.213***	(2.72)	0.209*	(1.82)
Lower Income (1=Yes)	-0.242*	(-1.73)	-0.113	(-0.89)
Higher Income (1=Yes)	-0.001	(-0.01)	0.176	(1.18)
Lower Income*Negative Price Difference			-1.210*	(-1.92)
Higher Income*Negative Price Difference			-0.726	(-1.29)
Lower Income*Positive Price Difference			0.307*	(1.87)
Higher Income*Positive Price Difference			-0.289	(-1.45)
Constant	-0.230**	(-2.11)	-0.306***	(-2.72)
Order-Dummies	Yes		Yes	
Observations	1500		1500	
Subjects	150		150	
R^2	0.33		0.32	
$Prob > \chi^2$	0.00		0.00	

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

Notes: This table contains OLS estimates, taking the difference between BL and ADD/EQU as the dependent variable. t-statistics based on robust standard errors are shown in parentheses. Standard errors are clustered at the subject level. Order effects are controlled for and are insignificant in all specifications. Using a Tobit Model as an alternative estimation that accounts for potential censoring in the contribution data does not change results. Similarly, including further individual controls (age, gender, points in real-effort task, stated risk aversion, number of experiments, beliefs regarding the existence and severity of climate change) does not affect results.

poorer decision-makers contribute slightly higher amounts in the outsider conditions. This could point at both spite and inequity aversion as potential motives. Specification (2) shows tentative evidence that the latter might be the more plausible explanation. The significant interaction term between the price and low-income variable indicates that low-income subjects reduce some of the initial inequality by selectively contributing more when the comparatively richer outsider has to pay for the largest share of this contribution. The main effect of the low-income variable, which would also point at pure spite, is now insignificant.

As discussed in Section 6.4, the presence of a harmed outsider, if at all, should only affect behavior among those subjects holding social preferences, whereas selfish types should be unaffected by its presence. Up to this point the analysis of the underlying motives has not taken account of this distinction. By fully exploiting the within-subjects design, I now turn to the question whether subjects displaying altruistic preferences in the BL are more strongly affected by the presence of a harmed outsider in the ADD and EQU conditions. This complementary approach rests on the observation that in simple distribution task, comparable with the BL in the present experiment, many subjects follow a consistent contribution rule (Andreoni and Miller, 2002; Fisman et al., 2007; Ubeda, 2014). Thus, decision-makers are categorized into different "types" according to their behavior in the BL. In a subsequent step, I then move on to ask if certain types are more likely to deviate from this consistent contribution rule as soon as they move to the ADD or EQU conditions. To study such heterogeneous responses, I discern between four BL types based on the theoretical considerations laid out in Section 6.4: The selfish type will always contribute zero at positive prices, and zero or a positive amount when the price drops to zero. The second type will conform to preferences for donation targeting by contributing a constant and positive amount in all five decisions. The third type shows preferences in accordance with impure altruism by contributing a positive amount in at least one decision and reacting to rebates by increasing contributions in a (weakly) monotonic way. The fourth type does not follow any of the three previous rules. Most of the decision-makers classified accordingly contribute positive amounts in most decisions but react to price changes non-monotonically. This can be seen as a violation of consistency in the sense that they make at least one pareto-dominated choice in five decisions. Based on this classification, 78 % of the decision-makers choose perfectly consistent in the BL. This is slightly below the rates of consistency reported in Andreoni and Miller (2002) and Fisman et al. (2007).[18] Table 6.6 compares types between the BL and the ADD condition. Column 1 displays the distribution of types in the BL: 26 % of subjects have purely selfish preferences, 4 % follow strict donation targeting mainly by contributing the maximum amount of 8 Euros in all five decisions, 48 % can be described as impure altruists and 22 % violate consistency.

Table 6.6: Conditional distribution of types for the ADD conditions

	Type ADD			
Type Baseline [Perc. of Total]	Type 1 (Selfish)	Type 2 (Constant)	Type 3 (Impure Altruist)	Type 4 (Inconsistent)
Type 1 (Selfish) [26%]	79.49%	0.00%	20.51%	0.00%
Type 2 (Constant) [4%]	0.00%	66.67%	0.00%	33.33%
Type 3 (Impure Altruist) [48%]	11.11%	2.78%	22.22%	63.89%
Type 4 (Inconsistent) [22%]	6.06%	9.09%	18.18%	66.67%

Notes: This table displays a cross tabulation of the different types found in the ADD condition for a given type assigned based on BL behavior.

The remaining columns display, for each of the four types, the fraction of decision-makers who stay within the same category when moving to the ADD condition and the fraction who switch to one of the other categories. Comparing values along the diagonal axis shows that for three out of four types at least two-thirds of decision-makers remain in the same category. As expected theoretically, the selfish type is least affected by the

[18] In their experiments an even larger fraction of subjects (95%) choose consistently. These findings are, however, based on a larger number of choices and a more lenient criterion that allows for a number of violations.

Table 6.7: Conditional distribution of types for the EQU conditions

	Type ADD			
	Type 1	Type 2	Type 3	Type 4
Type Baseline [Perc. of Total]	(Selfish)	(Constant)	(Impure Altruist)	(Inconsistent)
Type 1 (Selfish) [26%]	84.62%	0.00%	10.26%	5.13%
Type 2 (Constant) [4%]	0.00%	66.67%	33.33%	0.00%
Type 3 (Impure Altruist) [48%]	2.78%	0.00%	76.39%	20.83%
Type 4 (Inconsistent) [22%]	0.00%	0.00%	21.21%	78.79%

Notes: This table displays a cross tabulation of the different types found in the EQU condition for a given type assigned based on BL behavior.

presence of an outsider, and 79 % of decision-makers remain in the same category.[19] Similarly, 67 % of type 2 and type 4 subject remain within the same category. In line with theoretical consideration, the only type which displays a strong tendency to change its contribution rule in the ADD condition is constituted by type 3 decision-makers. A majority (77%) of subjects displaying preferences consistent with impure altruism in the BL follow a different rule when deciding in the ADD condition. As many as 11 % react strongly to the presence of a harmed outsider and switch to full free-riding. A larger fraction (64%) is now classified as inconsistent. Reflecting concerns for pay-off inequality, most of these subjects reduce contributions as P_s falls and consequently P_o rises, resulting in an inverse U-shaped relationship between contributions and price.

Table 6.7 contains the same analysis for the EQU condition. Here, in stark contrast to the previous analysis, most decision-makers remain in the same category when they move from the BL to the EQU condition. This holds also for most subjects who display preferences for impure altruism in the BL. The different results for the ADD and EQU conditions underline that (altruistic) subjects react toward the relative rather than the absolute level of harm imposed on an outsider.

Based on the combined evidence from analyzing behavior and the motivational fine structure at the aggregate and individual level, I formulate the following results. A majority of subjects does not change their behavior in the EQU condition. Therefore, I reject Hypothesis 3 and 5 and state:

[19]Unaccounted for by theoretical considerations, the remaining 21 % begin to contribute positive amounts as soon as they are assured that an outsider had to contribute as well. For those subjects free-riding could be an expression of a hurt sense of fairness rather than being motivated by pure payoff maximization

Result 3: ***A large fraction of subjects does not, in general, reduce contributions, despite imposing harm on a third party. This also holds when contributing decreases total efficiency in the EQU condition.***

In the ADD condition, a majority of subjects changes behavior when contributing is inequality increasing. As a consequence, average contributions are significantly lower in the ADD condition. Overall, I fail to reject Hypothesis 4a and 4b and state:

Result 4: ***A majority of subjects reduces contributions in the presence of an affected outsider, when doing so is inequality decreasing.***

6.6 Conclusion

In the past decades, experiments have amassed evidence on the ubiquity of social preferences: Subjects share money with strangers, give to charities, or cooperate in public good games, even at non-trivial personal cost. In contrast to these typically linear decision tasks, many real-world situations are morally more ambiguous because they allow individuals to take from one in order to give to another. When decision-makers can "do good with other people's money"(Friedman, 1962), they are given the opportunity to take a moral free-ride – as long as they are willing to ignore the negative welfare effects they impose on the other person. Based on a set of stylized allocation tasks, I have shown that in many cases decision-makers are indeed ready to ignore the welfare implications that they impose on an outsider while doing good. My data show only one exception to this rule. Contribution levels are significantly lower when the relative level of harm is high and contributions are therefore inequality increasing.

These finding have several implications. First, for standard models of altruism (e.g. Andreoni, 1989, 1990) and the associated experiments (e.g. Karlan and List, 2007; Huck and Rasul, 2011) they pose the, as yet, understudied question of how to incorporate the trade-off between decision-makers and affected outsiders. For instance, can decision-makers still be said to experience a warm glow of giving, although they only pass on money belonging to someone else who is potentially even hurt by their behavior? Second, my findings can be informative in terms of two related policy contexts. Many governments use tax rebates or similar instruments to subsidize charitable donations. While concerns regarding potential crowding out have been well-studied (e.g. Kingma, 1989; Manzoor and Straub, 2005), my results point toward distributional concerns as an additional channel through which rebates might fail to increase contributions. Of course, before readily extrapolating from simplified laboratory conditions to behavior beyond the laboratory, some important qualifications should be taken into account. It is, for instance, unclear whether potential contributors perceive money spent by their government as exogenous (comparable to the BL condition of the experiment) or as being taken from a fellow citizen (comparable to the ADD and EQU conditions). Moreover, the experimental design makes the harm imposed on the outsider very salient. For many contribution decisions beyond the laboratory, the positive effect on the beneficiaries is likely to be far more salient than the negative effect on a dispersed group of harmed outsiders as evidence on the "identifiable victims effect" (Small and Loewenstein, 2003; Small et al., 2007) suggests. This could be an interesting avenue for further research. Finally, as mentioned in Section 6.3, voluntary climate actions – the naturally occurring public good used in this experiment – have a pronounced intergenerational dimension. Policies to subsidize public good provision in this area are therefore often motivated by concerns regarding the welfare of future generations. My results highlight conditions under which such concerns can collide with intragenerational distributional concerns. Climate change policies that ignore this trade-off (Grösche and Schröder, 2014) might undermine their own effectiveness (Baumol and Oates, 1988, p.235). Again, this will depend on the saliency of the trade-off between intragenerational and intergenerational inequality.

6.7 Appendix

Instructions

Experiment	Laboratory: Random seat assignment Personal code for anonymity Tasks implemented in z-Tree • Real-effort task *(page 162)* • Contribution task *(page 165)* Payment according to personal code

General Instructions

Original instructions in German are available from the author on request.

[SCREEN 1]
Dear participant,

Thank you for supporting our research by your participation. On this screen you will find some general information regarding this study:

- In the course of this study, you will take part in different computerized tasks and questionnaires.
- Please follow the instructions on each screen.
- At the end of today's session you will receive a payment.
- The funds for your compensation have been provided by the Ministry of Education and Research.
- As a compensation for your participation you will receive 2 €.
- You can earn **further payments** in the different task of this experiment.
- These payments depend on your own choices and in some cases on the choices of other participants in this room.
- Each task will be explained to you. Please, read the instructions **thoroughly**.

Questionnaire

[SCREEN 2]
Dear participant,

before proceeding with the study, please provide some information regarding your person:

- How old are you? << Input: Age >>
- Please select your gender. << Radio Buttons: Male/Female >>
- Are you a student? << Radio Buttons: Phd/Master/Bachelor/No/Other >>
- Imagine how much money you can spend each month. What statement is correct, compared to your fellow students. I can spend ...
  << Radio Line: (6) Much More ... (0) Much Less >>

Introduction real-effort task

[SCREEN 3]
Task 1 begins now.

You will receive instructions on the next screen.

Instructions real-effort task

[SCREEN 4]

General Information:

- In this task you will be matched with one other participant here in this room. Thus, there will be two participants in each group.
- This participant will be selected randomly and anonymously. Therefore, you will not learn at any point of the experiment with whom you are interacting in this task.

Your Task:

- During task 1 you will see a succession of images on the **left part** of your screen.
- It will be your job to count, how often you see a certain object on each image.
- **Above each image**, there will be a description of the object you are supposed to count.
- As soon as you press "proceed" you will proceed to an example screen, which further illustrates the set-up.
- As soon as you finish one image by entering your count, a new image will appear automatically.
- In total, you will have **180 seconds** to finish as many screens as possible.
- You will receive one point per correct count.

Payment *(EI Condition)*:

- You and the other participant in your group will receive 8 € independent of who reaches a higher number of points.
- These earnings can be used in the subsequent decision task.

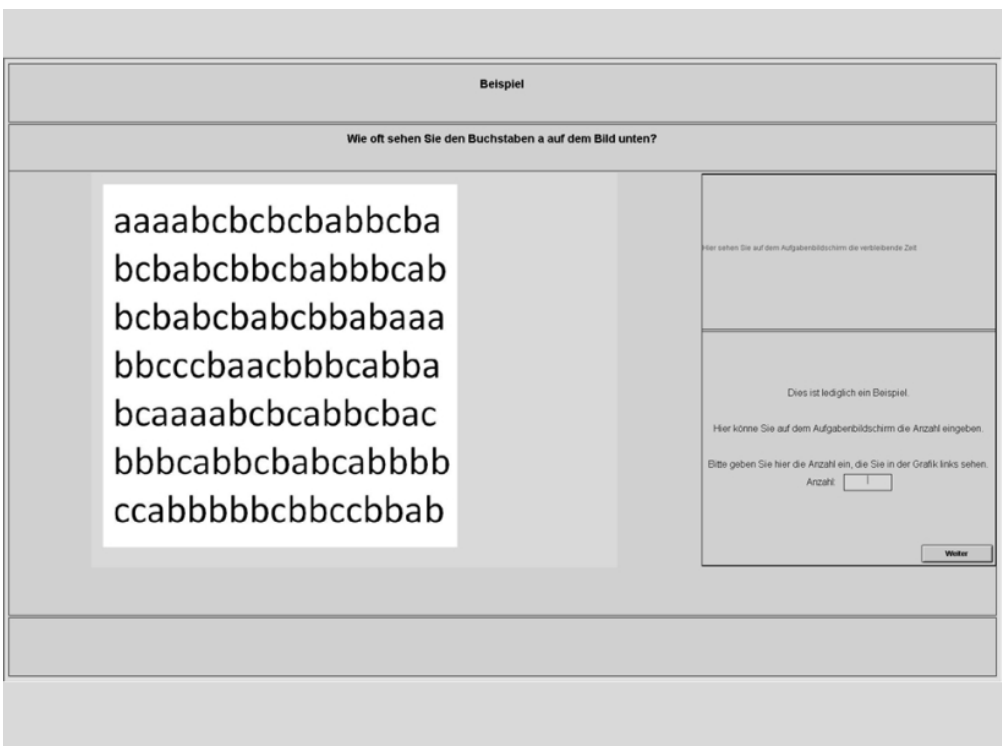
Payment *(DI Condition)*:

- Your payment in this task will depend on how many points you and the other participant reach.
- The person who reaches more points, will receive 12 €, the person who reaches less points will receive 8 €
- These earnings can be used in the subsequent decision task.

Example

[SCREEN 5]

FIGURE 6.4: Example screen real-effort task



Notes: German original. Translation: [Above] How often do you see the letter a on the image below. [Right Box] This is only an example. On the actual task screen you can enter your count here.

Start real-effort task

[SCREEN 6]

Task 1 will begin in *Counter:<< -10* >> seconds.

Real-effort task

[SCREEN 7]

Same set-up as example for 180 seconds (compare figure 6.4). For variability, on other images participants were supposed to count the appearance of the number 1 in a sequence of 0 and the appearance of a triangle of a certain color.

End real-effort task

[SCREEN 8]

Points:

- You have reached << x >> points.
- The other participant has reached << y >> points.

Points:
In EI Euroself = 8
In DI Euroself = 8 if $x<y$
In Di Euroself = 12 if $x>y$

- You will thus receive << Euroself >>.
- The other participant will thus receive << Euroother >>.
- You can use this money in the next task or can receive a cash payment at the end of the session.
- You will learn more about the next task on the following screens.

Instructions contribution task

[SCREEN 9]

General Information:

- In this task you will be matched with another participant in the room. Thus, there are two participants in each group.
- This participant will be selected randomly and anonymously. Therefore, you will not learn at any point of the experiment with whom you interact in this task.
- This participant will be a **different** participant than the one in task 1.

Roles:

- There will be **two different roles** in the second task. Which role you will assume, will be determined by a random draw at the end of the experiment. Independent of the actual role assignment you will answer all questions for the case that you will be assigned to role A.
- If you will be assigned to role A at the end of the experiment, the other participant in your group takes role B and vice versa. The probability that you take a certain role will be 50% (i.e., comparable to a coin toss).

Roles and Decisions:

- In role B you can make no active decision in task 2.
- In role A you can decide, if a certain fraction of the money which was earned in task 1 will be used to reduce the global CO_2 emissions.
- To do so you will see a list of different decision scenarios.
- In some of these scenarios, you can only decide about your own earning from task 1. In other scenarios you can also decide about the amount the other participant in your group (role B) earned in task 1.
- The decision scenarios will also differ by how costly it is to spend 1 € on reducing global CO_2 emissions.
- For each decision scenario you will see your own cost, and to what degree the other participant is affected by your choices (i.e., the cost of the other participant).
- As soon as you press “proceed” we will explain to you how we will reduce global carbon emissions.

Information screen real public good

[SCREEN 10]

Why reduce CO_2?
The more CO_2 is released to earth's atmosphere, the more likely is the occurrence of the environmental problem of 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 around the globe.
How is it possible to reduce CO_2 emissions?
The framework of the Kyoto Protocol has created a transparent possibility to support certified projects that aim at reducing CO_2 emissions (CDM). For instance, such projects are engaged in the global development and deployment of renewable energies in order to reduce CO_2 emissions. As the climate system reacts slowly to a change in CO_2 emissions, the effects of reducing CO_2 on climate change are not immediate. Contributions to CDM projects are furthermore traceable via the Federal Environmental Agency. For instance, when you are offered to compensate CO_2 during booking a flight or bus-ticket, your compensation payments often go to CDM projects.
What is your decision?
As soon as you press "proceed" you can decide (in different decision scenarios) whether you want to use some of the money that was earned in task 1 to support projects against climate change. Each Euro contributed, compensates emissions of approximately 70 kg CO_2. 70 kg roughly correspond to CO_2 emissions that would arise if driving from Frankfurt am Main to Hamburg by car. The average German citizen emits 9 tons of CO_2 each year.
How can you verify whether contributions were used to retire CO_2 permits?
At the end of this study a receipt over the total amount contributed to CDM projects will be posted at the notice board of the Chair of Behavioral Economics (Prof. Dr. C. Schwieren).

Instructions contribution task

[SCREEN 11]

General Information:

- In the following, you will proceed through five screens displaying three decision scenarios each (a total of 15 scenarios).
- In each scenario you can choose how much money you wish to pass on to climate protection projects. In each scenario you are free to choose any amount between 0 € and 8 €
- You will select that amount for each scenario.
- If your are randomly assigned to **role A** at the end of the experiment, a **randomly** selected scenario (with equal probability) will be executed according to your choice. I.e. your payment, the payment of the other participant (role B) and the amount of money that goes to climate protection projects can depend on your choices.

Information screen before contribution task

[SCREEN 12]

General Information:

- The decision task will begin on the next screen.
- You have been randomly matched with an anonymous participant.
- In task 1 the other participant has earned << Euroself >>.
- In task 1 you have earned << Eurother >>.

Decision scenario (example PS = 20)

[SCREEN 12]

All three decision scenarios next to each other on one screen. Equivalent description for other prices of giving.

EQU-Condition PS=0.20

- Scenario A
- Please enter a number into the blue field to indicate how many Euros are supposed to be passed on to climate protection projects.
- You are free to enter any number between 0 and 8.
- Each Euro costs **you** 0.20 Euro. I.e., for each Euro that goes to the climate protection project 0.20 Euro will be subtracted from your **own final earnings**.
- Each Euro costs **the other participant** 0.20 Euro. I.e., for each Euro that goes to the climate protection project 0.20 Euro will be subtracted from the **final earnings of the other participant**.

Sceanrio A	Laboratory: << Blue Field >> Please indicate how many Euros (0–8) are to be passed on to climate protection projects. Each Euro passed on costs you: **0.20** Each Euro passed on costs the other participant: **0.20**

ADD-Condition PS=0.20

- Scenario B
- Please enter a number into the blue field to indicate how many Euros are supposed to be passed on to climate protection projects.
- You are free to enter any number between 0 and 8.
- Each Euro costs **you** 0.20 Euro. I.e., for each Euro that goes to the climate protection project 0.20 Euro will be subtracted from your **own final earnings**.
- Each Euro costs **the other participant** 0.80 Euro. I.e., for each Euro that goes to the climate protection project 0.80 Euro will be subtracted from the **final earnings of the other participant** .

Sceanrio B	Laboratory: << Blue Field >> Please indicate how many Euros (0–8) are to be passed on to climate protection projects. Each Euro passed on costs you: **0.20** Each Euro passed on costs the other participant: **0.80**

BL-Condition PS=0.20

- Scenario C
- Please enter a number into the blue field to indicate how many Euros are supposed to be passed on to climate protection projects.
- You are free to enter any number between 0 and 8.
- Each Euro costs **you** 0.20 Euro. I.e., for each Euro that goes to the climate protection project 0.20 Euro will be subtracted from your **own final earnings**.
- The other participant is **not affected** by your decision.

Sceanrio C	Laboratory: << Blue Field >> Please indicate how many Euros (0–8) are to be passed on to climate protection projects. Each Euro passed on costs you: **0.20** The other participant is **not affected** by your decision.

References

Abdellaoui, M., Baillon, A., Placido, L., and Wakker, P. P. (2011). The rich domain of uncertainty: Source functions and their experimental implementation. *The American Economic Review*, 101(2):695–723.

Agranov, M., Caplin, A., and Tergiman, C. (2015). Naive play and the process of choice in guessing games. *Journal of the Economic Science Association*, 1:1–12.

Ahn, T.-K., Ostrom, E., and Walker, J. (2003). Incorporating motivational heterogeneity into game theoretic models of collective action. *Public Choice*, 117(3).

Al-Ubaydli, O., Jones, G., and Weel, J. (2014). Average player traits as predictors of cooperation in a repeated prisoner's dilemma. *MPRA Working Paper No. 55383*.

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

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

Allcott, H. and Mullainathan, S. (2010). Behavioral science and energy policy. *Science*, 327(5970):1204–1205.

Alós-Ferrer, C. and Strack, F. (2014). From dual processes to multiple selves: Implications for economic behavior. *Journal of Economic Psychology*, 41:1–11.

Alpizar, F., Carlsson, F., and Johansson-Stenman, O. (2008). Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica. *Journal of Public Economics*, 92(5–6):1047–1060.

Amir, O. and Rand, D. G. (2012). Economic games on the internet: The effect of $1 stakes. *PloS one*, 7(2).

Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M., and List, J. A. (2011). Stakes matter in ultimatum games. *The American Economic Review*, 101(7):3427–3439.

Anderson, J., Burks, S. V., Carpenter, J., Götte, L., Maurer, K., Nosenzo, D., Potter, R., Rocha, K., and Rustichini, A. (2013). 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.

Andreoni, J. (1988). Why free ride? Strategies and learning in public goods experiments. *Journal of Public Economics*, 37(3):291–304.

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

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

Andreoni, J. (1995). Cooperation in public-goods experiments: kindness or confusion? *The American Economic Review*, pages 891–904.

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

Andreoni, J., Rao, J. M., and Trachtman, H. (2011). Avoiding the ask: A field experiment on altruism, empathy, and charitable giving. *NBER Working Paper No. 17648*.

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

Ashley, R., Ball, S., and Eckel, C. (2010). Motives for Giving: A Reanalysis of Two Classic Public Goods Experiments. *Southern Economic Journal*, 77(1):15–26.

Ashraf, N., Camerer, C. F., and Loewenstein, G. (2005). Adam Smith, behavioral economist. *The Journal of Economic Perspectives*, 19(3):131–145.

Axelrod, R. (1984). *The evolution of cooperation*. Basic Books, Inc., New York.

Axelrod, R. and Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211(4489):1390–1396.

Baltussen, G., Post, G. T., Van Den Assem, M. J., and Wakker, P. P. (2012). Random incentive systems in a dynamic choice experiment. *Experimental Economics*, 15(3):418–443.

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

Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford Economic Papers*, 48:878–894.

Bartling, B., Weber, R. A., and Yao, L. (2014). Do markets erode social responsibility? *The Quarterly Journal of Economics*, 130(1):219–266.

Baumol, W. J. and Oates, W. E. (1988). *The theory of environmental policy*. Cambridge university press.

Bayer, R.-C., Renner, E., and Sausgruber, R. (2013). Confusion and learning in the voluntary contributions game. *Experimental Economics*, 16(4):1–19.

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

Benito-Ostolaza, J. M., Hernández, P., and Sanchis-Llopis, J. A. (2016). Do individuals with higher cognitive ability play more strategically? *Journal of Behavioral and Experimental Economics*, forthcoming.

Benjamin, D. J., Brown, S. A., and Shapiro, J. M. (2013). Who is 'behavioral'? Cognitive ability and anomalous preferences. *Journal of the European Economic Association*, 11(6):1231–1255.

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

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

Bernhard, H., Fischbacher, U., and Fehr, E. (2006). Parochial altruism in humans. *Nature*, 442(7105):912–915.

Beshears, J., Choi, J. J., Laibson, D., and Madrian, B. C. (2008). How are preferences revealed? *Journal of Public Economics*, 92(8–9):1787–1794.

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

Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review*, 8(2):225–246.

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

Bolton, G. E. and Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *The American Economic Review*, 90(1):166–193.

Börger, T. (2015). Are fast responses more random? Testing the effect of response time on scale in an online choice experiment. *Environmental and Resource Economics*, pages 1–25.

Bosch-Domènech, A., Brañas–Garza, P., and Espín, A. M. (2014). Can exposure to prenatal sex hormones (2d:4d) predict cognitive reflection? *Psychoneuroendocrinology*, 43(0):1–10.

Bowles, S. and Gintis, H. (2011). *A cooperative species: Human reciprocity and its evolution*. Princeton University Press, Princeton.

Brañas–Garza, P., García-Muñoz, T., and González, R. H. (2012). Cognitive effort in the beauty contest game. *Journal of Economic Behavior & Organization*, 83(2):254–260.

Brañas–Garza, P., Kujal, P., and Lenkei, B. (2015). Cognitive reflection test: Whom, how, and when. *mimeo*.

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

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

Brock, J. M., Lange, A., and Ozbay, E. Y. (2013). Dictating the risk: Experimental evidence on giving in risky environments. *The American Economic Review*, 103(1):415–437.

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

Buchan, N. R., Grimalda, G., Wilson, R., Brewer, M., Fatas, E., and Foddy, M. (2009). Globalization and human cooperation. *Proceedings of the National Academy of Sciences*, 106(11):4138–4142.

Buhrmester, M., Kwang, T., and Gosling, S. D. (2011). Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1):3–5.

Burks, S. V., Carpenter, J. P., Goette, L., and Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences*, 106(19):7745–7750.

Burnham, T. C., Cesarini, D., Johannesson, M., Lichtenstein, P., and Wallace, B. (2009). Higher cognitive ability is associated with lower entries in a p-beauty contest. *Journal of Economic Behavior & Organization*, 72(1):171–175.

Burton-Chellew, M. N. and West, S. A. (2013). Prosocial preferences do not explain human cooperation in public-goods games. *Proceedings of the National Academy of Sciences*, 110(1):216–221.

Buser, T. (2012). Digit ratios, the menstrual cycle and social preferences. *Games and Economic Behavior*, 76(2):457–470.

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

Cappelen, A. W., Nielsen, U. H., Tungodden, B., Tyran, J.-R., and Wengström, E. (2014). Fairness is intuitive. *NHH Dept. of Economics Discussion Paper 14-10*.

Cappelletti, D., Güth, W., and Ploner, M. (2011). Being of two minds: Ultimatum offers under cognitive constraints. *Journal of Economic Psychology*, 32(6):940–950.

Capraro, V. and Cococcioni, G. (2015). Social setting, intuition and experience in laboratory experiments interact to shape cooperative decision-making. *Proceedings of the Royal Society of London B: Biological Sciences*, 282(1811).

Capraro, V., Jordan, J. J., and Rand, D. G. (2014). Heuristics guide the implementation of social preferences in one-shot prisoner's dilemma experiments. *Nature Scientific Reports*, 4.

Carlsson, F. (2010). Design of stated preference surveys: Is there more to learn from behavioral economics? *Environmental and Resource Economics*, 46(2):167–177.

Carlsson, F., He, H., and Martinsson, P. (2013). Easy come, easy go: The role of windfall money in lab and field experiments. *Experimental Economics*, 16(2):190–207.

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

Carlsson, F., Johansson-Stenman, O., and Nam, P. K. (2014). Social preferences are stable over long periods of time. *Journal of Public Economics*, 117:104–114.

Carlsson, F., Kataria, M., Lampi, E., and Levati, M. V. (2011). Doing good with other people's money: A charitable giving experiment with students in environmental sciences and economics. Working Paper GSU Working Papers in Economics 487.

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

Carpenter, J., Graham, M., and Wolf, J. (2013). Cognitive ability and strategic sophistication. *Games and Economic Behavior*, 80:115–130.

Cesarini, D., Dawes, C. T., Fowler, J. H., Johannesson, M., Lichtenstein, P., and Wallace, B. (2008). Heritability of cooperative behavior in the trust game. *Proceedings of the National Academy of Sciences*, 105(10):3721–3726.

Chandler, J., Mueller, P., and Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, 46(1):112–130.

Charness, G., Gneezy, U., and Kuhn, M. A. (2013). Experimental methods: Extra-laboratory experiments-extending the reach of experimental economics. *Journal of Economic Behavior & Organization*, 91(0):93–100.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, 14(1):47–83.

Chavanne, D., McCabe, K., and Paganelli, M. P. (2011). Whose money is it anyway? Ingroups and distributive behavior. *Journal of Economic Behavior & Organization*, 77(1):31–39.

Chen, C.-C., Chiu, I., Smith, J., and Yamada, T. (2013). Too smart to be selfish? Measures of cognitive ability, social preferences, and consistency. *Journal of Economic Behavior & Organization*, 90:112–122.

Cherry, T. L., Frykblom, P., and Jason F. Shogren (2002). Hardnose the dictator. *The American Economic Review*, 92(4):1218–1221.

Cherry, T. L., Kroll, S., and Shogren, J. F. (2005). The impact of endowment heterogeneity and origin on public good contributions: Evidence from the lab. *Journal of Economic Behavior & Organization*, 57(3):357–365.

Cheung, S. N. (1973). Fable of the bees: An economic investigation. *Journal of Law and Economics*, 16(1):11–33.

Choi, J.-K. and Bowles, S. (2007). The coevolution of parochial altruism and war. *Science*, 318(5850):636–640.

Chou, E., McConnell, M., Nagel, R., and Plott, C. R. (2009). The control of game form recognition in experiments: Understanding dominant strategy failures in a simple two person "guessing" game. *Experimental Economics*, 12(2):159–179.

Christelis, D., Jappelli, T., and Padula, M. (2010). Cognitive abilities and portfolio choice. *European Economic Review*, 54(1):18–38.

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

Cohen, M. A., Fullerton, D., and Topel, R. H. (2013). *Distributional Aspects of Energy and Climate Policies*. Edward Elgar Publishing.

Cone, J. and Rand, D. G. (2014). Time pressure increases cooperation in competitively framed social dilemmas: a successful replication. *PloS one*, 9(12).

Cook, J., Jeuland, M., Maskery, B., and Whittington, D. (2012). Giving stated preference respondents "time to think": Results from four countries. *Environmental and Resource Economics*, 51(4):473–496.

Corgnet, B., Espín, A. M., and Hernán-González, R. (2015). The cognitive basis of social behavior: Cognitive reflection overrides antisocial but not always prosocial motives. Working Paper.

Costa-Gomes, M., Crawford, V. P., and Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, 69(5):1193–1235.

Croson, R., Fatas, E., and Neugebauer, T. (2005). Reciprocity, matching and conditional cooperation in two public goods games. *Economics Letters*, 87(1):95–101.

Croson, R. and Gächter, S. (2010). The science of experimental economics. *Journal of Economic Behavior & Organization*, 73(1):122–131.

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

Croson, R. and Konow, J. (2009). Social preferences and moral biases. *Journal of Economic Behavior & Organization*, 69(3):201–212.

Croson, R. and Treich, N. (2014). Behavioral environmental economics: Promises and challenges. *Environmental and Resource Economics*, 58(3):335–351.

Croson, R. T. (2007). Theories of commitment, altruism and reciprocity: Evidence from linear public goods games. *Economic Inquiry*, 45(2):199–216.

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

Cunha, F. and Heckman, J. J. (2009). The economics and psychology of inequality and human development. *Journal of the European Economic Association*, 7(2):320–364.

Dana, J., Weber, R. A., and Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1):67–80.

Davis, D. D., Millner, E. L., and Reilly, R. J. (2005). Subsidy schemes and charitable contributions: A closer look. *Experimental Economics*, 8(2):85–106.

de Oliveira, A. C., Croson, R. T., and Eckel, C. (2011). The giving type: Identifying donors. *Journal of Public Economics*, 95(5–6):428–435.

de Wit, H., Flory, J. D., Acheson, A., McCloskey, M., and Manuck, S. B. (2007). IQ and nonplanning impulsivity are independently associated with delay discounting in middle-aged adults. *Personality and Individual Differences*, 42(1):111–121.

Deck, C. and Jahedi, S. (2015). The effect of cognitive load on economic decision making: A survey and new experiments. *forthcoming: European Economic Review.*

DellaVigna, S., List, J. A., and Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *The Quarterly Journal of Economics*, 127(1):1–56.

Diederich, J. and Goeschl, T. (2013). To give or not to give: The price of contributing and the provision of public goods. *NBER Working Paper Series 19332.*

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

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100(3):1238–60.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Downs, A. (1957). An economic theory of political action in a democracy. *The Journal of Political Economy*, 65(2):135–150.

Dreber, A., Fudenberg, D., Levine, D. K., and Rand, D. G. (2014). Altruism and Self-Control. *Working Paper: SSRN 2477454.*

Duffy, S. and Smith, J. (2014). Cognitive load in the multi-player prisoner's dilemma game: Are there brains in games? *Journal of Behavioral and Experimental Economics*, 51:47–56.

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

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

Eckel, C. C. and Grossman, P. J. (2006). Subsidizing charitable giving with rebates or matching: Further laboratory evidence. *Southern Economic Journal*, 72(4):794–807.

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

Ein-Gar, D. and Levontin, L. (2013). Giving from a distance: Putting the charitable organization at the center of the donation appeal. *Journal of Consumer Psychology*, 23(2):197–211.

Engel, C. and Rockenbach, B. (2011). We are not alone: The impact of externalities on public good provision. *MPI Collective Goods Pre Print*, (29).

Engel, C. and Zhurakhovska, L. (2014). Conditional cooperation with negative externalities – An experiment. *Journal of Economic Behavior & Organization*, 108(0):252–260.

Engelmann, D. and Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *The American Economic Review*, 94(4):857–869.

Erkal, N., Gangadharan, L., and Nikiforakis, N. (2011). Relative earnings and giving in a real-effort experiment. *The American Economic Review*, 101(7):3330–3348.

Evans, A. M., Dillon, K. D., and Rand, D. G. (2014). Reaction times and reflection in social dilemmas: Extreme responses are fast, but not intuitive. *SSRN Working Paper 2436750.*

Evans, A. M., Dillon, K. D., and Rand, D. G. (2015). Reaction times and reflection in social dilemmas: Extreme responses are fast, but not intuitive. *Discussion Paper(Available at SSRN: 2436750).*

Evans, J. S. B. T. (2003). In two minds: dual-process accounts of reasoning. *Trends in cognitive sciences*, 7(10):454–459.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.*, 59:255–278.

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

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

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

Fehr, E. and Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425(6960):785–791.

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

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

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3):817–868.

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

Fiedler, S., Glöckner, A., Nicklisch, A., and Dickert, S. (2013). Social value orientation and information search in social dilemmas: An eye-tracking analysis. *Organizational Behavior and Human Decision Processes*, 120(2):272–284.

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

Fischbacher, U. and Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, 100(1):541–556.

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

Fisher, J., Isaac, R. M., Schatzberg, J. W., and Walker, J. M. (1995). Heterogenous demand for public goods: Behavior in the voluntary contributions mechanism. *Public Choice*, 85(3):249–266.

Fisman, R., Kariv, S., and Markovits, D. (2007). Individual preferences for giving. *The American Economic Review*, 97(5):1858–1876.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, 19(4):25–42.

Friedman, M. (1962). *Capitalism and freedom*. University of Chicago press.

Fudenberg, D. and Levine, D. K. (2006). A dual-self model of impulse control. *The American Economic Review*, 96(5):1449–1476.

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

Gächter, S. (2012). Human behaviour: A cooperative instinct. *Nature*, 489(7416):374–375.

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

Galizzi, M. M. and Navarro-Martinez, D. (2015). On the external validity of social-preference games: A systematic lab-field study. *Workin Paper Series*.

Gangadharan, L. and Nemes, V. (2009). Experimental analysis of risk and uncertainty in provisioning private and public goods. *Economic Inquiry*, 47(1):146–164.

Gilboa, I. (2009). *Theory of decision under uncertainty*, volume 1. Cambridge university press Cambridge.

Gill, D. and Prowse, V. (2015). Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. *forthcoming in Journal of Political Economy.*

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

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

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

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

Grimm, V. and Mengel, F. (2012). An experiment on learning in a multiple games environment. *Journal of Economic Theory*, 147(6):2220–2259.

Grösche, P. and Schröder, C. (2014). On the redistributive effects of Germany's feed-in tariff. *Empirical Economics*, 46(4):1339–1383.

Grossman, Z., Van der Weele, J. J., and Andrijevik, A. (2014). A test of dual-process reasoning in charitable giving. *Working Paper (Available at SSRN: 2520585).*

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

Hanley, N. and Shogren, J. F. (2005). Is cost-–benefit analysis anomaly–proof? *Environmental and Resource Economics*, 32(1):13–24.

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

Hauge, K. E., Brekke, K. A., Johansson, L.-O., Johansson-Stenman, O., and Svedsäter, H. (2009). Are social preferences skin deep? Dictators under cognitive load. *GUPEA Working Papers in Economics 371.*

Hauser, O. P., Rand, D. G., Peysakhovich, A., and Nowak, M. A. (2014). Cooperating with the future. *Nature*, 511(7508):220–223.

Heckman, J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–482.

Heckman, J. J. and Kautz, T. (2013). Fostering and measuring skills: Interventions that improve character and cognition. *NBER Workin Paper No. 19656.*

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

Heineck, G. and Anger, S. (2010). The returns to cognitive abilities and personality traits in Germany. *Labour Economics*, 17(3):535–546.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., and Gintis, H. (2004). *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies.* Oxford University Press.

Hoppe, E. I. and Kusterer, D. J. (2011). Behavioral biases and cognitive reflection. *Economics Letters*, 110(2):97–100.

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

Huck, S. and Rasul, I. (2011). Matched fundraising: Evidence from a natural field experiment. *Journal of Public Economics*, 95(5):351–362.

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

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

IPPC (2013). *Summary for policy makers, Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change.* Cambridge University Press, Cambridge.

Isaac, R. M., Norton, D. A., and Pevnitskaya, S. (2013). Polarized demands for public goods and the generalized voluntary contributions mechanism. *Working Paper Florida State University.*

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

Isaac, R. M., Walker, J. M., and Thomas, S. H. (1984). Divergent evidence on free riding: An experimental examination of possible explanations. *Public Choice*, 43(2):113–149.

Jones, G. (2008). Are smarter groups more cooperative? Evidence from prisoner's dilemma experiments, 1959 to 2003. *Journal of Economic Behavior & Organization*, 68(3):489–497.

Jones, G. R. and George, J. M. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review*, 23(3):531–546.

Jones, M. T. (2014). Strategic complexity and cooperation: An experimental study. *Journal of Economic Behavior & Organization*, 106:352–366.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review*, 93(5):1449–1475.

Kahneman, D. (2011). *Thinking, fast and slow.* Farrar, Straus and Giroux.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1990). Experimental tests of the endowment effect and the coase theorem. *Journal of Political Economy*, 98(6):1325–1348.

Kahneman, D. and Thaler, R. H. (2006). Anomalies: Utility Maximization and Experienced Utility. *The Journal of Economic Perspectives*, 20(1):221–234.

Kanazawa, S. and Fontaine, L. (2013). Intelligent people defect more in a one-shot prisoner's dilemma game. *Journal of Neuroscience, Psychology, and Economics*, 6(3):201–213.

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

Keser, C. (1996). Voluntary contributions to a public good when partial contribution is a dominant strategy. *Economics Letters*, 50(3):359–366.

Keser, C. and Winden, F. v. (2000). Conditional cooperation and voluntary contributions to public goods. *The Scandinavian Journal of Economics*, 102(1):23–39.

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

Kessler, J. B. and Meier, S. (2014). Learning from (failed) replications: Cognitive load manipulations and charitable giving. *Journal of Economic Behavior & Organization*, 102:10–13.

Kesternich, M., Lange, A., and Sturm, B. (2014a). The impact of burden sharing rules on the voluntary provision of public goods. *Journal of Economic Behavior & Organization*, 105:107–123.

Kesternich, M., Löschel, A., and Römer, D. (2014b). The long-term impact of matching and rebate subsidies when public goods are impure: Field experimental evidence from the carbon offsetting market. *ZEW-Centre for European Economic Research Discussion Paper*, (14-098).

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

Kiss, H., Rodriguez-Lara, I., and Rosa-García, A. (2015). Think twice before running! Bank runs and cognitive abilities. *fourthcoming: Journal of Behavioral and Experimental Economics*.

Knack, S. and Keefer, P. (1997). Does social capital have an economic payoff? A cross-country-investigation. *The Quarterly Journal of Economics*, 112(4):1251–1288.

Kocher, M., Myrseth, K., Martinsson, P., and Wollbrant, C. (2012). Strong, bold, and kind: Self-control and cooperation in social dilemmas. *Working Paper*.

Kocher, M. G., Martinsson, P., and Visser, M. (2008). Does stake size matter for cooperation and punishment? *Economics Letters*, 99(3):508–511.

Kocher, M. G., Pahlke, J., and Trautmann, S. T. (2013). Tempus fugit: Time pressure in risky decisions. *Management Science*, 59(10):2380–2391.

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

Krajbich, I., Armel, C., and Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, 13(10):1292–1298.

Krajbich, I., Bartling, B., Hare, T., and Fehr, E. (2015). Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications*, 6.

Krajbich, I., Oud, B., and Fehr, E. (2014). Benefits of neuroeconomic modeling: New policy interventions and predictors of preference. *The American Economic Review*, 104(5):501–506.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2):443–477.

Laury, S. K. and Taylor, L. O. (2008). 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.

Ledyard, J. O. (1995). Public Goods: A Survey of Experimental Research. In Kagel, J. and Roth, A., editors, *Handbook of experimental economics*. Princeton University Press, Princeton.

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

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

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

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

List, J. A. (2006). Friend or foe? A natural experiment of the prisoner's dilemma. *The Review of Economics and Statistics*, 88(3):463–471.

List, J. A. (2011). The market for charitable giving. *The Journal of Economic Perspectives*, 25(2):157–180.

Loewenstein, G. and O'Donoghue, T. (2007). Animal spirits: Affective and deliberative processes in economic behavior. *SSRN Working Paper 539843*.

Loewenstein, G., Rick, S., and Cohen, J. D. (2008). Neuroeconomics. *Annu. Rev. Psychol.*, 59:647–672.

Loewenstein, G. and Small, D. A. (2007). The scarecrow and the tin man: The vicissitudes of human sympathy and caring. *Review of General Psychology*, 11(2):112.

Lohse, J. (2014). Smart or selfish? When smart guys finish nice. *University of Heidelberg Department of Economics Discussion Paper Series*, 578.

Lohse, J., Goeschl, T., and Diederich, J. (2014). Giving is a question of time: Response times and contributions to a real world public good. *University of Heidelberg Department of Economics Discussion Paper Series*, (566).

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

Lotito, G., Migheli, M., and Ortona, G. (2012). Is cooperation instinctive? Evidence from the response times in a public goods game. *Journal of Bioeconomics*, 15(2):1–11.

Manzoor, S. H. and Straub, J. D. (2005). The robustness of kingma's crowd-out estimate: Evidence from new data on contributions to public radio. *Public Choice*, 123(3-4):463–476.

Markowitz, E. M. and Shariff, A. F. (2012). Climate change and moral judgement. *Nature Climate Change*, 2(4):243–247.

Martinsson, P., Myrseth, K. O. R., and Wollbrant, C. (2014). Social dilemmas: When self-control benefits cooperation. *Journal of Economic Psychology*, 45:213–236.

Mazzonna, F. and Peracchi, F. (2012). Ageing, cognitive abilities and retirement. *European Economic Review*, 56(4):691–710.

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

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

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

Milinski, M. and Wedekind, C. (1998). Working memory constrains human cooperation in the prisoner's dilemma. *Proceedings of the National Academy of Sciences*, 95(23):13755–13758.

Millet, K. and Dewitte, S. (2007). Altruistic behavior as a costly signal of general intelligence. *Journal of Research in Personality*, 41(2):316–326.

Mullainathan, S. and Shafir, E. (2013). *Scarcity: Why having too little means so much.* Macmillan.

Myrseth, K. and Wollbrant, C. (2015). Intuitive cooperation refuted: Commentary on Rand et al. (2012) and Rand et al. (2014). *GUEPA Working Papers in Economics*, 617.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review*, 85(5):1313–1326.

Nielsen, U. H., Tyran, J.-R., and Wengström, E. (2014). Second thoughts on free riding. *Economics Letters*, 122(2):136–139.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, 92(1):91–112.

Nordhaus, W. D. (1991). To slow or not to slow: The economics of the greenhouse effect. *The Economic Journal*, pages 920–937.

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

Nordhaus, W. D. (2015). Climate clubs: Overcoming free-riding in international climate policy. *American Economic Review*, 105(4):1339–70.

Nosenzo, D., Quercia, S., and Sefton, M. (2015). Cooperation in small groups: The effect of group size. *Experimental Economics*, 18(1):4–14.

Nowak, M. A. (2012). Evolving cooperation. *Journal of Theoretical Biology*, 299:1–8.

Oechssler, J., Roider, A., and Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, 72(1):147–152.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

Ostrom, E., Walker, J., and Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *The American Political Science Review*, 86(2):404–417.

Oxoby, R. J. and Spraggon, J. (2013). A clear and present minority: Heterogeneity in the source of endowments and the provision of public goods. *Economic Inquiry*, 51(4):2071–2082.

Paolacci, G., Chandler, J., and Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision making*, 5(5):411–419.

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

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

Piovesan, M. and Wengström, E. (2009). Fast or fair? A study of response times. *Economics Letters*, 105(2):193–196.

Pollak, R. A. (1998). Imagined risks and cost-benefit analysis. *American Economic Review*, 88(2):376–380.

Ponti, G. B. and Rodriguez-Lara, I. (2015). Social preferences and cognitive reflection: Evidence from dictator game experiment. *Frontiers in Behavioral Neuroscience*, 9:146.

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

Rand, D. G., Greene, J. D., and Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, 489(7416):427–430.

Rand, D. G., Newman, G. E., and Wurzbacher, O. M. (2015). Social context and the dynamics of cooperative choice. *Journal of behavioral decision making*, 28(2):159–166.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., and Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, 5.

Recalde, M. P., Riedl, A., and Vesterlund, L. (2014). Error prone inference from response time: The case of intuitive generosity. *CESifo Working Paper Series*, 4987.

Reinstein, D. and Riener, G. (2012). Decomposing desert and tangibility effects in a charitable giving experiment. *Experimental Economics*, 15(1):229–240.

Reuben, E. and Riedl, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, 77(1):122–137.

Roch, S. G., Lane, J. A., Samuelson, C. D., Allison, S. T., and Dent, J. L. (2000). Cognitive load and the equality heuristic: A two-stage model of resource overconsumption in small groups. *Organizational Behavior and Human Decision Processes*, 83(2):185–212.

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

Rubinstein, A. (2007). Instinctive and cognitive reasoning: A study of response times. *The Economic Journal*, 117(523):1243–1259.

Rubinstein, A. (2013). Response time and decision making: An experimental study. *Judgment and Decision Making*, 8(5):540–551.

Rustichini, A. (2015). The role of intelligence in economic decision making. 5:32–36.

Rydval, O., Ortmann, A., and Ostatnicky, M. (2009). Three very simple games and what it takes to solve them. *Journal of Economic Behavior & Organization*, 72(1):589–601.

Salanié, F. and Treich, N. (2009). Regulation in Happyville. *The Economic Journal*, 119(537):665–679.

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

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

Schotter, A. (2008). What's so informative about choice? In Schotter, A. and Caplin, A., editors, *The foundations of positive and normative economics: a handbook*, pages 70–94. Oxford University Press.

Schotter, A. and Trevino, I. (2014). Is response time predictive of choice? An experimental study of threshold strategies. *WZB Discussion Paper No. SP II 2014-305*.

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

Schulz, J. F., Fischbacher, U., Thöni, C., and Utikal, V. (2014). Affect and fairness: Dictator games under cognitive load. *Journal of Economic Psychology*, 41:77–87.

Schumacher, H., Kesternich, I., Kosfeld, M., and Winter, J. (2014). Us and them: Distributional preferences in small and large groups. *CESifo Working Paper Series*, (453).

Segal, N. L. and Hershberger, S. L. (1999). Cooperation and competition between twins: Findings from a prisoner's dilemma game. *Evolution and Human Behavior*, 20(1):29–51.

Shang, J. and Croson, R. (2009). 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.

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

Small, D. A. and Loewenstein, G. (2003). Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk and Uncertainty*, 26(1):5–16.

Small, D. A., Loewenstein, G., and Slovic, P. (2007). Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes*, 102(2):143–153.

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

Spiliopoulos, L. and Ortmann, A. (2014). The BCD of response time analysis in experimental economics. *SSRN Working Paper 2401325*.

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

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

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

Suter, R. S. and Hertwig, R. (2011). Time and moral judgment. *Cognition*, 119(3):454–458.

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

Terhune, K. W. (1974). 'Wash-In', 'Wash-Out', and systemic effects in extended prisoner's dilemma. *Journal of Conflict Resolution*, 18(4):656–685.

Teyssier, S. (2012). Inequity and risk aversion in sequential public good games. *Public Choice*, 151(1):91–119.

Thöni, C., Tyran, J.-R., and Wengström, E. (2012). Microfoundations of social capital. *Journal of Public Economics*, 96(7–8):635–643.

Tinghög, G., Andersson, D., Bonn, C., Böttiger, H., Josephson, C., Lundgren, G., Västfjäll, D., Kirchler, M., and Johannesson, M. (2013). Intuition and cooperation reconsidered. *Nature*, 498(7452):E1–E2.

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

Tonin, M. and Vlassopoulos, M. (2013). Sharing one's fortune? An experimental study on earned income and giving. *CesIfo Working Paper No. 4475*.

Toplak, M. E., West, R. F., and Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, 39(7):1275–1289.

Toplak, M. E., West, R. F., and Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & Reasoning*, 20(2):147–168.

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

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

Van Lange, P. A., Joireman, J., Parks, C. D., and Van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes*, 120(2):125–141.

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

Verkoeijen, P. P. and Bouwmeester, S. (2014). Does intuition cause cooperation? *PloS one*, 9(5).

Vesterlund, L. (2014). Voluntary giving to public goods: Moving beyond the linear VCM. In Kagel, J. and Roth, A., editors, *Handbook of experimental economics*. Priceton University Press, Priceton.

Volk, S., Thöni, C., and Ruigrok, W. (2011). Personality, personal values and cooperation preferences in public goods games: A longitudinal study. *Personality and individual Differences*, 50(6):810–815.

Volk, S., Thöni, C., and Ruigrok, W. (2012). Temporal stability and psychological foundations of cooperation preferences. *Journal of Economic Behavior & Organization*, 81(2):664–676.

Vollan, B. and Ostrom, E. (2010). Cooperation and the commons. *Science*, 330(6006):923–924.

Voors, M., Turley, T., Kontoleon, A., Bulte, E., and List, J. A. (2012). 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.

Wechsler, D. (1955). Wechsler Adult Intelligence Scale. *Psycholgical Corporation.*

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

Weitzman, M. L. (2009). On modeling and interpreting the economics of catastrophic climate change. *Review of Economics and Statistics*, 91(1):1–19.

West, S. A., El Mouden, C., and Gardner, A. (2011). Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior*, 32(4):231–262.

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

Zaki, J. and Mitchell, J. P. (2013). Intuitive prosociality. *Current Directions in Psychological Science*, 22(6):466–470.

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

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